Author:
Ruzzi, Dario

Title:
Flight to Safety in Financial Modelling

University of Bristol

School of Economics, Finance and Management

Flight to Safety in Financial Modelling

Dario Ruzzi

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Philosophy in the Faculty of Social Sciences and Law

June, 2019

Supervisors of the PhD Thesis:

Prof. Nick Taylor, School of Economics, Finance and Management, University of Bristol

Dr. Evarist Stoja, School of Economics, Finance and Management, University of Bristol

Abstract

Following a large body of literature on cross-market crisis linkages, this thesis investigates the potential effects of "flight-to-safety" phenomena in two strands of financial modelling: volatility forecasting and bond pricing. In particular, we consider the switching from equity to safe haven assets for modelling the dynamics of financial market volatility and US Treasury prices.

Chapter 2 examines the relationship between the occurrence of flight-to-safety (FTS) episodes and subsequent equity return volatility. We assign to each day in the sample the status of flight-to-safety based on either stock-bond or stock-gold return comovements. The link between flight-to-safety and future equity return volatility is assessed by including the lag of the FTS conditioning variable in a model for the daily realised variance of the S&P 500 index returns. The results are consistent with our expectations: there is a positive, statistically significant interaction between FTS and the next-day volatility of the stock market. Out-of-sample, we find that the one-day ahead volatility forecasts based on FTS information outperform those of some of the most common models proposed in the literature.

Chapter 3 extends the research scope by examining how flight-to-safety can improve the modelling of both equity and safe haven return volatility in a multi-period setting. To this end, we propose a new proxy for FTS based on the realised semi-covariance computed from negative intraday equity returns and positive intraday safe haven returns. We use the FTS proxy to model directly the conditional variance of equity returns, but we also allow for an effect on the safe haven by means of a volatility spillover from the stock market. The incremental value of employing flight-to-safety for volatility forecasting is not only statistically but also economically significant. When comparing with a benchmark, we find that the predictions of our approach are more accurate, and their application to asset allocation leads to better portfolio performance for an investor implementing volatility-timing strategies over horizons as long as one month.

Chapter 4 turns the attention to the role of equity jump tail risk in pricing government bonds. We estimate a term structure model for US interest rates where bond prices are determined also by an equity left tail factor that reflects investors' fear of abrupt negative return shocks to the international stock market. The model shows that equity tail risk is priced in the US term structure. Consistent with the theory of flight-to-safety, we find that the response of Treasury bond yields and future excess returns to a shock to the equity tail factor is negative and opposite to what happens in the stock market. The evidence of flight-to-safety is stronger at the short end of the US yield curve where equity tail risk has significant explanatory power for Treasury risk premia and is responsible for large term premium drops since the recent financial crisis.

Overall, the findings of this thesis contribute to the literature on flight-to-safety by showing that the switching from equity to safe haven assets in times of stress has important implications for volatility forecasting. Specifically, accounting for flight-to-safety allows for improved predictions of return volatility and better performance of a volatility-timing strategy. Furthermore, this thesis provides guidance on how the safety attribute of Treasuries varies with the maturity of the bonds. Lastly, this study sheds light on the existence of a common predictor between equity and bond markets, whose existence can be justified by the safe haven potential of US Treasuries.

Acknowledgements

I would like to express my deepest gratitude to my supervisors, Prof. Nick Taylor and Dr. Evarist Stoja. I am truly indebted for their expert guidance and invaluable support throughout my PhD.

I would like to thank Dr. Mirco Rubin for encouraging me in all my research endeavours and providing me with his most valuable comments on all the parts of my thesis.

Finally, I would like to thank Torben Andersen, Nicola Fusari and Viktor Todorov for providing precious feedback and advice on my job market paper, which is based on a chapter of this thesis.

Author's Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's *Regulations and Code of Practice for Research Degree Programmes* and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: .... DATE: ....

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In the last thirty years, extreme movements in one financial market have spilled over into others. Comovement has been observed not only across markets of the same asset class but also across assets of different classes. Motivated by the impact on portfolio diversification and the possible effects on the real economy, the transmission of shocks among markets lies at the heart of the analysis of international financial stability. Contagion phenomena have initially been studied in regard to stock markets, mainly using correlation analysis as in Hamao et al. (1990), King and Wadhwani (1990), Lin et al. (1994), Susmel and Engle (1994), and Forbes and Rigobon (2002). Empirical work has subsequently been undertaken in order to examine the spillover within other asset classes. Examples are the currency crises studied, among others, by Sachs et al. (1995), Eichengreen et al. (1996) and Kaminsky and Reinhart (2000). Cross-asset crisis linkages were first explored in the seminal paper of Hartmann et al. (2004), who focus on stock and government bond market crashes. The field has seen an enormous body of work since then. Undoubtedly, one topic that has received considerable attention within this literature is the switch in the sign of the relation between stocks and government bonds that took place at the beginning of the twenty-first century in the United States. Specifically, the correlation between these two asset classes turned from being mostly positive to being mostly negative. The predominant explanation for this change seems to remain that of David and Veronesi (2013): investors' fears of hyperinflation, which prevailed in the 1980s and 90s, switched to deflation fears in the 2000s. The stock-bond covariance was positive when agents were expecting high inflation because such expectations cause a drop in valuation for both assets. The relation turned negative when a too-low inflation became the agents' primary concern as stock prices fall but bond prices rise in a deflationary recession regime. Besides expected inflation, several other factors have been used to explain the time variation in the stock-bond covariance. David and Veronesi (2016) provide a literature survey of the work done on this subject.

Episodes of financial contagion are frequently accompanied by, and associated with, episodes of "flight-to-safety". This is market stress with large losses on the equity market and, simultaneously, positive price movements in a safe haven, namely an asset able to retain its value and reduce portfolio losses in crisis periods. A number of authors have described the role of US government securities as a shelter against stock market losses, see, for instance, Longstaff (2004), Connolly et al. (2005), Krishnamurthy and Vissing-Jorgensen (2012) and Adrian et al. (2018). However, several other assets have been considered for this role: Euro-area government bonds, see Beber et al. (2009); currencies, see Beber et al. (2014); commodities, see Baur and Lucey (2010), Baur and McDermott (2010), and Li and Lucey (2017). Furthermore, the implications of flight-to-safety and safe havens have been documented in wider contexts by Bekaert et al. (2009), Baele et al. (2015), Ghysels et al. (2016), Bekaert and Engstrom (2017), Caballero et al. (2017), Boudry et al. (2018), and Adrian et al. (2019).

This thesis considers flight-to-safety (FTS) for (i) more accurately modelling financial market volatility dynamics; and (ii) pricing government bonds with the downside tail risk of the stock market. Both of these applications have strong influences on important financial decisions, first and foremost portfolio diversification and asset allocation.

Regarding point (i), over the past fifteen years there has been a growing consensus that flight-to-safety episodes are triggered by high equity market volatility and low aggregate liquidity (see, e.g., Vayanos (2004); Caballero and Krishnamurthy (2008); Brunnermeier and Pedersen (2009)). The theory behind these phenomena is simple: seeking liquidity in times of heightened market uncertainty, investors stampede out of risky assets and move into safe havens. In order to understand the economic rationale for why flight-to-safety can be important for modelling the future dynamics of financial return volatility, we suggest the following explanation. In the seminal paper of Vayanos (2004), investors are all assumed to be fund managers that, fearing redemptions due to a bad performance, shift their holdings into instruments like cash and Treasuries in times of elevated risk. If we relax this assumption, then we should also consider those investors that, acting with a delay with respect to fund managers, rebalance their portfolio once the FTS spell has initiated or even terminated. For instance, many retail investors take some time to react to market events. According to Barber et al. (2001), this happens either because retail investors only gain access to consensus recommendation changes one or more days after the changes occur, or because they cannot engage in daily portfolio rebalancing with the same ease as institutional investors. Additionally, there could be a delayed portfolio rebalancing by the investors that hold their losing investments for longer than it is considered rational. This reluctance to sell losers is generally referred to as the disposition effect (see, e.g., Shefrin and Statman (1985); Odean (1998)). All these activities of fund allocation, which are not as timely as the switching from equity to safe haven assets by the fund managers of Vayanos (2004), can create further volatility in the market. The implications of flight-to-safety for future return volatility can also be understood in terms of countries' integration into world capital markets. With regard to this, Adrian et al. (2019) show that higher world capital market integration not only fuels growth but also exposes countries to greater macroeconomic and financial risks. This insight, together with the fact that flight-to-safety may be an episode of world market integration with strong comovement in the international financial markets (see, for instance, Baur and Lucey (2009)), is consistent with periods of high volatility induced by FTS events. In light of the considerations raised, therefore, there appears to be a considerable need to examine how flight-to-safety can be used to improve the forecasts of future return volatility.

Turning to point (ii), the literature on asset pricing has found that downside risk carries a significant premium: investors are willing to pay a price for securities that perform well in adverse economic conditions. In a recent paper, Farago and Tédongap (2018) find that a consumption-based general equilibrium model that includes disappointment-related factors (i.e. measures of downside risk) is very successful in pricing the cross-section of stocks, options, currencies, commodity futures, Treasury bonds and corporate bonds. Furthermore, it has been shown that the pricing of negative jump tail risk accounts for a significant fraction of the equity risk premium. Andersen et al. (2015b), Bollerslev et al. (2015), Andersen et al. (2017a,b) and

Li and Zinna (2018) are among the studies that find a good degree of stock return predictability in the investors' fear of a market crash. The aversion of investors to extreme downside events, which is stronger for the short run (Li and Zinna, 2018), has been shown to have significant interactions with measures of the monetary policy stance such as the real interest rate. Bekaert et al. (2013) find that a lax monetary policy decreases risk aversion (i.e. it increases risk appetite) in the future, and periods of high risk aversion are followed by a looser monetary policy stance. The relationship is also economically significant since monetary policy shocks account for a significant proportion of the variance of the risk aversion proxy. In line with these findings, Busetti and Caivano (2018) show that the long-term behaviour of the real interest rate has been determined also by investors' increased appetite for safe assets. These insights, together with the well-documented impact of monetary policy shocks on government bond rates (see, e.g., Gordon and Leeper (1994); Buraschi and Whelan (2016)) and the safe haven nature of top-tier government bonds, allow for a relation between the investors' fear of a market crash and the risk premium in both equity and bond markets. Therefore, there appears to be a role for the downside tail risk of the stock market in pricing government bonds during distress periods.

The objective of this thesis is the integration of stock and safe haven return comovements in financial modelling. The rest of the thesis is articulated in three additional chapters, followed by a concluding chapter that draws together the conclusions of this study and suggests directions for future research. Chapter 2 presents several measures that identify flight-to-safety days. These measures are based on US equity, Treasury bond and gold returns, and are shown to have predictive power for future stock return volatility. Chapter 3 proposes a new proxy for FTS computed from intraday equity and safe haven returns. An empirical application uncovers the statistical and economic relevance of flight-to-safety for predicting equity and safe haven return volatility up to one month ahead. Chapter 4 introduces the downside tail risk of the equity market into a term structure model, and considers the response of US government securities to extreme stock market events. The content of each chapter is outlined more extensively below.

Detailed outline of the chapters

Chapter 2: Flight to Safety: Does it matter for Volatility Forecasting?

The second chapter examines the relationship between the occurrence of flight-to-safety episodes and subsequent equity return volatility. It is well established in the literature that high equity market volatility and low aggregate liquidity are key factors that lead to flight-to-safety, see, e.g., Vayanos (2004); Caballero and Krishnamurthy (2008); Brunnermeier and Pedersen (2009). In this chapter, we make the conjecture that FTS events not only follow high equity market volatility but are also likely to precede periods of high volatility. Previous studies provide empirical support for our argument that a link exists between flight-to-safety and future equity return volatility. For instance, it is clear from the work of Anand et al. (2013) that the effects of some stress events - including flights - last well beyond the end of the spell. In addition, we know from the work of Baele et al. (2015) and Boudry et al. (2018) that FTS days signal an impeding downturn in economic activity. Combining this result with the considerable evidence that stock market volatility is always higher during recessions (see, e.g., Schwert (1989); Hamilton and Lin (1996)), we propose and test the predictive content of FTS for future equity market volatility. Although the literature seems to be in agreement as to what factors lead to flight-to-safety, identifying these stress episodes in the capital markets remains a non-trivial task due to their unpredictable nature and non-univocal definition. Baele et al. (2015) define and empirically characterise flight-to-safety with regard to the relation between stock and government bond returns. We broaden their definition to include the stock-gold return relation. Using the fact that gold and government bonds may act as safe haven against stock market losses at different times (see, e.g., Baur and Lucey (2010); Baur and McDermott (2010, 2012); Li and Lucey (2017)), we attempt a more thorough characterisation of flight-to-safety than Baele et al. (2015).

The empirical application assigns to each day in the sample the status of being a FTS day using the methodologies of Baele et al. (2015). We apply these techniques to: (i) US equity and Treasury bond returns; (ii) US equity and gold returns. The inclusion of the yellow metal in the analysis leads to additional flight-to-safety episodes other than those indicated by sovereign debt securities. To investigate the relation between flight-to-safety and future stock market volatility, we include the lag of the dummies flagging the occurrence of FTS days in the Corsi (2009) model for the daily realised variance of the S&P 500 equity index returns. We find a statistically significant interaction between FTS and the next-day volatility of the stock market. We follow the analysis of Patton (2011) and use the Reality Check test of White (2000) to compare the forecasting performance of the new model with that of some existing volatility models which are popular but ignore FTS. The fact that the one-day ahead forecasts of our approach significantly outperform those of the benchmarks is an indication that explicitly accounting for FTS in the model specification is beneficial for predicting future variation in the stock market.

Chapter 3: Forecasting Volatility with a Semi-Covariance-based FTS measure

The third chapter corresponds to the paper *Flight to Safety in Volatility Forecasting*, joint work with Nick Talyor. It introduces a new proxy for FTS based on the realised semi-covariance computed from negative intraday equity returns and positive intraday safe haven returns. As in Chapter 2, we treat US Treasury bonds and gold as two alternative safe haven assets between which to choose. Unlike the previous chapter, however, we assess the benefits of using flight-to-safety in volatility forecasting for: (i) not only equity but also safe haven returns; (ii) not only one-period but also multi-period horizons. For the advocated proxy we propose realised semi-covariance between falling equity and rising safe haven returns because its value is closely related to the symptoms of flight-to-safety described by Baele et al. (2015) and strongly correlates with the FTS dummies of Chapter 2. Realised semi-covariance belongs to the class of estimators that are computed from high-frequency data and are extremely popular in forecasting volatility as they are highly informative about the current level of risk (see, e.g., Andersen et al. (2003); Ghysels et al. (2006); Andersen et al. (2007); Corsi (2009); Shephard and Sheppard (2010); Hansen et al. (2012)). Recently, Bollerslev et al. (2017) have added realised semi-covariance to the more traditional realised measures of variance and semi-variance. While the approach of Bollerslev et al. (2017) to forecasting with realised semi-covariance has proven useful when the joint negative returns of a large number of stocks are considered, in this chapter we focus on just two assets - equity and the safe haven - and use their high-frequency returns of opposite sign to improve the predictions of the asset volatilities.

The empirical application revisits the GARCH model by Hansen et al. (2014) in order to assess the incremental value of employing flight-to-safety for predicting equity and safe haven return volatility. Specifically, we include the daily lag of the FTS proxy in the conditional variance equation of equity returns and we allow for a spillover effect from the stock market to the safe haven asset. We find that the FTS proxy has strong predictive power for higher equity return variation. We compare the performance of the proposed model with that of the Hansen et al. (2014) model. The FTS proxy offers substantial gains in terms of the out-of-sample forecasts of both equity and safe haven volatility, especially at long horizons. These statistically significant results are robust to the choice of forecast target and benchmark model. Although the benefits of using FTS are not constant over time, we note that the forecast losses are normally lower during the 2008-09 crisis. We conduct a portfolio exercise as in Fleming et al. (2001, 2003) in order to assess the economic value of using realised semi-covariance between falling equity and rising safe haven returns for volatility forecasting. We find that the more accurate predictions of the asset volatilities can lead to better portfolio performance for an investor implementing volatility-timing strategies over horizons as long as one month and subject to transaction costs. Of the safe havens that we consider it is US Treasuries that, when used to construct the proxy for FTS, yield the most significant improvements in forecasting and volatility-timing performance.

Chapter 4: The Impact of Equity Tail Risk on Bond Risk Premia: Evidence of FTS in the US Term Structure

The fourth and concluding chapter introduces the downside tail risk of the equity market into the pricing of bonds and, by doing so, offers evidence of flight-to-safety for the US government bond market. Over the last decade it has been clearly established that the variables determining the shape of today's term structure do not accurately predict future bond returns. This has led researchers to consider the role of additional factors besides the traditional level, slope and curvature. In this regard, not only higher order principal components (see, e.g., Adrian et al. (2013); Malik and Meldrum (2016)) and new linear combinations of interest rates (see, e.g., Cochrane and Piazzesi (2005, 2008)), but also factors of non-bond-market origin have been proposed. While there is a literature on the informative content of macroeconomic variables with respect to Treasury bond risk premia (see, e.g., Cooper and Priestley (2008); Ludvigson and Ng (2009); Duffee (2011); Joslin et al. (2014)), there does not exist a study examining the effects of equity tail risk on the US yield curve and the cross-section of future bond returns. We aim to fill this gap by considering the possibility that pricing factors of Treasury bonds originate also in the stock market.

To investigate the response of US government securities to extreme stock market events, we estimate a Gaussian affine term structure model in which the main drivers of US interest rates are the principal components of the yield curve and an equity left tail factor. The latter is based on the downside jump intensity factors extracted from US, UK and Euro-zone equity-index options using the Andersen et al. (2015b) model. Since the equity-index option surface embeds rich information about the pricing of extreme events and has proven useful for predicting future equity returns (see, e.g., Bollerslev et al. (2015); Andersen et al. (2017b)), we exploit the theory of flight-to-safety and consider its explanatory power for government bond risk premia. The term structure model, which is estimated as in Adrian et al. (2013), shows that equity tail risk is strongly priced in the US term structure and its pricing is significantly time-varying. Analysing the response of bond prices and future expected returns to a shock to the equity left tail factor, we find that the former increase and the latter shrink when the fear of negative return jumps in the stock market is higher. These observations confirm the role of US Treasuries as a safe haven and, when combined with the previously documented positive relationship between jump tail risk and future equity returns, indicate the presence of a common predictor across the two asset classes. The evidence of flight-to-safety is stronger at the short end of the US yield curve where the equity tail factor has significant explanatory power for future returns and where large drops in the term premia are attributable to equity tail risk. Thus, while the Fed asset purchase programmes have been a major force in lowering longer-term yields since the recent financial crisis (see, e.g., Kaminska and Zinna (2018)), the reduction in shorter-term yields is likely to have been caused by the investors' increased appetite for safe assets.

Chapter 2

Flight to Safety: Does it matter for Volatility Forecasting?

Abstract

We study the relation between the occurrence of flight-to-safety (FTS) episodes and subsequent equity return volatility. We assign to each day in the sample a probability of being an FTS day after observing (ab)normal movements in the US equity, US bond and gold markets. By allowing each FTS day to be an indicator of higher future volatility, we document statistically significant improvements in the accuracy of the 1-day ahead forecasts for the realised variance of the S&P 500 equity index. Superior model performance is found over some of the most common univariate volatility forecasting models proposed in the literature that do not specifically account for these, generally short-lived, market stress episodes.

JEL Codes: C22, C53, C58, G17.

Keywords: Flight to safety, safe haven, gold, realised variance, volatility forecasting.

2.1 Introduction

Top-tier government bonds have long been considered an effective safe haven that can help reduce portfolio losses during times of market uncertainty; see, e.g., Hartmann et al. (2004), Longstaff (2004), Baur and Lucey (2009) and Beber et al. (2009). When equity markets tumble, investors' safe-haven demand increases and a surge in sovereign bond prices is observed. In the present chapter, we ask whether this behaviour can help predict volatility in the equity market. In particular, we evaluate the accuracy of a volatility forecasting model after conditioning on the occurrence of a flight-to-safety (FTS) event. The non-trivial task of identifying FTS greatly depends on the definition used for this type of events and on the choice of the safe haven. The academic literature has already attempted to capture FTS days using Treasury bonds, see Baele et al. (2015). Following the existing literature for the definition of safe haven, we show that the consideration of a second safe asset may result in a more thorough identification of FTS days.

Recently, amid global financial market turbulence, geopolitical tensions, low interest rates, and the need for a safe haven, gold has received a great deal of attention from the financial media with the price of the precious metal reaching an all-time high in 2011. The peculiarities of gold as an investment and a store of wealth have been considered and discussed extensively in the literature; see, e.g., Lawrence (2003), Capie et al. (2005), and McCown and Zimmerman (2006). The safe haven status of gold is not new. In the past, the metal has always rallied during the most noticeable moments of panic such as the Black Monday (19^{th} October 1987), the collapse of LTCM (fall 1998), and the 09/11 terrorist attack in New York. Previous studies have tested the hypothesis of gold being a safe haven by examining its behaviour on all the occasions in which the equity markets were hit by heavy losses. For instance, Baur and Lucey (2010) demonstrate that gold is a safe haven for stocks in the US, UK, and Germany, while Baur and McDermott (2012) consider the role of uncertainty in investor behaviour and infer that gold and bonds are two different types of safe haven. Inspired by these results, we consider whether gold is able to capture additional flight-to-safety spells other than those indicated by sovereign debt securities as in Baele et al. (2015). Our work is complimentary to theirs also in that we are expanding the dataset to include recent episodes of market turbulence due to China's slowdown, cheap oil, terror attacks by ISIS and the UK's vote to leave the EU.

In line with the literature that explains FTS with a general increase in investors' risk aversion, we build a risk measure that captures such phenomena on a daily basis using both Treasuries and gold as safe haven. We then assess the significance of the risk measure within the context of future volatility forecasting. High equity market volatility, together with low aggregate liquidity, is suggested by the literature as a key factor that leads to flight-to-safety, see, e.g., Vayanos (2004); Caballero and Krishnamurthy (2008); Brunnermeier and Pedersen (2009). In this chapter, we make the conjecture that FTS events not only follow and reflect high equity market volatility but are also likely to precede periods of high volatility. There are a number of reasons why this could happen. For instance, after the investment managers described by Vayanos (2004) react to high volatility by shifting their holdings towards safe assets leading to FTS, there could be a portfolio rebalancing effect that creates further volatility in the market. In addition, it is clear from the work of Anand et al. (2013) and Boudry et al. (2018) that the effects of some FTS episodes last well beyond the end of the spell. In line with the study by Baele et al. (2015), Boudry et al. (2018) find that clusters of FTS days predict a decline in future economic activity. Based on this result and the considerable evidence that stock market volatility is always higher during recessions (Schwert, 1989; Hamilton and Lin, 1996), we propose and test the predictive content of FTS for future equity market volatility. A link between FTS and future equity variance is also implied by Baele et al. (2015) who document contemporaneous increases in the VIX during flight-to-safety. Because the conditional expectation of future equity return variance composes the VIX, we find it natural to examine in detail the relationship between the occurrence of FTS episodes and subsequent equity return volatility. In this regard, our study can contribute to existing literature investigating the FTS-related responses in stock volatility and liquidity (Hameed et al., 2010; Greenwood and Thesmar, 2011).[1] Lastly, motivation for our study also stems from several results in the literature, which conclude that risk aversion and market uncertainty are negatively correlated with future stock-bond return correlation and

[1]Hameed et al. (2010) document an interindustry spillover effect in liquidity following periods of large equity market declines. They find this liquidity commonality to be positively related to market volatility but unrelated to idiosyncratic volatility. These results can therefore support the idea of high volatility following FTS episodes. Greenwood and Thesmar (2011) build a fragility measure that is similar to our FTS measure in the sense that it increases when liquidity deteriorates. They find that the fragility of an asset increases when investors experience correlated liquidity shocks and the fragility measure can forecast return volatility. These results can therefore support the idea of forecasting future equity return variance using FTS information.

drive fluctuations in the conditional volatility of returns (Connolly et al., 2005; Bekaert et al., 2009; Baele et al., 2010; Bansal et al., 2010). Since equity volatility is itself a component of the stock-bond return correlation, and since flight-to-safety occurs when risk aversion and market uncertainty are higher, it follows that FTS may have some predictive power for future equity return volatility. Moreover, since lower-than-average and negative stock-bond return correlations are observed in high-volatility stock market regimes (Ilmanen, 2003; Bansal et al., 2010), we can expect FTS days to be predictors of higher equity variance in the future.

Motivated by the above considerations, we extend the Heterogeneous Autoregressive model of Realized Volatility (HAR-RV) proposed by Corsi (2009) with an FTS measure, that is a dummy variable that flags days in which investors flock to safe haven assets. We thus name the extended model "HAR-RV FTS", and examine its out-of-sample performance for the one-day ahead forecasts of S&P 500 realised variance (RV). Following the analysis of Patton (2011), we evaluate the forecasting performance of the HAR-RV FTS model on the basis of mean squared error (MSE) and quasi-likelihood (QLIKE) loss functions.[2] We compare the new model with some existing volatility forecasting models which are popular but ignore FTS. Using the same modification of the Reality Check test proposed by Bollerslev et al. (2016), we formally test and find predictive superiority of the HAR-RV FTS model over the benchmark models under both MSE and QLIKE loss functions.[3] Overall, the results show that explicitly accounting for FTS episodes in the model specification can be beneficial to the forecast of future realised variance.

The remainder of the chapter is structured as follows. Section 2.2 offers a review of the literature related to flight-to-safety, gold, and volatility forecasting. Section 2.3 contains a description of the methodologies used to identify FTS days and presents the empirical results with both a full-sample and a recursive computation of the statistics. Section 2.4 proposes a new model for realised variance based on extending the standard HAR-RV with an FTS measure. Section 2.5 concludes.

[2]When comparing the in-sample and out-of-sample performances of the one-day ahead forecasts from the various volatility models we also report the mean absolute error (MAE), the mean absolute percentage error (MAPE), and the results of the Mincer-Zarnowitz regression. However, as shown by the study of Patton (2011), MAE and MAPE are not part of the family of loss functions able to *yield rankings of volatility forecasts robust to noise in the proxy* (Patton, 2011). Indeed, we assess the forecasts by comparing them with realised variance, which is only an imperfect estimator of the true unobservable volatility.

[3]The Reality Check statistical test in its original specification is formulated in White (2000).

2.2 Related Literature

This chapter relates to at least three literatures. First, there is a literature that describes financial stress events, including flight-to-safety, and the role played by government bonds. Second, there is a literature focused on the characteristics of gold as a financial and potentially safe asset. Third, there is a vast literature for estimating and forecasting financial market volatility and for the stylised facts that should be incorporated in a forecasting model.

2.2.1 Literature on Flights

A number of authors have described the role of government securities during episodes of severe market stress characterised by flights to liquidity, flights to quality, flights to safety, contagion and asset interdependence. For instance, Beber et al. (2009) show that, in times of heightened market uncertainty, Euro-area government bonds benefit from the investors' demand for liquidity. Focusing on the collapse of LTCM hedge fund and the Russian financial crisis that took place in the second half of 1998, Upper (2000) demonstrates that market turmoil does not prevent investors from trading, albeit with higher costs due to worsened liquidity. He finds that on that occasion German government yields fell as investors were flocking to the safe haven asset, but debt market liquidity strongly deteriorated and yield volatility spiked. These findings about bond liquidity in times of stress are consistent with those of Chordia et al. (2005), which demonstrate the systemic nature of liquidity shocks. They identify volatility linkages, in addition to monetary policy decisions and variations in trading activity, as a major determinant of the liquidity comovements in the US equity and Treasury markets. They show that an increase in volatility predicts a negative future shock in the liquidity of both markets. In turn, the illiquidity of Treasury bonds is found, by Goyenko and Sarkissian (2014), to predict stock returns around the world and command a premium in global equity markets.

Using data on the US government bond market, Longstaff (2004) documents a significant liquidity premium, which accounts for up to 15% of the bond value during flight-to-safety. In a similar vein, Vayanos (2004) shows that in periods of high volatility investors are willing to pay a bigger premium for liquid assets and this has direct consequences for pricing. He analyses flight-to-liquidity phenomena by assuming investors are all fund managers that, fearing redemptions due to a bad performance, shift their holdings into instruments like cash and Treasuries in times of elevated risk. He notes that a greater compensation required by investors to accept risk when uncertainty is high provides evidence for increasing risk-aversion and flight-to-quality. His empirical analysis demonstrates that aggregate volatility strongly affects a number of quantities. In times of stress the price differential between government securities of different liquidity and similar other characteristics widen, assets become more negatively correlated with volatility, and illiquid assets become more sensitive to market movements, hence riskier. As a final remark, Vayanos (2004) notes that it is worth studying liquidity "both as an asset characteristic and as a risk factor, in explaining cross-sectional expected returns".

Financial market liquidity is key also in Caballero and Krishnamurthy (2008) and Brunnermeier and Pedersen (2009). The explanation offered by Brunnermeier and Pedersen (2009) for FTS is centered on the role of speculators that, when they stop trading high-risk assets in times of high volatility, cause a deterioration in aggregate liquidity. Caballero and Krishnamurthy (2008) note that investor behaviour is strongly affected by Knightian uncertainty (Knight, 1921). When liquidity is scarce and agents fear that it will eventually dry up, they stampede out of risky assets and move into safe havens. The analysis of Caballero and Krishnamurthy (2008) shows that liquidity injections by central banks are crucial to stabilise the financial system.

Since black-swan type of events have become more and more numerous, attention has focused on cross-asset linkages. Hartmann et al. (2004), after defining flight-to-quality as "a crash in stock markets accompanied by a boom in government bond markets", find evidence of limits to the propagation of a crisis from equity to bonds. Instead of following common practice in the literature and using conditional correlation analysis, they rely on Extreme Value Theory to estimate in a non-parametric fashion the likelihood of a crash spillover from one market to another. They argue that single market crashes are rare events that "happen once or twice per lifetime". These findings strongly depend on the severity of the crashes considered; in their study the authors talk of a 20% and a 8% weekly loss for stocks and bonds, respectively. They find that simultaneous crashes of equity markets are more likely to be seen than for bond markets, while cross-asset co-crashes are quite infrequent.[4] Based on these results, the authors claim that

[4]Regarding equity market co-crashes, Forbes and Rigobon (2002) suggest that it is more appropriate to think

contagion between different assets "cannot be a prevalent phenomenon among G-5 countries", which is what provides motivation for researching the existence of a safe haven.

The conclusions drawn by Hartmann et al. (2004) with respect to the infrequency of equity-bond co-crashes are consistent with the findings of Ilmanen (2003), Connolly et al. (2005), Baele et al. (2010) and Bansal et al. (2010). These studies provide strong evidence for lower and negative stock-bond return correlations during and following periods of elevated stock market uncertainty. For instance, negative correlations between the two assets are attributed by Ilmanen (2003) to deflationary recessions, equity weaknesses and high-volatility stock market regimes. Further, Connolly et al. (2005) note that a flight-to-quality occurs at the same time as the implied volatility from equity index options, measured by the VIX, reaches relatively high levels. In addition, they find that the correlation between stock and bond returns is lower over the month following the increase in the VIX. Similarly, by using a bivariate regime-switching model on US stock and bond returns, Bansal et al. (2010) show that a higher lagged VIX indicates a higher probability of transitioning and staying in a high-stress regime where stock market volatility is high, stock-bond return correlation is lower and bond return mean is higher. Lastly, with the use of a dynamic factor model, Baele et al. (2010) demonstrate the importance of the equity variance risk premium, which represents the compensation demanded for bearing variance risk, in explaining time variation and negative values in the stock-bond return relation.

Baur and Lucey (2009) add to the literature that deals with comovements in the equity and bond markets by saying that contagion (flight-to-quality) occurs when, following a crisis, there is a significant increase (decrease) in the correlation coefficient and a positive (negative) correlation level as well. Their econometric analysis, based on the time-varying relationships between stocks and bonds of a number of countries, finds that a global nature characterises the flights between the two assets. They argue that simultaneous flights to quality around the world, like those during the Russian crisis in 1998, and simultaneous flights from quality, like those during the Enron crisis in 2001, provide indirect evidence for cross-country contagion.

of some market comovements as simple interdependence rather than contagion. Their statistical work, which accounts for the effect of market volatility on correlation across asset classes, finds no evidence of contagion in 1987 (the US stock market crash), 1994 (Mexican devaluation), and 1997 (Asian crisis). They claim that markets strongly comoved but cross-market linkages remained the same as during non-crisis periods.

The study of Baele et al. (2015) came to quite different conclusions on the nature of flights to safety as only one out of four flights can be characterised as global. They define flight-to-safety as a period that features high equity volatility, large and negative equity returns, large and positive bond returns, and a correlation that temporarily turns negative.[5] With a plethora of econometric techniques, which make use of stock and bond returns from over 20 countries, they attempt a day-to-day identification of FTS episodes from 1980 to 2012.[6] They suggest that on average 7 working days in a year can be categorised as FTS, with the bond market outperforming the equity market by 2.5 to 4% on those days. They note that in nearly 90% of the cases the safe haven property of Treasuries disappears within 3 days but it can also lasts for more than 10 days. Interestingly, their statistical model identifies very few episodes before 1995. We use this result to support a claim that something is missing in the way FTS episodes are identified. Therefore, in our study we propose to append gold to the analysis in order to capture also those flight-to-safety spells in which investors used the precious metal, rather than the US Treasuries, as a shelter against equity market losses.

In a recent study, Boudry et al. (2018) use the econometric techniques proposed by Baele et al. (2015) for FTS identification in order to investigate the effects of flight-to-safety on the real estate industry. Their analysis shows that the real estate market provides a partial hedge against FTS, with daily returns and liquidity being less affected than in many other industries. Moreover, by studying the long-run implications of flight-to-safety for economic fundamentals, Boudry et al. (2018) claim that FTS days have effects beyond the actual event days and signal an impeding downturn in economic activity. In particular, they find that realised GDP growth is lower in quarters following an FTS cluster. These findings add to the studies of Bekaert et al. (2009) and Bekaert and Engstrom (2017), that explain the effect of FTS on asset valuation in terms of the investors' reaction to changes in the real economy.

The spirit of Baele et al. (2015) for the empirical characterisation of flight-to-safety is also

[5]Baele et al. (2015) expect the (unconditional) correlation between equity and bond returns to be positive in normal times because, they claim, both are long duration assets.

[6]More will be said later about the statistical techniques implemented in Baele et al. (2015) as we follow their approach in identifying FTS events with some modifications. In particular, to ensure that our FTS measure does not incorporate future information at any point in time, we build on the method of Hollo et al. (2012), whose real-time systemic stress index is produced using recursive computation on expanding samples.

followed by Ghysels et al. (2016), who study time variation in the risk-return relationship and pricing of volatility risk. Ghysels et al. (2016) find that the expected excess equity market return is positively related to its conditional variance only over samples that exclude financial crises. Moreover, using an FTS conditioning variable, they show that the traditional risk-return trade-off does not hold for those months or quarters that follow an FTS episode. This result clearly suggests that investors are not willing to buy volatility during flight-to-safety.

2.2.2 Literature on Gold as a financial (safe) asset

Our motivation of including gold in the study follows from the strand of literature that describes the peculiarities of the precious metal. Lawrence (2003) talks of "insulation" of gold from the business cycle. No significant linear relationship is found between the metal and changes in inflation, GDP, and interest rates. On the other hand, core macroeconomic variables are found to be strongly correlated with equity and bond returns. Several properties of gold, including its homogeneous, fungible and indestructible nature, are pointed out in the study. The author argues that the "insulation" of the precious metal can be explained with the unique supply and demand dynamics of its market. Indeed, a positive demand shock in the gold market can be faced not only with newly-mined product but also mostly with gold recovered from scrap, thus limiting the pressure on price. Noting that the yellow metal is less correlated with equity and bond indices than other commodities, the study concludes that gold may be an effective portfolio diversifier when it comes to managing portfolio risk.

Mills (2004), among others, offers a detailed analysis of the statistical behaviour of gold price from 1971 to 2002. The start point, 1971, represents an important date in the history of gold. In August 1971, the US President Nixon ended the dollar convertibility to gold and the price of the yellow metal, close to $30 per ounce at that time, started being determined by the market. The study attributes the first spikes in the time series of gold price mainly to the collapse of the Bretton Woods system, while the rapid surge in the second half of the 70's, which culminated in a peak between $700 and $800 in 1980, could be the result of high inflation and USD depreciation. The study notices that, after a downward trend that lasted until mid-1982, the precious metal started trading in a much more stable range of $250-$500.

The statistical analysis addresses the fluctuations in the non-stationary time series of gold price and finds evidence of autocorrelation up to lag 15 days. The paper also describes gold returns, which feature two stylised facts of most financial assets, i.e. daily returns have mean close to zero and are highly leptokurtic. The analysis demonstrate that Gaussianity is recovered in the return distribution approximately every eleven months. The final section of the study considers the volatility of gold and finds that long-run correlation is significantly different from zero.

Capie et al. (2005) focus on one specific attribute of the precious metal, that is protection against weakening of the US Dollar. They argue that this property has changed over time. In the nineteenth century, during the gold standard, "gold was inevitably a good hedge". They study if and to what extent gold has remained a hedge against currency fluctuations since it stopped being the basis of the monetary system. In their data, which, for the same reasons as given above, starts in 1971, they find significant and negative correlation between contemporaneous changes of the variables, confirming the hedging property of gold for both the GBP/USD and the YEN/USD exchange rates. Their statistical work also shows that positive price shocks have greater impact on the conditional variance of gold returns than negative shocks of the same magnitude. This is in contrast to the asymmetric effect of other financial assets such as equity, for which negative shocks have greater impact.

Further analysis of the unique characteristics of gold is provided by McCown and Zimmerman (2006). Using a data set spanning almost the same time frame as in Mills (2004), they find that returns on gold are slightly higher than those on Treasuries and lower than those on equity, but with higher volatility. To explain these findings, the article refers to Jaffe (1989), which suggests that gold offers lower returns than other risky assets because it has "liquidity value", "consumption value", and "convenience value". Indeed, gold is the most liquid asset during times of financial distress. When used in jewellery and as a means of adornment, especially in Asia and the Middle East, gold gives its owner some sort of utility that other financial assets do not provide. Finally, when used in production, gold is stored and thus has a convenience yield. McCown and Zimmerman (2006) analyse the investment performance of gold using standard asset pricing models. They demonstrate that gold does not bear any systematic risk using the CAPM, and show that gold is able to hedge against inflation risk using the arbitrage pricing model. The second finding is corroborated by cointegration found between the metal and consumer prices. Based on these results, they conclude that "gold can be a useful addition to many investment portfolios", which is in line with the final remarks of Lawrence (2003).

The response of gold to macroeconomic announcements is examined by Tully and Lucey (2007). After a review of the contradictory results present in the literature that focuses on this topic, they show that the US Dollar is in many cases the only macro factor influencing the price of the yellow metal. Due to a negative and significant relationship with the USD, they label gold "anti-dollar". Confirming the findings of Lawrence (2003), the relationship with equity is negative but insignificant. They also examine the variance of gold returns and conclude that it is not affected by macroeconomic variables. As it is the case for several financial time series, they find that gold volatility clusters and innovations have asymmetric effects on it.

More recently, the literature has focused on the safe haven potential of gold. For instance, Baur and Lucey (2010) are the first to build an econometric model able to test this property. Within their study, an asset is defined as a safe haven with respect to a reference asset if correlation between the two is zero or negative when the reference asset is doing badly. If instead the non-positive correlation holds on average then the authors talk of hedge, whilst if correlation on average is positive but not perfect then they talk of diversifier. They consider the relation of gold with both stock and bond returns in the US, in the UK, and in Germany. They find that the yellow metal is a safe haven for stocks in all three markets but is nowhere a safe haven for bonds. They also observe that the safe haven property dissipates within 15 days from the equity market shock and is only triggered by extreme losses exceeding the 2.5%-ile and the 1%-ile of the equity return distribution. Analysing the potential of gold to be a hedge, they note a non-positive correlation on average only with stocks of the US and the UK.

Baur and McDermott (2010) extends the analysis to global level. They consider the safe haven property of gold for stock market indices of both developed and emerging countries. From the results of their work it emerges that the relation between gold and equity is not constant over time (hence gold is not always a hedge) and is strikingly different across the world. They show that gold is generally a safe haven in Europe and in the US but not in Canada, Australia, Japan and the largest emerging markets. The authors suggest that the findings for Australia, Canada and Japan might be explained by the singular nature of their stock markets, whilst in the developing countries an exposure to developed world markets rather than gold could be the preferred choice of jittery investors. They add to the definitions of Baur and Lucey (2010) by distinguishing between a weak and a strong safe haven, with only the latter able to act as a stabilising force of the financial system. Evidence of gold being a strong safe haven that reduces investor losses is found during the peak of the last financial crisis in 2008 and for the Black Monday, October 19, 1987. When conditioning on financial market uncertainty (proxied by conditional variance of the world index) instead of stock market losses, they note that at extreme levels of global uncertainty the safe haven effect stops and the precious metal starts moving down as the rest of the market.

After evidence was found for the safe haven property of gold, Baur and McDermott (2012) investigate the differences between the yellow metal and the traditional safe haven US Treasuries. They justify the fact that both bonds and gold are commonly referred to as safe haven by claiming that the two assets offer, respectively, a fixed rate of return if held until maturity and protection against inflation, currency, and default risk. Starting from the observation that it is not efficient to invest in gold since the same return level can be achieved at a lower risk with other assets, they consider the role of uncertainty in spreading fear and panic in the market and driving investors to make a seemingly irrational choice. Their main finding is that different degrees of uncertainty trigger the safe haven status of one or the other asset. They describe bonds as more responsive since they react first to increased uncertainty and show that investors flee to the perceived safety of Treasuries simultaneously with the stock market's fall. They note, on the other hand, a delayed response from gold, whose reaction is stronger than the one of bonds at extreme levels of ambiguity. On those occasions, the worst-case scenario receives, the authors claim, great attention from the investors, which stop looking at gold as an inefficient investment and consider the importance of holding a physical asset in order to avoid a total wipeout of their investment. Their conclusion, *investors generally use bonds to hedge against stock market losses and under certain circumstances they use gold as a hedge against bonds* (Baur and McDermott, 2012), is in line with the findings of Adrian et al. (2018), who have shown that, at extreme levels of stock market volatility, flights to the safety of bonds turn into flights from bonds.[7] In their statistical analysis, Baur and McDermott (2012) find that on average correlation in 1980-2010 is negative but weak between stocks and bonds, positive between stocks and gold, and positive but weak between gold and bonds. Conditioning on stock market losses leads to different considerations as they show that bonds and gold are positively correlated between each other and are both negatively correlated with stocks, revealing flights to safety.

An extensive investigation of the dynamics between gold and a variety of asset classes is offered by Ciner et al. (2013). They consider stocks, bonds, gold, oil, and exchange rates and test the safe haven and hedge hypotheses for each of those assets. Their results contradict earlier work by showing that gold is not a safe haven for US equity but it can be for bonds. However, in accordance with the literature, they find that the precious metal provides shelter against currency depreciations (particularly against fluctuations in the foreign exchange value of US Dollar and GBP) and that the bond market plays a safe haven role for when the stock market plunges in the US and the UK.

The determinants of the safe haven nature of gold are analysed in a recent paper by Li and Lucey (2017). They find that a higher level of Economic Policy Uncertainty increases the odds of gold being a safe haven against stock market falls in France, Japan and India, and against bond market extreme events in US, UK, Germany, France, Italy and India. In the US, the VIX appears to be a significant determinant of the safe haven status of gold against the S&P 500. Moreover, by comparing the safe haven property of gold with that of silver, platinum and palladium, Li and Lucey (2017) observe that, on average, gold is not the most frequent safe haven across eleven countries. For instance, silver is found to be the best safe haven against S&P 500 and bond market falls. However gold is the best safe haven in UK, against both stock and bond market events, and in Canada, against bond market declines. Interestingly, the authors note, gold is the least common safe haven for both stock and bond market declines in China.

[7] In their article, Adrian et al. (2018) suggest that when VIX is above 50 (the 99.3rd-percentile) it is not infrequent that stock markets strongly bounce back and central banks intervene with interest-rate cuts, which reduce the attractiveness of bonds. See, for example, Shiller and Beltratti (1992), for practical considerations on the relation between long-term interest rates and both stock and bond prices.

2.2.3 Literature on Volatility Forecasting

Moving on to the last literature that is of relevance to this chapter, Boudry et al. (2018) suggest that flight-to-safety events signal a future deterioration in economic fundamentals. The relationship between the business cycle and stock market volatility is well documented in the literature. For instance, Schwert (1989) notes that stock market volatility is always higher during recessions and claims that the level of real economic activity is the most important determinant of the conditional variance of stock returns. The same conclusion is reached by Hamilton and Lin (1996) who find that economic recessions are the primary factor that drives fluctuations in the volatility of stock returns. Motivated by these considerations, we investigate the contribution of a {0,1} FTS dummy variable to the forecasts of future realised variance of the S&P 500 index.

The literature on volatility forecasting is so vast that we can discuss only the relevant papers to this study and we refer to other works for further details. Engle and Patton (2001) suggest that a good volatility model is one that can make good forecasts and capture the following stylised facts about volatility: persistence (volatility shocks die out very slowly), mean reversion (the shocks will eventually die), asymmetric response to return innovations of different sign, and influence of exogenous variables like, for instance, the 3-m US Treasury bill rate.

The HAR-RV model proposed by Corsi (2009) has become very popular in this field due to its simplicity of construction, its ability to reproduce key features of financial time series, such as fat-tailed return distributions and long memory behaviour of volatility, and, not least, its good forecasting power. Instead of specifying an equation for the mean and an equation for the unobserved conditional variance of returns as in the generalized autoregressive conditional heteroskedasticity (GARCH) model, Corsi (2009) fits an autoregressive (AR) process to an observable proxy, the so-called realised volatility.[8] With only three factors, which reflect volatility realised over a short-term, a medium-term, and a long-term respectively, the HAR-RV model includes information up to lag 22 days (approximately the number of working days per month) and, in contrast to standard GARCH models, is able to preserve the long memory property of volatility. Typically, autoregressive models, from the simple AR(1) to the more sophisticated

[8]For a more comprehensive description of ARCH-GARCH models and their use in finance applications, see Bollerslev (1986) and Engle (2004). For a detailed study of stock return realised volatility, including techniques for its computation and descriptive analysis, refer to Andersen et al. (2001).

HAR-RV model, are fitted to the log-transformed time series of daily realised variance. Indeed, it was observed in previous studies that *the logarithms of the realised variances are approximately normal* (Andersen et al., 2001).

In addition to GARCH, HAR-RV, and AR(1), the other volatility models that we consider are the RiskMetrics, introduced by J.P. Morgan in 1996 for risk management purposes, and the high frequency based volatility (HEAVY) models proposed by Shephard and Sheppard (2010).

The RiskMetrics (RM) model is an exponentially smoothed estimator for latent variance which assigns declining weights to past squared returns. The use of RM for Value-at-Risk calculation is analysed in more detail in the literature; see, e.g., Christoffersen et al. (2001). With a decay factor set to 0.94, variance estimation using RM is a lot easier than traditional GARCH. Although the RiskMetrics is a nested model, its forecasts for volatility are often found to be at least as good as those of a standard GARCH model.

Shephard and Sheppard (2010) add the high frequency information contained in daily realised measures, such as realised variance (RV), to GARCH-type specifications but, as the authors claim, the models obtained are not nested by Bollerslev's (1986) model. HEAVY models consist of two equations, one for the conditional variance of returns, and one for the conditional expected value of the realised measure. The authors name the two equations HEAVY-r model and HEAVY-RM model, respectively. They show that the new models have less memory than GARCH but allow for momentum in addition to mean reversion, and react more quickly to breaks in the volatility level. In the out-of-sample analysis the authors find predictive superiority of the HEAVY models over GARCH models at all horizons considered.

The attractive features of the HAR-RV model listed above motivate our choice to include the FTS dummy in the model of Corsi (2009). The potential benefits of accounting for flight-to-safety events could also be studied by applying the dummy to a different model. The focus of this work, however, is on the FTS measure and its use for improved forecasts of realised variance and not on the choice of the best volatility model.

2.3 FTS Identification

We proceed now to the description of the data and techniques we use and the results we obtain for the identification of flight-to-safety events when both US Treasuries and gold can represent a safe haven against the US stock market.

2.3.1 Data

The data set consists of daily observations for the US equity index, the US 10-year Treasury bonds, and the price of gold on the London Bullion Market. For the sake of comparison, we use the same source of data for stock and government bonds as in Baele et al. (2015). Data come from Datastream. We use the Total Return Index (RI) for the United States Benchmark 10 Year Datastream Government Index and for the United States-Datastream Market to calculate Treasury bond and stock returns, respectively.[9] We use the (Adjusted - Default) Price (P), expressed in USD per Troy Ounce, for the gold bullion. Our sample starts on January 1, 1980 and ends on June 30, 2016, yielding a total of 9,523 observations. Table 2.1 identifies the data used in the study and Figure 2.1 shows their respective time series.

[Insert Table 2.1 here]

[Insert Figure 2.1 here]

The daily returns used in the FTS study are discretely compounded rates expressed in percentages. Table 2.2 and 2.3 report the summary statistics and correlation matrix for gold, equity and bond returns. The mean and standard deviation are expressed in percentages on a yearly basis. The unconditional correlation is expressed in percentages.

[Insert Table 2.2 here]

[Insert Table 2.3 here]

[9] The dataset also includes the Interest Yield (IY) time series for the 10-year benchmark bond index, which is used in the Bivariate Regime-Switching model for flight-to-safety identification

Consistent with previous findings, Table 2.2 shows that volatility of gold returns is higher than volatility of the other two assets. However, in contrast with the results reported by McCown and Zimmerman (2006), we find that the average daily return of gold is lower than the return of Treasuries. The distributions of the two riskiest assets are skewed to the left, whilst a slightly positive skewness is found for government bonds. Evidence of fat-tailedness for all three assets is provided by the higher-than-normal kurtosis of their return time series. The results of correlation analysis are generally consistent with those of Baur and McDermott (2012). We find that on average bond returns are positively correlated with gold returns and negatively correlated with equity returns. This last finding contradicts the intuition of Baele et al. (2015), which states that a positive correlation is expected between bonds and stocks since they both are long-duration assets. Lastly, a negative but not significant relationship exists on average between stock and gold returns.

2.3.2 Econometric Framework

In this section, we present a number of models that make use of daily stock, bond, and gold returns in order to distinguish between FTS days and non-FTS days. Baele et al. (2015) provide the methodologies to identify flight-to-safety episodes, and thus we refer to their paper for a more detailed description of the econometric techniques. Here, we limit ourselves to giving a brief summary of the method and pointing out the FTS spells captured by the two safe haven assets. As in their study, we use a Threshold model, an Ordinal Index model, a Univariate and a Bivariate Regime-Switching (RS) model to produce flight-to-safety daily probabilities that are converted into $\{0, 1\}$ dummy variables to signal the occurrence of flights. Four individual FTS measures are derived using equity and bond returns, and four more are derived using equity and gold returns within the models mentioned above. A Joint model aggregates into a single statistic the four FTS measures that use bonds as safe haven asset, and into a second statistic the four FTS measures that use gold as safe haven. Overall, the results suggest the importance of gold for the identification of non-overlapping episodes with those discovered by bonds. Indeed, despite a similar behaviour of government bonds and gold, the safe haven status is sometimes asynchronous across assets.

2.3.2.1 The Threshold model

The first technique implemented by Baele et al. (2015) is the so-called Threshold model. The name comes from the fact that the daily returns of the safe haven and equity need to exceed a positive and a negative threshold, respectively, for a day to be an FTS day. In other words, an extreme negative stock return and an extreme positive return of the potential safe asset are the only two variables that matter for identifying FTS episodes.

Let H denote the potential safe haven, either US government bonds or gold. The model recognises a flight-to-safety at time t if the return of the safe haven on that day $r_{t,H}$ is above a positive threshold $z_{t,H}$ and the stock return $r_{t,s}$ is below a negative threshold $z_{t,s}$:

$$FTS_t = \begin{cases} 1 & \text{if } (r_{t,H} > z_{t,H}) \text{ AND } (r_{t,s} < z_{t,s}) \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

where FTS_t is both the $\{0,1\}$ dummy and probability of a flight-to-safety on day t according to the Threshold model. The time-t thresholds for the safe haven and equity are calculated as:

$$z_{t,H} = k \times \sigma_{t,H} \qquad z_{t,s} = -k \times \sigma_{t,s} \tag{2.2}$$

where k is a constant threshold parameter that indicates how many standard deviations the return of equity (safe haven asset) must be below (above) 0 to identify a day as an FTS day.

$z_{t,H}$ and $z_{t,s}$ change over time because they depend on time-varying volatilities, which are found by using a two-sided Gaussian Kernel. Gradually decreasing weights are assigned to distant observations, both in the past and in the future. The kernel of this study has a bandwidth h of 250 days expressed as a fraction of sample size $T = 9,522$ observations. The time-varying variance at time τ (normalised date t/T) is:

$$\sigma^2_{\tau,i} = \sum_{t=1}^{T} K_h\left(\frac{t}{T} - \tau\right) r^2_{t,i} \qquad i = H, s \qquad h = \frac{250}{T} \tag{2.3}$$

where the kernel function is $K(z) = \frac{1}{\sqrt{2\pi}} \exp^{(-\frac{z^2}{2})}$.

Baele et al. (2015) conduct a simulation experiment that shows how the magnitude of the threshold parameter k influences the incidence of flights, i.e. the number of days identified as FTS days expressed as a fraction of the total sample size. In their experiment stock and bond returns are assumed to be normally distributed with their standard deviations and correlations estimated with the simple kernel method described above; the means are set at their full-sample values. Since stock and bond returns are drawn from a bivariate normal distribution, it is sensible to expect fewer FTS events in the simulation experiment than in the real world where fat-tailedness and negative skewness are stylised facts for financial returns. In other words, the value of k needs to be chosen high enough to ensure that the actual data generate more FTS episodes than the simulated data in the bivariate normal world.

In our work we replicate the analysis of Baele et al. (2015) that considers changes of k using bond and stock returns, and then we repeat the experiment on the threshold parameter by replacing bond with gold returns. Figure 2.2 and Figure 2.3 show the percentage of FTS events both in the actual data and in the simulated data as a function of k for US bonds and gold, respectively. The figures also plot the average return difference between the safe haven and equity on the days identified as FTS according to the value used for k.

[Insert Figure 2.2 here]

[Insert Figure 2.3 here]

We note that the FTS incidence decreases as k increases. This holds true for both simulated and actual data, in both gold and bond analysis. The real-world data deliver more flight-to-safety episodes than the bivariate normal data when k is greater than 1.108 in the Threshold model based on bond returns, and when k is greater than 1.369 in the Threshold model based on gold returns. In line with the choice of Baele et al. (2015), we set $k = 1.5$ for both US Treasuries and gold. The results for the Threshold FTS measure are reported in Table 2.4 and are graphically represented in Figure 2.4 and Figure 2.5.

[Insert Table 2.4 here]

[Insert Figure 2.4 here]

[Insert Figure 2.5 here]

The furthest right column of Table 2.4, which represents the incidence of flights when at least one of the two assets is flagging an FTS day within the Threshold model, suggests that we are able to capture additional episodes by considering the safe haven property of bonds as well as of gold. This is also evident from Figure 2.5, which plots more FTS spells in the first part of the sample than Figure 2.4.

2.3.2.2 The Ordinal Index model

The second econometric approach is based on the work of Hollo et al. (2012) and tries to capture FTS episodes by combining multiple variables into an ordinal index. Baele et al. (2015) study the symptoms of a flight-to-safety in order to identify what variables are important for this purpose. They look into the difference between stock and bond returns, their correlation, and equity market volatility. On this basis they propose six variables ($var^{(1)} - var^{(6)}$) that are either positively or negatively correlated with the likelihood of an FTS episode, and they determine variable-specific boundary values beyond which "mild FTS-symptoms" are triggered.

We adopt this setup but replace bonds with the more general term safe haven, which we denote by H. The precise definitions of the variables, their correlation with FTS occurrence ($\pm$), and their boundary values for symptoms are as follows:

1. The difference between the return of the safe haven and the return of equity. (Positive correlation +, FTS symptom if $var^{(1)} \geq 0$)

$$var_t^{(1)} = r_{t,H} - r_{t,s} \tag{2.4}$$

2. The difference between the return of the safe haven and the return of equity, relative to its long-term moving average. (Positive correlation +, FTS symptom if $var^{(2)} \geq 0$)

$$var_t^{(2)} = (r_{t,H} - r_{t,s}) - \overline{(r_{\tau,H} - r_{\tau,s})} \tag{2.5}$$

where

$$\overline{(r_{\tau,H} - r_{\tau,s})} = \sum_{t=1}^{T} K_h(\frac{t}{T} - \tau)[r_{t,H} - r_{t,s}] \qquad h = \frac{250}{T} \tag{2.6}$$

3. The short-term stock-safe haven asset return correlation. (Negative correlation −, FTS symptom if $var^{(3)} \leq 0$)

$$var_t^{(3)} = \rho_{\tau,[H,s],short} = \frac{\sigma_{\tau,[H,s],short}}{\sqrt{\sigma^2_{\tau,s,short}} \times \sqrt{\sigma^2_{\tau,H,short}}} \tag{2.7}$$

where

$$\sigma_{\tau,[H,s],short} = \sum_{t=1}^{T} K_h(\frac{t}{T} - \tau)[r_{t,H} \times r_{t,s}] \qquad h = \frac{5}{T} \tag{2.8}$$

and

$$\sigma^2_{\tau,i,short} = \sum_{t=1}^{T} K_h(\frac{t}{T} - \tau) r^2_{t,i} \qquad h = \frac{5}{T} \qquad i = H, s \tag{2.9}$$

4. The difference between the short-term and the long-term stock-safe haven asset return correlation. (Negative correlation −, FTS symptom if $var^{(4)} \leq 0$)

$$var_t^{(4)} = \rho_{\tau,[H,s],short} - \rho_{\tau,[H,s],long} \tag{2.10}$$

where

$$\rho_{\tau,[H,s],long} = \frac{\sigma_{\tau,[H,s],long}}{\sqrt{\sigma^2_{\tau,s,long}} \times \sqrt{\sigma^2_{\tau,H,long}}} \tag{2.11}$$

$$\sigma_{\tau,[H,s],long} = \sum_{t=1}^{T} K_h(\frac{t}{T} - \tau)[r_{t,H} \times r_{t,s}] \qquad h = \frac{250}{T} \tag{2.12}$$

and

$$\sigma^2_{\tau,i,long} = \sum_{t=1}^{T} K_h(\frac{t}{T} - \tau) r^2_{t,i} \qquad h = \frac{250}{T} \qquad i = H, s \tag{2.13}$$

5. The short-term equity return volatility. (Positive correlation +, FTS symptom if more than double the unconditional standard deviation, i.e. $var^{(5)} \geq 2 \times \sigma_s$)

$$var_t^{(5)} = \sqrt{\sigma^2_{\tau,s,short}} \tag{2.14}$$

where

$$\sigma^2_{\tau,s,short} = \sum_{t=1}^{T} K_h(\frac{t}{T} - \tau) r^2_{t,s} \qquad h = \frac{5}{T} \tag{2.15}$$

and

$$\sigma_s = \sqrt{\sum_{t=1}^{T} \frac{r^2_{t,s}}{T}} \tag{2.16}$$

6. The difference between the short-term and long-term equity return volatility. (Positive correlation +, FTS symptom if $var^{(6)} \geq 0$)

$$var^{(6)}_t = \sqrt{\sigma^2_{\tau,s,short}} - \sqrt{\sigma^2_{\tau,s,long}} \tag{2.17}$$

where

$$\sigma^2_{\tau,s,long} = \sum_{t=1}^{T} K_h(\frac{t}{T} - \tau) r^2_{t,s} \qquad h = \frac{250}{T} \tag{2.18}$$

$var^{(2)} - var^{(6)}$ use the same kernel method that was described for the time-varying volatilities of the Threshold model. The bandwidth h is 250 days for long-term variables and 5 days for short-term ones.

The approach used for the construction of the ordinal index and its subsequent transformation into a flight-to-safety probability is described in detail in Baele et al. (2015). Here, we illustrate the main idea in the following steps:

i The observations of variables that correlate positively (+) with FTS occurrence are ranked in ascending order, whilst those of variables that correlate negatively (−) are ranked in descending order.

ii Normalised ordinal numbers are obtained by replacing each observation with its ranking number divided by the total number of observations. Values are bounded between 0 and 1, with values close to 1 associated with larger probability of FTS.

iii For each time t, the normalised ordinal numbers are averaged across the 6 FTS variables to produce a time-t value of the ordinal index.

iv The average ordinal numbers of the days that satisfy all the "mild FTS-symptoms" are collected and the minimum of those numbers is set as a threshold.

v The Ordinal FTS probability is set equal to 0 for all the observations with an average ordinal number below the threshold. The $\{0, 1\}$ dummy is also 0 on those days.

vi For observations above the threshold the Ordinal FTS probability is set equal to *one minus the percentage of "false positives", calculated as the percentage of observations with an ordinal number above the observed ordinal number that do not match our FTS criteria* (Baele et al., 2015). When this probability is larger than 0.5 a flight-to-safety day is identified by the Ordinal model and the $\{0, 1\}$ dummy is set equal to 1.

$$FTS_t = \begin{cases} 1 & \text{if Ordinal FTS probability} > 0.5 \\ 0 & \text{otherwise} \end{cases} \tag{2.19}$$

where FTS_t is the $\{0, 1\}$ dummy of the Ordinal model that takes value 1 on those days when the Ordinal FTS probability is larger than 50%.

We report the results of the Ordinal FTS measure in Table 2.5 and we show the corresponding dummies in Figure 2.6 and Figure 2.7.

[Insert Table 2.5 here]

[Insert Figure 2.6 here]

[Insert Figure 2.7 here]

As is the case with the Threshold model, the identification of FTS with this second econometric technique depends on which asset is chosen as safe haven against stocks. Last column in Table 2.5 and visual inspection of the dummies in Figure 2.6 and 2.7 suggest that gold and US Treasuries do not always act as safe haven at the same time.

2.3.2.3 The Univariate Regime-Switching model

Another technique of FTS identification comprises a 3-state Markov-switching model for the difference between the return of the safe haven $r_{t,H}$ and the stock return $r_{t,s}$:

$$y_t = r_{t,H} - r_{t,s} \tag{2.20}$$

Following the extant literature on state-dependent equity returns, Baele et al. (2015) suggest that, in addition to a regime that functions as the FTS regime, two more are needed for a low and a high volatility environment, respectively. Based on these assumptions, we use Regime 1 for the low volatility regime, Regime 2 for the high volatility regime, and Regime 3 for the flight-to-safety regime. The univariate RS model is characterised by regime shifts for both the intercept and volatility parameter:

$$y_t = \mu_v + \sigma_v \epsilon_t \qquad \epsilon_t \sim N(0,1) \tag{2.21}$$

where μ and σ are the mean and volatility of y_t, respectively, and v is the index used for the time-t regime.

As in Baele et al. (2015), we assume that the regime variable follows a Markov Chain with constant transition probabilities, and we estimate the model parameters under the following constraints:

- positive mean for y_t in the flight-to-safety regime, $\mu_3 > 0$.
- the mean of y_t in the flight-to-safety regime is higher than in the low volatility regime, $\mu_3 > \mu_1$.
- the mean of y_t in the flight-to-safety regime is higher than in the high volatility regime, $\mu_3 > \mu_2$.

Table 2.6 reports the estimation results for both the case in which expression in (2.20) is based on US Treasuries and stock returns, and the case in which gold and stock returns are used. The parameter estimates we obtain when we consider US government bonds as the safe haven are roughly consistent with those reported in Baele et al. (2015).

[Insert Table 2.6 here]

Following the intuition of Baele et al. (2015), we use the smoothed regime probabilities, which are based on full sample information, to determine whether a flight-to-safety event takes place at a particular point in time:

$$FTS_t = \begin{cases} 1 & \text{if Regime 3 smoothed probability} > 0.5 \\ 0 & \text{otherwise} \end{cases} \tag{2.22}$$

where FTS_t is the $\{0, 1\}$ dummy of the Univariate RS model that takes value 1 on those days when the smoothed probability of the FTS regime, Regime 3, is larger than 50%.

The results we obtain for the Univariate RS FTS measure are listed in Table 2.7 and presented in Figure 2.8 and 2.9:

[Insert Table 2.7 here]

[Insert Figure 2.8 here]

[Insert Figure 2.9 here]

Overall, the results generally support the idea that non-overlapping FTS spells can be captured by including the potential of gold as a safe haven in the analysis of Baele et al. (2015). This is in line with what the Threshold and Ordinal models reveal about the occurrence of flights.

2.3.2.4 The Bivariate Regime-Switching model

As an extension of the multi-regime framework described above, Baele et al. (2015) propose a more sophisticated Markov-switching model to capture not only the large positive difference between bond and stock returns but also the negative correlation and high equity volatility observed during flights. Therefore, our last FTS measure is derived by assuming a regime-switching behaviour for both equity and safe haven asset. Two equations, one for stock returns

$r_{t,s}$ and one for either bond or gold returns $r_{t,H}$ are defined as follows:

$$\begin{gathered} r_{t,s} = \alpha_0 + \alpha_1 J_{t,s}^{lh} + \alpha_2 J_{t,s}^{hl} + \alpha_3 (J_t^{FTS} + \upsilon S_t^{FTS}) + \epsilon_{t,s} \\ \epsilon_{t,s} \sim N(0, h_s(S_t^s)) \end{gathered} \tag{2.23}$$

$$\begin{gathered} r_{t,H} = \beta_0 + \beta_1 J_{t,H}^{lh} + \beta_2 J_{t,H}^{hl} + \beta_3 (J_t^{FTS} + \upsilon S_t^{FTS}) + (\beta_4 + \beta_5 S_t^{FTS}) r_{t,s} + \epsilon_{t,H} \\ \epsilon_{t,H} \sim N(0, \theta_{t-1} h_H(S_t^H)) \end{gathered} \tag{2.24}$$

The unobservable Markovian state variables S_t^s and S_t^H determine for the variance of the stock return and for the variance of the safe haven return, respectively, shifts between two regimes. By including the stock return $r_{t,s}$ in (2.24), a volatility spillover from the equity market is considered. Baele et al. (2015) scale the bond return variance by the lagged interest rate level as they note large bond price fluctuations in the first part of the sample when interest rates were high. In our model, θ_{t-1} is the lagged bond yield when US Treasuries are the safe haven and a constant equal to 1 when gold is the safe haven asset H. The jump terms $J_{t,s}^{lh}$ ($J_{t,H}^{lh}$) and $J_{t,s}^{hl}$ ($J_{t,H}^{hl}$) are dummy variables that signal for the equity (safe haven) return variance a regime shift from low to high and from high to low, respectively. The latent Markovian state process S_t^{FTS} and the jump term J_t^{FTS} are used to identify FTS episodes along with parameter constraints which ensure that during flights the stock market drops ($\alpha_3 < 0$), gold or bond prices rise ($\beta_3 > 0$), correlation between equity and the safe haven decreases ($\beta_5 < 0$), and the FTS effect is maximal on the first day ($\upsilon > 0$).

In this setup, the Bivariate RS model comprises 3 two-sate Markov chain processes. S_t^H is assumed to be independent of both S_t^s and S_t^{FTS}. However, we follow Baele et al. (2015) and rule out the possibility that equity volatility might be in the low regime during a fligh-to-safety episode:

$$Pr(S_t^s = 1 | S_{t-1}^s, S_t^{FTS} = 1) = 1 \tag{2.25}$$

This assumption reduces the number of regimes from eight to six. We use Regime 1 for the low equity volatility, low bond volatility, no-FTS regime; Regime 2 for the low equity volatility, high bond volatility, no-FTS regime; Regime 3 for the high equity volatility, low bond volatility, no-FTS regime; Regime 4 for the high equity volatility, low bond volatility, FTS regime; Regime 5 for the high equity volatility, high bond volatility, no-FTS regime; and finally Regime 6 for the high equity volatility, high bond volatility, FTS regime.

In Table 2.8, we report estimation results for the Bivariate RS model that uses US Treasury Bond returns in (2.24) and the one that uses gold returns instead.

[Insert Table 2.8 here]

As in the univariate case, we use the smoothed probabilities of the regimes in which a flight-to-safety is present to determine whether an FTS event takes place at a particular point in time:

$$FTS_t = \begin{cases} 1 & \text{if Regime 4 or Regime 6 smoothed probability} > 0.5 \\ 0 & \text{otherwise} \end{cases} \tag{2.26}$$

where FTS_t is the $\{0,1\}$ dummy of the Bivariate RS model that takes value 1 on those days when the smoothed probability that Regime 4 or Regime 6 occurs is larger than 50%.

We list the results for the Bivariate RS FTS measure in Table 2.9 and we graphically show the dummies for US Treasuries and gold in Figure 2.10 and 2.11, respectively.

[Insert Table 2.9 here]

[Insert Figure 2.10 here]

[Insert Figure 2.11 here]

Although the incidence of flights according to the Bivariate RS model is significantly higher than the models described above, we find again that, when gold and government bonds are jointly considered, FTS days are more numerous than when only one safe haven is allowed to exist.

2.3.2.5 The Joint model

Baele et al. (2015) propose a Joint model for summarising into a single measure the information provided by the Threshold, the Ordinal Index, and the two Regime-Switching models about the occurrence of flights. The aggregation method is based on previous studies that document *an algebraically convenient representation for the multivariate Bernoulli distribution* (Teugels, 1990). In this context, the 4 FTS measures presented above (the daily probabilities of flight-to-safety and respective $\{0, 1\}$ dummies) form a sequence of Bernoulli random variables. The probability of FTS that each of those variables assigns to each day in the sample and the full-sample covariances between dummies enter the Kronecker product that delivers the multivariate Bernoulli. The time-t Joint FTS probability is obtained by adding up the time-t probabilities for all the cases in which at least 3 of the 4 variables flag a flight-to-safety episode.

$$FTS_t = \begin{cases} 1 & \text{if Joint FTS probability} > 0.5 \\ 0 & \text{otherwise} \end{cases} \tag{2.27}$$

where FTS_t is the $\{0, 1\}$ dummy of the Joint model that takes value 1 on those days when at least 3 of the 4 econometric techniques indicate a flight-to-safety.

We list the results for the Joint FTS measure in Table 2.10 and show the corresponding FTS dummies in Figure 2.12, 2.13, and 2.14.

[Insert Table 2.10 here]

[Insert Figure 2.12 here]

[Insert Figure 2.13 here]

[Insert Figure 2.14 here]

Consistent with Baele et al. (2015), the incidence of flights in the aggregate framework is lower than in the RS models and between the figures produced by the Threshold and Ordinal models. We find that the percentage of days in the US that can be classified as FTS days is 2.668%

when bonds are the safe haven for the stock market, 1.218% when gold replaces bonds in the models, and 3.245% when both assets offer protection to investors. Our results based on bond returns fall within the interquartile range of 1.29%-3.55% reported in the multi-country analysis of Baele et al. (2015). Full details about the persistence of spells and contribution of the different methodologies to the joint measure are provided in their study on the characteristics of FTS episodes. Figure 2.12-2.14 corroborate the findings of Baele et al. (2015) of much more frequent FTS instances in the second half of the sample. However, we continue to observe in the aggregate framework that additional flight-to-safety episodes can be identified by considering gold the main beneficiary of the safe haven dash before 1995.

2.3.3 Empirical Results

All the FTS measures that we have discussed in our econometric framework are based on full-sample computation. In other words, information from the entire data set is used to assign a time-t probability of a flight-to-safety, and corresponding $\{0, 1\}$ dummy, in each model. For instance, the Threshold model produces a generic time-t volatility with a two-sided kernel that considers returns both in the past and in the future. The same kernel method is applied to the Ordinal Index. Besides that, in this second model the original values of asset returns are arranged in ascending or descending order on the basis of the total empirical cumulative distribution function, spanning years 1980-2016. Lastly, in the RS models we adopt the algorithm of Kim (1994) to produce the smoothed regime probabilities that identify FTS days. The smoothing method starts from the end of the sample and iterates backwards the equations for filtering and prediction probabilities, in which estimates of the parameters are used. See Hamilton (1994) for details on how to estimate Markov-switching models.

Following this approach, despite the benefits of using more information, we find flight-to-safety spells whose occurrence may be signalled by future events. An alternative way that ensures time-consistency and no look ahead bias of the FTS statistics, i.e. no contribution of future data to identifying flights, is the recursive methodology described in Hollo et al. (2012). We apply the Threshold, Ordinal, and Univariate RS models recursively over expanding samples while keeping their econometric specification unchanged. The Bivariate RS model is excluded from the analysis for two reasons. First, the cumbersome nature of its estimation makes the recursive methodology unfeasible in a sensible amount of time. Second, in our view, the Bivariate RS model seems to overestimate the actual incidence of flights.

In order to obtain the FTS measures with the new method, we first set a pre-recursion period that starts in January 1980 and ends in December 1999. For all subsequent days until 30 June 2016, we derive an FTS probability by running the models recursively with one new observation added at a time. From each model run, we save the results about the occurrence of flights only for the date of the last observation that was added. Our choice of pre-recursion period is related to the availability of data on realised variance from the Oxford-Man Institute's library, and more will be said about this later. In our opinion, switching from full-sample to recursive computation is essential to obtain an FTS measure eligible for a volatility forecasting model.

Table 2.11 reports the results of the various models when computed recursively and non-recursively. Start Date and End Date denote, respectively, the first and the last date for which a probability of FTS and the $\{0,1\}$ dummy are defined. The number of observations between the two dates is given by N^o Days, while the last column recognises as time-consistent only the recursively computed statistics. We report the FTS incidence measured over two different time periods. The incidence of flights over the period 1980-2016 is only available in case of full-sample computation, and figures are the same as discussed in the previous section. The second measure of FTS incidence, which runs from 2000 to 2016, allows for comparison since dummies and daily probabilities of a flight-to-safety are available for both recursive and full-sample computation over that time span.

[Insert Table 2.11 here]

The FTS measures in Panel C (Panel F) are based on the results of Panel A and B (Panel D and E). As mentioned previously, when bonds and gold are jointly considered, the US Treasuries and Gold $\{0,1\}$ FTS dummy takes value 1 at time t if at least one of the two assets classify that day as an FTS event.[10] The aggregation method used to find the recursive Joint FTS measures

[10]We also define US Treasuries and Gold probabilities of a flight-to-safety despite the fact that we only use

of Panel D and E is the same as in Section 2.3.2.5 with the only difference that we now set the time-t Joint FTS probability equal to the sum of the time-t probabilities of all the cases in which at least 2 of the 3 total variables flag a flight-to-safety episode.

The comparison of the FTS incidence measured over the two time periods reinforces our earlier inference that flights have become more frequent in the second half of the sample. Results differ across models when switching from full-sample to recursive methodology. We note that the Threshold models of Treasuries and gold deliver a bigger number of spells when the FTS statistics are recursively computed. This is also true of the Univariate RS model applied to bonds. Within the Ordinal model the FTS signal is "stronger" for US Treasuries and "weaker" for gold when the statistics are based on the full data sample. The overall effect is such that the Joint FTS measure, in both gold and Treasuries analyses, identifies slightly more episodes in the full-sample case.

2.3.3.1 Robustness of the FTS measures

In the following, we attempt to assess the robustness of the Threshold, Ordinal and Univariate RS FTS measures. Hollo et al. (2012) suggest that a financial stress indicator can be used for real-time applications only if it is stable over time and avoid the so-called "event reclassification" problem. In our context, this would mean that if a model flags an FTS episode at a particular point in time, then it is optimal that this period is still classified as such when new data is added to the sample and the model is run again to produce the next value of the FTS measure. The robustness property is tested for both US Treasuries- and gold-based measures by comparing the respective dummies and probabilities when computed recursively or computed only once using the full data sample. Table 2.12 reports the FTS % incidence of different methodologies as in Table 2.11. The table also shows the percentage of non-matching days, i.e. the percentage of days from 03/01/2000 to 30/06/2016 on which the dummies of full-sample and recursive computation do not agree in value. The furthest right column shows the percentage of days on

the FTS dummies in our subsequent work. When neither gold nor bonds flag an FTS day, the US Treasuries and Gold probability of a flight is set equal to the minimum of the probabilities given by the two assets. If both signal an FTS day then the US Treasuries and Gold probability is set equal to the maximum of the two probabilities. When bonds (gold) but not gold (bonds) classify a day as FTS day, the US Treasuries and Gold probability is set equal to the probability yielded by bonds (gold).

which the recursively computed FTS measure is in "great" disagreement with the full-sample measure. This happens when the difference between the two probabilities of a flight-to-safety is larger than 0.5. As we only have two cases, either FTS day or non-FTS day, 0.5 is chosen as a threshold beyond which extremely diverging results are found between full-sample and recursive statistics.

[Insert Table 2.12 here]

We believe that these results are most likely explained by changes in model parameters. Overall, the discrepancies between full-sample and recursive methods seem to be minor, supporting the stability and robustness of the FTS measures over time.

Henceforth we shall deal only with the recursively computed statistics, simply referred to as the Threshold, Ordinal, Univariate RS, and Joint FTS measures. As already mentioned, these do not incorporate future information and thus can be used to correctly assess the out-of-sample performance of a volatility forecasting model.

2.4 Volatility Forecasting with FTS

We now describe an application of the foregoing FTS measures in forecasting the volatility of the US S&P 500 equity index. After an overview of the data and models considered in this study, we analyse the contribution of flight-to-safety events to the in-sample and, more importantly, out-of-sample predictive power of the original HAR-RV model introduced by Corsi (2009).

2.4.1 Data

To produce the second data set that we use in this chapter, we collect daily prices, returns, and realised volatility measures for the S&P 500 index. These are provided by the Oxford-Man Institute's "realised library". For our analysis, we consider the standard realised variance estimator based on 5-minute intraday returns. The use of five-minute returns for volatility calculation is discussed in detail by Andersen et al. (2001).

We obtain a time series of daily returns and realised variances of the US stock index from January 3, 2000 to June 30, 2016, i.e. a total of 4,121 observations. Table 2.13 identifies the data collected for this study and Figure 2.15 shows their respective time-series. The historical prices of the S&P 500 are reported and graphically shown here, but are not used in the rest of the study.

[Insert Table 2.13 here]

[Insert Figure 2.15 here]

The descriptive statistics for returns and realised variances of the S&P 500 are summarised in Table 2.14. All values shown are based on daily returns expressed in percentages. This implies that the statistics of raw realised variance data, with the exception of skewness and kurtosis, must be multiplied by 10,000 to produce the results reported in Table 2.14.

[Insert Table 2.14 here]

Consistent with previous results, daily stock returns have mean close to zero, are skewed to the left, and are highly leptokurtic, as is indicated by a negative skewness and a positive kurtosis, respectively. An examination of the summary statistics for the realised daily variances in the above table shows that on average the annualised standard deviation of the index is approximately 18%. Just like in the case of returns, the distribution of realised variances has fatter tails than those predicted by the normal distribution, but is bounded by zero with a long right tail.

In our modelling, the realised variances haven been annualised and, in some cases, also log transformed, whilst the raw series of daily returns has been used in latent volatility models, such as the GARCH type of models.

2.4.2 Econometric Framework

In this subsection, we formalise the notion of improved forecasts of future equity return volatility by accounting for the flight-to-safety phenomena identified, on a daily basis, in Section

2.3. The intuition is related to the analysis of Boudry et al. (2018), which finds that clusters of FTS days predict a decline in future economic activity, and the fact that economic recessions are the primary factor that drives fluctuations in the variance of stock returns (Schwert, 1989; Hamilton and Lin, 1996). Further, it adds to the results reported by Baele et al. (2015) regarding the effect of flights on the US VIX. While their work and ours are similar in considering flight-to-safety in the context of equity volatility, they also differ in significant ways. In their study, Baele et al. (2015) investigate contemporaneous relationships by regressing daily changes in the implied volatility of the S&P 500 on their FTS dummies.[11] In our study we focus on the benefits, in terms of more accurate performance, of using an FTS dummy to forecast the one-day ahead realised volatility of the S&P 500.

We start out by examining the predictive ability of FTS for future stock volatility using the Granger (1969) causality test. To this end, we specify a vector autoregressive (VAR) process for the realised volatility of the S&P 500 and the $\{0,1\}$ dummy that identifies FTS episodes according to the Joint model for US Treasuries and gold. The Akaike information criterion (AIC) and Bayesian information criterion (BIC) select a VAR process of order 21 and 9, respectively. Regardless of the order of the process, we strongly reject the hypothesis that the FTS dummy is not Granger-causing realised equity volatility (p-value is essentially 0). A significant causal relationship is also found from equity volatility to FTS. However, this comes as no surprise since high equity market volatility is identified in the literature as a key factor that leads to flight-to-safety (Vayanos, 2004; Caballero and Krishnamurthy, 2008; Brunnermeier and Pedersen, 2009).[12]

After finding statistical evidence in favour of a causality from FTS to equity return volatility, we show, in Figure 2.16, the realised daily variance of the S&P 500 index (red line) along with the one-day lagged value of the $\{0,1\}$ dummy that identifies FTS episodes according to the Joint model for US Treasuries and gold (black line).

[Insert Figure 2.16 here]

[11] Baele et al. (2015) also use the decomposition of the squared VIX, proposed in Bekaert and Hoerova (2014), to regress the expected conditional volatility of the stock market on their FTS dummies. However, we are not aware of any study that actually uses an FTS dummy in a volatility model and quantifies the improved forecasts.

[12] The results of this analysis are available upon request.

Not surprisingly, we observe that the pronounced spikes in realised equity volatility, appearing in the early and late 2000s and, more sparsely, in the last few years, are all closely related to the flight-to-safety phenomena that took place on the day before. These relationships are corroborated by the crisis periods defined in previous literature. For instance, in Baur and Lucey (2009), the terrorist attack of September 11, 2001, the Enron scandal in December 2001, and the WorldCom's bankruptcy in July 2002 are considered times of market stress. Hollo et al. (2012) depict also the subprime mortgage crisis starting in August 2007, the collapse of Lehman Brothers on September 15, 2008, and the Greek sovereign debt crisis that began in 2009/10 as major financial stress events. Our results add to this list the most recent episodes of market turmoil due to a number of factors, including the UK's vote to leave the EU on June 23, 2016. Lastly, visual inspection of Figure 2.16 shows that flight-to-safety episodes tend to cluster in the same way as large or small movements in the stock index are followed by changes of similar magnitude (see, e.g., Engle and Patton (2001) for a discussion on volatility clustering).

Based on the above observations, we now attempt to predict the future realised variance of the stock index by exploiting past information on the occurrence of flights. While the extant literature offers a variety of methods to predict equity market volatility, in this article we assess the effect on forecast accuracy of using a $\{0, 1\}$ FTS dummy in the HAR-RV model of Corsi (2009). The extended version of the popular HAR-RV is henceforth referred to as the "HAR-RV FTS" model. For comparison purposes, we also produce the forecasts from various benchmark models, including standard HAR-RV, simple AR(1), basic GARCH, RiskMetrics, HEAVY-r, and HEAVY-RM. We refer the reader to Section 2.2 for a review of all these models.

At this point, it is worth noting that although the methodologies differ in the use of raw or transformed data, the forecasting performance is always assessed by comparing the actual with the predicted annualised figures for the realised daily variance of the S&P 500 index. Whenever the forecast produced is expressed in daily terms, this must be multiplied by 252 (the number of days per year that we assume) to obtain the corresponding annualised figure.

2.4.2.1 The HAR-RV FTS model and its competitors

Before introducing our own approach, it seems logical to specify first the HAR-RV model since we build on the method pioneered by Corsi (2009). Using only a daily, a weekly, and a monthly component, this model for realised daily volatility represents a parsimonious alternative to the autoregressive process of order 22. The intuition of using both high- and low-frequency variables comes from observing that long-term volatility influences the short-term one. Corsi (2009) argues that the Heterogeneous Market Hypothesis of Müller et al. (1997) helps explain the asymmetric propagation of volatility. The heterogeneity in the time horizon of traders generate different types of volatility components. However, while short-term volatility does not markedly affect long-term traders, the long-term volatility is an important matter also for those that, operating at the highest dealing frequencies, generate short-term volatility. Therefore, the observed (log) variance realised over a time interval of one day, which is denoted by RV_t^d, becomes a function of not only its own lag RV_{t-1}^d but also the weekly and monthly volatility components of the previous day, which are denoted, respectively, by RV_{t-1}^w and RV_{t-1}^m.

Exploiting the approximate normality of the log transformed data series, which is described in detail by Andersen et al. (2001), we apply the HAR-RV model to the logarithm of the realised daily variances and we write it as:

$$RV_t^d = \beta_0 + \beta_1 RV_{t-1}^d + \beta_2 RV_{t-1}^w + \beta_3 RV_{t-1}^m + \epsilon_t \tag{2.28}$$

where RV_{t-1}^d , RV_{t-1}^w , and RV_{t-1}^m take the following form:

$$\begin{cases} RV_{t-1}^d = RV_{t-1} & (2.29) \\ RV_{t-1}^w = \dfrac{\sum_{j=1}^{5} RV_{t-j}}{5} & (2.30) \\ RV_{t-1}^m = \dfrac{\sum_{j=1}^{22} RV_{t-j}}{22} & (2.31) \end{cases}$$

The methodological novelty of our work is that we add the one-day lag of the FTS measures of Section 2.3 to the autoregressive structure for daily realised volatility proposed by Corsi (2009).[13]

[13] In results available upon request, we have considered additional specifications of the Corsi (2009) model that

We examine whether equity volatility is high only on the occasion of flight-to-safety episodes or remains high for a while after the safe-haven demand ceases. We expect the lagged FTS dummies, which take the value 1 when a flight is happening, to show a positive sign coefficient and deliver 1-day ahead volatility forecasts improved relative to those of the standard HAR-RV model. If we take the same approach as before, the extended version of the model is fitted to log transformed data and the HAR-RV FTS is specified as follows:

$$RV_t^d = \beta_0 + \beta_1 RV_{t-1}^d + \beta_2 RV_{t-1}^w + \beta_3 RV_{t-1}^m + \beta_4 \mathbb{1}_{t-1}^{FTS} + \epsilon_t \tag{2.32}$$

where RV_{t-1}^d , RV_{t-1}^w , and RV_{t-1}^m use the expressions in (2.29), (2.30), and (2.31), respectively, while the $\mathbb{1}_{t-1}^{FTS}$ is defined as in (2.1), (2.19), (2.22), or (2.27). In particular, we assess the forecasting performance of the HAR-RV FTS model using the broad range of dummies that are provided by the recursively computed FTS measures listed in Panel D, E, and F of Table 2.11.

The third model considered is a first-order AR process for the logarithm of the variance realised over a time interval of one day. In this case RV_t^d is regressed only on a constant and its own lag. Then the AR(1) model simply reads as:

$$RV_t^d = \beta_0 + \beta_1 RV_{t-1}^d + \epsilon_t \tag{2.33}$$

All models that have been presented thus far share the common feature that some sort of autoregressive structure is assigned to the log realised variance.[14] As an alternative to these, we next present some of the most widely used models that specify an equation for the true latent volatility instead of for *a conditionally unbiased, but imperfect, volatility proxy* (Patton, 2011). We consider a standard GARCH(1,1), its nested version known as RiskMetrics, and two

include different lags of the FTS measures. Overall, the results suggest a small number of lags of the FTS dummy in the regression equations. For instance, the BIC criterion selects only the first two lags of the $\{0, 1\}$ dummy that identifies FTS episodes according to the Joint model for US Treasuries and gold.

[14] As a competitor of our model, we also considered a version of the extended HAR proposed by Corsi and Renó (2012) that allows volatility to increase more after a negative shock than a positive one of the same magnitude. This model was found to perform better than ours in terms of forecast accuracy. However, after accounting for the leverage effect also in the HAR-RV FTS, the most accurate results came, once again, from our specified model with the FTS dummies. In Appendix A, we provide the in-sample estimates and measures of model fit for a number of specifications of the Corsi (2009) model in which the coefficients can switch based on the leverage effect only, the FTS effect only, and both effects. Overall, the results show that the explanatory power of FTS for future equity return volatility is still significant after controlling for the leverage effect.

HEAVY-type models, one for the conditional variance of returns and one for the conditional mean of the realised variance. Note that, unlike stochastic volatility models, all these methods formulate the conditional variance in a completely deterministic way, with no error term in the equation. In practice, however, to obtain an estimate of the conditional variance from these approaches, we must specify an equation also for the conditional mean of returns. Since we use data of daily frequency and the average return is close to zero, it is reasonable to assume the following return process:

$$r_t = u_t \ , \quad u_t \sim N(0, \sigma_t^2) \tag{2.34}$$

Given the expression in (2.34), which represents the conditional mean equation used in this study, modelling the conditional variance of the error term u_t is equivalent to modelling the variance of stock return r_t.[15] We assume that u_t is a zero mean normally distributed random variable with time-varying volatility σ_t^2. In this chapter, for the conditional variance we use the notation h_t, which is common in the literature.

In the classic GARCH(1,1) model of Bollerslev (1986), the time-t conditional variance is a function of a long-term mean and the immediately previous value of both squared error (or, equivalently, the squared return) and the fitted variance from the model:

$$h_t = \omega_G + \alpha_G r_{t-1}^2 + \beta_G h_{t-1} \tag{2.35}$$

Besides the non-negativity constraints imposed on all the coefficients to ensure positive values of h_t, the sum of α_G and β_G is required to be less than one so that the unconditional variance of u_t is well defined. In doing so, the GARCH model allows volatility not only to cluster but also to mean-revert to its historic average. The latter property is lost in the RiskMetrics.

The J.P. Morgan's RiskMetrics is an exponentially weighted moving average model in which a decay factor (λ) of 0.94 determines how much weight the latest returns carry compared to the observations further in the past. In the RM model the estimate of latent volatility for period t is given by:

$$h_t = \lambda h_{t-1} + (1 - \lambda) r_{t-1}^2 \ , \quad \lambda = 0.94 \tag{2.36}$$

[15] Note that u_t represents the error term of the conditional mean, not variance, of stock returns.

It now becomes apparent why the RM is nested by the autoregressive conditionally heteroskedastic class of models. Indeed, when $\omega_G = 0$, $\beta_G = 0.94$, and $\alpha_G = 1-\beta_G$, the GARCH(1,1) reduces to the RM model.

The remaining two volatility forecasting methods that appear in this chapter derive from a decomposition of the system of equations that defines the HEAVY models proposed by Shephard and Sheppard (2010). In their most basic form, the two equations can be written as:

$$\begin{cases} h_t = \omega + \alpha RV_{t-1} + \beta h_{t-1} & (2.37) \\ \mu_t = \omega_R + \alpha_R RV_{t-1} + \beta_R \mu_{t-1} & (2.38) \end{cases}$$

where h_t denotes, as before, the unobservable conditional variance of stock returns, while μ_t represents the conditional mean of a daily realised measure, which is here the standard RV estimator described in Section 2.4.1. Despite the similarities with expression in (2.35), the conditional variance of returns in (2.37) is entirely determined by the high-frequency information included in the realised variance. Based on this important remark, Shephard and Sheppard (2010) conclude that GARCH and HEAVY models are non-nested.

In our empirical analysis we impose the "Targeting Reparameterisation" suggested by the authors to relate the intercepts of the equations to the unconditional mean of squared returns and realised variances. However, we do not consider the HEAVY models in their "integrated" version since we focus here on one-step ahead, not multi-period, forecasts of volatility. After reparameterising (2.37) and (2.38), the two equations take the form:

$$\begin{cases} h_t = \mu(1 - \alpha\kappa - \beta) + \alpha RV_{t-1} + \beta h_{t-1} & (2.39) \\ \mu_t = \mu_R(1 - \alpha_R - \beta_R) + \alpha_R RV_{t-1} + \beta_R \mu_{t-1} & (2.40) \end{cases}$$

Here expressions in (2.39) and (2.40) represent, respectively, the HEAVY-r and HEAVY-RM model. α, β, α_R, and β_R are the only parameters that need to be estimated using a Gaussian quasi-likelihood with constraints $\alpha_R, \beta_R \geq 0$, $\alpha_R + \beta_R < 1$, and $k = \frac{\mu_R}{\mu} \leq 1$. For μ_R, μ, and k we use the estimators proposed by Shephard and Sheppard (2010).

2.4.2.2 Estimation Details

In this subsection we present the results of the estimated models using the full sample of data. In the HAR-RV FTS, HAR-RV, and AR(1) cases the variable of interest, RV_t^d, is observed and is a function of observables. Estimation of these models makes use of log transformed annualised data and is obtained by minimising the sum of squared errors and correcting the standard errors for possible heteroscedasticity and autocorrelation.[16] It is worth noting that coefficient estimates in the HAR-RV FTS model change with the $\{0, 1\}$ FTS dummy that is used within the model. In Table 2.15 we only provide values for the US Treasuries and Gold $\{0, 1\}$ Joint FTS dummy. Estimation results with the use of other dummies are qualitatively similar to those we present here and are available upon request.

In the GARCH and HEAVY cases, the variable of interest, denoted by, respectively, h_t and μ_t, is unobserved and formulated as function of observables. The parameters associated with these models are estimated such that the Gaussian quasi-likelihoods are maximised. Quasi-maximum likelihood (QML) is applied to raw data for returns and realised daily variances. As for the decay factor in the RM model, no estimation technique is here required as we rely on the value of 0.94 found by the RiskMetrics Group for daily financial data.

[Insert Table 2.15 here]

Table 2.15 reports parameter estimates, with robust standard errors in parentheses, for the models considered in this article. Consistent with previous findings, our results indicate that the coefficients of all three volatility components in the model of Corsi (2009) are highly significant. The same holds true for our extended model with the FTS dummy as regressor. Moreover, the β_4 coefficient in equation (2.32) is positive and strongly statistically significant. Therefore, as expected, the model suggests that an FTS day is followed by an increase in equity market volatility. As for the remaining models, we find that, in line with the literature, the intercept of the GARCH model is small and insignificant whilst the β_G coefficient is highly significant with value between 0.8 and 0.9. Lastly, we observe that the estimation results for the HEAVY

[16] For the HAR-RV FTS and HAR-RV models we employ Newey-West standard error correction with 22 lags.

models are similar to those reported by Shephard and Sheppard (2010) for the S&P 500 index.

2.4.3 Empirical Results

As an important test of the adequacy of the HAR-RV FTS introduced in Section 2.3.2, the forecasting accuracy of the new method needs to be considered and compared with that of various benchmark models. We evaluate the one-day ahead forecasts for the realised variance of the S&P 500 index made by the different models, both in-sample and out-of-sample, using the following loss functions:

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(RV_t - F_t)^2}$$

$$MAE = \frac{1}{T}\sum_{t=1}^{T} \mid RV_t - F_t \mid$$

$$MAPE(\%) = \frac{100}{T}\sum_{t=1}^{T}\left|\frac{RV_t - F_t}{RV_t}\right|$$

$$QLIKE = \frac{1}{T}\sum_{t=1}^{T}\left[\frac{RV_t}{F_t} - \log\left(\frac{RV_t}{F_t}\right) - 1\right]$$

where RV_t and F_t are, respectively, the actual and predicted annualised values of the realised daily variance at time t. Here T denotes the total number of forecasts available for the analysis and is different for the in-sample and out-of-sample cases. To determine whether the forecasts are biased or unbiased, we also estimate the Mincer-Zarnowitz (MZ) regression equation:

$$RV_t = b_0 + b_1 F_t + \epsilon_t \tag{2.41}$$

where RV_t and F_t are defined as above. According to the MZ test, unbiasedness is violated if the confidence intervals of b_0 and b_1 do not include the values of zero and one respectively.

In the out-of-sample context, we begin by estimating the models using the first 1,000 observations and holding back all subsequent observations. The holdout sample thus starts on February 27, 2004. The one-step ahead forecasts for all other days until the end of the sample are generated by re-estimating the models every time with a new set of data and one observation added at a time. In this work we use both a rolling window (RW), in which the length of the in-sample period is fixed to 1,000 observations, and an increasing window (IW), in which the initial estimation date is fixed to February 02, 2000.[17] Once the out-of-sample forecasts from each model have been obtained, we calculate the loss functions defined before and we also specify the Reality Check of White (2000) in the same way as Bollerslev et al. (2016) do to *test whether the loss of a given model (HARQ) is lower than that from the best model among a set of benchmark models* (Bollerslev et al., 2016). Under this modification of the test, we write the hypotheses as:

$$H_0 : \min_{k=1,...K} \mathbb{E}[L^k(RV, F) - L^0(RV, F)] \leq 0$$
$$H_1 : \min_{k=1,...K} \mathbb{E}[L^k(RV, F) - L^0(RV, F)] > 0$$

where L^0 and L^k indicate the losses associated, respectively, with the HAR-RV FTS and with the benchmark models, i.e. HAR-RV, AR(1), GARCH(1,1), HEAVY-r, HEAVY-RM, and RM. Significant predictive superiority of our model is found in case of rejection of the null. We perform the test using a stationary bootstrap with 999 re-samplings and an average block length of five, as in Bollerslev et al. (2016). Although each of the loss functions we presented above can be used to implement the test, we follow the literature and consider only QLIKE and mean squared error (MSE) losses.[18] This choice is in line with previous studies demonstrating that those two loss functions are robust to the noise present in the adopted proxy for volatility (realised variance, in our specific case) and thus can be used for comparing forecasts of conditional variance (see, e.g., Patton (2011)).

Before presenting in detail the in- and out-of sample performance of the various models, we summarise in Table 2.16 the RMSE and QLIKE losses of the proposed HAR-RV FTS for different $\{0, 1\}$ FTS dummies. We consider each of the recursively computed FTS measures that we discussed earlier. In spite of their intrinsic limitations when compared with the more thorough Joint FTS measure, the Threshold, Ordinal, and Univariate RS techniques are here treated

[17] Note that the very first observations of our sample, between January 03 and February 01, 2000, are lost in calculating the weekly and monthly components of realised volatility, and therefore forecasts cannot be produced.

[18] The mean squared error loss function is defined as $MSE = \frac{1}{T}\sum_{t=1}^{T}(RV_t - F_t)^2$. The RMSE is simply the square root of the MSE and it also can be used as a robust loss function for volatility forecast comparison.

as stand-alone methods for capturing flight-to-safety episodes, and therefore their individual dummies, as well as the joint ones, are incorporated into the model specified in equation (2.32). Table 2.17 shows the same results as Table 2.16, but calculated in proportion to the losses associated with the HAR-RV model. In this manner, we can directly compare the model of Corsi (2009) with ours and determine which FTS measures, when used to extend the standard HAR-RV, lead to improved forecasts of future realised variance.

[Insert Table 2.16 here]

[Insert Table 2.17 here]

Since we are interested in model evaluation, we now consider only the out-of-sample forecasts and leave the others for later discussion.

If we first look at the results in Table 2.16, we see that a superior performance is achieved with IW rather than RW estimation. The only exceptions are a few RMSE losses in each panel, which are lower when rolling windows of past data are used in the estimation method.

From an examination of Table 2.17, a number of remarks can be made concerning the use of flight-to-safety information for volatility forecasting. These results suggest that the mean losses of the HAR-RV FTS can be, but not necessarily are, lower than those associated with the standard HAR-RV. The forecasting accuracy of the new model depends on both the safe haven and the econometric technique that are used to identify flight-to-safety days. When either bonds or gold and bonds are the chosen safe haven, the new model systematically outperforms the HAR-RV model. Indeed, under both loss functions and for both window estimations, the performance of our model is superior regardless of the technique for FTS identification. In case of gold being the only safe haven, the HAR-RV FTS improves on the forecasts from the standard Corsi (2009) model only when the Ordinal and Joint FTS measures are used. Mixed results are found for the other two FTS measures.

A common feature that is shared by Panel B and C, across both losses and both window estimations, is the delivery of the most accurate results from the Ordinal and Joint FTS measures, among all techniques considered for FTS identification. This is true for Panel A only when the assessment is based on the QLIKE loss function.

Lastly, it is important to note that the gains in forecast accuracy due to the FTS measures of Panel C tend to be larger than the gains obtained in Panel A and B. In other words, the forecasts associated with the proposed model benefit more from capturing flights with both US Treasuries and gold than with only one of the two assets.

In the rest of this section we explicitly discuss the in- and out-of-sample performance of the new HAR-RV FTS and all competitor models. For presentation purposes, we select the Ordinal FTS measure of Panel C in Table 2.16 and 2.17 that seems to give the best overall results. The complete set of performance measures for every $\{0, 1\}$ FTS dummy is available upon request.

2.4.3.1 In-Sample Performance

We begin by assessing the performance of the models in-sample where we expect relatively good results since the forecasts refer to the same data that is used to estimate the parameters. In the analysis of forecast errors, all models are ranked on the basis of their losses relative to those of the HAR-RV model. We restrict the discussion to the robust RMSE and QLIKE, although the means of all loss functions defined earlier are provided in Table 2.18. As we have already remarked, the rankings of volatility forecasts by mean absolute error, MAE, and mean absolute percentage error, MAPE, can be misleading. These two loss functions are not robust to the noise that exists in the proxy used for volatility, i.e. the realised daily variances. Patton (2011) finds that the MAE loss function leads to an optimal forecast that is markedly biased downwards. Indeed, it has been argued that MSE penalises large errors more heavily than MAE does (for a review, see, e.g., Brooks (2008)).

[Insert Table 2.18 here]

The results in Table 2.18 indicate that across all loss functions the HAR-RV FTS model outperforms its competitors. The improvements offered by our model in-sample are around 3 and 3.5% for QLIKE and RMSE, respectively. Here and henceforth we quantify the improvements by comparing the forecasting accuracy (mean forecast loss) of the HAR-RV FTS to that of the second best performing model. The model of Corsi (2009) is found to be second only to to our model under the RMSE loss, but delivers on average higher QLIKE losses than the HEAVY-RM model as well. The forecast losses associated with the simple AR(1), GARCH (1,1), HEAVY-r, and RM models are always greater than those associated with the HAR-RV.

As for the Mincer-Zarnowitz regression, we use this test to determine whether the forecasts of the various models are biased or unbiased. In doing so, we calculate the confidence intervals for the intercept and slope parameter in the equation and check whether they include the value of 0 and 1, respectively. Table 2.19 shows the results of the regression analysis along with the 95% confidence intervals. It is worth remembering that for all models, regardless of any log-transformation of the data in the estimation procedure, the actual annualised value for realised daily variance at time t is regressed on a constant and the annualised value predicted from each model based on information available at the immediately previous time step.

[Insert Table 2.19 here]

The Mincer-Zarnowitz tests presented in Table 2.19 provide evidence that, in-sample, the forecasts associated with all models that specify an autoregressive structure for realised volatility are unbiased. The same is true of the simple RM model but not of the more theoretically sound GARCH (1,1). Lastly, we note that HEAVY-r and HEAVY-RM display some of the highest Mincer-Zarnowitz R^2, although the test indicates that both high frequency volatility based models produce biased in-sample forecasts.

2.4.3.2 Out-of-Sample Performance

Now we turn to the question of how the HAR-RV FTS model performs out-of-sample when compared with a number of widely used volatility forecasting models. We follow the same approach as in the in-sample context. First, we discuss the mean losses associated with each model relative to the HAR-RV, focusing predominantly on the two robust loss functions. The results are provided in Table 2.20. Then, to ensure unbiasedness of the out-of-sample forecasts, we calculate the confidence interval of the parameters in the MZ regression equation applied to each model. The results of the regression analysis are given in Table 2.21 and 2.22. Convincing proof of the adequacy of the proposed HAR-RV FTS model requires a statistical test that, after controlling for data snooping, ascribes superior results to inherent merits of the new methodology rather than pure chance. For this purpose, we rely on the modified Reality Check test that we described above and whose results are reported in Table 2.23.

[Insert Table 2.20 here]

We begin by observing that the ability of the HAR-RV FTS model to deliver superior forecasts is confirmed by the out-of-sample analysis. Indeed, our model exhibits the lowest values for all mean losses considered in Table 2.20. The out-of-sample improvements are around 2 and 3.5% for QLIKE and RMSE, respectively. In addition, we note that with RW estimation the gains in forecast accuracy are greater than when IWs of past data generate the forecasts. As we have seen in Table 2.18, the standard HAR-RV comes second in the ranking of out-of-sample forecasts for the RMSE but not for the QLIKE loss, in which case the model of Corsi (2009) is outperformed also by the HEAVY-RM. The other high frequency volatility based model fails to improve on the standard HAR-RV under all losses. Comparing the GARCH (1,1) and the nested RM, we find that the former can perform better under the RMSE but not the QLIKE loss. Overall, the results of these two models, as well as of the simple AR(1) process, are substantially inferior to those of the best performing methods.

As part of our model evaluation exercise, we now consider the results of the Mincer-Zarnowitz tests provided in Table 2.21 and 2.22 for, respectively, RW and IW methodologies.

[Insert Table 2.21 here]

[Insert Table 2.22 here]

The results in Table 2.21 and 2.22 indicate that the out-of-sample forecasts of the HAR-RV FTS model are unbiased across both window lengths. The model of Corsi (2009) and the RM are the only other two cases for which the same conclusion can be drawn. Examining the results for the HEAVY models, we notice again that a relatively high Mincer-Zarnowitz R^2 does not imply unbiased forecasts under this test.

The last step of this out-of-sample analysis comprises a formal test for the significance of the improvements offered by the HAR-RV FTS model in volatility forecasting. It is important to remember that the Reality Check test is specified such that under the null the forecast losses of the HAR-RV FTS are not significantly lower than those of all benchmark models. The p-values of the test performed for the MSE and QLIKE loss functions, using both RW and IW estimation methods, are provide in Table 2.23.

[Insert Table 2.23 here]

The results in Table 2.23 provide evidence about the significance of the improvements offered by the HAR-RV FTS. Indeed, we reject the null hypothesis of no predictive superiority of the new model at the 1% significance level in the vast majority of cases. Only under the MSE loss function and IW estimation method the p-value is above 1% but still less than 2%. Interestingly, the p-values are lower under the QLIKE loss whilst we have shown in Table 2.20 that the gains in forecast accuracy, in terms of minor mean losses, are higher under the RMSE loss.

In conclusion, we can state that our volatility forecasting model that explicitly accounts for FTS events delivers more accurate forecasts than a set of benchmark models, and the magnitude of the improvements is significant in a statistical sense.

2.5 Conclusions

In this chapter we show that a method for identifying flight-to-safety events can significantly improve the accuracy of a volatility forecasting model. We focus on the role of US Treasuries and gold as safe havens against the US stock market. The non-overlapping FTS episodes detected by government bonds and gold can be interpreted as evidence that both assets are safe haven but with different properties. We calculate a number of statistics that signal the day-to-day occurrence of flight-to-safety events. The key feature of our FTS measures is that they are time-consistent in the sense that they do not depend on future information and thus can be readily applied in forecasting the future. We demonstrate that explicitly accounting for flight-to-safety in the context of volatility modelling can lead to significant improvements in the one-day ahead forecasts of realised variance of the S&P 500 equity index returns.

The analytical framework that we derive in this chapter to forecast future volatility using flight-to-safety comprises two parts. In the first part we create several FTS measures that assign daily probabilities of flight-to-safety on the basis of (ab)normal movements in the equity, bond, and gold markets. In the second part we introduce the FTS measures of the first part into a volatility forecasting model. Incorporating the symptoms of flight-to-safety directly into a volatility model would allow us to swap from a two-stage to a single-stage framework. This problem is addressed in the next chapter of this series. In Chapter 3, we propose realised semi-covariance between negative intraday equity returns and positive intraday safe haven returns as a proxy measure of flight-to-safety and a direct predictor of equity return variance in a multivariate GARCH model. The nature of the model is such that the benefits of using flight-to-safety in volatility forecasting can be assessed for both equity and safe haven returns, not only for one-period but also for multi-period horizons.

As a final remark concerning the detection of FTS episodes, we would like to point out that the sample of our study could be expanded to pursue a multi-country analysis and include additional safe haven assets. We leave investigation of such possibilities to future research.

2.6 Tables of Chapter 2

Table 2.1 – Datastream data

	Series (Datatypes)	Description
US Equity	TOTMKUS (RI)	US-DS Market - TOT RETURN IND
US Bonds	BMUS10Y (RI)	US BENCHMARK 10 YEAR DS GOVT. INDEX - TOT RETURN IND
Gold	GOLDBLN (P)	Gold Bullion LBM U$/Troy Ounce

Notes: Data used in the study to identify FTS events. Sample starts on January 1, 1980 and ends on June 30, 2016. Source: Datastream.

Table 2.2 – Summary statistics of asset returns

	N^o Obs	Mean	Std Dev	Min	Max	Skewness	Kurtosis
US Equity returns	9522	12.366	17.191	-18.691	11.531	-0.642	17.582
US Bond returns	9522	7.704	7.991	-3.590	4.134	0.173	4.548
Gold returns	9522	4.221	18.879	-16.364	12.984	-0.138	14.321

Notes: Summary statistics for daily stock, bond and gold returns. Returns are discretely compounded rates expressed in percentages. The mean and standard deviation are expressed on a yearly basis. Excess Kurtosis is assumed.

Table 2.3 – Correlation matrix of asset returns

	US Equity returns	US Bond returns	Gold returns
US Equity returns	100.000		
US Bond returns	-3.609	100.000	
Gold returns	-1.227	3.288	100.000

Notes: Unconditional correlation matrix for daily stock, bond and gold returns.

Table 2.4 – The Threshold FTS measure

$k = 1.5$	US Treasury Bonds	Gold	US Treasury Bonds and Gold
FTS % Incidence	0.893%	0.567%	1.302%

Notes: Results of the Threshold FTS measure for both US Treasuries and gold. The threshold parameter k is fixed to 1.5. Last column represents the incidence of FTS events when either asset is flagging a flight-to-safety day according to the model.

Table 2.5 – The Ordinal FTS measure

	US Treasury Bonds	Gold	US Treasury Bonds and Gold
FTS % Incidence	5.692%	6.700%	9.368%

Notes: Results of the Ordinal FTS measure for both US Treasuries and gold. Last column represents the incidence of FTS events when either asset is flagging a flight-to-safety day according to the model.

Table 2.6 – Estimation results of the Univariate RS model

	US Treasury Bonds	Gold
Regime-dependent Mean (expressed in daily %)		
μ_1	-0.053	-0.084
μ_2	-0.012	0.014
μ_3	0.192	0.119
Regime-dependent Volatility (expressed in annualised terms)		
σ_1	0.098	0.151
σ_2	0.197	0.265
σ_3	0.473	0.656

Notes: In-sample estimation results of the Univariate RS model for US Treasuries and gold. The model is characterised by regime shifts for both the intercept and volatility parameter. We use Regime 1 for the low volatility regime, Regime 2 for the high volatility regime, and Regime 3 for the flight-to-safety regime.

Table 2.7 – The Univariate RS FTS measure

	US Treasury Bonds	Gold	US Treasury Bonds and Gold
FTS % Incidence	6.858%	5.083%	8.843%

Notes: Results of the Univariate RS FTS measure for both US Treasuries and gold. Last column represents the incidence of FTS events when either asset is flagging a flight-to-safety day according to the model.

Table 2.8 – Estimation results of the Bivariate RS model

	US Treasury Bonds	Gold
Regime-dependent Mean and Jump Terms (expressed in daily %)		
α_0	0.085	0.080
α_1	-0.416	-0.042
α_2	-0.009	0.004
β_0	0.008	0.020
β_1	-0.113	-0.095
β_2	0.031	-0.060
FTS Estimates (expressed in in daily %) and Beta Estimates		
α_3	-2.317	-14.426
β_3	0.800	9.263
υ	0.064	0.012
β_4	0.138	-0.033
β_5	-0.311	-0.016
Regime-dependent Volatility (expressed in annualised terms)		
$h_s(S_t^s = 1)$	0.100	0.107
$h_s(S_t^s = 2)$	0.263	0.284
$h_s(S_t^H = 1)$	0.021	0.106
$h_s(S_t^H = 2)$	0.050	0.353

Notes: In-sample estimation results of the Bivariate RS model for US Treasuries and gold. The model comprises 3 two-sate Markov chain processes and 6 possible regimes.

Table 2.9 – The Bivariate RS FTS measure

	US Treasury Bonds	Gold	US Treasury Bonds and Gold
FTS % Incidence	23.251%	10.218%	25.856%

Notes: Results of the Bivariate RS FTS measure for both US Treasuries and gold. Last column represents the incidence of FTS when at least one of the two assets is flagging a flight-to-safety day according to the model.

Table 2.10 – The Joint FTS measure

	US Treasury Bonds	Gold	US Treasury Bonds and Gold
FTS % Incidence	2.668%	1.218%	3.245%

Notes: Results of the Joint FTS measure for both US Treasuries and gold. Last column represents the incidence of FTS events when at least one of the two assets is flagging a flight-to-safety day according to the model.

Table 2.11 – Full sample VS Recursive computation of FTS measures

	FTS % Inc. 1980-2016	FTS % Inc. 2000-2016	Start Date	End Date	N^o Days	Time Consistent
Panel A: Safe haven is US Treasuries - Full Sample						
Threshold FTS measure	0.893%	1.603%	02/01/1980	30/06/2016	9522	NO
Ordinal FTS measure	5.692%	9.921%	02/01/1980	30/06/2016	9522	NO
Univariate RS FTS measure	6.858%	13.336%	02/01/1980	30/06/2016	9522	NO
Bivariate RS FTS measure	23.251%	44.122%	02/01/1980	30/06/2016	9522	NO
Joint FTS measure	2.668%	5.181%	02/01/1980	30/06/2016	9522	NO
Panel B: Safe haven is Gold - Full Sample						
Threshold FTS measure	0.567%	0.697%	02/01/1980	30/06/2016	9522	NO
Ordinal FTS measure	6.700%	9.108%	02/01/1980	30/06/2016	9522	NO
Univariate RS FTS measure	5.083%	7.110%	02/01/1980	30/06/2016	9522	NO
Bivariate RS FTS measure	10.218%	14.684%	02/01/1980	30/06/2016	9522	NO
Joint FTS measure	1.218%	2.370%	02/01/1980	30/06/2016	9522	NO
Panel C: Safe haven is US Treasuries and Gold - Full Sample						
Threshold FTS measure	1.302%	2.021%	02/01/1980	30/06/2016	9522	NO
Ordinal FTS measure	9.368%	13.313%	02/01/1980	30/06/2016	9522	NO
Univariate RS FTS measure	8.843%	14.289%	02/01/1980	30/06/2016	9522	NO
Bivariate RS FTS measure	25.856%	44.424%	02/01/1980	30/06/2016	9522	NO
Joint FTS measure	3.245%	6.273%	02/01/1980	30/06/2016	9522	NO
Panel D: Safe haven is US Treasuries - Recursive						
Threshold FTS measure	NA	1.952%	03/01/2000	30/06/2016	4304	YES
Ordinal FTS measure	NA	9.410%	03/01/2000	30/06/2016	4304	YES
Univariate RS FTS measure	NA	14.893%	03/01/2000	30/06/2016	4304	YES
Joint FTS measure	NA	5.088%	03/01/2000	30/06/2016	4304	YES
Panel E: Safe haven is Gold - Recursive						
Threshold FTS measure	NA	0.860%	03/01/2000	30/06/2016	4304	YES
Ordinal FTS measure	NA	9.526%	03/01/2000	30/06/2016	4304	YES
Univariate RS FTS measure	NA	5.158%	03/01/2000	30/06/2016	4304	YES
Joint FTS measure	NA	2.207%	03/01/2000	30/06/2016	4304	YES
Panel F: Safe haven is US Treasuries and Gold - Recursive						
Threshold FTS measure	NA	2.486%	03/01/2000	30/06/2016	4304	YES
Ordinal FTS measure	NA	13.941%	03/01/2000	30/06/2016	4304	YES
Univariate RS FTS measure	NA	15.544%	03/01/2000	30/06/2016	4304	YES
Joint FTS measure	NA	6.180%	03/01/2000	30/06/2016	4304	YES

Table 2.12 – Robustness of the FTS measures

	FTS % Inc. FS	FTS % Inc. Rec	Start Date	End Date	Non-match days %(1)	Non-match days %(2)
Panel A: Safe haven is US Treasuries						
Threshold FTS measure	1.603%	1.952%	03/01/2000	30/06/2016	0.767%	0.767%
Ordinal FTS measure	9.921%	9.410%	03/01/2000	30/06/2016	4.647%	2.742%
Univariate RS FTS measure	13.336%	14.893%	03/01/2000	30/06/2016	6.947%	5.135%
Panel B: Safe haven is Gold						
Threshold FTS measure	0.697%	0.860%	03/01/2000	30/06/2016	0.302%	0.302%
Ordinal FTS measure	9.108%	9.526%	03/01/2000	30/06/2016	5.855%	3.950%
Univariate RS FTS measure	7.110%	5.158%	03/01/2000	30/06/2016	3.253%	1.975%

Notes: This table contains the results of the robustness test for the Threshold, Ordinal and Univariate RS FTS measures. The table shows the FTS % incidence with recursive and full-sample computation, and also the percentage of non-matching days, i.e. the percentage of days from 03/01/2000 to 30/06/2016 on which the dummies of full-sample and recursive computation do not agree in value. The furthest right column shows the percentage of days on which the recursively computed FTS measure is in "great" disagreement with the full-sample measure. This happens when the difference between the two probabilities of a flight-to-safety is larger than 0.5. Beyond this threshold, extremely diverging results are found between full-sample and recursive statistics.

Table 2.13 – OMI realised library data

	Series (Datatypes)	Description
US Equity	S&P 500 LIVE (SPX2.closeprice)	S&P 500 - CLOSING PRICE
	S&P 500 LIVE (SPX2.r)	S&P 500 - RETURN
	S&P 500 LIVE (SPX2.rv)	S&P 500 - REALISED VARIANCE (5-minute)

Notes: Data used in the study to forecast the volatility of the S&P 500 Index. Sample starts on January 3, 2000 and ends on June 30, 2016. Source: Oxford-Man Institute.

Table 2.14 – Summary statistics of S&P 500 returns and RV

S&P500	N^o Obs	Mean	Std Dev	Min	Max	Skewness	Kurtosis
Returns	4121	0.008	1.202	-9.351	10.220	-0.166	7.460
Realised Variance	4121	1.231	2.633	0.016	77.477	10.984	216.096

Notes: Summary statistics for daily returns and realised daily variances of the S&P 500 index. Statistics for both series refer to returns expressed in percentages. Excess Kurtosis is assumed.

Table 2.15 – In-sample estimation of volatility models

HAR-RV FTS	$RV_t^d = \beta_0 + \beta_1 RV_{t-1}^d + \beta_2 RV_{t-1}^w + \beta_3 RV_{t-1}^m + \beta_4 \mathbb{1}_{t-1}^{FTS} + \epsilon_t$
HAR-RV	$RV_t^d = \beta_0 + \beta_1 RV_{t-1}^d + \beta_2 RV_{t-1}^w + \beta_3 RV_{t-1}^m + \epsilon_t$
AR(1)	$RV_t^d = \beta_0 + \beta_1 RV_{t-1}^d + \epsilon_t$
GARCH (1,1)	$h_t = \omega_G + \alpha_G r_{t-1}^2 + \beta_G h_{t-1}$
HEAVY-r	$h_t = \mu(1 - \alpha\kappa - \beta) + \alpha RV_{t-1} + \beta h_{t-1}$
HEAVY-RM	$\mu_t = \mu_R(1 - \alpha_R - \beta_R) + \alpha_R RV_{t-1} + \beta_R \mu_{t-1}$
RM	$h_t = \lambda h_{t-1} + (1 - \lambda) r_{t-1}^2$

	HAR-RV FTS	HAR-RV	AR(1)	GARCH (1,1)	HEAVY-r	HEAVY-RM	RM
β_0	-0.375***	-0.236***	-0.879***				
	(0.045)	(0.048)	(0.081)				
β_1	0.306***	0.324***	0.789***				
	(0.029)	(0.029)	(0.018)				
β_2	0.423***	0.432***					
	(0.037)	(0.037)					
β_3	0.186***	0.188***					
	(0.025)	(0.025)					
β_4	0.320***						
	(0.042)						
ω_G				0.218e−5			
				(0.398e−6)			
α_G				0.129*			
				(0.013)			
β_G				0.861***			
				(0.013)			
μ					0.145e−3		
α					0.389*		
					(0.039)		
κ					0.851		
β					0.657**		
					(0.035)		
μ_R						0.123e−3	
α_R						0.418*	
						(0.042)	
β_R						0.557*	
						(0.046)	
λ							0.94

Notes: This table provides in-sample parameter estimates for the HAR-RV FTS model and its competitors. The HAR-RV FTS, HAR-RV, and AR(1) models are estimated by OLS using log transformed annualised data. Robust standard errors are reported in parentheses and significance of regression coefficients is assessed by means of usual inference. Estimation results for the HAR-RV FTS are based on the US Treasuries and Gold $\{0, 1\}$ Joint FTS dummy. The GARCH and HEAVY models are estimated by QML using raw data for both returns and realised daily variances. Significance of the coefficients is assessed with Likelihood ratio (LR) tests.

* Statistical significance at the 10% level.
** Statistical significance at the 5% level.
*** Statistical significance at the 1% level.

Table 2.16 – Recursively computed FTS dummies in volatility forecasting (1)

HAR-RV FTS Model		In-Sample		Out-of-Sample RW		Out-of-Sample IW	
	FTS % Inc.	QLIKE	RMSE	QLIKE	RMSE	QLIKE	RMSE
Panel A: Safe haven is US Treasuries							
Threshold FTS measure	1.952%	0.194	0.045	0.215	0.049	0.213	0.049
Ordinal FTS measure	9.410%	0.190	0.045	0.209	0.049	0.208	0.049
Univariate RS FTS measure	14.893%	0.195	0.045	0.214	0.049	0.213	0.049
Joint FTS measure	5.088%	0.194	0.045	0.214	0.049	0.212	0.049
Panel B: Safe haven is Gold							
Threshold FTS measure	0.860%	0.195	0.045	0.215	0.049	0.213	0.049
Ordinal FTS measure	9.526%	0.192	0.043	0.210	0.047	0.210	0.047
Univariate RS FTS measure	5.158%	0.195	0.045	0.215	0.049	0.214	0.049
Joint FTS measure	2.207%	0.195	0.043	0.214	0.047	0.213	0.047
Panel C: Safe haven is US Treasuries and Gold							
Threshold FTS measure	2.486%	0.194	0.044	0.214	0.049	0.213	0.049
Ordinal FTS measure	13.941%	0.189	0.043	0.209	0.047	0.208	0.047
Univariate RS FTS measure	15.544%	0.195	0.045	0.214	0.049	0.213	0.049
Joint FTS measure	6.180%	0.193	0.043	0.213	0.047	0.211	0.047

Notes: This table contains the in- and out-of-sample RMSE and QLIKE losses of the HAR-RV FTS model for different FTS dummies. We consider the Threshold, Ordinal, Univariate RS, and Joint FTS measures recursively computed using either US Treasuries, or gold, or both assets as safe havens. The incidence of flight-to-safety events according to each measure is also shown.

Table 2.17 – Recursively computed FTS dummies in volatility forecasting (2)

HAR-RV FTS Model		In-Sample		Out-of-Sample RW		Out-of-Sample IW	
	FTS % Inc.	QLIKE	RMSE	QLIKE	RMSE	QLIKE	RMSE
Panel A: Safe haven is US Treasuries							
Threshold FTS measure	1.952%	0.995	0.993	0.998	0.993	0.996	0.991
Ordinal FTS measure	9.410%	0.972	0.995	0.973	0.996	0.974	0.998
Univariate RS FTS measure	14.893%	0.998	0.998	0.996	0.996	0.998	0.998
Joint FTS measure	5.088%	0.990	0.991	0.994	0.994	0.992	0.992
Panel B: Safe haven is Gold							
Threshold FTS measure	0.860%	0.998	0.997	0.999	1.002	0.999	0.999
Ordinal FTS measure	9.526%	0.979	0.967	0.977	0.959	0.981	0.968
Univariate RS FTS measure	5.158%	0.999	0.997	1.001	1.004	1.000	0.999
Joint FTS measure	2.207%	0.995	0.967	0.995	0.958	0.995	0.961
Panel C: Safe haven is US Treasuries and Gold							
Threshold FTS measure	2.486%	0.993	0.989	0.996	0.994	0.995	0.990
Ordinal FTS measure	13.941%	0.965	0.964	0.970	0.964	0.971	0.966
Univariate RS FTS measure	15.544%	0.998	0.998	0.996	0.997	0.998	0.998
Joint FTS measure	6.180%	0.986	0.963	0.989	0.954	0.988	0.959

Notes: This table contains the in- and out-of-sample RMSE and QLIKE losses of the HAR-RV FTS model for different FTS dummies, divided by those of HAR-RV model. We consider the Threshold, Ordinal, Univariate RS, and Joint FTS measures recursively computed using either US Treasuries, or gold, or both assets as safe havens. The incidence of flight-to-safety events according to each measure is also shown.

Table 2.18 – In-sample performance - Mean forecast losses

	HAR-RV FTS	HAR-RV	AR(1)	GARCH (1,1)	HEAVY-r	HEAVY-RM	RM
RMSE	0.965	1.000	1.075	1.079	1.060	1.015	1.157
MAE	0.966	1.000	1.090	1.310	1.165	1.025	1.351
MAPE (%)	0.973	1.000	1.199	1.662	1.299	1.038	1.504
QLIKE	0.966	1.000	1.158	1.278	1.046	0.995	1.389

Notes: This table contains the in-sample mean forecast losses for each model relative to the HAR-RV. Results for the HAR-RV FTS are based on the US Treasuries and Gold $\{0,1\}$ Ordinal FTS dummy.

Table 2.19 – In-sample performance - Mincer-Zarnowitz regression

	b_0	b_1	R^2
HAR-RV FTS	-0.003 (-0.007 , 0.001)	1.138 (0.970 , 1.306)	0.584
HAR-RV	-0.001 (-0.004 , 0.001)	1.078 (0.950 , 1.206)	0.547
AR(1)	-0.003 (-0.007 , 0.000)	1.170 (0.998 , 1.342)	0.484
GARCH(1,1)	0.001 (-0.004 , 0.006)	0.802 (0.630 , 0.974)	0.507
HEAVY-r	0.003 (0.001 , 0.004)	0.778 (0.717 , 0.839)	0.535
HEAVY-RM	0.003 (0.001 , 0.005)	0.905 (0.842 , 0.967)	0.534
RM	0.003 (-0.003 , 0.010)	0.768 (0.521 , 1.015)	0.433

Notes: This table contains the estimated parameters, their respective 95% confidence interval (in parentheses), and R^2 of the Mincer-Zarnowitz regressions for the in-sample forecasts of each model. Results for the HAR-RV FTS are based on the US Treasuries and Gold $\{0,1\}$ Ordinal FTS dummy.

Table 2.20 – Out-of-sample performance - Mean forecast losses

		HAR-RV FTS	HAR-RV	AR(1)	GARCH(1,1)	HEAVY-r	HEAVY-RM	RM
RMSE	RW	0.964	1.000	1.083	1.013	1.048	1.014	1.145
	IW	0.966	1.000	1.074	1.157	1.058	1.010	1.146
MAE	RW	0.962	1.000	1.105	1.303	1.156	1.024	1.326
	IW	0.967	1.000	1.112	1.412	1.207	1.036	1.341
MAPE (%)	RW	0.988	1.000	1.270	1.790	1.280	1.073	1.495
	IW	0.990	1.000	1.299	1.818	1.395	1.091	1.527
QLIKE	RW	0.970	1.000	1.190	1.433	1.037	0.991	1.302
	IW	0.971	1.000	1.142	1.499	1.064	0.990	1.311

Notes: This table contains the out-of-sample mean forecast losses for each model relative to the HAR-RV. Results for the HAR-RV FTS are based on the US Treasuries and Gold $\{0, 1\}$ Ordinal FTS dummy.

Table 2.21 – Out-of-sample performance - Mincer-Zarnowitz regression (RW)

Rolling Window	b_0	b_1	R^2
HAR-RV FTS	-0.002 (-0.005 , 0.001)	1.128 (0.943 , 1.312)	0.588
HAR-RV	0.001 (-0.001 , 0.004)	0.996 (0.874 , 1.118)	0.548
AR(1)	-0.003 (-0.006 , 0.000)	1.188 (0.997 , 1.378)	0.483
GARCH(1,1)	-0.001 (-0.007 , 0.005)	0.881 (0.635 , 1.127)	0.550
HEAVY-r	0.002 (-0.001 , 0.004)	0.818 (0.702 , 0.933)	0.533
HEAVY-RM	0.003 (0.001 , 0.005)	0.903 (0.831 , 0.976)	0.542
RM	0.003 (-0.003 , 0.009)	0.778 (0.506 , 1.051)	0.447

Notes: This table contains the estimated parameters, their respective 95% confidence interval (in parentheses), and R^2 of the Mincer-Zarnowitz regressions. The out-of-sample forecasts delivered by each model with a rollwing window are assessed. Results for the HAR-RV FTS are based on the US Treasuries and Gold $\{0, 1\}$ Ordinal FTS dummy.

Table 2.22 – Out-of-sample performance - Mincer-Zarnowitz regression (IW)

Increasing Window	b_0	b_1	R^2
HAR-RV FTS	-0.003 (-0.006 , 0.001)	1.167 (0.980 , 1.354)	0.591
HAR-RV	0.000 (-0.003 , 0.003)	1.064 (0.920 , 1.208)	0.551
AR(1)	-0.004 (-0.007 , 0.000)	1.191 (1.009 , 1.373)	0.493
GARCH(1,1)	0.003 (-0.001 , 0.007)	0.740 (0.553 , 0.927)	0.460
HEAVY-r	0.002 (0.000 , 0.004)	0.789 (0.695 , 0.884)	0.540
HEAVY-RM	0.002 (0.001 , 0.004)	0.914 (0.838 , 0.991)	0.544
RM	0.003 (-0.003 , 0.009)	0.778 (0.506 , 1.051)	0.447

Notes: This table contains the estimated parameters, their respective 95% confidence interval (in parentheses), and R^2 of the Mincer-Zarnowitz regressions. The out-of-sample forecasts delivered by each model with an increasing window are assessed. Results for the HAR-RV FTS are based on the US Treasuries and Gold $\{0, 1\}$ Ordinal FTS dummy.

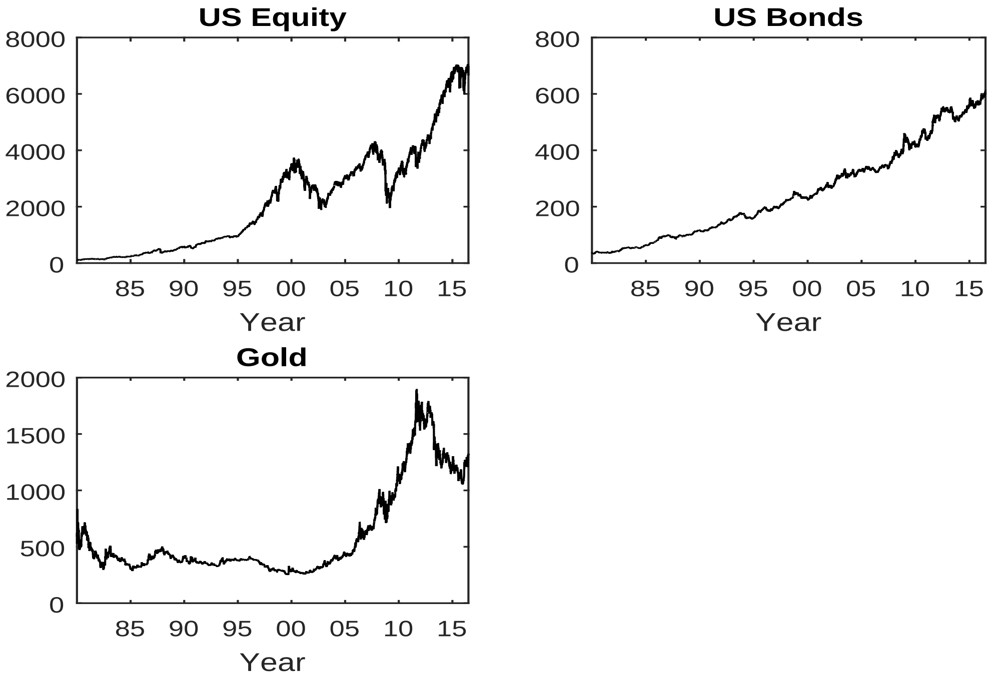
Table 2.23 – Out-of-sample performance - Reality Check test

		p-value
MSE	RW IW	0.005 0.018
QLIKE	RW IW	0.000 0.000

Notes: This table contains the p-values of the Reality Check test. Under the null hypothesis the out-of-sample forecast losses of the HAR-RV FTS are not significantly lower than those of all benchmark models. Results are based on the US Treasuries and Gold $\{0, 1\}$ Ordinal FTS dummy.

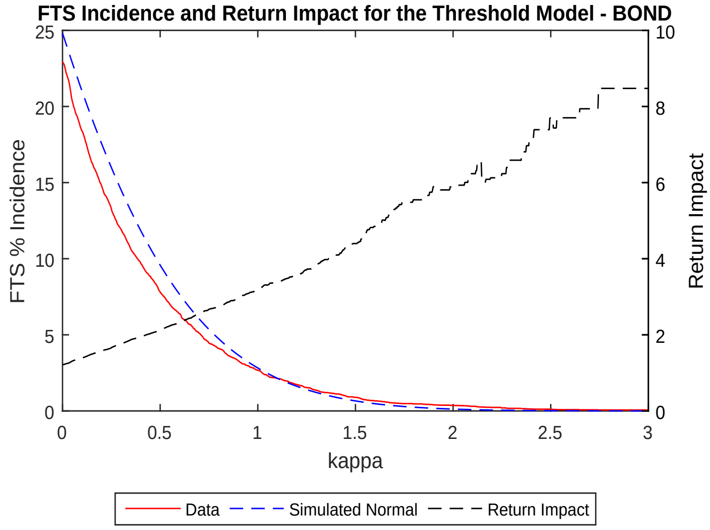
2.7 Figures of Chapter 2

Figure 2.1 – Time series of US equity, US 10-y Treasury bonds and gold



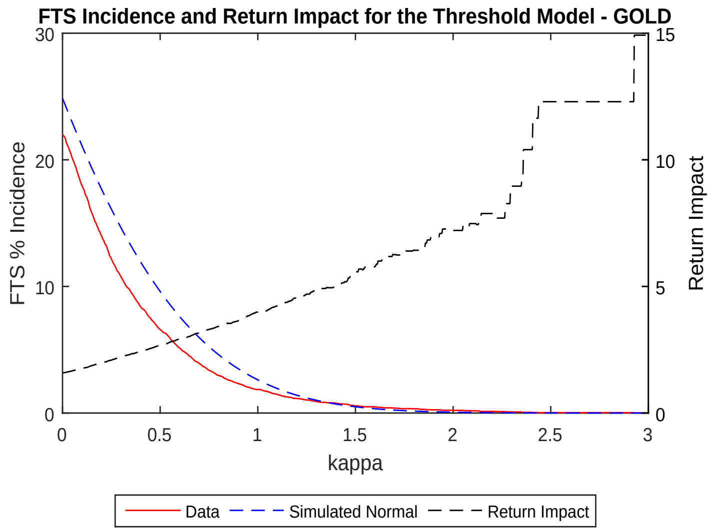
The figure displays the daily time series of the United States-Datastream Market (top left panel), the United States Benchmark 10 Year Datastream Government Index (top right panel), and the gold bullion (bottom left panel) from January 1980 to June 2016. The Total Return Index (RI) is used for the first two assets, while for gold the (Adjusted - Default) Price (P), expressed in USD per Troy Ounce, is used. Source: Datastream.

Figure 2.2 – Simulation experiment - US Treasury bonds



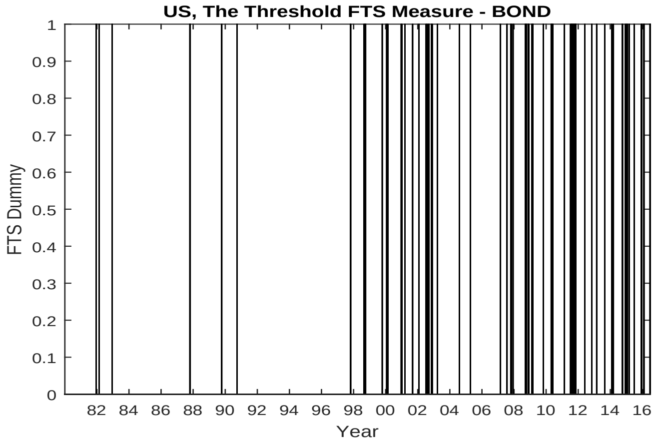
The figure displays the results of the simulation experiment when US Treasury bonds are the safe haven in the Threshold model. FTS Incidence (FTS days as a % of total sample size) and Return Impact (average difference between bond and equity returns on FTS days) are computed as a function of the threshold parameter k ranging from 0 to 3. The curves of the actual and bivariate normal data intersect at $k = 1.108$, where FTS incidence is equal to 2.111%. For $k = 1.5$, 0.893% of all days are FTS days compared with 0.661% in the bivariate normal world.

Figure 2.3 – Simulation experiment - Gold



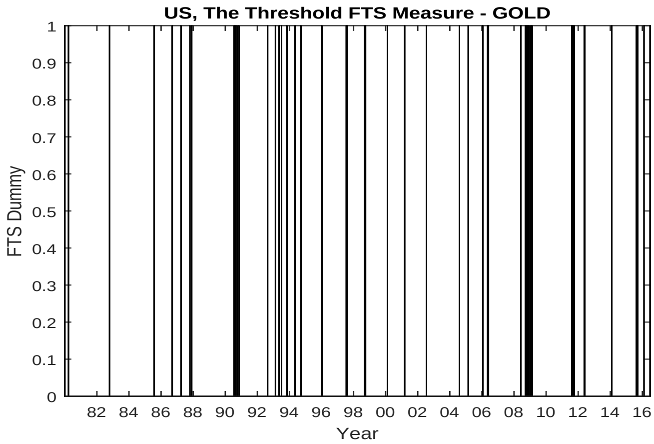
The figure displays the results of the simulation experiment when gold is the safe haven in the Threshold model. FTS Incidence (FTS days as a % of total sample size) and Return Impact (average difference between gold and equity returns on FTS days) are computed as a function of the threshold parameter k ranging from 0 to 3. The curves of the actual and bivariate normal data intersect at $k = 1.369$, where FTS incidence is equal to 0.809%. For $k = 1.5$, 0.567% of all days are FTS days compared with 0.499% in the bivariate normal world.

Figure 2.4 – The Threshold FTS dummy - US Treasury bonds



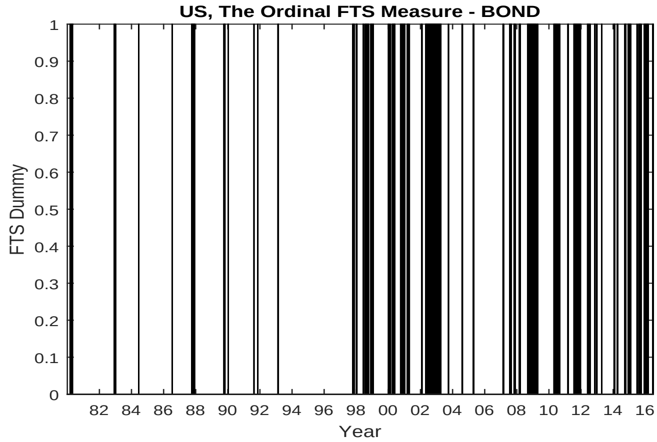
The figure displays the $\{0, 1\}$ FTS dummy according to the Threshold model in which the US Treasury bonds act as safe haven. 0.893% of all days are classified as FTS days.

Figure 2.5 – The Threshold FTS dummy - Gold



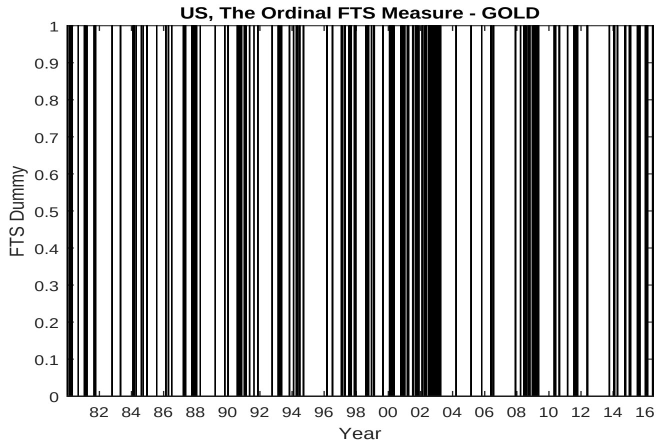
The figure displays the $\{0, 1\}$ FTS dummy according to the Threshold model in which gold acts as safe haven. 0.567% of all days are classified as FTS days.

Figure 2.6 – The Ordinal FTS dummy - US Treasury bonds



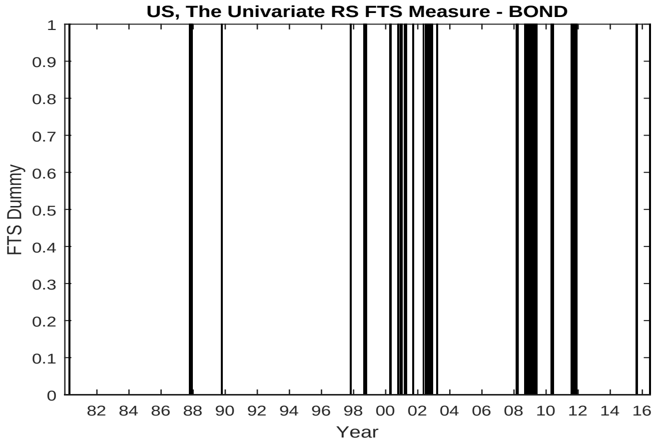
The figure displays the $\{0, 1\}$ FTS dummy according to the Ordinal Index model in which the US Treasury bonds act as safe haven. 5.692% of all days are classified as FTS days.

Figure 2.7 – The Ordinal FTS dummy - Gold



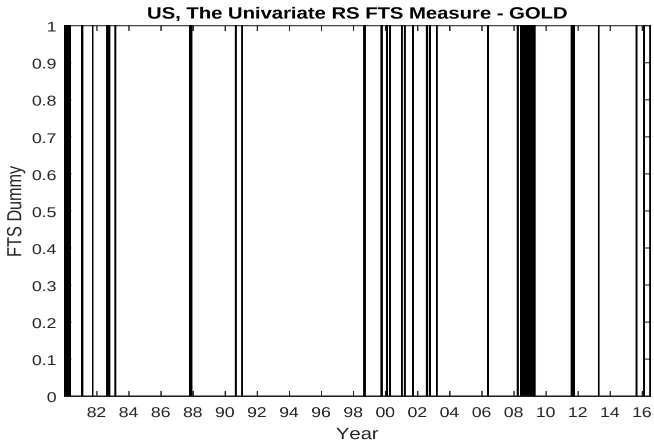
The figure displays the $\{0, 1\}$ FTS dummy according to the Ordinal Index model in which gold acts as safe haven. 6.700% of all days are classified as FTS days.

Figure 2.8 – The Univariate RS FTS dummy - US Treasury bonds



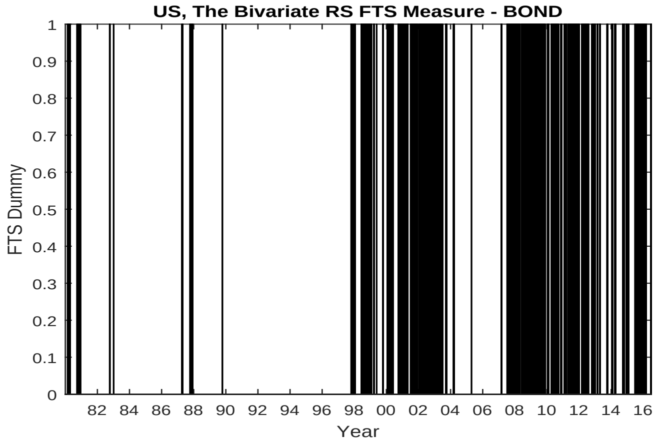
The figure displays the $\{0, 1\}$ FTS dummy according to the Univariate RS model in which the US Treasury bonds act as safe haven. 6.858% of all days are classified as FTS days.

Figure 2.9 – The Univariate RS FTS dummy - Gold



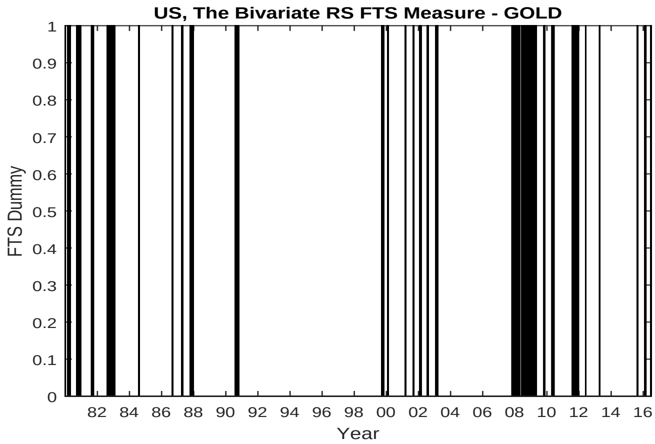
The figure displays the $\{0, 1\}$ FTS dummy according to the Univariate RS model in which gold acts as safe haven. 5.083% of all days are classified as FTS days.

Figure 2.10 – The Bivariate RS FTS dummy - US Treasury bonds



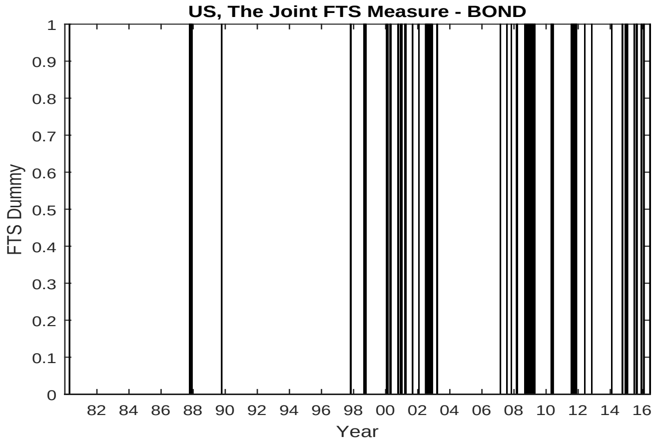
The figure displays the $\{0,1\}$ FTS dummy according to the Bivariate RS model in which the US Treasury bonds act as safe haven. 23.251% of all days are classified as FTS days.

Figure 2.11 – The Bivariate RS FTS dummy - Gold



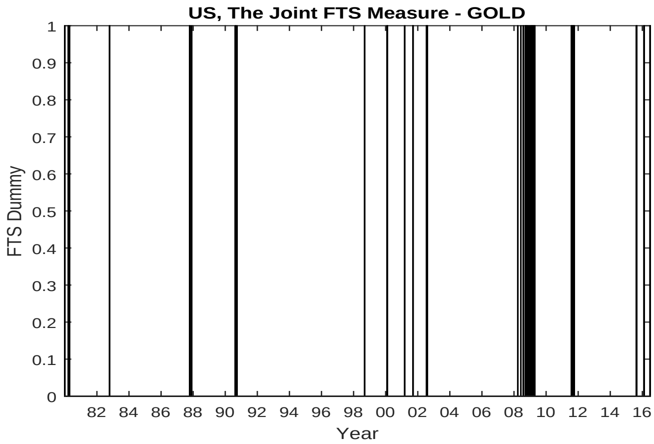
The figure displays the $\{0,1\}$ FTS dummy according to the Bivariate RS model in which gold acts as safe haven. 10.218% of all days are classified as FTS days.

Figure 2.12 – The Joint FTS dummy - US Treasury bonds



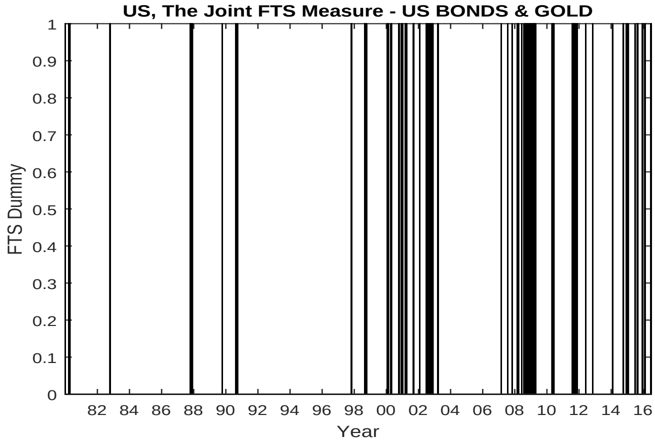
The figure displays the $\{0, 1\}$ FTS dummy that aggregates the 4 individual models in which the US Treasury bonds act as safe haven. 2.668% of all days are classified as FTS days.

Figure 2.13 – The Joint FTS dummy - Gold



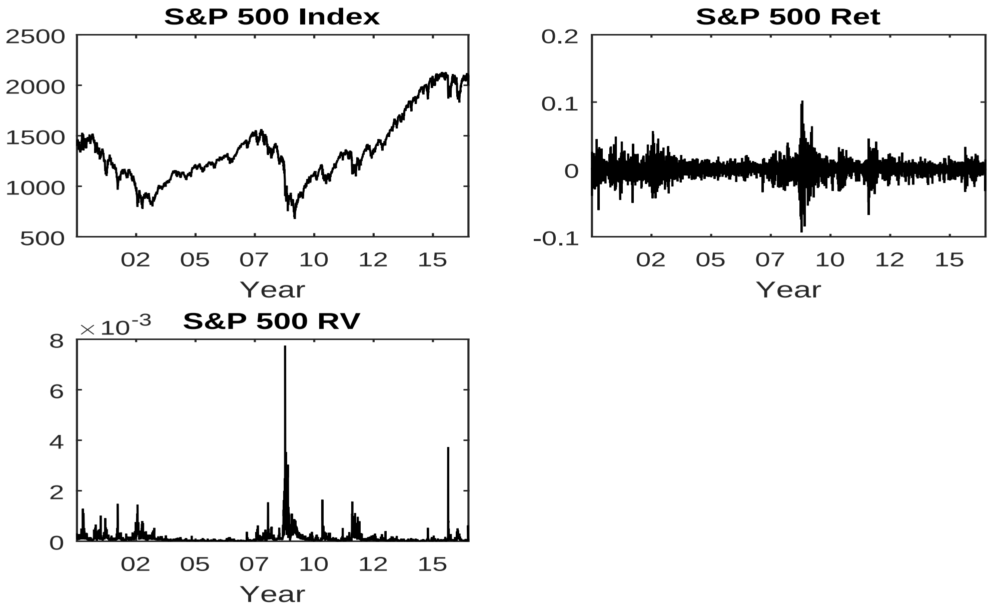
The figure displays the $\{0, 1\}$ FTS dummy that aggregates the 4 individual models in which gold acts as safe haven. 1.218% of all days are classified as FTS days.

Figure 2.14 – The Joint FTS dummy - US Treasury bonds & Gold



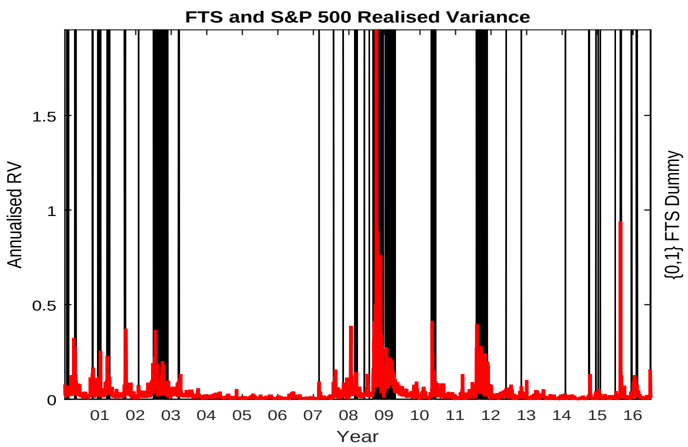
The figure displays the $\{0,1\}$ FTS dummy derived from the union of the Joint FTS dummies of US Treasury bonds and gold. 3.245% of all days are classified as FTS days.

Figure 2.15 – Time series of S&P 500 Index values, returns and RV



The figure displays the daily time series of the S&P 500 Index values (top left panel), returns (top right panel), and realised variances (bottom left panel) from January 2000 to June 2016. Realised variance is based on 5-min intraday returns. Source: OMI realised library.

Figure 2.16 – Time series of S&P 500 RV and FTS dummy



The figure displays the daily time series of the S&P 500 annualised realised variance (red line) and the US Treasuries & Gold $\{0, 1\}$ Joint FTS dummy (black line) lagged by one day.

2.8 Appendix A: FTS and Leverage effect

In this appendix we compare the standard Corsi (2009) model with several specifications in which the intercept and the coefficients of lagged daily, weekly and monthly variance can switch based on the leverage effect, the FTS effect or both effects. The FTS effect is captured by $\mathbb{1}_{t-1}^{FTS}$, which represents the daily lag of the $\{0, 1\}$ dummy that identifies FTS days according to the Ordinal model for US Treasuries and gold. The leverage effect is captured by $\mathbb{1}_{t-1}^{-}$, which represents the daily lag of the $\{0, 1\}$ dummy that identifies days of negative returns on the S&P 500 index. The Corsi (2009) model is specified as in equation (2.28) and is here denoted by M1. Models M2-M5 denote extended versions of the HAR-RV model in which all or some coefficients switch when equity returns are negative and/or flight-to-safety is taking place. In Table 2.24 we provide the in-sample parameter estimates, with robust standard errors in parentheses, for all model specifications considered. Table 2.24 also reports the AIC and BIC information criteria measures for the M1-M5 models, together with the results of two F-tests. We use $F\text{-test}^{(1)}$ to check whether each of the extended models gives a significantly better fit to the data than does M1. We use $F\text{-test}^{(2)}$ to check whether an extended HAR-RV model that accounts for both the leverage and FTS effect (M4) gives a significantly better fit to the data than does an extended HAR-RV model that only accounts for the leverage effect (M2).

From the results in Table 2.24, it is clear that the explanatory power of FTS for future equity return volatility is significant even after controlling for the leverage effect. In model M4, where all coefficients can switch with both the leverage and FTS effect, the intercept and the coefficient of lagged daily variance conditioned on FTS days are highly statistically significant, whilst no coefficient conditioned on the leverage effect is significant. Moreover, model M4 provides a significantly better fit to the data than do the standard Corsi (2009) model and its extended version that only accounts for the effect of negative equity returns. Lastly, we note that, although the results of $F\text{-test}^{(1)}$ are significant for all extended models, the information criteria clearly select a model that accounts for both effects. In particular, we find that the AIC information criterion selects a model where all coefficients switch with the two effects (M4), while on the basis of the BIC information criterion the HAR-RV model extended only for the effects on the intercept and the lagged daily variance (M5) is preferred.

Table 2.24 – In-sample estimates and goodness-of-fit tests

	M1	M2	M3	M4	M5
β_0	−0.236***	−0.373***	−0.470***	−0.406***	−0.503***
	(0.048)	(0.061)	(0.050)	(0.060)	(0.060)
β_1	0.324***	0.205***	0.270***	0.202***	0.257***
	(0.029)	(0.035)	(0.029)	(0.035)	(0.030)
β_2	0.432***	0.486***	0.411***	0.481***	0.456***
	(0.037)	(0.047)	(0.042)	(0.048)	(0.037)
β_3	0.188***	0.251***	0.225***	0.254***	0.198***
	(0.025)	(0.035)	(0.030)	(0.035)	(0.024)
β_0^-		0.255***		−0.046	0.097
		(0.089)		(0.101)	(0.088)
β_1^-		0.125***		0.049	−0.028
		(0.039)		(0.041)	(0.020)
β_2^-		−0.007		−0.038	
		(0.061)		(0.065)	
β_3^-		−0.121		−0.074	
		(0.056)		(0.056)	
β_0^{FTS}			0.384***	0.370**	0.525***
			(0.134)	(0.149)	(0.126)
β_1^{FTS}			0.163***	0.164***	0.092***
			(0.058)	(0.061)	(0.034)
β_2^{FTS}			−0.014	−0.033	
			(0.082)	(0.087)	
β_3^{FTS}			−0.120*	−0.080	
			(0.069)	(0.072)	
F-test$^{(1)}$		0.000	0.000	0.000	0.000
F-test$^{(2)}$				0.000	
AIC	7.401	7.186	7.263	7.138	7.150
BIC	7.426	7.237	7.313	7.213	7.201

Notes: This table provides the in-sample parameter estimates and standard errors (in brackets) for the HAR-RV model in its original specification (M1) and when augmented to account for the leverage and FTS effects on its coefficients (M2-M5). The estimated model equations are:

a) M1 : $RV_t^d = \beta_0 + \beta_1 RV_{t-1}^d + \beta_2 RV_{t-1}^w + \beta_3 RV_{t-1}^m + \epsilon_t$,
b) M2 : $\text{M1} + \mathbb{1}_{t-1}^-(\beta_0^- + \beta_1^- RV_{t-1}^d + \beta_2^- RV_{t-1}^w + \beta_3^- RV_{t-1}^m)$,
c) M3 : $\text{M1} + \mathbb{1}_{t-1}^{FTS}(\beta_0^{FTS} + \beta_1^{FTS} RV_{t-1}^d + \beta_2^{FTS} RV_{t-1}^w + \beta_3^{FTS} RV_{t-1}^m)$,
d) M4 : $\text{M2} + \mathbb{1}_{t-1}^{FTS}(\beta_0^{FTS} + \beta_1^{FTS} RV_{t-1}^d + \beta_2^{FTS} RV_{t-1}^w + \beta_3^{FTS} RV_{t-1}^m)$,
e) M5 : $\text{M1} + \mathbb{1}_{t-1}^-(\beta_0^- + \beta_1^- RV_{t-1}^d) + \mathbb{1}_{t-1}^{FTS}(\beta_0^{FTS} + \beta_1^{FTS} RV_{t-1}^d)$,

where $\mathbb{1}_{t-1}^{FTS}$ denotes the daily lag of the $\{0, 1\}$ dummy that identifies FTS days according to the Ordinal model for US Treasuries and gold, and $\mathbb{1}_{t-1}^-$ denotes the daily lag of the $\{0, 1\}$ dummy that identifies days of negative returns on the S&P 500 index. RV_t^d, RV_t^w and RV_t^m denote, respectively, the daily, weekly and monthly components of the log transformed realised variance of the S&P 500 returns. We use a window of 22 lags to compute the Newey-West robust standard errors of the models. We also report the AIC and BIC information criteria measures for the M1-M5 models, together with the results of two F-tests. We use F-test$^{(1)}$ to check whether each of the extended models gives a significantly better fit to the data than does M1. We use F-test$^{(2)}$ to check whether M4 gives a significantly better fit to the data than does M2.
* (resp. **, and ***) denote statistical significance at the 10% (resp. 5%, and 1%) level.

Chapter 3

Forecasting Volatility with a Semi-Covariance-based FTS measure

Abstract

We propose a proxy for flight-to-safety (FTS) based on the realised semi-covariance between falling equity and rising safe haven returns. The benefits of modelling the conditional distribution of asset returns with this FTS proxy in a multivariate and multi-period setting are considered. An empirical application based on US stock, US government bond and gold market data shows that our realised measure of semi-covariance is a statistically significant predictor of future equity volatility. Moreover, when comparing with a benchmark and controlling for data snooping, we find that the proposed model yields superior out-of-sample forecasts of the variance of both equity and safe haven returns. The more accurate forecasts are of economic importance for an investor facing portfolio decisions. However, the performance can be sensitive to the choice of the safe haven asset with the US Treasuries being the most effective in this regard.

JEL Codes: C22, C53, C58, G17.

Keywords: Flight to safety, realised semi-covariance, volatility forecasting.

3.1 Introduction

Flight-to-safety (FTS) may be defined as a market stress event characterised by heavy losses on the equity market and, simultaneously, positive returns on a safe haven asset, see Baele et al. (2015). FTS has previously been used to explain time-varying liquidity and risk premia within the context of asset pricing models, see Vayanos (2004), Chordia et al. (2005), Ghysels et al. (2016) and Adrian et al. (2018), among others. Moreover, the safe haven potential of specific assets has been extensively discussed in the literature. For instance, Hartmann et al. (2004) recognise the safe haven status of top-tier government bonds based on the infrequency of equity-bond co-crashes. More recently, Baur and Lucey (2010), Baur and McDermott (2010), Ciner et al. (2013) and Li and Lucey (2017) demonstrate that gold can also have a safe haven role.

In contrast to the above-mentioned studies, we focus on whether flight-to-safety can be used to forecast the conditional distribution of asset returns and, more specifically, their volatility. Anticipating future volatility is particularly useful for portfolio selection, derivatives pricing and risk management applications. The link between FTS and future equity market volatility has been motivated and described in the previous chapter of this series (see Section 2.1 and 2.2.3 in Chapter 2). In that study, we have identified FTS days via the methods proposed by Baele et al. (2015) and used the resulting dummy variable to augment the univariate HAR model of Corsi (2009). In the empirical analysis, significant improvements in the accuracy of the 1-day ahead forecasts of the realised variance of S&P 500 equity index returns were documented with the use of the $\{0, 1\}$ FTS dummy. Inspired by these results, we propose here a multivariate framework that relies on a high-frequency measure of semi-covariance to track the likelihood of FTS events and make multi-step ahead predictions of return volatility. Hence, the main contribution of this chapter is to address whether the realised semi-covariance of falling equity and a rising safe haven may be used to more accurately model financial market volatility dynamics.

The use of realised measures, i.e. estimators that are computed from high-frequency data, is common practice in the context of forecasting volatility as they are highly informative about the current level of risk. Indeed, several applications have shown that more accurate forecasts can be obtained with realised measures of variance, and more generally, with intraday data, see, e.g., Andersen et al. (2003), Ghysels et al. (2006), Andersen et al. (2007), Corsi (2009), Shephard and Sheppard (2010) and Hansen et al. (2012). More recently, Bollerslev et al. (2017) add realised semi-covariance to the context of high-frequency data. In their empirical analysis based on vector HAR-type models, Bollerslev et al. (2017) find that lagged semi-covariances computed for more than ten stocks have strong predictive power for the total covariance matrix and aggregate portfolio volatility. They explain the superior forecasting performance of models that incorporate realised semi-covariance in terms of time-varying parameters which allow the model-implied volatility to adjust more quickly following a shock. We add to this literature by proposing the realised semi-covariance between negative equity returns and positive safe haven returns as a proxy measure of flight-to-safety and a direct predictor of equity return variance.[1] After the explanatory power of the FTS measure is tested across a range of univariate volatility models, we examine its contribution in a multivariate generalized autoregressive conditional heteroskedasticity (GARCH) model. To do this, we extend the model of Hansen et al. (2014) by including the daily lag of the FTS measure in the conditional variance equation of equity returns. Hence, we name the proposed framework: Realised FTS GARCH. In order to allow for multi-step-ahead volatility forecasts, dynamics must be imposed on the new FTS measure. We follow the results of Bollerslev et al. (2017) for mixed-sign semi-covariances and choose an autoregressive structure composed of daily, weekly and monthly lagged components.

Our empirical analysis relies on high-frequency data for the S&P 500 index futures, the 10-Year US Treasury Note futures and the gold futures over the period 2000 to 2016. We consider the safe haven property of bonds separately from that of gold. Indeed, we apply the Realised FTS GARCH model first to equity and bonds and then to equity and gold. Using the same data we also estimate the Realized Beta GARCH model proposed by Hansen et al. (2014), referred to as the benchmark model henceforth. To answer the question of whether flight-to-safety can yield significant gains in forecasting volatility, we draw on the existing literature of forecast evaluation. We consider mean squared error (MSE) and quasi-likelihood (QLIKE) loss functions as measures of performance as they yield robust rankings of models in the presence of measurement error in the realised variance measures (Patton, 2011). When testing predictive ability we control for data mining issues by using the Reality Check test formulated by White

[1]Throughout the rest of the chapter, we refer to this realised semi-covariance as the FTS measure.

(2000). Overall, the results show that our proxy of FTS is a strong predictor of future equity volatility. Moreover, we find that the realised semi-covariance of falling equity and a rising safe haven can significantly improve the out-of-sample forecasts of both equity and safe haven volatility, especially for long horizons. The gains of predictability are not only statistically but also economically significant for investors facing portfolio decisions. Finally, we observe that the benefits of introducing FTS information depend on the choice of the safe haven. Of the assets that we consider it is US Treasuries that, when used to construct the FTS measure, yield the most significant improvements in forecasting and volatility-timing performance.

The remainder of the chapter is structured as follows. In Section 3.2 we introduce the FTS measure and conduct preliminary analysis in the univariate volatility framework. In Section 3.3 we present the multivariate model and discuss its estimation. Section 3.4 covers the empirical application. Section 3.5 concludes.

3.2 Realised Semi-Covariance and Equity Volatility Dynamics

To construct our FTS measure, we build on the definition of a flight-to-safety event by Baele et al. (2015) as “a day on which bond returns are positive, equity returns are negative, the stock bond return correlation is negative, and there is market stress as reflected in elevated equity return volatility”. In view of earlier studies indicating that gold is also a safe haven with respect to the stock market, we consider the precious metal an alternative asset to government bonds when constructing the FTS measure. Furthermore, unlike Baele et al. (2015), where the highest frequency of the data is daily, we exploit the superior informative content of intraday data to proxy the likelihood of FTS events. To this end, we collect intraday returns to equity ($r_{0,i}$) and to either safe haven ($r_{1,i}$), and we compute realised semi-covariance as follows:

$$m_t^- = \sum_{i=1}^{n} \left(r_{0,i} \mathbb{1}_{r_{0,i}<0} \right) \times \left(r_{1,i} \mathbb{1}_{r_{1,i}>0} \right), \tag{3.1}$$

where n refers to the total number of high-frequency observations on day t. We propose the realised semi-covariance of equation (3.1) as the FTS measure of this study because its (absolute) value is closely related to the symptoms of flight-to-safety described by Baele et al. (2015).

Indeed, large and negative values of m_t^- occur in the presence of frequent and extreme negative stock returns and positive safe haven returns, which relate to a high probability of an FTS event. Moreover, we find a noticeable (absolute) correlation of 0.44 (0.39) between the realised semi-covariance of equation (3.1), computed with intraday returns to equity and US Treasuries (gold), and the $\{0, 1\}$ dummy that was defined in Chapter 2 and identifies FTS days according to the Joint model for US Treasuries and gold.

Now that we have defined our proxy measure of flight-to-safety, we provide evidence of its explanatory power for future equity volatility. In doing this preliminary analysis, we let the US Treasuries be the safe haven in equation (3.1) and we make use of a number of simple yet successful univariate forecasting methods to explain equity return variance.[2] All the results that we present in this section are based on the full sample of US stock and bond market data described in Section 3.4.1. The volatility forecasting methods that we consider are the heterogeneous autoregressive (HAR) model proposed by Corsi (2009), the classic GARCH (1,1) model of Bollerslev (1986) and the Integrated HEAVY models introduced by Shephard and Sheppard (2010). The models are estimated in their original specification and when augmented with the daily lag of the FTS measure. Therefore, the effect of flight-to-safety on future equity volatility is captured by the coefficient on this measure (denoted γ) in the aforementioned models. The estimated coefficients and robust standard errors are provided in Table 3.1.

[Insert Table 3.1 here]

As can be seen from an examination of Table 3.1, our FTS measure has high predictive power for the equity return variation. Indeed, the estimated coefficient γ is statistically significant at the 5% level or lower across all volatility models considered here. Furthermore, we note that the sign of the coefficient is negative. Therefore, the models indicate that, when the realised semi-covariance in (3.1) becomes larger in absolute value, and consequently the FTS signal is stronger, the volatility of the stock market tends to increase the next day.

[2] In results available upon request, we find that the safe haven property of gold leads to similar conclusions to those drawn here using US Treasuries.

This preliminary analysis has revealed that realised semi-covariance between falling equity and a rising safe haven might be a valuable addition to a volatility forecasting model. It has also provided indirect evidence for a relationship between the occurrence of flight-to-safety events and stock market volatility dynamics, supporting the conclusions drawn in Chapter 2. In the next section, we extend the study to a multivariate framework, in which our FTS measure can impact the future return variation of multiple assets. By means of a more sophisticated model that controls for several stylised facts of volatility, we will be able to explore in greater detail the benefits of using flight-to-safety in forecasting the volatility dynamics for both stock and safe haven returns.

3.3 Realised FTS GARCH

In this section we introduce the Realised FTS GARCH model as an extension to the Realized Beta GARCH model proposed by Hansen et al. (2014). The methodological novelty of this work is that realised semi-covariance is attached to the framework to forecast, with flight-to-safety information, the conditional distribution of a vector of returns. We present the model in the two-dimension case, though extending to higher dimensions is a trivial extension.

3.3.1 Notation

The Realised FTS GARCH model is a multivariate GARCH model that uses realised measures of volatility, correlation and semi-covariance. In this study, the multivariate realised kernel by Barndorff-Nielsen et al. (2011) represents the realised measures of volatility and correlation. This class of estimators, which can be used in the presence of non-synchronous trading, is robust to market microstructure noise and the Epps (1979) effect. For other possible approaches to estimate the integrated covariance of two assets, see, amongst others, Hayashi and Yoshida (2005), Aït-Sahalia et al. (2010) and Zhang (2011).[3] As for the realised measure of semi-covariance, we use m_t^- from equation (3.1), which corresponds to the off-diagonal elements of the 'mixed' semi-covariance matrix, $\mathcal{M}_t^-$, in Bollerslev et al. (2017). In order to balance efficiency and biases

[3]See, Aït-Sahalia et al. (2005), Zhang et al. (2005), Barndorff-Nielsen et al. (2008), Barndorff-Nielsen et al. (2009) and Aït-Sahalia et al. (2011) for the univariate case of volatility estimation.

associated with the Epps (1979) effects, we follow Bollerslev et al. (2017) and use a 15-minute sampling frequency to construct m_t^- from falling equity and rising safe haven returns.

Most variables presented thus far are constrained to lie in a particular range of values (e.g., correlation can only take values inside $(-1,1)$ and our semi-covariance measure is forced to be non-positive). To ease the impact of this restriction, we follow Hansen et al. (2012) and employ the following transformations of the latent conditional variance (h_t), latent correlation (ρ_t), the observed realised variance (x_t), realised correlation (y_t) and realised semi-covariance (m_t^-):

$$\begin{aligned}
\tilde{h}_{0,t} &= \log(h_{0,t}), \\
\tilde{h}_{1,t} &= \log(h_{1,t}), \\
\tilde{\rho}_{1,t} &= F(\rho_{1,t}) = \frac{1}{2}\log\left(\frac{1+\rho_{1,t}}{1-\rho_{1,t}}\right), \\
\tilde{m}_t^- &= \log(-m_t^-),
\end{aligned}
\qquad
\begin{aligned}
\tilde{x}_{0,t} &= \log(x_{0,t}), \\
\tilde{x}_{1,t} &= \log(x_{1,t}), \\
\tilde{y}_{1,t} &= F(y_{1,t}) = \frac{1}{2}\log\left(\frac{1+y_{1,t}}{1-y_{1,t}}\right),
\end{aligned}$$

where $r_{0,t}$ is the daily return on the equity market, $r_{1,t}$ is the daily return on the safe haven asset, $x_{0,t}$ is the realised volatility of the equity return, $x_{1,t}$ is the realised volatility of the safe haven return, $h_{0,t}$ is the latent volatility of the equity return (conditional on past information), $h_{1,t}$ is the latent volatility of the safe haven return (conditional on past information), $y_{1,t}$ is the realised correlation of asset returns, $\rho_{1,t}$ is the latent correlation of asset returns (conditional on past information), m_t^- is the realised semi-covariance between falling equity and rising safe haven returns, and F is the Fisher transformation that offers a 1-to-1 mapping from $(-1,1)$ into $\mathbb{R}$.

3.3.2 Model Specification

The model we analyse is a development of that by Hansen et al. (2014), thus the reader should consult their paper for a more detailed description of the econometric techniques. Here, we limit ourselves to highlighting some of the basic ideas behind its design. The first distinctive feature of the Realized Beta GARCH model is that the dynamic modelling of realised volatility and correlation is based on a “measurement equation”. This approach has been put forward by Hansen et al. (2012) and ties the realised measures to their latent equivalent. For instance, realised variance at time t is just a linear function of the contemporaneous conditional variance.[4] The second characteristic feature of the multivariate GARCH model proposed by Hansen et al. (2014) is its hierarchical structure. The authors propose the 'market' as the main driver of the correlation structure. In doing so, only the realised correlations between each asset return and the 'market' return are needed as these measures also capture variation in the correlations between the individual asset returns. The hierarchical structure implies that the 'market' is modelled in the univariate setting of a realised EGARCH model by Hansen and Huang (2016) while the model of the individual assets is conditional not only on the past but also on contemporary 'market' variables. The authors refer to the former as the "marginal model" and the latter as the "conditional model".

We now proceed to the specification of the Realised FTS GARCH model. The measurement equations and the factor structure of volatility, which includes a spillover effect from the market, are preserved in our version of the model. However, the FTS measure that we introduce to forecast the conditional variance of equity returns is not modelled like realised volatility and correlation as we do not have a latent measure available for semi-covariance. Therefore, we choose for its dynamics an autoregressive structure composed of daily, weekly and monthly lagged components. This is in line with the findings reported by Bollerslev et al. (2017), who have shown that the semi-covariance derived from returns with opposite sign has long memory and its main drivers are its own lags.

We impose the following dynamics on equity returns and variances:

$$r_{0,t} = \sqrt{h_{0,t}}\; z_{0,t}, \tag{3.2a}$$

$$\tilde{h}_{0,t} = a_0 + b_0\tilde{h}_{0,t-1} + c_0\tilde{x}_{0,t-1} + g(z_{0,t-1};\tau_{01},\tau_{02}) + e_0 m^{-}_{t-1}, \tag{3.2b}$$

$$\tilde{x}_{0,t} = \xi_0 + \varphi_0\tilde{h}_{0,t} + g(z_{0,t};\delta_{01},\delta_{02}) + u_{0,t}. \tag{3.2c}$$

Here, the leverage function $g(.)$ allows volatility to increase more after a negative (positive)

[4] An alternative approach to modelling realised measures of volatility that allow multi-period ahead forecasting is described in Engle and Gallo (2006) and Shephard and Sheppard (2010).

shock to returns than a positive (negative) one, and is given by

$$g(z_{0,t};\tau_{01},\tau_{02}) = \tau_{01}z_{0,t} + \tau_{02}(z_{0,t}^2 - 1),$$

where $z_{0,t}$ is the standardised return given by $r_{0,t}/\sqrt{h_{0,t}}$. In a similar vein, safe haven returns and variances are given by

$$r_{1,t} = \sqrt{h_{1,t}}\, z_{1,t}, \tag{3.3a}$$

$$\tilde{h}_{1,t} = a_1 + b_1\tilde{h}_{1,t-1} + c_1\tilde{x}_{1,t-1} + d_1\tilde{h}_{0,t} + g(z_{1,t-1};\tau_{11},\tau_{12}), \tag{3.3b}$$

$$\tilde{x}_{1,t} = \xi_1 + \varphi_1\tilde{h}_{1,t} + g(z_{1,t};\delta_{11},\delta_{12}) + u_{1,t}. \tag{3.3c}$$

Finally, correlations and semi-covariance are assumed to evolve as follows:

$$\tilde{\rho}_{1,t} = a_{10} + b_{10}\tilde{\rho}_{1,t-1} + c_{10}\tilde{y}_{1,t-1}, \tag{3.4a}$$

$$\tilde{y}_{1,t} = \xi_{10} + \varphi_{10}\tilde{\rho}_{1,t} + v_{1,t}, \tag{3.4b}$$

$$\tilde{m}_t^- = \zeta + \phi_d\tilde{m}_{t-1}^- + \phi_w\tilde{m}_{t-1|t-5}^- + \phi_m\tilde{m}_{t-1|t-22}^- + \psi_0 z_{0,t} + \psi_1 z_{1,t} + w_{1,t}. \tag{3.4c}$$

Without the last covariate in equation (3.2b) and its dynamics illustrated in equation (3.4c), the framework reduces to the Realized Beta GARCH model proposed by Hansen et al. (2014).

In equation (3.4c), $\tilde{m}_{t-1}^-$, $\tilde{m}_{t-1|t-5}^-$ and $\tilde{m}_{t-1|t-22}^-$ represent the daily, weekly and monthly lag of the modified semi-covariance and are defined as follows:

$$\tilde{m}_{t-1}^- = \log(-m_{t-1}^-),$$

$$\tilde{m}_{t-1|t-5}^- = \frac{\sum_{j=1}^{5}\tilde{m}_{t-j}^-}{5},$$

$$\tilde{m}_{t-1|t-22}^- = \frac{\sum_{j=1}^{22}\tilde{m}_{t-j}^-}{22}.$$

The standard HAR model of Corsi (2009) that we adopt for realised semi-covariance in equation (3.4c) is augmented by the contemporaneous effect of the standardised returns on equity and the safe haven. In doing so, we capture the dependence between shocks to the asset returns and shocks to their semi-covariance with $(\psi_0 z_{0,t}, \psi_1 z_{1,t})$ and assume that $w_{1,t}$ is independent of $(z_{0,t}, z_{1,t})$.

As for the error terms in equation (3.2a) to (3.4c), we assume the following:

Assumption 1. *The error components are independent and identically distributed (i.i.d.) normal random variables:* $z_{0,t} \sim i.i.d.\ \mathcal{N}(0,1)$, $u_{0,t} \sim i.i.d.\ \mathcal{N}(0,\sigma^2_{u_0})$, $z_{1,t} \sim i.i.d.\ \mathcal{N}(0,1)$, $u_{1,t} \sim i.i.d.\ \mathcal{N}(0,\sigma^2_{u_1})$, $v_{1,t} \sim i.i.d.\ \mathcal{N}(0,\sigma^2_{v_1})$ *and* $w_{1,t} \sim i.i.d.\ \mathcal{N}(0,\sigma^2_{w_1})$.

The Gaussian specification for standardised returns and volatility measurement errors is supported by Andersen et al. (2001) and Andersen et al. (2003) who find that standardised returns and realised volatilities are, respectively, normal and log-normal.

Assumption 2. *The standardised returns,* $(z_{0,t}, z_{1,t})$*, are correlated between each other with covariance (and correlation)* $\rho_{1,t} = corr(r_{0,t}, r_{1,t}) = \frac{cov(\sqrt{h_{0,t}}\ z_{0,t}\ ,\ \sqrt{h_{1,t}}\ z_{1,t})}{\sqrt{h_{0,t}}\ \sqrt{h_{1,t}}} = cov(z_{0,t}, z_{1,t})$.

Assumption 3. *The standardised returns,* $(z_{0,t}, z_{1,t})$*, are independent of the error terms in the measurement equations of volatility and correlation,* $(u_{0,t}, u_{1,t}, v_{1,t})$.

Hansen et al. (2012) show that the independence between z_t and u_t can be met by use of the log-linear specification with a leverage function, which is what we consider in this chapter.

Assumption 4. *The three measurement errors,* $(u_{0,t}, u_{1,t}, v_{1,t})$*, are correlated between each other with variance-covariance matrix* Σ.

In their study, Hansen et al. (2014) find significant correlation across all measurement errors.

Assumption 5. *The error in the dynamics of realised semi-covariance,* $w_{1,t}$*, is correlated with the measurement errors* $(u_{0,t}, u_{1,t}, v_{1,t})$ *but independent of the standardised returns* $(z_{0,t}, z_{1,t})$.

While Assumption 1-4 are the same as in Hansen et al. (2014), Assumption 5 is new. As mentioned previously, the inclusion of $(z_{0,t}, z_{1,t})$ in the equation of $\tilde{m}_t^-$ should leave no unmodelled dependence between shocks to the asset returns and shocks to their semi-covariance. However, we allow $w_{1,t}$ to be correlated with the measurement errors of volatility and correlation.

3.3.3 Estimation

We denote by θ the full set of parameters, and by θ_0, θ_1, θ_2 and θ_3 the following subsets,

$$\begin{aligned}
\theta &= (\theta_0', \theta_1', \theta_2', \theta_3')', \\
\theta_0 &= (a_0, b_0, c_0, \tau_{01}, \tau_{02}, e_0, \xi_0, \varphi_0, \delta_{01}, \delta_{02}, h_{0,1})', \\
\theta_1 &= (a_1, b_1, c_1, d_1, \tau_{11}, \tau_{12}, \xi_1, \varphi_1, \delta_{11}, \delta_{12}, h_{1,1})', \\
\theta_2 &= (a_{10}, b_{10}, c_{10}, \xi_{10}, \varphi_{10}, \rho_{11})', \\
\theta_3 &= (\zeta, \phi_d, \phi_w, \phi_m, \psi_0, \psi_1)'.
\end{aligned}$$

The model parameters are estimated such that the joint likelihood function, $\mathcal{L}$, or, equivalently, the joint log-likelihood function, L, is maximised:

$$L(\theta) = \log \mathcal{L}(\theta) = \sum_{t=1}^{T} \log f(r_{0,t}, x_{0,t}, r_{1,t}, x_{1,t}, y_{1,t}, m_t^- \mid \mathcal{F}_{t-1}). \tag{3.5}$$

For the purpose of estimation, we decompose the time-t joint density, conditional on past information $\mathcal{F}_{t-1}$, as follows,

$$\begin{aligned}
f(r_{0,t}, x_{0,t}, r_{1,t}, x_{1,t}, y_{1,t}, m_t^- \mid \mathcal{F}_{t-1}) &= f(r_{0,t}, x_{0,t} \mid \mathcal{F}_{t-1}) \times f(r_{1,t}, x_{1,t}, y_{1,t} \mid r_{0,t}, x_{0,t}, \mathcal{F}_{t-1}) \\
&\quad \times f(m_t^- \mid r_{0,t}, x_{0,t}, r_{1,t}, x_{1,t}, y_{1,t}, \mathcal{F}_{t-1}). \qquad (3.6)
\end{aligned}$$

The first two terms on the right side of equation (3.6) represent, respectively, the marginal and conditional model of Hansen et al. (2014). Apart from the lagged realised semi-covariance that appears as an additional explanatory variable in $\tilde{h}_{0,t}$, the Gaussian specification and further decomposition of $f(r_{0,t}, x_{0,t} \mid \mathcal{F}_{t-1})$ and $f(r_{1,t}, x_{1,t}, y_{1,t} \mid r_{0,t}, x_{0,t}, \mathcal{F}_{t-1})$ are the same as in Hansen et al. (2014),

$$f(r_{0,t}, x_{0,t} \mid \mathcal{F}_{t-1}) = f_{r_0}(r_{0,t} \mid \mathcal{F}_{t-1}) \times f_{x_0|r_0}(x_{0,t} \mid r_{0,t}, \mathcal{F}_{t-1}), \tag{3.7}$$

$$f(r_{1,t}, x_{1,t}, y_{1,t} \mid r_{0,t}, x_{0,t}, \mathcal{F}_{t-1}) = f_{r_1|r_0,x_0}(r_{1,t} \mid r_{0,t}, x_{0,t}, \mathcal{F}_{t-1})$$
$$\times f_{x_1,y_1|r_1,r_0,x_0}(x_{1,t}, y_{1,t} \mid r_{1,t}, r_{0,t}, x_{0,t}, \mathcal{F}_{t-1}). \quad (3.8)$$

The two densities on the RHS of equation (3.7), which are modelled with $r_{0,t} \sim \mathcal{N}(0, h_{0,t})$ and $\log x_{0,t} \sim \mathcal{N}(\xi_0 + \varphi_0 \log h_{0,t} + g(z_{0,t}; \delta_{01}, \delta_{02}), \sigma^2_{u_0})$, contribute to $L(\theta)$ in equation (3.5) with, respectively, ℓ_{z_0} and ℓ_{u_0},

$$\ell_{z_0} = -\frac{1}{2}\sum_{t=1}^{T} \log h_{0,t} + \frac{r_{0,t}^2}{h_{0,t}} = -\frac{1}{2}\sum_{t=1}^{T} \log h_{0,t} + z_{0,t}^2, \quad (3.9)$$

$$\ell_{u_0} = -\frac{1}{2}\sum_{t=1}^{T} \log \sigma^2_{u_0} + \frac{\overbrace{(\log x_{0,t} - \xi_0 - \varphi_0 \log h_{0,t} - g(z_{0,t}; \delta_{01}, \delta_{02}))^2}^{u_{0,t}^2}}{\sigma^2_{u_0}}. \quad (3.10)$$

The conditional density $f_{r_1|r_0,x_0}(r_{1,t} \mid r_{0,t}, x_{0,t}, \mathcal{F}_{t-1})$ on the right side of equation (3.8) is modelled with $r_{1,t} \mid z_{0,t} \sim \mathcal{N}(\rho_{1,t}\sqrt{h_{1,t}}z_{0,t}, (1-\rho^2_{1,t})h_{1,t})$. Its contribution to the log-likelihood function in equation (3.5) is $\ell_{z_1|z_0}$,

$$\ell_{z_1|z_0} = -\frac{1}{2}\sum_{t=1}^{T} \log[(1-\rho^2_{1,t})h_{1,t}] + \frac{(r_{1,t} - \rho_{1,t}\sqrt{h_{1,t}}z_{0,t})^2}{(1-\rho^2_{1,t})h_{1,t}}. \quad (3.11)$$

The conditional density $f_{x_1,y_1|r_1,r_0,x_0}(x_{1,t}, y_{1,t} \mid r_{1,t}, r_{0,t}, x_{0,t}, \mathcal{F}_{t-1})$ results from the decomposition in equation (3.8) and is modelled with $u_{1,t}, v_{1,t} \mid u_{0,t} \sim \mathcal{N}(\boldsymbol{\mu}_t, \boldsymbol{\Omega})$. The multivariate Normal distribution of $(u_{1,t},\ v_{1,t})$ is conditional only on $u_{0,t}$ due to the assumption of independence from $(z_{0,t}, z_{1,t})$. The mean and variance of this distribution are defined as follows,

$$\boldsymbol{\mu}_t = \begin{pmatrix} \sigma_{u_1,u_0}/\sigma^2_{u_0} \\ \sigma_{v_1,u_0}/\sigma^2_{u_0} \end{pmatrix} u_{0,t},$$
$$\boldsymbol{\Omega} = \begin{bmatrix} \sigma^2_{u_1} & \sigma_{u_1,v_1} \\ \bullet & \sigma^2_{v_1} \end{bmatrix} - \begin{bmatrix} \sigma_{u_1,u_0} \\ \sigma_{v_1,u_0} \end{bmatrix} \frac{1}{\sigma^2_{u_0}} \begin{bmatrix} \sigma_{u_0,u_1} & \sigma_{u_0,v_1} \end{bmatrix},$$

The contribution of $f_{x_1,y_1|r_1,r_0,x_0}(x_{1,t}, y_{1,t} \mid r_{1,t}, r_{0,t}, x_{0,t}, \mathcal{F}_{t-1})$ to the log-likelihood function in equation (3.5) is measured by $\ell_{u_1,v_1|u_0}$,

$$\ell_{u_1,v_1|u_0} = -\frac{1}{2}\sum_{t=1}^{T} \log\det \boldsymbol{\Omega} + \bar{\boldsymbol{\mu}}_t' \boldsymbol{\Omega}^{-1} \bar{\boldsymbol{\mu}}_t, \tag{3.12}$$

where $\bar{\boldsymbol{\mu}}_t = \begin{pmatrix} u_{1,t} \\ v_{1,t} \end{pmatrix} - \boldsymbol{\mu}_t$.

The last term on the RHS of equation (3.6), $f(m_t^- \mid r_{0,t}, x_{0,t}, r_{1,t}, x_{1,t}, y_{1,t}, \mathcal{F}_{t-1})$, is the conditional density that relates to the dynamics of realised semi-covariance for which we propose a variation of the HAR model. The assumption of independence of $w_{1,t}$ from $(z_{0,t}, z_{1,t})$ implies that the conditioning on $(r_{0,t}, r_{1,t})$ is captured by $(\psi_0 z_{0,t}, \psi_1 z_{1,t})$. The conditional density is therefore modelled with $w_{1,t} \mid u_{0,t}, u_{1,t}, v_{1,t} \sim \mathcal{N}(\eta_t, \Gamma)$, where the mean and variance of the Normal distribution are defined as follows,

$$\eta_t = \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \begin{bmatrix} u_{0,t} \\ u_{1,t} \\ v_{1,t} \end{bmatrix}, \qquad \Gamma = \sigma_{w_1}^2 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21},$$

where we have defined

$$\boldsymbol{\Sigma}_{12} = \begin{bmatrix} \sigma_{w_1,u_0} & \sigma_{w_1,u_1} & \sigma_{w_1,v_1} \end{bmatrix}, \qquad \boldsymbol{\Sigma}_{22} = \begin{bmatrix} \sigma_{u_0}^2 & \sigma_{u_1,u_0} & \sigma_{v_1,u_0}, \\ \bullet & \sigma_{u_1}^2 & \sigma_{u_1,v_1} \\ \bullet & \bullet & \sigma_{v_1}^2 \end{bmatrix}, \qquad \boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}_{12}^{\top}.$$

The conditional density $f(m_t^- \mid r_{0,t}, x_{0,t}, r_{1,t}, x_{1,t}, y_{1,t}, \mathcal{F}_{t-1})$ contributes to the log-likelihood function in (3.5) with $\ell_{w_1|u_0,u_1,v_1}$,

$$\ell_{w_1|u_0,u_1,v_1} = -\frac{1}{2}\sum_{t=1}^{T} \log\det\Gamma + \bar{\eta}_t'\Gamma^{-1}\bar{\eta}_t = -\frac{1}{2}\sum_{t=1}^{T} \log\Gamma + \frac{\bar{\eta}_t^2}{\Gamma}, \tag{3.13}$$

where $\bar{\eta}_t = w_{1,t} - \eta_t$. In order to reduce the number of free parameters to estimate with maximum likelihood, we follow Hansen et al. (2014) and we compute $\sigma_{u_0}^2$, σ_{u_1,u_0}, σ_{v_1,u_0}, $\sigma_{u_1}^2$, σ_{u_1,v_1}, σ_{w_1,u_1}, $\sigma_{v_1}^2$, σ_{w_1,v_1}, σ_{w_1,u_0}, $\bar{\boldsymbol{\mu}}_t$, $\bar{\eta}_t$, $\boldsymbol{\Omega}$ and Γ using the residuals of the measurement equations and modified

HAR equation, i.e. $\hat{u}_{0,t}$, $\hat{u}_{1,t}$, $\hat{v}_{1,t}$ and $\hat{w}_{1,t}$.

The parameter vector θ is then estimated as the vector value that maximises the log-likelihood function $L(\theta)$,

$$L(\theta) = \ell_{z_0}(\theta_0) + \ell_{u_0}(\theta_0) + \ell_{z_1|z_0}(\theta_1, \theta_2) + \ell_{u_1,v_1|u_0}(\theta_1, \theta_2) + \ell_{w_1|u_0,u_1,v_1}(\theta_3). \tag{3.14}$$

In order to account for possible misspecification of the probability density, we compute quasi-maximum likelihood standard errors of the estimated parameter vector $\hat{\theta}$ based on the following approximation of the variance-covariance matrix,

$$E(\hat{\theta} - \theta)(\hat{\theta} - \theta)' \cong T^{-1}\{\mathcal{J}_{2D}\mathcal{J}_{OP}^{-1}\mathcal{J}_{2D}\}^{-1},$$

where $\mathcal{J}_{2D}$ and $\mathcal{J}_{OP}$ denote, respectively, the second-derivative estimate and the outer-product estimate of the information matrix. See Hamilton (1994) for formulas and details on how to calculate these measures.

3.4 Empirical Application

An application of the Realised FTS GARCH model to data is provided in this section. We present empirical results using the US stock and US Treasury markets in one instance and the US stock and gold markets in another. We discuss the results relative to the benchmark model by Hansen et al. (2014). Models are estimated using open-to-close returns to preserve consistency with the time frame over which high-frequency prices are available. We report the estimation results using the full sample of data and claim that our FTS measure is a good predictor of equity return volatility. We further examine the quality of volatility forecasts up to one month ahead in- and out-of-sample. Finally, we conduct a series of robustness checks and evaluate the economic implications of our findings.

3.4.1 Data Description

We consider data for the S&P 500 index futures (SP), the 10-Year US Treasury Note futures (TY) and the gold futures (GC) traded on a 24-hour clock via the Chicago Mercantile

Exchange (CME) GLOBEX trading platform. The dataset covers the tick-by-tick transaction prices recorded in Central Time from January 1, 2000 to October 31, 2016. Cleaned data is provided by Tick Data, Inc. High-frequency prices from both pit and electronic trading are used, therefore we define the day based on the electronic trading system: each day starts at 5 p.m. on the previous day and ends at 4 p.m. on the current day. For each asset, we take the average price when multiple transactions have the same time stamp. Before constructing our realised measures, we remove from the sample half trading days and other days where the prices did not fully span the official trading hours. These exclusions leave us with 4,259 complete trading days for which we apply the 'refresh time' scheme proposed by Barndorff-Nielsen et al. (2011) to synchronise the high-frequency prices of the three assets.

As mentioned above, realised volatility and correlation in this study are based on the multivariate realised kernel introduced by Barndorff-Nielsen et al. (2011). This technique makes use of all the synchronised prices. In contrast, estimation of realised semi-covariance is based on only a smaller part of the available data since it is the sum of the products of mixed sign intraday returns sampled every 15 minutes, as in Bollerslev et al. (2017). We multiply realised variances and semi-covariances by 10,000 to be on the same scale as the daily log-returns expressed in percentages. When estimating the Realised FTS GARCH model, we deal with zero values of realised semi-covariance in $\log(-m_t^-)$ by means of $\min(m_t^-, -10^{-5})$. Figure 3.1 displays the realised variances and correlations of the three markets and the realised semi-covariances between equity and either safe haven. The descriptive statistics for open-to-close returns and realised measures are summarised in Table 3.2.

[Insert Figure 3.1 here]

[Insert Table 3.2 here]

Consistent with previous findings, Table 3.2 shows that daily stock and gold returns have mean close to zero, are skewed to the left, and are highly leptokurtic. The same properties apply to daily Treasury bond returns, with the exception of a slightly positive instead of negative skewness. Examining the summary statistics for the realised daily variances, we observe an average annualised volatility of approximately 18%, 6% and 17% in, respectively, the US stock, US bond and gold market. The distributions of these realised variances have fatter tails than those predicted by the normal distribution, and are bounded by zero with a long right tail. With regard to the interdependence between equity and the two safe haven assets, we note that on average the realised correlation is negative with US Treasury bonds and positive with gold. However, we find that the realised semi-covariance between the falling equity market and the rising gold market has a deeper minimum and a more negative mean value than that between the falling equity market and the rising bond market. Very high kurtosis is evident in both realised semi-covariances. Finally, it appears that the realised correlation between the two safe haven assets is positive on average and skewed to the right.

3.4.2 In-Sample Estimation

The estimated coefficients and robust standard errors associated with the Realised FTS GARCH model are provided in Tables 3.3 and 3.4. For the sake of comparison we also include values for the Realized Beta GARCH model. Results in Table 3.3 are based on the full sample of equity and Treasuries data, whilst those in Table 3.4 are based on the full sample of equity and gold data.

[Insert Table 3.3 here]

[Insert Table 3.4 here]

From an examination of the parameters that are common to both methods, a number of remarks can be made. Loosely speaking, we find close agreement with the estimates of the benchmark model. However, there are some notable exceptions, especially in the stock-gold case. For instance, parameter d_1, which measures the volatility spillover effect from the equity market to the precious metal market, is small and insignificant in our model. In addition, we note that our parameters b_i and c_i tend to be, respectively, higher and smaller, regardless of which asset is used as the safe haven. A final remark concerns the parameter φ in the measurement equations. All estimates of φ are not too far from unity, which is to be expected since the returns used in the estimation procedure are computed over the same time span as the realised measures.

Examining the new parameter e_0, which captures the effect of flight-to-safety on the conditional variance of equity returns, we observe that the estimate is negative both in Table 3.3 and in Table 3.4. This indicates that, in agreement with the findings of Section 3.2, when realised semi-covariance between falling equity and rising safe haven returns becomes more negative, the model predicts an increase in the stock market volatility for the next day. The parameter is larger in the case of US government bonds being the safe haven. It follows that volatility spikes tend to be more pronounced when FTS is measured in terms of the stock-bond relation than when it is measured in terms of the stock-gold one. Moreover, the robust standard errors suggest that our FTS measure is a more significant predictor of equity volatility when US Treasuries, and not gold, are used in the Realised FTS GARCH model.

The results in Panel D of Table 3.3 and 3.4 indicate that the coefficients of daily, weekly and monthly lagged components in the HAR-type specification for $\tilde{m}_t^-$ are all statistically significant. This highlights the persistent nature of realised semi-covariance, which was first documented by Bollerslev et al. (2017). Not surprisingly, the contemporaneous effect of standardised returns on $\tilde{m}_t^-$ is positive for the safe haven and negative for the equity return. This is simply related to the sign of the underlying high-frequency returns used to compute realised semi-covariance. It is important to remember that, for modelling purposes, we are flipping the sign of m_t^- and then taking the logarithm of it. Therefore, ψ_0 and ψ_1 would have to be, respectively, negative and positive in oder for the model-implied realised semi-covariance to become more negative when the stock market falls and the investors' safe-haven demand increases.

As a final remark on the empirical fit of the Realised FTS GARCH model, we follow Hansen et al. (2012) and consider the in-sample partial log-likelihood function for equity returns, see equation (3.9). There are two motivations for using ℓ_{z_0} instead of the joint log-likelihood function. First, our primary focus is on the stock market since we introduce the lagged FTS measure in the equity conditional variance equation. Second, the benchmark model does not use realised semi-covariance, so its joint log-likelihood is not comparable with that of the Realised FTS GARCH model. When US Treasuries are used in conjunction with the US stock market, the partial log-likelihood function is $\ell_{z_0}^{Beta} = -1512.71$ for the Realized Beta GARCH. This value rises to $\ell_{z_0}^{FTS} = -1507.98$ in our model. Similar results hold for the stock-gold combination. In this case the Realised FTS GARCH model dominates the benchmark model with a log-likelihood function of $\ell_{z_0}^{FTS} = -1505.44$, which is greater than $\ell_{z_0}^{Beta} = -1516.06$. Based on these figures, we can conclude that using realised semi-covariance as a proxy of flight-to-safety to forecast volatility provides a better fit to the data, at least for the equity market.

3.4.3 Forecasting

In this subsection the predictive performance of the Realised FTS GARCH model is assessed against that of the benchmark model. We consider in- and out-of-sample 1 to 22-day ahead forecasts generated by the two models for the volatility of equity and the safe haven asset returns. We conduct both pointwise and cumulative comparisons to examine the quality of the prediction at a specific point in time as well as the performance over the whole forecast horizon. We evaluate the accuracy of the forecasts under the MSE and QLIKE loss functions,

$$MSE(RV_{t+h}, F_{t+h|t}) = (RV_{t+h} - F_{t+h|t})^2,$$
$$QLIKE(RV_{t+h}, F_{t+h|t}) = \frac{RV_{t+h}}{F_{t+h|t}} - \log\left(\frac{RV_{t+h}}{F_{t+h|t}}\right) - 1,$$

where RV_{t+h} and $F_{t+h|t}$ are, respectively, the realised measure of variance that proxies the actual volatility at time $t+h$, and its value predicted at time t. We denote the forecast horizon by h, which takes values in the range 1 to 22.

In the out-of-sample analysis, we adopt a rolling forecasting scheme in which models are re-estimated daily by recursively adding one observation to the end of the sample and removing one from the start. We follow Shephard and Sheppard (2010) and adopt a size of 1,008 observations for the rolling window.[5] The holdout sample thus starts on March 5, 2004. We calculate MSE and QLIKE for the out-of-sample forecasts and then use these loss function values in the Reality Check of White (2000). This is a statistical test for the predictive superiority of a model that

[5]All the code has been written in Python to benefit from its computational speed. Moreover, since the estimation problem in the out-of-sample exercise is inherently independent from one day to another, we relied on BlueCrystal, the High Performance Computing (HPC) machine provided by the Advanced Computing Research Centre at University of Bristol, in order to exploit the power of multiple CPUs at the same time.

accounts for the effect of data snooping. Loosely speaking, we want to show that if the Realised FTS GARCH model produces more accurate forecasts than the benchmark, then it is due to its inherent merits rather than pure chance. To do so, we consider the following hypotheses:

$$H_0 : \mathbb{E}[L^{Beta}(RV, F) - L^{FTS}(RV, F)] \leq 0,$$
$$H_1 : \mathbb{E}[L^{Beta}(RV, F) - L^{FTS}(RV, F)] > 0,$$

where $L^{Beta}(RV, F)$ and $L^{FTS}(RV, F)$ indicate the losses associated, respectively, with the Realised FTS GARCH and the Realized Beta GARCH model. The test is in favor of our model in case of rejection of the null hypothesis, under which the out-of-sample forecast losses of the Realised FTS GARCH are not significantly lower than those of the benchmark model. We implement the Reality Check test using the stationary bootstrap proposed by Politis and Romano (1994) and we follow Bollerslev et al. (2016) in choosing an average block length of five and 999 re-samplings.

As mentioned earlier, we consider the joint dynamics of equity and Treasuries separate from that of equity and gold. We begin by examining the stock-bond pair, for which the average MSE and QLIKE losses of the Realised FTS GARCH model are reported in Table 3.5. The loss measures have been divided by those of the benchmark model. Since the model under consideration is one for forecasting, we should heavily focus on its out-of-sample performance and treat the lower in-sample losses simply as preliminary evidence of the gains from the FTS measure. Overall, the results in Table 3.5 suggest that the FTS information contained in the realised semi-covariance between US equity and government bonds can be beneficial to the forecasts of the asset volatilities. When looking at the out-of-sample losses for equity volatility (Panel A), we note that the results under the MSE are somewhat mixed with improvements of more than 10% for the very short and long forecast horizons and a slightly worse performance than the benchmark for the weekly horizon. On the other hand, the gains in QLIKE are more stable and typically of the order of 3.5%. From an examination of Panel B in Table 3.5, we note that the Realised FTS GARCH model systematically outperforms the benchmark model in predicting the volatility of US Treasuries. The gains under the QLIKE loss are in the range of 4% to 8%, with the most accurate forecasts delivered for longer horizons. The use of MSE results in even larger improvements with a peak of nearly 30% for the 2-week ahead volatility.

[Insert Table 3.5 here]

In Table 3.6, we report the p-values of the Reality Check test on the forecast improvements for US equity and bond volatility. Under the QLIKE loss, we reject, at the 5% significance level, the null of no predictive superiority of the Realised FTS GARCH model for both assets at any forecast horizon considered. Similar results are observed in Panel B with the assumption that the forecasts of the safe haven variance are assessed using the MSE loss function. When the same function is used to test the predictions on the equity asset, we observe that the magnitude of the gains in predicting short-term variance is not statistically significant. However, statistically superior performance emerge for the monthly horizon. In the pointwise comparison, the significance of the gains at the 10% level is already found for the two-week ahead forecasts of equity volatility.

[Insert Table 3.6 here]

The potential gains in the US equity and bond volatility forecasts can be seen more clearly in Figure 3.2, panel (a), where the relative losses of the Realised FTS GARCH model are plotted against the forecast horizon. To go further in the analysis, we consider how the performance changes over time. We follow Taylor (2017) and use a Gaussian kernel with bandwidths of 3 months to calculate the time-varying average MSE and QLIKE values. Figure 3.2 shows the time series of the relative mean losses for the daily horizon in panel (b) and (c), for the weekly (5-day ahead) horizon in panel (d) and (e), and for the monthly (22-day ahead) horizon in panel (f) and (g). We note that the benefits of using flight-to-safety are not constant over time. For instance, an inferior performance of the Realised FTS GARCH model in predicting equity volatility at any forecast horizon is observed in the the first couple of years of the holdout period. Evidence of less accurate predictions from our model is also found in year 2007, especially at the daily horizon. Since then, however, the predictive power of our model for short- and long-term volatility has long been dominant, with gains of over 10% during the 2008-09 financial crisis, when the FTS events were more likely to have occurred. Not surprisingly, such benefits are more evident under the MSE. In fact, when volatility abruptly jumps up we frequently encounter under-predictions, which are penalised more by the QLIKE function. Looking at the weekly horizon gives less clear-cut results: under the MSE loss function we observe alternate periods of superior and inferior forecats, while on the basis of the QLIKE loss function the Realised FTS GARCH model greatly outperforms the benchmark in predicting equity return volatility. Interestingly, the Realised FTS GARCH model is consistently better than the benchmark at forecasting bond variance in the short-, medium- and long-run. Although there is considerable time variation, the average MSE and QLIKE losses associated with our model have hardly been higher since 2004.

[Insert Figure 3.2 here]

We now discuss the results of applying the models to US equity and gold data. The average MSE and QLIKE loss values of the Realised FTS GARCH model divided by those of the Realized Beta GARCH model are provided in Table 3.7. These results reinforce the idea that the use of flight-to-safety, when proxied by realised semi-covariance of falling equity and rising safe haven returns, can improve the performance of a volatility forecasting model. Indeed, from Panel A of Table 3.7, it appears that the proposed model delivers on average lower QLIKE losses than the benchmark for US equity volatility. The improvements start at approximately 4% for the next-day volatility forecasts and increase with the horizon. In contrast, when the MSE loss function is assumed, the vast benefits offered in-sample are preserved out-of-sample only for forecast horizons longer than a week. The forecasts delivered by our model for the 22-day ahead (cumulative) variance of equity are more than 10% more accurate than those from the benchmark. When looking at the out-of-sample losses for gold volatility (Panel B), we note that the results are not as good as those for equity. Under the MSE loss function assumption, the Realised FTS GARCH model fails to improve on the benhcmark model over all horizons, except for the pointwise one-month ahead. On a positive note, we do find evidence of a small reduction in the average QLIKE losses, but these gains never exceed 2% and are strongly reduced with respect to those obtained in-sample.

[Insert Table 3.7 here]

In Table 3.8, we report the p-values of the Reality Check test when US stock and gold data are used. The results associated with equity volatility demonstrate that when using the QLIKE loss function to assess the predictions, the gains in forecast accuracy offered by our model are statistically significant at the 5% level. Under the MSE loss function, however, superior predictive ability is only found for the monthly horizon, similarly to what we have shown in Table 3.6. With regard to the safe haven volatility, we need only note that the improvements offered by the Realised FTS GARCH model, if any, are not significant in a statistical sense. This holds true for all cases, with the only exception being the long-term forecasts assessed with the QLIKE loss function.

[Insert Table 3.8 here]

Finally, we illustrate the effects of the FTS measure on the forecasts of equity and gold volatility in Figure 3.3. The top panel (a) shows for each forecast horizon the relative mean losses of the Realised FTS GARCH model, which we have already discussed earlier. More interesting is to look at the time variation in the relative performance of the new model. We plot the time series of the relative mean losses for the daily horizon in panels (b) and (c), for the weekly horizon in panels (d) and (e), and for the monthly horizon in panels (f) and (g). At the shorter horizon we note that the benefits of using flight-to-safety in forecasting the volatility of the precious metal are very small and do not vary substantially over time. For the next-day forecasts of equity variance, however, the loss functions recognise a significantly superior predictive power of our model in 2007 and 2013, and inferior performance in 2011. While no noticeable benefits are observed at the daily horizon, we find at the weekly horizon multiple periods of superior forecasts delivered by the Realised FTS GARCH model for gold volatility. As for the medium-term equity volatility, we observe gains of over 20% in 2007 under both MSE and QLIKE loss functions. Moreover, panels (f) and (g) show that the monthly forecast losses associated with the Realised FTS GARCH model are lower during the 2008-09 financial crisis. This holds true for both equity and gold volatility. Generally, more variability is found in the relative performance at the monthly horizon. Here, the average forecast losses of our model for the volatility of gold are lower than those of the benchmark in the first part of the sample and higher in the last part. The opposite is seen for equity variance with more accurate forecasts delivered by the benchmark at the start of the sample and by the Realised FTS GARCH model in the later years.

[Insert Figure 3.3 here]

3.4.4 Robustness

In this subsection, we assess the robustness of our findings. To do so, we first examine to what extent the predictive superiority of the Realised FTS GARCH model over the benchmark model is driven by the choice of realised measures used. As a further robustness check, we consider the forecasting performance of the proposed model against that of alternative benchmark models.

3.4.4.1 Alternative Realised Variance Estimators

The Realised FTS GARCH model is an extension of the multivariate GARCH model of Hansen et al. (2014) that uses a realised measure of semi-covariance in addition to the more traditional measures of volatility and correlation. In the analysis presented so far, we have used the multivariate realised kernel by Barndorff-Nielsen et al. (2011) to obtain realised variance and covariance, and the sum of the products of mixed sign intraday returns sampled every 15 minutes to obtain realised semi-covariance (Bollerslev et al., 2017). The forecasts of conditional variance delivered by the Realised FTS GARCH and Realized Beta GARCH models have then been assessed by comparing them with the elements on the main diagonal of the covariance matrix produced by the kernel estimator of Barndorff-Nielsen et al. (2011). The elements of this matrix, like all other realised variance measures, are contaminated by measurement errors and, as such, provide an imperfect proxy for the 'true' unobservable volatility. To determine whether our results are influenced by the noise that exists in the kernel estimator proposed by Barndorff-Nielsen et al. (2011), we now consider variation in the proxy for volatility that we choose as the forecast target. In particular, we compare the forecasts of conditional variance delivered by the two GARCH models with some of the most commonly used realised variance measures, namely: 5-minute (RV), sub-sampled (SSRV), two-scales (TSRV) and maximum likelihood (MLE). However, to permit direct comparison with results obtained in Section 3.4.3, we continue to estimate the models with the same realised measures of volatility, correlation and semi-covariance as before.[6]

RV is the standard realised variance estimator obtained from the sum of squared 5-minute intraday returns as discussed by Andersen et al. (2001). Introduced by Zhang et al. (2005), SSRV and TSRV are two realised measures of variance robust to market microstructure noise. The former represents the sub-sampled version of RV that averages over multiple time grids with 5-minute intervals, while the latter bias-corrects this average using all of the data. Finally, MLE is the maximum likelihood estimator of volatility proposed by Aït-Sahalia et al. (2005). Unlike RV, SSRV and TSRV, MLE is derived within a fully parametric framework for the distribution of market microstructure noise. In practice, MLE estimates the daily return variance by assuming that the intraday returns follow an MA(1) process and that a fixed time interval separates successive high-frequency return observations.

The implementation of the aforementioned alternative realised variance estimators is based on the high-frequency prices of equity, bond and gold synchronised with the 'refresh time' scheme proposed by Barndorff-Nielsen et al. (2011). The estimation of sub-sampled and two-scale realised variances is done by sampling sparsely at the 5-minute frequency and averaging over multiple time grids spaced by 1-minute intervals. As for the parametric estimator of volatility, we use a fixed time interval of 1 minute between consecutive observations, which is reasonably consistent with the optimal sampling frequency of *as often as possible* (Aït-Sahalia et al., 2005).

We begin by examining the robustness of the out-of-sample results obtained when the models are applied to US equity and bond futures data. For each of the alternative realised variance estimators, we report the mean losses associated with the Realised FTS GARCH relative to (divided by) those of the Realized Beta GARCH model in Table 3.9, and the p-values of the Reality Check test in Table 3.10. For ease of comparison, we also include the previous results

[6] We also investigated the performance of models estimated with the alternative realised variance measures used as forecast target and with their covariance equivalent, resulting in qualitatively similar findings to those discussed here.

that used the realised kernel (RK) estimator of Barndorff-Nielsen et al. (2011) as forecast target. Overall, the results in Table 3.9 confirm that realised semi-covariance between US equity and government bonds can be beneficial to the forecasts of the asset volatilities. Indeed, the Realised FTS GARCH model continues to outperform the benchmark across a wide range of forecast horizons, even when the predictions are assessed against the alternative realised variance estimators. Compared with the results reported in Section 3.4.3, we note a slightly worse relative performance of our model in predicting equity volatility when the forecasts are evaluated against the TSRV estimator. On the other hand, we observe a slightly better relative performance of our model in predicting bond volatility when the forecasts are evaluated against any of the alternative estimators considered. From Panel A of Table 3.10, it appears that the forecast improvements for equity volatility offered by the Realised FTS GARCH model relative to the benchmark are statistically significant at the monthly horizon across all proxies used for volatility. Panel B of Table 3.10 shows that the same holds true for bond volatility, but in this case the losses associated with our model are significantly lower than those of the benchmark at any forecast horizon. We find that the p-value associated with the alternative realised variance measures is generally consistent with that associated with the RK estimator.

[Insert Table 3.9 here]

[Insert Table 3.10 here]

We now discuss the robustness of the out-of-sample results obtained when the models are applied to US equity and gold data. For RK and each of the alternative realised variance estimators used as forecast target, we report the mean losses associated with the Realised FTS GARCH relative to those of the Realized Beta GARCH model in Table 3.11, and the p-values of the Reality Check test in Table 3.12. From an examination of the tables, it may be seen that the alternative realised variance measures lead to similar considerations to those detailed in Section 3.4.3 about the accuracy of the proposed model in forecasting equity volatility. On the other hand, we note that the relative performance of the models in forecasting gold volatility is more sensitive to the choice of forecast target. Indeed, when the predictions for gold volatility are compared with some of the alternative realised variance estimators, we find evidence of forecast improvements offered by our model, which we failed to find with the RK estimator.

[Insert Table 3.11 here]

[Insert Table 3.12 here]

In conclusion, we can state that our previous findings on volatility forecasting using realised semi-covariance of falling equity and a rising safe haven are robust to the presence of measurement errors in the adopted proxy for 'true' volatility. Indeed, regardless of the realised variance estimator used as forecast target, the Realised FTS GARCH model offers clear improvements over the benchmark in terms of more accurate out-of-sample forecasts of both equity and safe haven volatility.

3.4.4.2 Alternative Benchmark Models

In order to further examine the robustness of our results, we consider a variation of the benchmark against which the proposed model is compared. Specifically, we evaluate the out-of-sample forecasting performance of the highly parameterised Realised FTS GARCH model as compared to that of a number of univariate volatility models with low dimensional parameters. The new set of benchmark models comprises a first-order autoregressive process (AR), the HAR-RV model proposed by Corsi (2009), the GARCH (1,1) model of Bollerslev (1986) specified with a targeting parameterisation, and the HEAVY models of Shephard and Sheppard (2010) formulated in both targeting and integrated version.[7] The latter version of the Shephard and Sheppard (2010) models is henceforth referred to as the IHEAVY model. Using the notation introduced in Section 3.3.1, we fit the aforementioned models to US equity, US bond and gold market data using the following functional forms,

[7]We refer the reader to Section 2.2.3 in Chapter 2 for a review of all these models.

$$\begin{aligned}
&\text{a) AR} &&: \tilde{x}_{i,t} = \phi_0 + \phi_1 \tilde{x}_{i,t-1} + \epsilon_t \ , \\
&\text{b) HAR-RV} &&: \tilde{x}_{i,t} = \phi_0 + \phi_1 \tilde{x}_{i,t-1} + \phi_2 \tfrac{1}{5} \sum_{j=1}^{5} \tilde{x}_{i,t-j} + \phi_3 \tfrac{1}{22} \sum_{j=1}^{22} \tilde{x}_{i,t-j} + \epsilon_t \ , \\
&\text{c) GARCH} &&: h_{i,t} = \mu_{\text{G}}(1 - \alpha_{\text{G}} - \beta_{\text{G}}) + \alpha_{\text{G}} r^2_{i,t-1} + \beta_{\text{G}} h_{i,t-1} \ , \\
&\text{d) HEAVY} &&: h_{i,t} = \mu_{\text{r}}(1 - \alpha_{\text{r}}\kappa - \beta_{\text{r}}) + \alpha_{\text{r}} x_{i,t-1} + \beta_{\text{r}} h_{i,t} \ , \\
& && \quad \mu_{i,t} = \mu_{\text{RM}}(1 - \alpha_{\text{RM}} - \beta_{\text{RM}}) + \alpha_{\text{RM}} x_{i,t-1} + \beta_{\text{RM}} \mu_{i,t-1} \ , \\
&\text{e) IHEAVY} &&: h_{i,t} = \omega_{\text{r}} + \alpha_{\text{r}} x_{i,t-1} + \beta_{\text{r}} h_{i,t-1} \ , \\
& && \quad \mu_{i,t} = \alpha_{\text{RM}} x_{i,t-1} + (1 - \alpha_{\text{RM}}) \mu_{i,t-1} \ ,
\end{aligned}$$

where $i = 0, 1$.[8] As before, we consider the safe haven property of bonds separately from that of gold. Therefore, we first compare the forecasts for equity ($i = 0$) and bond ($i = 1$) volatility delivered by the univariate benchmark models with those delivered by the Realised FTS GARCH model that uses realised semi-covariance between negative equity and positive bond returns. Second, we repeat the analysis for equity ($i = 0$) and gold ($i = 1$) volatility and compare the forecasts from the benchmark models with those from the Realised FTS GARCH model that uses realised semi-covariance between negative equity and positive gold returns. Using the same forecasting scheme of Section 3.4.3, we calculate MSE and QLIKE for the out-of-sample forecasts delivered by all models. We then use these loss function values in the modified version of Reality Check test proposed by Bollerslev et al. (2016) and already used in Section 2.4.3 in Chapter 2 to formally test whether the loss of a given model is lower than that from the best model among a set of benchmark models. We continue to implement the test as described in Section 3.4.3, but we now write the following hypotheses:

$$H_0 : \min_{k=1,\dots K} \mathbb{E}[L^k(RV, F) - L^0(RV, F)] \leq 0$$

$$H_1 : \min_{k=1,\dots K} \mathbb{E}[L^k(RV, F) - L^0(RV, F)] > 0$$

where L^0 and L^k indicate the losses associated, respectively, with the Realised FTS GARCH and with the benchmark models, i.e. AR, HAR-RV, GARCH(1,1), HEAVY and IHEAVY. Significant

[8] The performance of the two HEAVY models d) and e) is assessed based on forecasts for latent volatility, $h_{i,t}$.

predictive superiority of the Realised FTS GARCH is found in case of rejection of the null.

We begin by examining the stock-bond pair, for which we report the mean losses associated with each model relative to (divided by) that associated with the HAR-RV in Table 3.13, and the p-values of the Reality Check test in Table 3.14. Overall, the results confirm the ability of the Realised FTS GARCH model to deliver improved out-of-sample forecasts of future equity return volatility, but also provide evidence against its accuracy in forecasting bond volatility. Indeed, the proposed model exhibits the lowest values for most mean losses presented in Panel A of Table 3.13, but it never comes first in the ranking of bond volatility forecasts presented in Panel B. The improvements in the 1-day ahead forecasts of equity volatility are approximately 25% and 7% under the MSE and QLIKE loss, respectively. Looking at the weekly and monthly pointwise performance gives less clear-cut results as the forecasts of equity volatility delivered by the Realised FTS GARCH are the most accurate only under one of the two loss functions. However, when the cumulative performance is considered, the Realised FTS GARCH always outperforms the benchmark models, with gains of over 5%. From Panel A of Table 3.14, it appears that the forecast improvements for equity volatility offered by the Realised FTS GARCH tend to be statistically significant. Indeed, we reject, at the 10% significance level, the null hypothesis of no predictive superiority of the proposed model in the vast majority of cases. The poor results of the Reality Check test presented in Panel B of Table 3.14 come as no surprise since the Realised FTS GARCH always fails to improve on the set of benchmark models for bond volatility forecasting.

[Insert Table 3.13 here]

[Insert Table 3.14 here]

We now compare the forecasts of equity and gold volatility delivered by the benchmark models with those delivered by the Realised FTS GARCH that uses realised semi-covariance between negative equity and positive gold returns. We report the mean losses associated with each model relative to that associated with the HAR-RV in Table 3.15, and the p-values of the Reality Check test in Table 3.16. The results for the stock-gold pair are qualitatively similar to those for the stock-bond pair. Once again, we find that the Realised FTS GARCH model outperforms its competitors in predicting equity return volatility, but is outperformed in predicting volatility of the safe haven asset. From Panel A of Table 3.15, we note that the improvements in the 1-day ahead forecasts of equity volatility offered by the proposed model are approximately 13% and 8% under the MSE and QLIKE loss, respectively. When the cumulative performance is considered, the Realised FTS GARCH always outperforms the benchmark models, with gains as large as 10% under the MSE loss and gains typically of the order of 4% under the QLIKE loss. These improvements in equity volatility forecasting offered by the Realised FTS GARCH model are statistically significant. Indeed, Panel A of Table 3.16 shows that we reject, at the 10% significance level, the null hypothesis of no predictive superiority of the proposed model over its competitors in all cases, with the only exception being the 22-day ahead pointwise forecast assessed under the QLIKE loss.

[Insert Table 3.15 here]

[Insert Table 3.16 here]

To sum up, the Realised FTS GARCH model, which accounts for the FTS information contained in the realised semi-covariance between falling equity and rising safe haven returns, produces superior out-of-sample forecasts of future equity return volatility. These results are robust to the choice of benchmark used for comparison. On the other hand, when it comes to the volatility forecasts of a second asset, which in our study is the safe haven asset, the multivariate Realised FTS GARCH model delivers less accurate predictions than univariate volatility models that do not account for FTS information. We believe that this can be due to the great dimensionality of parameters involved in the Realised FTS GARCH model since evidence of outperformance was found in Section 3.4.3, where we considered a benchmark with similar dimensional parameters to ours.

3.4.5 Economic Value

In this subsection, we explore the usefulness of the predictive ability of the Realised FTS GARCH model from the perspective of market investors. To evaluate the economic significance of our results, we apply the framework developed by Fleming et al. (2001, 2003) and conduct a portfolio exercise where the out-of-sample forecasts of conditional return variance and covariance are the key determinant of portfolio optimisation. In particular, we consider a mean-variance investor who allocates funds between cash and two risky assets - equity and the safe haven. We continue to assess the safe haven property of bonds separately from that of gold, implying that, in one case, the fund allocation is across equity, bonds and cash, while, in the other case, it is across equity, gold and cash. Consistent with the data used so far, the investor implements his or her investment decisions by trading futures contracts on the risky assets. The investor follows a volatility-timing strategy where his or her portfolio is rebalanced based only on estimates of the conditional covariance matrix of asset returns. Therefore, portfolio rebalancing does not depend on the conditional expectations of returns, which are held constant over time. Portfolio weights are dynamically determined by minimising conditional volatility subject to a given target expected return. For horizons ranging from one day to one month, we assess the empirical performance of a strategy based on the forecasts of the Realised FTS GARCH model against that of a benchmark strategy based on the forecasts of the Realized Beta GARCH model. In doing so, we answer the question of whether the gains in volatility predictability offered by the proposed model are not only statistically, but also economically significant.

We introduce the notation for the one-day performance horizon where the investor makes allocation decisions and rebalances his or her portfolio daily. The same notation applies to the case of longer horizons with the difference that index t no longer denotes days but refers to other time intervals, such as weeks or months.[9] Letting μ_p denote the fixed target expected portfolio return, the day-t optimal risky asset weights of the portfolio that has minimum conditional variance for any choice of μ_p are obtained as follows,

$$\mathbf{w}_t = \frac{\mu_p \boldsymbol{\Sigma}_{t|t-1}^{-1} \boldsymbol{\mu}_{t|t-1}}{\boldsymbol{\mu}'_{t|t-1} \boldsymbol{\Sigma}_{t|t-1}^{-1} \boldsymbol{\mu}_{t|t-1}}, \tag{3.15}$$

where $\boldsymbol{\mu}_{t|t-1}$ and $\boldsymbol{\Sigma}_{t|t-1}$ denote, respectively, the expectation of daily returns on the risky assets

[9] When assessing the performance of the volatility-timing strategies over the h-day horizon, for $h > 1$, we use the h-day ahead cumulative forecasts of return variance and covariance resampled so that they refer to non-overlapping periods. The same is done for actual return data, meaning that we aggregate the daily return data to obtain h-day returns referring to non-overlapping periods.

and their covariance matrix, conditional on past information $\mathcal{F}_{t-1}$. We set the elements of $\boldsymbol{\mu}_{t|t-1}$ equal to the sample mean of $r_{0,t}$ and $r_{1,t}$ because, as we mentioned before, we treat expected returns as constant. On the other hand, we populate the conditional covariance matrix of daily returns with the out-of-sample estimates of $h_{0,t}$ and $h_{1,t}$ on the main diagonal, and with the out-of-sample estimates of return covariance, $\rho_{1,t}\sqrt{h_{0,t}}\sqrt{h_{1,t}}$, outside the main diagonal. As for the target expected portfolio return, we use $\mu_p = 10\%$ in portfolio analysis. Although arbitrary, this target is in line with the choice of Fleming et al. (2003), who noted that different levels of μ_p would simply alter the allocation of funds with little effect on the results. Once the portfolio optimisation problem has been solved, the weight in cash, which plays the role of the risk-free asset, is obtained as one minus the sum of the elements of $\mathbf{w}_t$.[10]

Within the framework of Fleming et al. (2001, 2003), the daily utility from investing in the portfolio with weights $\mathbf{w}_t$ is:

$$U(R_{p,t}) = (1 + R_{f,t} + R_{p,t}) - \frac{\gamma}{2(1+\gamma)}(1 + R_{f,t} + R_{p,t})^2, \tag{3.16}$$

where $R_{p,t} = \mathbf{w}_t' \begin{bmatrix} r_{0,t} \\ r_{1,t} \end{bmatrix}$ is the portfolio return, $R_{f,t}$ is the risk-free rate, and γ is the investor's relative risk aversion. Following Fleming et al. (2003), we use $\gamma = 1$ and 10 in our analysis. Further, we let $R_{p1,t}$ and $R_{p2,t}$ denote the returns on the volatility-timing strategies based on the estimates of $\boldsymbol{\Sigma}_{t|t-1}$ delivered by, respectively, the Realized Beta GARCH and Realised FTS GARCH model. By finding the value of Δ such that $\sum_{t=1}^{T} U(R_{p1,t}) = \sum_{t=1}^{T} U(R_{p2,t} - \Delta)$, we measure the maximum return an investor in the benchmark strategy would be willing to sacrifice each day in order to capture the performance gains associated with the strategy behind $R_{p2,t}$.

We start by assessing the empirical performance of volatility timing at the daily level and with mean returns on risky assets calculated for the full sample period 2000 – 2016. Thereafter, we will review the method used by Fleming et al. (2001, 2003) to consider different estimates of the mean returns and, hence, control for estimation risk regarding expected returns. Finally, we will evaluate the performance of the volatility-timing strategies over longer horizons in the

[10]In our analysis we use the risk-free rate available at a daily frequency from the Kenneth R. French - Data Library http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data˙library.html.

presence of estimation risk. Table 3.17 summarises the daily performance of volatility-timing strategies for the case of no estimation risk regarding expected returns. The two strategies in Panel A are based on the next-day forecasts related to US equity and bond returns, while those in Panel B are based on the next-day forecasts related to US equity and gold returns. We report results for the entire out-of-sample period and for a range of subperiods.

[Insert Table 3.17 here]

Panel A of Table 3.17 shows that the estimates for the conditional covariance matrix of equity and bond returns produced by the Realised FTS GARCH model are economically significant. Indeed, using these estimates results in a volatility-timing strategy that, compared with the benchmark, offers a more attractive Sharpe ratio (0.854 versus 0.844) in the form of higher average return (9.215% versus 9.199%) and lower volatility (10.792% versus 10.905%). Further, we observe that an investor with high (low) relative risk aversion would be willing to pay approximately 14 (3) basis points per year to switch from the volatility-timing strategy based on the Realized Beta GARCH model to that based on the Realised FTS GARCH model. Examining the results in the subperiods considered, we find that the superior portfolio performance offered by the proposed model is particularly evident in the first half of the sample (including the 2008-09 financial crisis), while the benchmark strategy dominates in the last years with a higher Sharpe ratio. Turning to the stock-gold pair presented in Panel B of Table 3.17, it is apparent that the performance of volatility timing is greatly reduced by our model if compared to the benchmark. In all periods, the strategy based on the Realised FTS GARCH model offers a lower Sharpe ratio and investors are reluctant to shift their portfolios towards it.

To illustrate graphically the performance of volatility timing over time, we plot the cumulative returns of stock and bond portfolios in Figure 3.4, and the cumulative returns of stock and gold portfolios in Figure 3.5. Consistent with the performance gains discussed above, a portfolio of equity and bonds formed using the conditional covariance matrix estimates delivered by the Realised FTS GARCH model accumulates higher returns than the benchmark strategy. At the end of the sample, both dynamically rebalanced portfolios have greater cumulative returns than buy-and-hold strategies on the individual risky assets. In contrast, when equity and gold portfolios are considered, the volatility-timing strategy based on the Realised FTS GARCH model is clearly outperformed by the benchmark since the 2008-09 crisis. On a positive note, we observe that its cumulative return at the end of the sample is higher than those on single-asset portfolios.

[Insert Figure 3.4 here]

[Insert Figure 3.5 here]

We now turn to the economic significance of our results when uncertainty about expected returns is accounted for. The implementation of volatility-timing strategies requires an estimate of $\boldsymbol{\mu}_{t|t-1}$ to use in equation (3.15). The analysis presented so far has been carried out with a single estimate of this vector, i.e. the full sample mean of risky asset returns. Fleming et al. (2001, 2003) propose a bootstrap experiment that controls for estimation risk by simulating a range of possible mean returns for $\boldsymbol{\mu}_{t|t-1}$. The algorithm works as follows. (i) Generate an artificial sample of size k by randomly sampling, with replacement, blocks of observations from the series of actual returns. (ii) Set the elements of $\boldsymbol{\mu}_{t|t-1}$ equal to the mean returns from this artificial sample and compute the portfolio weights according to equation (3.15). (iii) Apply the weights to the actual returns and evaluate the performance of the strategies. The resulting performance measures represent a single trial. (iv) Repeat step (i), (ii) and (iii) in a large number of trials and compute summary performance measures. We implement this procedure using the stationary bootstrap proposed by Politis and Romano (1994) with 1000 trials and an average block length of 15.[11] Table 3.18 summarises the daily performance of volatility-timing strategies for various levels of estimation risk quantified by the size, k, of the bootstrap samples.

[Insert Table 3.18 here]

[11] Following Fleming et al. (2003), we implement the bootstrap procedure with exclusionary criteria. Specifically, we check whether each artificial sample generates asset returns that satisfy the following criteria: the mean stock return is larger than the mean safe haven return, both mean returns are positive, and, when the safe haven is US bonds, the unconditional stock volatility is larger than the safe haven volatility. If any of these criteria is violated by the artificial sample, a new random sample is generated. This process is repeated until we have a total of 1000 trials in which the exclusionary criteria are satisfied.

From Panel A of Table 3.18, we note that the more precise forecasts produced by the Realised FTS GARCH model for US equity and bonds lead to better portfolio performance. For instance, using a sample size comparable to the actual data, $k = 3000$, the strategy based on the proposed model outperforms the benchmark with higher Sharpe ratio (0.834 versus 0.817), higher average return (9.402% versus 9.314%) and lower volatility (11.323% versus 11.452%). The p-value of 0.982 indicates that in 98.2% of the bootstrap trials using the forecasts of the Realised FTS GARCH earns a higher Sharpe ratio than using the forecasts of the benchmark model. As for the performance fee, we find that an investor with high (low) relative risk aversion would be willing to pay approximately 31 (11) basis points per year to switch from the volatility-timing strategy based on the Realized Beta GARCH model to that based on the Realised FTS GARCH model. The economic significance of these results also holds for higher (lower) levels of estimation risk obtained by reducing (increasing) the sample size k. A quite different situation from that just discussed can be observed in Panel B of Table 3.18. The volatility-timing strategy based on the forecasts delivered by the Realised FTS GARCH model for US equity and gold never outperforms the benchmark (p-value is 0 for all levels of estimation risk) and investors are reluctant to shift their portfolios towards it (Δ is negative for all levels of estimation risk and relative risk aversion).

Relying on the multi-period ahead forecasting ability of the two competing models, we now evaluate the performance of volatility timing over horizons as long as one month. Table 3.19 summarises the performance of portfolios that are rebalanced daily, weekly, biweekly and monthly based, respectively, on 1-day, 5-day, 10-day and 22-day ahead cumulative forecasts of the conditional covariance matrix of returns. We control for estimation risk by bootstrapping samples of $k = 4000$ observations at the different frequencies, from which we estimate the corresponding expected returns. We implement the bootstrap experiment with 1000 trials and an average block length of 1, 2, 3, and 15 observations for the monthly, biweekly, weekly and daily horizon, respectively. For the daily horizon, the means, volatilities, Sharpe ratios, p-values, and Δs reported in Table 3.19 are the same as those in Table 3.18. Examining Panel A in Table 3.19, we note that the economic gains seen in Table 3.18 are preserved when portfolios are rebalanced at lower frequencies. Indeed, the multi-step ahead predictions of equity and bond return volatility produced by the Realised FTS GARCH model result in better volatility-timing performance. For example, using our model to estimate the one-week ahead conditional covariance matrix of returns leads to a volatility-timing strategy that, compared with the benchmark, offers a higher Sharpe ratio (0.927 versus 0.898) in all bootstrap trials (p-value is 1.00). Moreover, an investor with high (low) relative risk aversion would be willing to pay a fee of approximately 33 (25) basis points per year to switch to that strategy. Lastly, we note that even when the 22-day ahead cumulative forecasts are used, the volatility-timing strategy based on the Realised FTS GARCH model outperforms that based on the Realized Beta GARCH model with a higher Sharpe ratio in 100% of the bootstrap trials. The performance measures in Panel B, suggest, on the other hand, that the estimates of the conditional covariance matrix produced by the Realised FTS GARCH model for equity and gold returns are economically insignificant at any of the forecast horizons considered. Indeed, when volatility timing is based on those forecasts, portfolio performance is always inferior to the benchmark. Only at the monthly horizon, we observe minor differences in the risk-adjusted returns earned by the competing strategies.

[Insert Table 3.19 here]

As a final exercise, we investigate the effects of transaction costs on the economic value of volatility timing. The severity of the impact of transaction costs on the profitability of volatility-timing strategies varies with both the frequency and the extent of portfolio rebalancing. In fact, the performance is reduced the most when portfolio weights are updated as frequently as possible (daily in our case) and the strategy involves high turnover, i.e. large changes in the portfolio weights between two consecutive time periods. To check whether the inclusion of transaction costs substantially changes the results obtained through the covariance matrix forecasts of the Realised FTS GARCH and Realized Beta GARCH model, we follow the approach of de Pooter et al. (2008). By assuming that transaction costs amount to a fixed percentage c on each traded dollar for any risky asset, the portfolio return net of transaction costs is:

$$R_{p,t}^{\text{NET}} = R_{p,t} - c||\mathbf{w}_{t+1} - \tilde{\mathbf{w}}_t||_1, \tag{3.17}$$

where $\tilde{\mathbf{w}}_t = \mathbf{w}_t \circ \begin{bmatrix} \frac{1+r_{0,t}}{1+R_{p,t}} \\ \frac{1+r_{1,t}}{1+R_{p,t}} \end{bmatrix}$ and $||\cdot||_1$ denotes the Manhattan norm.[12] Table 3.20 (Table 3.21) summarises the performance, net of transaction costs, of volatility-timing strategies based on the forecasts of the conditional covariance matrix of equity and bond (equity and gold) returns. We report results for portfolios that are rebalanced daily, weekly and monthly, and subject to transaction costs ranging from 0 to 20%.[13] These values of c represent the annualised level of costs paid if the entire portfolio is traded at every time period. We control for estimation risk regarding expected returns in the same way as we did for the results in Table 3.19.

[Insert Table 3.20 here]

[Insert Table 3.21 here]

As can be seen from Table 3.20, the estimates for the conditional covariance matrix of equity and bond returns produced by the Realised FTS GARCH model are still economically significant, even after transaction costs are taken into account. Indeed, compared to the benchmark, the volatility-timing strategy associated with the forecasts of the Realised FTS GARCH model achieves a better performance, net of transaction costs, at all rebalancing frequencies considered. For instance, using a relatively high level of transaction costs, $c = 10\%$, on daily rebalancing, the strategy based on the proposed model outperforms the benchmark with higher Sharpe ratio (0.724 versus 0.671), higher average return (8.012% versus 7.512%) and lower volatility (11.089% versus 11.220%). The p-value of 1.00 indicates that in all bootstrap trials using the forecasts of the Realised FTS GARCH earns a higher Sharpe ratio than using the forecasts of the benchmark model. As for the performance fee, we find that an investor with high (low) relative risk aversion would be willing to pay approximately 67 (52) basis points per year to switch from the volatility-timing strategy based on the Realized Beta GARCH model to that based on the Realised FTS GARCH model. This fee is over 40 basis points higher than in the case of zero transaction costs. As the level of transaction costs increases, we note that the investor is willing to pay even more

[12] For any N-dimensional vector $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ the Manhattan norm is defined as $||\boldsymbol{x}||_1 = \sum_{i=1}^{N} |x_i|$.

[13] Fleming et al. (2003) suggest that a reasonable level of transaction costs for the S&P 500 futures contract is 256 basis points, annualised.

for the forecasts produced by the Realised FTS GARCH model. For the extremely high level of $c = 20\%$, the fee amounts to approximately 110 (94) basis points per year for an investor with high (low) relative risk aversion.

Turning to the results presented in Table 3.21, it is evident that the inclusion of transaction costs does not greatly alter the conclusions drawn earlier about the stock-gold pair. Indeed, the estimates of the conditional covariance matrix produced by the Realised FTS GARCH model for equity and gold returns continue to be economically insignificant in most of the cases considered. For example, when the performance of volatility timing is assessed at the daily level, we observe that the strategy based on the Realised FTS GARCH model offers a lower Sharpe ratio than the benchmark and investors are reluctant to shift their portfolios towards it for all transaction cost levels. However, lowering the rebalancing frequency, which reduces transaction costs, can change the relative performance in favour of our implementation. For portfolios that are rebalanced monthly subject to transaction costs of 10% or higher, we find that in the majority of the bootstrap trials the volatility-timing strategy associated with the Realised FTS GARCH model earns a higher Sharpe ratio than the benchmark. Moreover, an investor with low risk aversion would be willing to pay a positive fee for the forecasts produced by the proposed model.

The analysis in this subsection has provided some insight into the economic value of using flight-to-safety information for volatility forecasting. Realised semi-covariance between falling equity and a rising safe haven can lead to more accurate predictions of the asset volatilities, which in turn can lead to better portfolio performance. Of the safe haven assets that we consider it is US Treasuries that, when used to construct the proxy for FTS, yield significant gains in volatility-timing performance. We find that the improved forecasts of equity and bond return volatility yield substantial economic benefits to volatility-timing strategies implemented over horizons as long as one month and subject to transaction costs.

3.5 Conclusion

In this chapter, we have presented the benefits of using flight-to-safety in the context of volatility forecasting. We have proposed the realised semi-covariance computed from negative intraday returns on equity and positive intraday returns on a safe haven, as a proxy measure of

FTS and a direct predictor of equity volatility. We have revised the GARCH model by Hansen et al. (2014) to accommodate time variations in the parameters that are specifically induced by flight-to-safety. While earlier approaches to forecasting with realised semi-covariances have proven useful when the joint negative returns of a large number of stocks are considered, we have focused here on just two assets and used their high-frequency returns of opposite sign to improve the forecasts of the asset volatilities.

The results of an application to the US stock, US bond and gold markets are summarised as follows. First, our FTS proxy has high predictive power for stock return variation in a wide range of volatility forecasting models. The coefficient is negative indicating that when the FTS signal is stronger the equity variance will increase more the next day. Second, when introduced in a multivariate GARCH framework, the FTS measure offers substantial gains in terms of the out-of-sample forecasts of both equity and safe haven volatility, especially at long horizons. These statistically significant results are robust to the choice of forecast target and benchmark model. Third, although the benefits of using FTS are not constant over time, the forecast losses are normally lower during the 2008-09 crisis. Fourth, the more accurate predictions of the asset volatilities can lead to better portfolio performance. Finally, when used to construct the FTS measure, the US Treasuries seem to be more effective than gold at improving the predictions and implementing successful volatility-timing strategies.

Given our specific findings regarding the use of long-term Treasuries, it will be of interest, in future work, to determine whether the predictive content of the FTS measure varies with the maturity of the bonds. In fact, it might be claimed that short-dated bonds are more sensitive to changes in market sentiment and hence more appropriate to capture flight-to-safety. This is indeed what we find in the next chapter of this series. Another possible extension is to introduce the Japanese Yen - U.S. Dollar exchange rate as a 'global' safe haven that captures financial stress across different countries. We leave investigation of such possibilities to future research.

3.6 Tables of Chapter 3

Table 3.1 – Preliminary Analysis

	HAR-RV	HAR-RV w/ FTS	GARCH	GARCH w/ FTS	IHEAVY	IHEAVY w/ FTS
ϕ_0	−0.025**	−0.087***				
	(0.011)	(0.015)				
ϕ_1	0.428***	0.376***				
	(0.025)	(0.025)				
ϕ_2	0.346***	0.349***				
	(0.033)	(0.032)				
ϕ_3	0.171***	0.168***				
	(0.023)	(0.022)				
ω_{G}			0.029***	0.031***		
			(0.003)	(0.010)		
α_{G}			0.116***	0.048***		
			(0.017)	(0.014)		
β_{G}			0.856***	0.724***		
			(0.017)	(0.032)		
ω_{r}					0.037**	0.032**
					(0.019)	(0.018)
α_{r}					0.406***	0.326***
					(0.069)	(0.067)
β_{r}					0.559***	0.580***
					(0.072)	(0.066)
α_{RM}					0.446***	0.446***
					(0.032)	(0.032)
γ		−0.397***		−2.263***		−0.719**
		(0.066)		(0.389)		(0.339)

Notes: This table provides the in-sample parameter estimates and standard errors (in brackets) for the HAR-RV, GARCH (1,1) and Integrated HEAVY models in their original specification and when augmented with the daily lag of the FTS measure defined in equation (3.1). The estimated model equations are:

a) HAR-RV : $RV_t^d = \phi_0 + \phi_1 RV_{t-1}^d + \phi_2 RV_{t-1}^w + \phi_3 RV_{t-1}^m + \epsilon_t$,

b) HAR-RV w/ FTS : $RV_t^d = \phi_0 + \phi_1 RV_{t-1}^d + \phi_2 RV_{t-1}^w + \phi_3 RV_{t-1}^m + \gamma m_{t-1}^- + \epsilon_t$,

c) GARCH : $h_t = \omega_{\mathrm{G}} + \alpha_{\mathrm{G}} r_{t-1}^2 + \beta_{\mathrm{G}} h_{t-1}$,

d) GARCH w/ FTS : $h_t = \omega_{\mathrm{G}} + \alpha_{\mathrm{G}} r_{t-1}^2 + \beta_{\mathrm{G}} h_{t-1} + \gamma m_{t-1}^-$,

e) IHEAVY : $h_t = \omega_{\mathrm{r}} + \alpha_{\mathrm{r}} RV_{t-1} + \beta_{\mathrm{r}} h_{t-1}$,
$\mu_t = \alpha_{\mathrm{RM}} RV_{t-1} + (1 - \alpha_{\mathrm{RM}})\mu_{t-1}$,

f) IHEAVY w/ FTS : $h_t = \omega_{\mathrm{r}} + \alpha_{\mathrm{r}} RV_{t-1} + \beta_{\mathrm{r}} h_{t-1} + \gamma m_{t-1}^-$,
$\mu_t = \alpha_{\mathrm{RM}} RV_{t-1} + (1 - \alpha_{\mathrm{RM}})\mu_{t-1}$,

where r_t is the daily equity return, RV_t is the observed realised variance of equity returns, h_t is the latent variance of equity returns conditional on past information, μ_t represents the expectation of RV_t conditional on past information, RV_t^d, RV_t^w and RV_t^m denote, respectively, the daily, weekly and monthly components of the log transformed realised variance of equity returns, and m_t^- is the realised semi-covariance computed from negative intraday returns on equity and positive intraday returns on US government bonds. We use a window of 22 lags to compute the Newey-West robust standard errors of HAR-RV models, whilst we use a sandwich estimator of the asymptotic covariance matrix to compute the standard errors of GARCH and HEAVY models. * (resp. **, and ***) denote statistical significance at the 10% (resp. 5%, and 1%) level.

Table 3.2 – Summary Statistics

	Mean	Std Dev	Min	Max	Skewness	Kurtosis
SP Returns	0.022	1.130	-10.916	11.654	-0.226	11.466
TY Returns	0.015	0.372	-2.015	3.275	0.037	3.161
GC Returns	0.009	1.099	-9.199	10.582	-0.348	7.123
SP Realised Variance	1.265	3.201	0.012	104.758	13.662	316.210
TY Realised Variance	0.135	0.232	0.004	11.179	26.891	1215.512
GC Realised Variance	1.159	1.754	0.018	39.590	7.611	101.756
SP-TY Realised Correlation	-0.359	0.273	-0.922	0.776	0.629	0.020
SP-GC Realised Correlation	0.012	0.281	-0.852	0.692	-0.107	-0.162
TY-GC Realised Correlation	0.085	0.235	-0.578	0.863	0.167	-0.017
SP-TY Realised Semi-Covariance	-0.101	0.191	-3.960	0.000	-7.821	98.710
SP-GC Realised Semi-Covariance	-0.148	0.363	-8.768	0.000	-12.539	218.008

Notes: Summary statistics for daily returns and realised variances, correlations and semi-covariances of the S&P 500 index futures (SP), the 10-Year US Treasury Note futures (TY) and the gold (GC) futures. Excess Kurtosis is assumed. Returns are expressed in percentages while realised variances and semi-covariances are times 10,000.

Table 3.3 – In-Sample Estimation Results: SP - TY

Panel A: Equity volatility parameters (θ_0)

	a_0	b_0	c_0	τ_{01}	τ_{02}	e_0	ξ_0	φ_0	δ_{01}	δ_{02}
Realised FTS GARCH	0.030***	0.749***	0.223***	−0.107***	0.005*	−0.061***	−0.189***	0.958***	−0.062***	0.056***
	(0.008)	(0.013)	(0.015)	(0.006)	(0.003)	(0.019)	(0.027)	(0.037)	(0.007)	(0.006)
Realized Beta GARCH	0.039***	0.677***	0.279***	−0.117***	0.013***		−0.161***	1.024***	−0.088***	0.059***
	(0.010)	(0.019)	(0.023)	(0.007)	(0.004)		(0.030)	(0.049)	(0.007)	(0.007)

Panel B: Safe haven volatility parameters (θ_1)

	a_1	b_1	c_1	d_1	τ_{11}	τ_{12}	ξ_1	φ_1	δ_{11}	δ_{12}
Realised FTS GARCH	−0.043**	0.819***	0.142***	0.009***	−0.012***	0.014***	−0.136	1.060***	−0.039***	0.123***
	(0.018)	(0.013)	(0.009)	(0.003)	(0.005)	(0.003)	(0.088)	(0.040)	(0.007)	(0.005)
Realized Beta GARCH	−0.062***	0.778***	0.171***	0.022***	−0.011**	0.028***	−0.165*	1.037***	−0.026***	0.119***
	(0.024)	(0.019)	(0.014)	(0.006)	(0.006)	(0.005)	(0.099)	(0.042)	(0.007)	(0.005)

Panel C: Correlation parameters (θ_2)

	a_{10}	b_{10}	c_{10}	ξ_{10}	φ_{10}
Realised FTS GARCH	0.034**	0.717***	0.323***	−0.137***	0.784***
	(0.016)	(0.023)	(0.043)	(0.033)	(0.067)
Realized Beta GARCH	0.023**	0.622***	0.370***	−0.122***	0.866***
	(0.011)	(0.041)	(0.032)	(0.023)	(0.043)

Panel D: Semi-covariance parameters (θ_3)

	ζ	ϕ_d	ϕ_w	ϕ_m	ψ_0	ψ_1
Realised FTS GARCH	−0.345***	0.049***	0.505***	0.333***	−0.275***	0.209***
	(0.038)	(0.013)	(0.024)	(0.026)	(0.011)	(0.010)

Notes: This table provides the in-sample estimates of parameter vector $\theta = (\theta_0', \theta_1', \theta_2', \theta_3')'$ for both the Realised FTS GARCH model and the Realized Beta GARCH model. Safe Haven against the stock market is the US 10Y Treasury Notes. Models are estimated by Quasi-Maximum Likelihood. Standard errors are based on the sandwich estimator $\{\mathcal{J}_{2D}\mathcal{J}_{OP}^{-1}\mathcal{J}_{2D}\}^{-1}$ and are presented in parentheses below the estimated parameters. The parameter on the extra predictor m_{t-1}^{-} used in equation (3.2b) is found to be significant at the 1% level when the safe asset is the US bond market.

* Statistical significance at the 10% level.
** Statistical significance at the 5% level.
*** Statistical significance at the 1% level.

Table 3.4 – In-Sample Estimation Results: SP - GC

Panel A: Equity volatility parameters (θ_0)

	a_0	b_0	c_0	τ_{01}	τ_{02}	e_0	ξ_0	φ_0	δ_{01}	δ_{02}
Realised FTS GARCH	0.035***	0.700***	0.264***	−0.120***	0.002	−0.018*	−0.167***	0.998***	−0.091***	0.060***
	(0.009)	(0.017)	(0.020)	(0.007)	(0.003)	(0.011)	(0.028)	(0.042)	(0.007)	(0.007)
Realized Beta GARCH	0.011	0.648***	0.291***	−0.122***	0.007**		−0.069**	1.089***	−0.095***	0.047***
	(0.011)	(0.021)	(0.023)	(0.008)	(0.003)		(0.033)	(0.056)	(0.007)	(0.005)

Panel B: Safe haven volatility parameters (θ_1)

	a_1	b_1	c_1	d_1	τ_{11}	τ_{12}	ξ_1	φ_1	δ_{11}	δ_{12}
Realised FTS GARCH	0.056***	0.795***	0.169***	0.001	0.000	0.019***	−0.333***	1.097***	−0.038***	0.099***
	(0.010)	(0.019)	(0.019)	(0.003)	(0.007)	(0.006)	(0.037)	(0.088)	(0.011)	(0.009)
Realized Beta GARCH	0.093***	0.665***	0.266***	0.017**	0.001	0.017***	−0.320***	1.118***	−0.042***	0.119***
	(0.024)	(0.042)	(0.034)	(0.008)	(0.008)	(0.006)	(0.042)	(0.077)	(0.011)	(0.009)

Panel C: Correlation parameters (θ_2)

	a_{10}	b_{10}	c_{10}	ξ_{10}	φ_{10}
Realised FTS GARCH	0.011*	0.741***	0.247***	−0.041*	0.976***
	(0.006)	(0.020)	(0.036)	(0.022)	(0.103)
Realized Beta GARCH	0.011	0.717***	0.230***	−0.041	1.163***
	(0.009)	(0.027)	(0.060)	(0.030)	(0.229)

Panel D: Semi-covariance parameters (θ_3)

	ζ	ϕ_d	ϕ_w	ϕ_m	ψ_0	ψ_1
Realised FTS GARCH	−0.289***	0.089***	0.461***	0.339***	−0.319***	0.220***
	(0.039)	(0.014)	(0.025)	(0.027)	(0.011)	(0.012)

Notes: This table provides the in-sample estimates of parameter vector $\theta = (\theta_0', \theta_1', \theta_2', \theta_3')'$ for both the Realised FTS GARCH model and the Realized Beta GARCH model. Safe Haven against the stock market is gold. Models are estimated by Quasi-Maximum Likelihood. Standard errors are based on the sandwich estimator $\{\mathcal{J}_{2D}\mathcal{J}_{OP}^{-1}\mathcal{J}_{2D}\}^{-1}$ and are presented in parentheses below the estimated parameters. The parameter on the extra predictor m_{t-1}^{-} used in equation (3.2b) is found to be significant at the 10% level when the safe asset is the gold market.

* Statistical significance at the 10% level.
** Statistical significance at the 5% level.
*** Statistical significance at the 1% level.

Table 3.5 – Forecast Losses: SP - TY

	In-Sample				Out-of-Sample			
	Pointwise		Cumulative		Pointwise		Cumulative	
	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE
Panel A: Equity Volatility Forecasting								
1-d ahead	0.968	1.017	0.968	1.017	0.867	0.974	0.867	0.974
5-d ahead	0.915	0.967	0.863	0.966	1.013	0.973	1.032	0.966
10-d ahead	0.899	0.956	0.800	0.949	0.919	0.965	0.948	0.972
22-d ahead	0.946	0.938	0.792	0.930	0.958	0.935	0.877	0.950
Panel B: Safe Haven Volatility Forecasting								
1-d ahead	0.995	1.006	0.995	1.006	0.945	0.961	0.945	0.961
5-d ahead	0.994	0.989	0.973	0.986	0.932	0.962	0.802	0.928
10-d ahead	0.996	0.985	0.972	0.974	0.962	0.963	0.732	0.916
22-d ahead	0.989	0.965	0.964	0.948	0.982	0.948	0.772	0.924

Notes: This table contains the mean forecast losses of the Realised FTS GARCH model applied to US equity (SP) and bond (TY) futures data. We report results relative to the Realized Beta GARCH model. Entries below unity favor the former model. A rolling window estimation method is assumed for the out-of-sample results in the right panel.

Table 3.6 – Reality Check Test: SP - TY

	Pointwise Performance		Cumulative Performance	
	MSE	QLIKE	MSE	QLIKE
Panel A: Equity Volatility Forecasting				
1-d ahead	0.147	0.028	0.147	0.028
5-d ahead	0.628	0.003	0.676	0.004
10-d ahead	0.082	0.022	0.243	0.035
22-d ahead	0.064	0.004	0.037	0.007
Panel B: Safe Haven Volatility Forecasting				
1-d ahead	0.013	0.000	0.013	0.000
5-d ahead	0.035	0.002	0.032	0.000
10-d ahead	0.117	0.001	0.039	0.000
22-d ahead	0.000	0.000	0.043	0.000

Notes: This table provides the p-values of the Reality Check test conducted on the out-of-sample forecast losses for US equity (SP) and bond (TY) volatility. Small p-values indicate significant predictive superiority of the Realised FTS GARCH model over the benchmark model. We perform the test using a stationary bootstrap with 999 re-samplings and an average block length of five.

Table 3.7 – Forecast Losses: SP - GC

	In-Sample				Out-of-Sample			
	Pointwise		Cumulative		Pointwise		Cumulative	
	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE
Panel A: Equity Volatility Forecasting								
1-d ahead	0.840	0.993	0.840	0.993	1.048	0.957	1.048	0.957
5-d ahead	0.898	0.942	0.777	0.928	0.971	0.960	1.021	0.948
10-d ahead	0.905	0.926	0.763	0.901	0.931	0.945	0.948	0.938
22-d ahead	0.958	0.914	0.810	0.877	0.961	0.934	0.893	0.915
Panel B: Safe Haven Volatility Forecasting								
1-d ahead	1.033	0.992	1.033	0.992	1.022	1.010	1.022	1.010
5-d ahead	0.989	0.970	0.994	0.953	1.002	0.991	1.011	0.992
10-d ahead	1.000	0.956	1.000	0.929	1.003	1.000	1.011	0.996
22-d ahead	0.968	0.912	0.970	0.869	0.991	0.981	1.025	0.986

Notes: This table contains the mean forecast losses of the Realised FTS GARCH model applied to US equity (SP) and gold (GC) futures data. We report results relative to the Realized Beta GARCH model. Entries below unity favor the former model. A rolling window estimation method is assumed for the out-of-sample results in the right panel.

Table 3.8 – Reality Check Test: SP - GC

	Pointwise Performance		Cumulative Performance	
	MSE	QLIKE	MSE	QLIKE
Panel A: Equity Volatility Forecasting				
1-d ahead	0.770	0.002	0.770	0.002
5-d ahead	0.155	0.007	0.689	0.002
10-d ahead	0.066	0.021	0.152	0.011
22-d ahead	0.037	0.011	0.058	0.007
Panel B: Safe Haven Volatility Forecasting				
1-d ahead	0.804	0.872	0.804	0.872
5-d ahead	0.524	0.192	0.583	0.269
10-d ahead	0.529	0.539	0.554	0.435
22-d ahead	0.402	0.076	0.662	0.292

Notes: This table provides the p-values of the Reality Check test conducted on the out-of-sample forecast losses for US equity (SP) and gold (GC) volatility. Small p-values indicate significant predictive superiority of the Realised FTS GARCH model over the benchmark model. We perform the test using a stationary bootstrap with 999 re-samplings and an average block length of five.

Table 3.9 – Robustness - Forecast Losses: SP - TY

Panel A: Equity Volatility Forecasting

		RK	RV	SSRV	TSRV	MLE
Panel A1: Pointwise Performance						
1-d ahead	MSE	0.867	0.872	0.865	0.923	0.854
	QLIKE	0.974	0.992	0.988	0.982	0.980
5-d ahead	MSE	1.013	0.997	1.004	1.043	1.010
	QLIKE	0.973	0.975	0.976	0.976	0.975
10-d ahead	MSE	0.919	0.902	0.916	0.923	0.915
	QLIKE	0.965	0.968	0.967	0.969	0.967
22-d ahead	MSE	0.958	0.954	0.959	0.957	0.958
	QLIKE	0.935	0.932	0.932	0.936	0.931
Panel A2: Cumulative Performance						
5-d ahead	MSE	1.032	1.016	1.015	1.125	1.024
	QLIKE	0.966	0.978	0.977	0.973	0.971
10-d ahead	MSE	0.948	0.923	0.933	1.010	0.937
	QLIKE	0.972	0.978	0.979	0.977	0.976
22-d ahead	MSE	0.877	0.867	0.873	0.892	0.871
	QLIKE	0.950	0.954	0.955	0.956	0.952

Panel B: Safe Haven Volatility Forecasting

		RK	RV	SSRV	TSRV	MLE
Panel B1: Pointwise Performance						
1-d ahead	MSE	0.945	0.925	0.941	0.891	0.929
	QLIKE	0.961	0.923	0.926	0.943	0.941
5-d ahead	MSE	0.932	0.910	0.923	0.881	0.913
	QLIKE	0.962	0.948	0.944	0.954	0.949
10-d ahead	MSE	0.962	0.938	0.950	0.919	0.941
	QLIKE	0.963	0.944	0.942	0.949	0.948
22-d ahead	MSE	0.982	0.975	0.978	0.969	0.975
	QLIKE	0.948	0.939	0.936	0.943	0.941
Panel B2: Cumulative Performance						
5-d ahead	MSE	0.802	0.775	0.794	0.709	0.769
	QLIKE	0.928	0.881	0.884	0.902	0.892
10-d ahead	MSE	0.732	0.710	0.725	0.648	0.702
	QLIKE	0.916	0.872	0.877	0.889	0.880
22-d ahead	MSE	0.772	0.775	0.784	0.716	0.765
	QLIKE	0.924	0.903	0.907	0.905	0.904

Notes: This table provides, for each of the alternative realised variance estimators, the out-of-sample mean forecast losses of the Realised FTS GARCH model applied to US equity (SP) and bond (TY) futures data. We report results relative to the Realized Beta GARCH model. A rolling window estimation method is assumed in the out-of-sample exercise.

Table 3.10 – Robustness - Reality Check Test: SP - TY

Panel A: Equity Volatility Forecasting

		RK	RV	SSRV	TSRV	MLE
Panel A1: Pointwise Performance						
1-d ahead	MSE	0.147	0.121	0.135	0.229	0.141
	QLIKE	0.028	0.350	0.243	0.128	0.123
5-d ahead	MSE	0.628	0.496	0.557	0.790	0.610
	QLIKE	0.003	0.001	0.003	0.001	0.002
10-d ahead	MSE	0.082	0.080	0.084	0.094	0.086
	QLIKE	0.022	0.045	0.040	0.040	0.044
22-d ahead	MSE	0.064	0.070	0.064	0.065	0.057
	QLIKE	0.004	0.004	0.004	0.004	0.005
Panel A2: Cumulative Performance						
5-d ahead	MSE	0.676	0.617	0.607	0.859	0.636
	QLIKE	0.004	0.046	0.041	0.012	0.020
10-d ahead	MSE	0.243	0.172	0.186	0.590	0.194
	QLIKE	0.035	0.098	0.124	0.080	0.082
22-d ahead	MSE	0.037	0.037	0.038	0.034	0.036
	QLIKE	0.007	0.012	0.014	0.011	0.012

Panel B: Safe Haven Volatility Forecasting

		RK	RV	SSRV	TSRV	MLE
Panel B1: Pointwise Performance						
1-d ahead	MSE	0.013	0.033	0.032	0.007	0.034
	QLIKE	0.000	0.000	0.000	0.000	0.000
5-d ahead	MSE	0.035	0.062	0.059	0.028	0.063
	QLIKE	0.002	0.000	0.000	0.000	0.000
10-d ahead	MSE	0.117	0.073	0.102	0.047	0.085
	QLIKE	0.001	0.000	0.000	0.000	0.000
22-d ahead	MSE	0.000	0.003	0.002	0.000	0.000
	QLIKE	0.000	0.000	0.000	0.000	0.000
Panel B2: Cumulative Performance						
5-d ahead	MSE	0.032	0.045	0.046	0.025	0.043
	QLIKE	0.000	0.000	0.000	0.000	0.000
10-d ahead	MSE	0.039	0.044	0.045	0.025	0.043
	QLIKE	0.000	0.000	0.000	0.000	0.000
22-d ahead	MSE	0.043	0.046	0.048	0.027	0.044
	QLIKE	0.000	0.000	0.000	0.000	0.000

Notes: This table gives, for each of the alternative realised variance estimators, the p-values of the Reality Check test conducted on the out-of-sample forecast losses for US equity (SP) and bond (TY) volatility. Small p-values indicate significant predictive superiority of the Realised FTS GARCH model over the benchmark model. We perform the test using a stationary bootstrap with 999 re-samplings and an average block length of five.

Table 3.11 – Robustness - Forecast Losses: SP - GC

Panel A: Equity Volatility Forecasting

		RK	RV	SSRV	TSRV	MLE
Panel A1: Pointwise Performance						
1-d ahead	MSE	1.048	1.060	1.049	1.075	1.058
	QLIKE	0.957	0.985	0.989	0.964	0.980
5-d ahead	MSE	0.971	0.975	0.980	0.992	0.981
	QLIKE	0.960	0.974	0.977	0.965	0.972
10-d ahead	MSE	0.931	0.922	0.933	0.933	0.930
	QLIKE	0.945	0.954	0.956	0.948	0.951
22-d ahead	MSE	0.961	0.957	0.962	0.959	0.961
	QLIKE	0.934	0.937	0.940	0.934	0.936
Panel A2: Cumulative Performance						
5-d ahead	MSE	1.021	1.024	1.026	1.067	1.035
	QLIKE	0.948	0.979	0.984	0.957	0.975
10-d ahead	MSE	0.948	0.945	0.953	0.977	0.957
	QLIKE	0.938	0.964	0.969	0.945	0.960
22-d ahead	MSE	0.893	0.893	0.898	0.896	0.897
	QLIKE	0.915	0.934	0.937	0.919	0.929

Panel B: Safe Haven Volatility Forecasting

		RK	RV	SSRV	TSRV	MLE
Panel B1: Pointwise Performance						
1-d ahead	MSE	1.022	1.009	1.011	1.033	1.016
	QLIKE	1.010	0.998	0.998	1.014	1.000
5-d ahead	MSE	1.002	0.993	0.993	1.011	0.992
	QLIKE	0.991	0.976	0.974	0.991	0.974
10-d ahead	MSE	1.003	0.994	0.995	1.012	1.000
	QLIKE	1.000	0.981	0.985	1.000	0.987
22-d ahead	MSE	0.991	0.982	0.982	0.997	0.983
	QLIKE	0.981	0.959	0.960	0.976	0.962
Panel B2: Cumulative Performance						
5-d ahead	MSE	1.011	0.988	0.996	1.029	0.998
	QLIKE	0.992	0.962	0.964	0.995	0.964
10-d ahead	MSE	1.011	0.988	0.991	1.031	0.998
	QLIKE	0.996	0.963	0.965	0.998	0.965
22-d ahead	MSE	1.025	0.997	1.002	1.046	1.011
	QLIKE	0.986	0.947	0.950	0.986	0.953

Notes: This table provides, for each of the alternative realised variance estimators, the out-of-sample mean forecast losses of the Realised FTS GARCH model applied to US equity (SP) and gold (GC) futures data. We report results relative to the Realized Beta GARCH model. A rolling window estimation method is assumed in the out-of-sample exercise.

Table 3.12 – Robustness - Reality Check Test: SP - GC

Panel A: Equity Volatility Forecasting

		RK	RV	SSRV	TSRV	MLE
Panel A1: Pointwise Performance						
1-d ahead	MSE	0.770	0.790	0.784	0.791	0.794
	QLIKE	0.002	0.175	0.274	0.003	0.116
5-d ahead	MSE	0.155	0.195	0.209	0.381	0.235
	QLIKE	0.007	0.047	0.075	0.010	0.037
10-d ahead	MSE	0.066	0.072	0.073	0.064	0.075
	QLIKE	0.021	0.037	0.057	0.025	0.040
22-d ahead	MSE	0.037	0.043	0.044	0.039	0.041
	QLIKE	0.011	0.017	0.018	0.010	0.016
Panel A2: Cumulative Performance						
5-d ahead	MSE	0.689	0.718	0.726	0.783	0.740
	QLIKE	0.002	0.155	0.228	0.007	0.107
10-d ahead	MSE	0.152	0.137	0.160	0.359	0.190
	QLIKE	0.011	0.078	0.110	0.013	0.064
22-d ahead	MSE	0.058	0.054	0.060	0.062	0.064
	QLIKE	0.007	0.020	0.025	0.010	0.019

Panel B: Safe Haven Volatility Forecasting

		RK	RV	SSRV	TSRV	MLE
Panel B1: Pointwise Performance						
1-d ahead	MSE	0.804	0.642	0.656	0.838	0.686
	QLIKE	0.872	0.442	0.428	0.934	0.508
5-d ahead	MSE	0.524	0.398	0.393	0.632	0.398
	QLIKE	0.192	0.020	0.015	0.210	0.016
10-d ahead	MSE	0.529	0.408	0.421	0.634	0.499
	QLIKE	0.539	0.132	0.216	0.524	0.255
22-d ahead	MSE	0.402	0.284	0.290	0.477	0.337
	QLIKE	0.076	0.004	0.004	0.048	0.007
Panel B2: Cumulative Performance						
5-d ahead	MSE	0.583	0.378	0.448	0.696	0.466
	QLIKE	0.269	0.005	0.005	0.361	0.008
10-d ahead	MSE	0.554	0.404	0.419	0.678	0.467
	QLIKE	0.435	0.032	0.041	0.490	0.041
22-d ahead	MSE	0.662	0.487	0.520	0.744	0.565
	QLIKE	0.292	0.015	0.022	0.272	0.029

Notes: This table gives, for each of the alternative realised variance estimators, the p-values of the Reality Check test conducted on the out-of-sample forecast losses for US equity (SP) and gold (GC) volatility. Small p-values indicate significant predictive superiority of the Realised FTS GARCH model over the benchmark model. We perform the test using a stationary bootstrap with 999 re-samplings and an average block length of five.

Table 3.13 – Robustness - Forecast Losses: SP - TY

Panel A: Equity Volatility Forecasting

		Real FTS GARCH	AR	HAR-RV	GARCH	HEAVY	IHEAVY
Panel A1: Pointwise Performance							
1-d ahead	MSE	**0.751**	1.125	1.000	1.267	1.011	1.024
	QLIKE	**0.932**	1.155	1.000	1.385	1.026	1.027
5-d ahead	MSE	1.011	1.404	**1.000**	1.070	1.056	1.073
	QLIKE	**0.946**	1.667	1.000	1.098	0.966	0.993
10-d ahead	MSE	**0.972**	1.380	1.000	1.092	1.051	1.079
	QLIKE	**0.961**	2.004	1.000	1.021	0.987	1.018
22-d ahead	MSE	**0.913**	1.071	1.000	0.951	0.933	1.032
	QLIKE	0.953	1.857	1.000	**0.894**	0.952	0.975
Panel A2: Cumulative Performance							
5-d ahead	MSE	**0.935**	1.534	1.000	1.276	1.033	1.049
	QLIKE	**0.897**	1.516	1.000	1.267	0.955	0.988
10-d ahead	MSE	**0.924**	1.759	1.000	1.212	1.059	1.094
	QLIKE	**0.908**	1.802	1.000	1.171	0.952	0.992
22-d ahead	MSE	**0.976**	1.774	1.000	1.223	1.095	1.217
	QLIKE	**0.923**	1.989	1.000	1.040	0.974	1.022

Panel B: Bond Volatility Forecasting

		Real FTS GARCH	AR	HAR-RV	GARCH	HEAVY	IHEAVY
Panel B1: Pointwise Performance							
1-d ahead	MSE	1.020	1.138	**1.000**	1.042	1.043	1.054
	QLIKE	1.002	1.196	**1.000**	1.090	1.067	1.092
5-d ahead	MSE	1.018	1.170	**1.000**	1.019	1.036	1.051
	QLIKE	1.005	1.692	**1.000**	1.038	1.022	1.052
10-d ahead	MSE	1.022	1.177	**1.000**	1.018	1.041	1.059
	QLIKE	1.009	1.737	**1.000**	1.020	1.009	1.050
22-d ahead	MSE	0.935	1.021	1.000	**0.892**	0.944	0.97
	QLIKE	0.210	0.326	1.000	**0.200**	0.203	0.210
Panel B2: Cumulative Performance							
5-d ahead	MSE	1.089	1.498	**1.000**	1.107	1.140	1.191
	QLIKE	1.016	1.872	**1.000**	1.109	1.069	1.127
10-d ahead	MSE	1.125	1.953	**1.000**	1.149	1.229	1.325
	QLIKE	1.021	2.478	**1.000**	1.104	1.046	1.132
22-d ahead	MSE	1.207	2.379	**1.000**	1.119	1.336	1.518
	QLIKE	1.032	2.950	1.000	1.049	**0.988**	1.097

Notes: This table contains the out-of-sample mean forecast losses of the Realised FTS GARCH model applied to US equity (SP) and bond (TY) futures data, along with losses of the benchmark models. We report results relative to the HAR-RV model. A rolling window estimation method is assumed in the out-of-sample exercise.

Table 3.14 – Robustness - Reality Check Test: SP - TY

	Pointwise Performance		Cumulative Performance	
	MSE	QLIKE	MSE	QLIKE
Panel A: Equity Volatility Forecasting				
1-d ahead	0.076	0.000	0.076	0.000
5-d ahead	0.171	0.009	0.080	0.003
10-d ahead	0.071	0.002	0.023	0.002
22-d ahead	0.045	0.275	0.072	0.000
Panel B: Bond Volatility Forecasting				
1-d ahead	0.437	0.124	0.437	0.124
5-d ahead	0.420	0.180	0.627	0.240
10-d ahead	0.405	0.183	0.494	0.230
22-d ahead	0.707	0.342	0.520	0.287

Notes: This table provides the p-values of the Reality Check test conducted on the out-of-sample forecast losses for US equity (SP) and bond (TY) volatility. Small p-values indicate significant predictive superiority of the Realised FTS GARCH over the AR, HAR-RV, Targeting GARCH (1,1), Targeting HEAVY and Integrated HEAVY models. The test uses a stationary bootstrap with 999 re-samplings and an average block length of five.

Table 3.15 – Robustness - Forecast Losses: SP - GC

Panel A: Equity Volatility Forecasting

		Real FTS GARCH	AR	HAR-RV	GARCH	HEAVY	IHEAVY
Panel A1: Pointwise Performance							
1-d ahead	MSE	**0.871**	1.125	1.000	1.267	1.011	1.024
	QLIKE	**0.920**	1.155	1.000	1.385	1.026	1.027
5-d ahead	MSE	**0.964**	1.404	1.000	1.070	1.056	1.073
	QLIKE	**0.960**	1.667	1.000	1.098	0.966	0.993
10-d ahead	MSE	**0.973**	1.380	1.000	1.092	1.051	1.079
	QLIKE	**0.962**	2.004	1.000	1.021	0.987	1.018
22-d ahead	MSE	**0.909**	1.071	1.000	0.951	0.933	1.032
	QLIKE	0.960	1.857	1.000	**0.894**	0.952	0.975
Panel A2: Cumulative Performance							
5-d ahead	MSE	**0.900**	1.534	1.000	1.276	1.033	1.049
	QLIKE	**0.908**	1.516	1.000	1.267	0.955	0.988
10-d ahead	MSE	**0.913**	1.759	1.000	1.212	1.059	1.094
	QLIKE	**0.919**	1.802	1.000	1.171	0.952	0.992
22-d ahead	MSE	**0.977**	1.774	1.000	1.223	1.095	1.217
	QLIKE	**0.932**	1.989	1.000	1.040	0.974	1.022

Panel B: Gold Volatility Forecasting

		Real FTS GARCH	AR	HAR-RV	GARCH	HEAVY	IHEAVY
Panel B1: Pointwise Performance							
1-d ahead	MSE	1.134	1.052	**1.000**	1.115	1.079	1.113
	QLIKE	1.039	1.164	**1.000**	1.132	1.064	1.070
5-d ahead	MSE	1.150	1.457	**1.000**	1.077	1.015	1.052
	QLIKE	1.025	1.608	**1.000**	1.097	1.024	1.017
10-d ahead	MSE	1.089	1.508	1.000	1.065	**0.994**	1.054
	QLIKE	1.018	1.773	**1.000**	1.085	1.017	1.033
22-d ahead	MSE	0.540	0.657	1.000	0.512	**0.500**	0.558
	QLIKE	**0.934**	1.558	1.000	0.978	0.941	0.940
Panel B2: Cumulative Performance							
5-d ahead	MSE	1.367	1.627	**1.000**	1.178	1.027	1.105
	QLIKE	1.066	1.636	**1.000**	1.211	1.056	1.047
10-d ahead	MSE	1.394	2.158	**1.000**	1.229	1.015	1.145
	QLIKE	1.079	2.113	**1.000**	1.247	1.057	1.051
22-d ahead	MSE	1.223	2.168	1.000	1.103	**0.932**	1.141
	QLIKE	1.072	2.591	**1.000**	1.258	1.070	1.076

Notes: This table contains the out-of-sample mean forecast losses of the Realised FTS GARCH model applied to US equity (SP) and gold (GC) futures data, along with losses of the benchmark models. We report results relative to the HAR-RV model. A rolling window estimation method is assumed in the out-of-sample exercise.

Table 3.16 – Robustness - Reality Check Test: SP - GC

	Pointwise Performance		Cumulative Performance	
	MSE	QLIKE	MSE	QLIKE
Panel A: Equity Volatility Forecasting				
1-d ahead	0.044	0.000	0.044	0.000
5-d ahead	0.069	0.066	0.019	0.004
10-d ahead	0.069	0.003	0.017	0.005
22-d ahead	0.028	0.337	0.059	0.001
Panel B: Gold Volatility Forecasting				
1-d ahead	0.856	0.856	0.856	0.856
5-d ahead	0.788	0.413	0.945	0.659
10-d ahead	0.624	0.202	0.833	0.590
22-d ahead	0.285	0.048	0.708	0.454

Notes: This table provides the p-values of the Reality Check test conducted on the out-of-sample forecast losses for US equity (SP) and gold (GC) volatility. Small p-values indicate significant predictive superiority of the Realised FTS GARCH over the AR, HAR-RV, Targeting GARCH (1,1), Targeting HEAVY and Integrated HEAVY models. The test uses a stationary bootstrap with 999 re-samplings and an average block length of five.

Table 3.17 – Economic Value - Daily (no estimation risk)

	Realized Beta GARCH			Realised FTS GARCH				
Period	μ	σ	SR	μ	σ	SR	Δ_1	Δ_{10}
Panel A: Equity and Bond								
Entire Sample	9.199	10.905	0.844	9.215	10.792	0.854	2.793	13.867
2004 – 2007	5.886	9.945	0.592	5.985	9.896	0.605	6.559	14.287
2008 – 2010	6.787	15.305	0.443	7.648	15.002	0.510	34.582	53.529
2011 – 2013	14.045	8.604	1.632	13.444	8.628	1.558	10.845	24.586
2014 – 2016	11.003	8.329	1.321	10.658	8.317	1.281	1.795	12.957
Panel B: Equity and Gold								
Entire Sample	10.464	30.078	0.348	7.771	29.994	0.259	-266.728	-243.939
2004 – 2007	8.094	20.344	0.398	6.729	20.242	0.332	-200.576	-179.817
2008 – 2010	-9.293	47.106	-0.197	-16.762	46.903	-0.357	-379.514	-337.201
2011 – 2013	26.617	25.646	1.038	25.850	25.680	1.007	-304.019	-274.256
2014 – 2016	17.614	20.015	0.880	16.255	20.025	0.812	-270.439	-246.989

Notes: This table summarises the performance of daily volatility-timing strategies for the case of no estimation risk regarding expected returns. The two volatility-timing strategies in Panel A are based on the next-day forecasts of US equity (SP) and bond (TY) return variance and covariance, while those in Panel B are based on the next-day forecasts of US equity (SP) and gold (GC) return variance and covariance. Each day, we solve a portfolio optimisation problem in which the expected return for each asset equals its in-sample mean return and the conditional covariance matrix is estimated out-of-sample using the Realized Beta GARCH model for one strategy, and the Realised FTS GARCH model for the second strategy. In each case, we solve for the portfolio weights that minimise conditional volatility subject to a target expected return of 10%. We report the annualised (percentage) mean return (μ), volatility (σ), and Sharpe ratio (SR) for each strategy, and the annualised basis point fees (Δ_γ) that an investor with quadratic utility and constant relative risk aversion of γ would be willing to pay to switch from the volatility-timing strategy based on the forecasts offered by the Realized Beta GARCH model to that based on the forecasts offered by the Realised FTS GARCH model. We report results for the entire out-of-sample period, that is from March 5, 2004 to September 30, 2016, and for a range of subperiods.

Table 3.18 – Economic Value - Daily

	Realized Beta GARCH			Realised FTS GARCH					
k	μ	σ	SR	μ	σ	SR	p-val	Δ_1	Δ_{10}
Panel A: Equity and Bond									
1000	8.465	10.573	0.804	8.553	10.459	0.822	0.979	10.512	25.451
2000	9.064	11.231	0.811	9.154	11.106	0.828	0.984	10.670	25.653
3000	9.314	11.452	0.817	9.402	11.323	0.834	0.982	11.018	31.026
4000	9.350	11.416	0.821	9.431	11.285	0.837	0.986	9.770	24.846
5000	9.393	11.412	0.824	9.468	11.281	0.840	0.974	9.159	23.764
10000	9.172	11.020	0.832	9.227	10.897	0.847	0.984	6.894	19.628
Panel B: Equity and Gold									
1000	6.816	19.481	0.348	5.469	19.446	0.280	0.000	-133.690	-124.713
2000	8.686	24.469	0.352	6.991	24.433	0.283	0.000	-170.605	-185.643
3000	9.378	26.620	0.351	7.559	26.575	0.284	0.000	-180.355	-167.000
4000	9.678	27.809	0.345	7.755	27.756	0.276	0.000	-189.745	-166.026
5000	9.481	26.875	0.353	7.597	26.830	0.283	0.000	-186.887	-173.229
10000	10.019	29.178	0.345	8.003	29.117	0.277	0.000	-199.258	-178.228

Notes: This table summarises the performance of daily volatility-timing strategies for the case of estimation risk regarding expected returns. The two volatility-timing strategies in Panel A are based on the next-day forecasts of US equity (SP) and bond (TY) return variance and covariance, while those in Panel B are based on the next-day forecasts of US equity (SP) and gold (GC) return variance and covariance. The results for each line in the table are based on 1000 simulation trials using a bootstrap sample of k daily returns to estimate the unconditional expected returns. For each trial, we solve a day-to-day portfolio optimisation problem in which the expected return for each asset equals the mean return from the bootstrap sample and the conditional covariance matrix is estimated out-of-sample using the Realized Beta GARCH model for one strategy, and the Realised FTS GARCH model for the second strategy. In each case, we solve for the portfolio weights that minimise conditional volatility subject to a target expected return of 10%. We report the average, over the number of trials, of the annualised (percentage) mean return (μ), volatility (σ), and Sharpe ratio (SR) for each strategy, the proportion of trials (p-val) in which using the forecasts of the Realised FTS GARCH earns a higher Sharpe ratio than using the forecasts of the benchmark model, and the average annualised basis point fees (Δ_γ) that an investor with quadratic utility and constant relative risk aversion of γ would be willing to pay to switch from the volatility-timing strategy based on the benchmark model to that based on the Realised FTS GARCH model.

Table 3.19 – Economic Value - Multi-horizon

h	Realized Beta GARCH μ	Realized Beta GARCH σ	Realized Beta GARCH SR	Realised FTS GARCH μ	Realised FTS GARCH σ	Realised FTS GARCH SR	p-val	Δ_1	Δ_{10}
Panel A: Equity and Bond									
1-d ahead	9.350	11.416	0.821	9.431	11.285	0.837	0.986	9.770	24.846
5-d ahead	9.819	10.931	0.898	10.060	10.851	0.927	1.000	24.948	33.219
10-d ahead	10.039	10.919	0.919	10.327	10.823	0.954	1.000	29.920	40.085
22-d ahead	10.284	12.481	0.824	10.318	12.377	0.834	1.000	4.674	17.882
Panel B: Equity and Gold									
1-d ahead	9.678	27.809	0.345	7.755	27.756	0.276	0.000	-189.745	-166.026
5-d ahead	11.240	27.073	0.413	9.823	27.306	0.357	0.053	-148.360	-209.968
10-d ahead	10.633	27.564	0.385	9.697	27.783	0.348	0.038	-99.854	-159.213
22-d ahead	10.971	27.682	0.396	10.894	27.864	0.391	0.194	-12.941	-66.644

Notes: This table summarises the performance of volatility-timing strategies measured over various non-overlapping horizons. The portfolios are rebalanced daily (resp. weekly, biweekly, and monthly) based on 1-day (resp. 5-day, 10-day, and 22-day) ahead (cumulative) forecasts of the conditional covariance matrix of returns. The two volatility-timing strategies in Panel A are based on the forecasts of US equity (SP) and bond (TY) return variance and covariance, while those in Panel B are based on the forecasts of US equity (SP) and gold (GC) return variance and covariance. The results in the table are based on 1000 simulation trials using a bootstrap sample of $k = 4000$ daily (resp. weekly, biweekly, and monthly) returns to estimate the unconditional daily (resp. weekly, biweekly, and monthly) expected returns. For each trial, we solve at the rebalancing frequency a portfolio optimisation problem in which the expected return for each asset equals the mean return from the bootstrap sample and the conditional covariance matrix is estimated out-of-sample using the Realized Beta GARCH model for one strategy, and the Realised FTS GARCH model for the second strategy. In each case, we solve for the portfolio weights that minimise conditional volatility subject to a target expected return of 10%. For each horizon, we report the average, over the number of trials, of the annualised (percentage) mean return (μ), volatility (σ), and Sharpe ratio (SR) for each strategy, the proportion of trials (p-val) in which using the forecasts of the Realised FTS GARCH earns a higher Sharpe ratio than using the forecasts of the benchmark model, and the average annualised basis point fees (Δ_γ) that an investor with quadratic utility and constant relative risk aversion of γ would be willing to pay to switch from the volatility-timing strategy based on the forecasts of the benchmark model to that based on the forecasts of the Realised FTS GARCH model.

Table 3.20 – Economic Value - Transaction Costs: SP - TY

	Realized Beta GARCH			Realised FTS GARCH					
c	μ	σ	SR	μ	σ	SR	p-val	Δ_1	Δ_{10}
Panel A: Daily Rebalancing									
0.0%	9.350	11.416	0.821	9.431	11.285	0.837	0.986	9.770	24.846
2.5%	8.786	11.211	0.785	8.967	11.082	0.810	1.000	19.728	34.742
5.0%	8.362	11.214	0.747	8.649	11.085	0.782	1.000	30.363	45.482
7.5%	7.937	11.217	0.709	8.330	11.087	0.753	1.000	40.998	56.225
10.0%	7.512	11.220	0.671	8.012	11.089	0.724	1.000	51.633	66.970
12.5%	7.088	11.223	0.633	7.693	11.091	0.696	1.000	62.268	77.717
15.0%	6.663	11.227	0.596	7.375	11.094	0.667	1.000	72.904	88.467
17.5%	6.239	11.230	0.558	7.057	11.096	0.638	1.000	83.540	99.220
20.0%	5.814	11.233	0.520	6.738	11.098	0.609	1.000	94.176	109.975
Panel B: Weekly Rebalancing									
0.0%	9.819	10.931	0.898	10.060	10.851	0.927	1.000	24.948	33.219
2.5%	9.232	10.941	0.844	9.523	10.860	0.877	1.000	30.000	38.406
5.0%	8.645	10.952	0.789	8.986	10.869	0.827	1.000	35.052	43.599
7.5%	8.057	10.963	0.735	8.449	10.879	0.777	1.000	40.105	48.798
10.0%	7.470	10.975	0.681	7.912	10.889	0.727	1.000	45.159	54.003
12.5%	6.883	10.987	0.627	7.375	10.900	0.677	1.000	50.213	59.215
15.0%	6.296	11.000	0.572	6.838	10.911	0.627	1.000	55.268	64.433
17.5%	5.709	11.014	0.518	6.302	10.923	0.577	1.000	60.323	69.657
20.0%	5.121	11.028	0.465	5.765	10.935	0.527	1.000	65.379	74.888
Panel C: Monthly Rebalancing									
0.0%	10.284	12.481	0.824	10.318	12.377	0.834	1.000	4.674	17.882
2.5%	9.706	12.504	0.776	9.805	12.389	0.791	1.000	11.367	25.702
5.0%	9.126	12.530	0.728	9.291	12.405	0.749	1.000	18.063	33.536
7.5%	8.547	12.559	0.681	8.778	12.423	0.707	1.000	24.760	41.384
10.0%	7.968	12.591	0.633	8.264	12.444	0.664	1.000	31.459	49.246
12.5%	7.388	12.626	0.585	7.750	12.469	0.622	1.000	38.159	57.121
15.0%	6.809	12.664	0.538	7.236	12.496	0.579	1.000	44.861	65.008
17.5%	6.229	12.706	0.490	6.722	12.526	0.537	1.000	51.564	72.908
20.0%	5.649	12.750	0.443	6.208	12.560	0.494	1.000	58.269	80.820

Notes: This table summarises the performance, net of transaction costs, of volatility-timing strategies based on the forecasts of US equity (SP) and bond (TY) return variance delivered by the Realized Beta GARCH and Realised FTS GARCH model. The portfolios are rebalanced daily (resp. weekly and monthly) based on 1-day (resp. 5-day and 22-day) ahead (cumulative) forecasts of the conditional covariance matrix. The c column indicates the annualised transaction costs paid if the entire portfolio is traded every day (resp. week and month) during the whole year. The results in the table are based on 1000 simulation trials using a bootstrap sample of $k = 4000$ daily (resp. weekly and monthly) returns to estimate the unconditional daily (resp. weekly and monthly) expected returns. For each trial, we solve for the portfolio weights that minimise conditional volatility subject to a target return of 10%. For each level of transaction costs, we report the average, over the number of trials, of the annualised (percentage) mean return (μ), volatility (σ), and Sharpe ratio (SR) for each strategy, the proportion of trials (p-val) in which using the forecasts of the Realised FTS GARCH earns a higher Sharpe ratio than using the forecasts of the benchmark model, and the average annualised basis point fees (Δ_γ) that an investor with quadratic utility and constant relative risk aversion of γ would be willing to pay to switch from the volatility-timing strategy based on the benchmark model to that based on the Realised FTS GARCH model.

Table 3.21 – Economic Value - Transaction Costs: SP - GC

	Realized Beta GARCH			Realised FTS GARCH					
c	μ	σ	SR	μ	σ	SR	p-val	Δ_1	Δ_{10}
Panel A: Daily Rebalancing									
0.0%	9.678	27.809	0.345	7.755	27.756	0.276	0.000	-189.745	-166.026
2.5%	8.903	27.007	0.334	7.122	26.956	0.268	0.000	-175.188	-149.170
5.0%	8.493	27.015	0.320	6.796	26.963	0.257	0.000	-166.872	-140.825
7.5%	8.083	27.022	0.305	6.469	26.971	0.246	0.000	-158.583	-132.785
10.0%	7.673	27.030	0.291	6.142	26.979	0.234	0.000	-150.321	-125.065
12.5%	7.264	27.038	0.276	5.815	26.988	0.223	0.000	-142.086	-117.686
15.0%	6.854	27.047	0.262	5.489	26.997	0.212	0.000	-133.879	-110.673
17.5%	6.444	27.056	0.248	5.162	27.006	0.200	0.000	-125.699	-104.055
20.0%	6.034	27.065	0.233	4.835	27.016	0.189	0.000	-117.546	-97.868
Panel B: Weekly Rebalancing									
0.0%	11.240	27.073	0.413	9.823	27.306	0.357	0.053	-148.360	-209.968
2.5%	10.620	27.084	0.391	9.251	27.322	0.337	0.062	-143.700	-206.579
5.0%	10.000	27.096	0.368	8.679	27.339	0.316	0.071	-139.040	-203.188
7.5%	9.380	27.109	0.346	8.107	27.355	0.296	0.082	-134.380	-199.795
10.0%	8.761	27.121	0.323	7.536	27.372	0.275	0.087	-129.720	-196.398
12.5%	8.141	27.134	0.301	6.964	27.389	0.255	0.090	-125.061	-192.999
15.0%	7.521	27.147	0.278	6.392	27.406	0.234	0.098	-120.401	-189.597
17.5%	6.901	27.160	0.256	5.820	27.424	0.214	0.103	-115.742	-186.192
20.0%	6.281	27.174	0.234	5.248	27.442	0.193	0.113	-111.083	-182.784
Panel C: Monthly Rebalancing									
0.0%	10.971	27.682	0.396	10.894	27.864	0.391	0.194	-12.941	-66.644
2.5%	10.249	27.729	0.370	10.223	27.923	0.366	0.261	-8.206	-65.326
5.0%	9.526	27.778	0.343	9.552	27.984	0.342	0.351	-3.464	-63.991
7.5%	8.804	27.829	0.317	8.881	28.046	0.317	0.474	1.292	-62.631
10.0%	8.082	27.881	0.291	8.210	28.110	0.293	0.639	6.076	-61.231
12.5%	7.359	27.934	0.264	7.540	28.175	0.269	0.776	10.908	-59.771
15.0%	6.636	27.990	0.238	6.869	28.242	0.244	0.851	15.760	-58.280
17.5%	5.912	28.047	0.212	6.198	28.311	0.220	0.934	20.603	-56.795
20.0%	5.189	28.105	0.186	5.526	28.381	0.196	0.996	25.369	-55.404

Notes: This table summarises the performance, net of transaction costs, of volatility-timing strategies based on the forecasts of US equity (SP) and gold (GC) return variance delivered by the Realized Beta GARCH and Realised FTS GARCH model. The portfolios are rebalanced daily (resp. weekly and monthly) based on 1-day (resp. 5-day and 22-day) ahead (cumulative) forecasts of the conditional covariance matrix. The c column indicates the annualised transaction costs paid if the entire portfolio is traded every day (resp. week and month) during the whole year. The results in the table are based on 1000 simulation trials using a bootstrap sample of $k = 4000$ daily (resp. weekly and monthly) returns to estimate the unconditional daily (resp. weekly and monthly) expected returns. For each trial, we solve for the portfolio weights that minimise conditional volatility subject to a target return of 10%. For each level of transaction costs, we report the average, over the number of trials, of the annualised (percentage) mean return (μ), volatility (σ), and Sharpe ratio (SR) for each strategy, the proportion of trials (p-val) in which using the forecasts of the Realised FTS GARCH earns a higher Sharpe ratio than using the forecasts of the benchmark model, and the average annualised basis point fees (Δ_γ) that an investor with quadratic utility and constant relative risk aversion of γ would be willing to pay to switch from the volatility-timing strategy based on the benchmark model to that based on the Realised FTS GARCH model.

3.7 Figures of Chapter 3

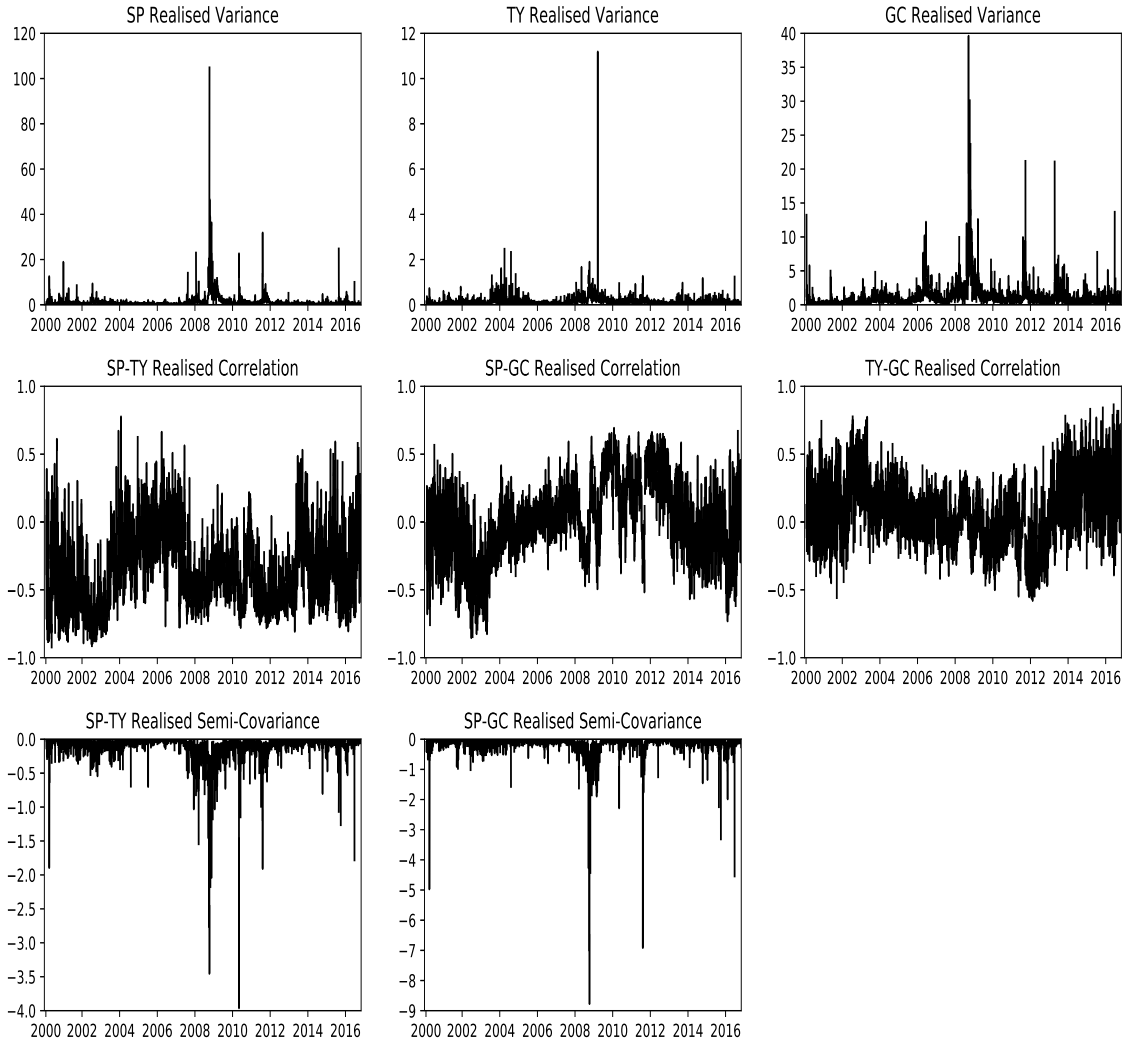


Figure 3.1 – Time series plots for daily realised measures of the S&P 500 index futures (SP), the 10-Year US Treasury Note futures (TY) and the gold futures (GC). Realised variances and semi-covariances are multiplied by 10,000. Top panel: realised variance of the stock, bond and gold markets. Middle panel: realised correlation between the stock and the bond market, between the stock and the gold market, and between the bond and the gold market. Bottom panel: realised semi-covariance between the falling stock market and the rising bond market, and between the falling stock market and the rising gold market.

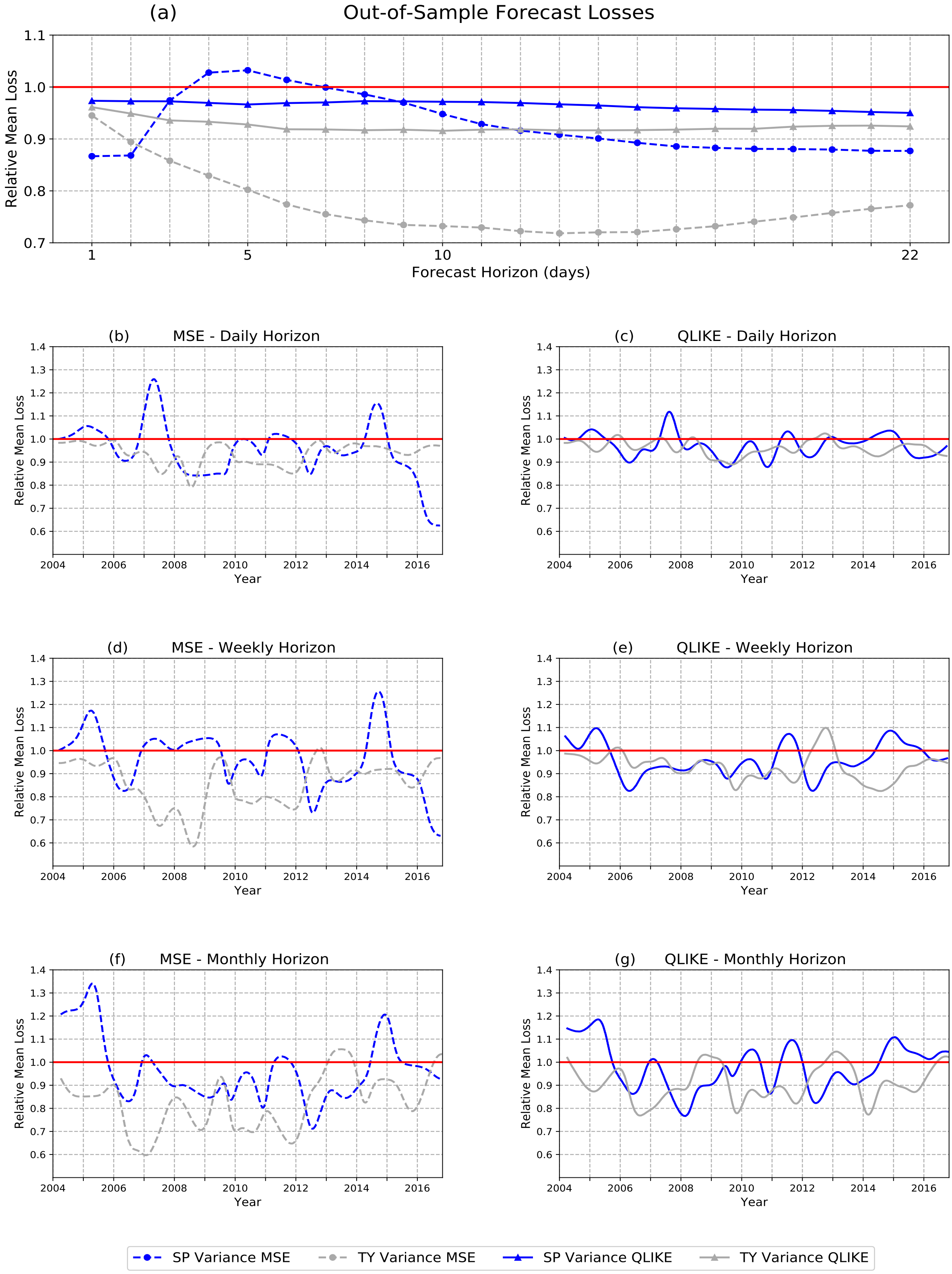


Figure 3.2 – This figure illustrates the forecast losses for the Realised FTS GARCH model relative to the Realized Beta GARCH model (red line). Results are based on the out-of-sample cumulative forecasts of US equity (SP, blue line) and bond (TY, grey line) return volatility. Panel (a) plots the relative mean losses against the forecast horizons. The remaining panels depict the time-varying mean forecast losses for the daily, weekly and monthly horizon. The solid lines refer to QLIKE loss function, while the dashed lines refer to MSE function.

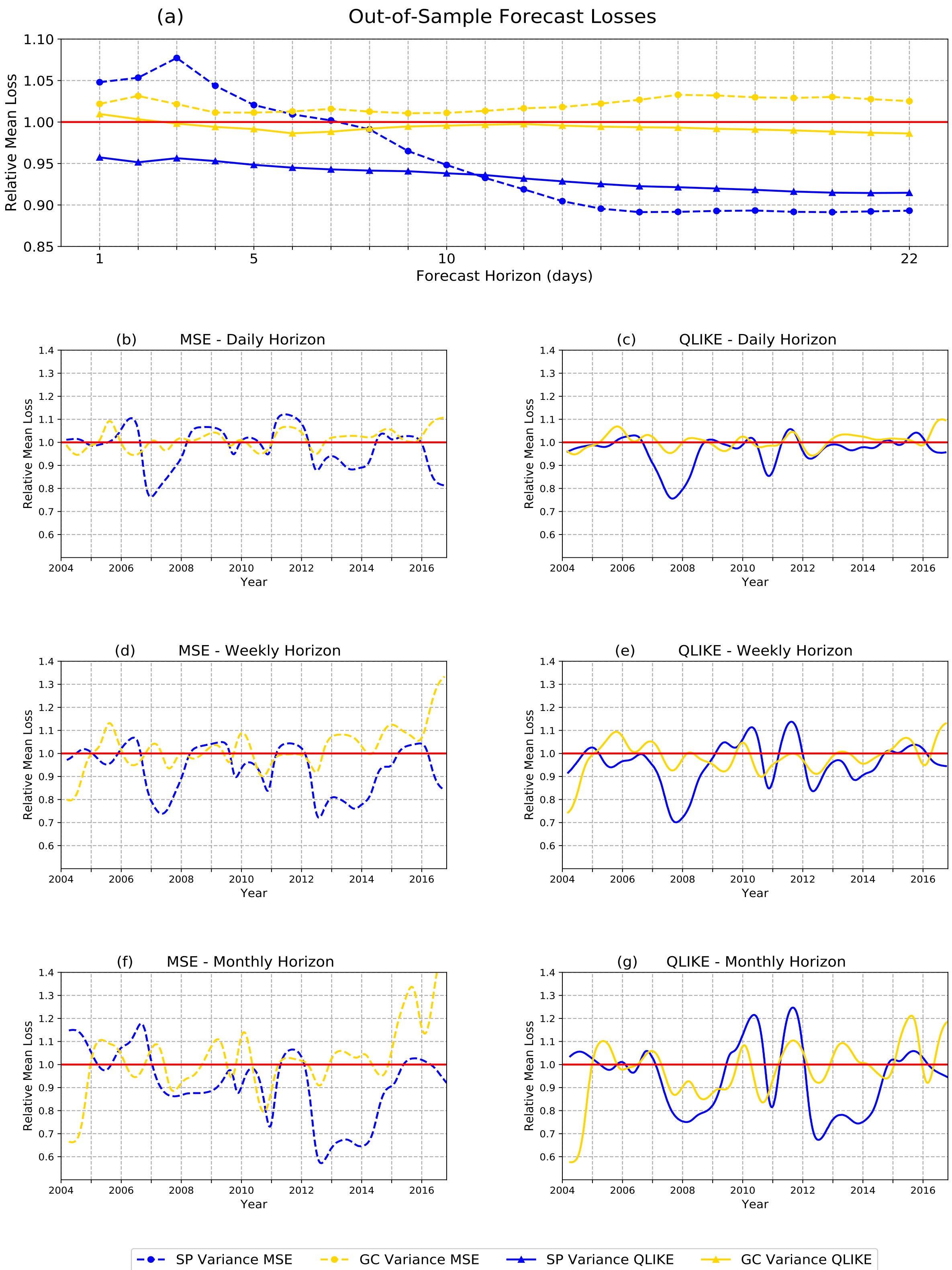


Figure 3.3 – This figure illustrates the forecast losses for the Realised FTS GARCH model relative to the Realized Beta GARCH model (red line). Results are based on the out-of-sample cumulative forecasts of US equity (SP, blue line) and gold (GC, yellow line) return volatility. Panel (a) plots the relative mean losses against the forecast horizons. The remaining panels depict the time-varying mean forecast losses for the daily, weekly and monthly horizon. The solid lines refer to QLIKE loss function, while the dashed lines refer to MSE function.

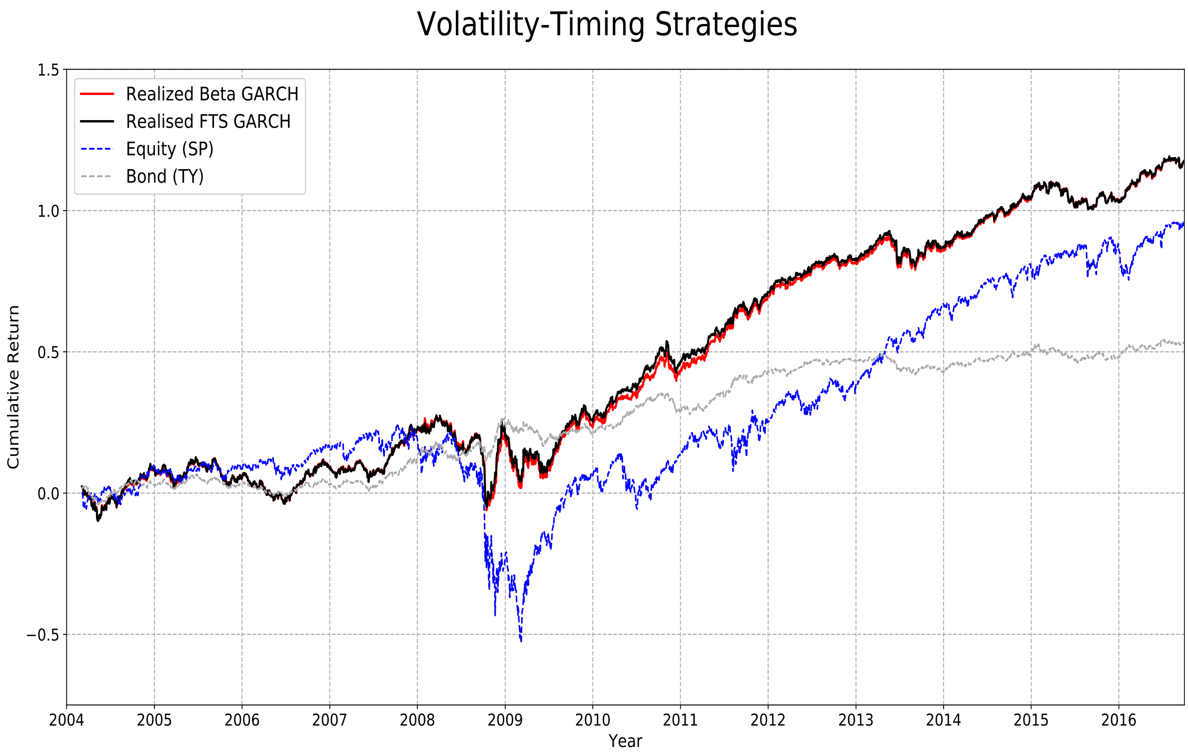


Figure 3.4 – This figure illustrates the cumulative returns of the S&P 500 index futures (SP) portfolio (dashed blue line), the 10-Year US Treasury Note futures (TY) portfolio (dashed grey line), a portfolio of SP and TY formed using the estimates of the conditional covariance matrix delivered by the Realized Beta GARCH model (solid red line), and a portfolio of SP and TY formed using the estimates of the conditional covariance matrix delivered by the Realised FTS GARCH model (solid black line). The portfolios of the volatility-timing strategies are rebalanced daily based on 1-day ahead forecasts of SP and TY return variance and covariance. Each day, we solve a portfolio optimisation problem in which the expected return for each asset equals its in-sample mean return and the conditional covariance matrix is estimated out-of-sample using the Realized Beta GARCH model for one strategy, and the Realised FTS GARCH model for the second strategy. In each case, we solve for the portfolio weights that minimise conditional volatility subject to a target expected return of 10%.

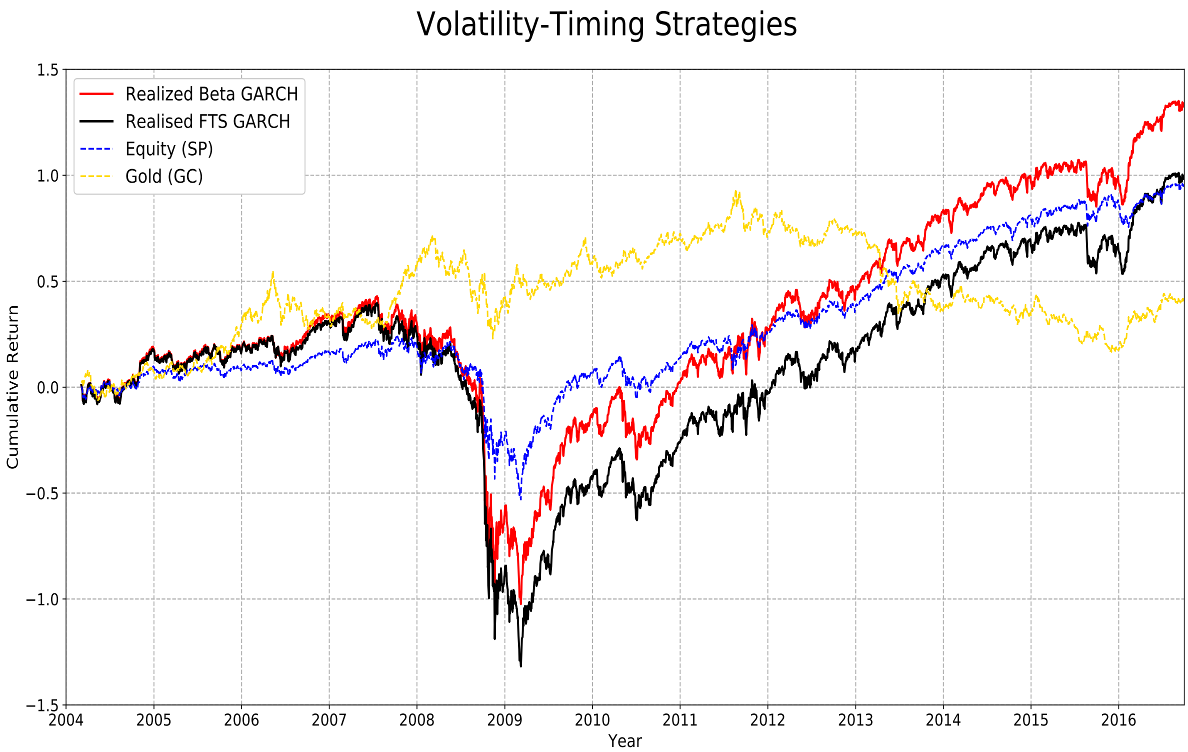


Figure 3.5 – This figure illustrates the cumulative returns of the S&P 500 index futures (SP) portfolio (dashed blue line), the gold futures (GC) portfolio (dashed yellow line), a portfolio of SP and GC formed using the estimates of the conditional covariance matrix delivered by the Realized Beta GARCH model (solid red line), and a portfolio of SP and GC formed using the estimates of the conditional covariance matrix delivered by the Realised FTS GARCH model (solid black line). The portfolios of the volatility-timing strategies are rebalanced daily based on 1-day ahead forecasts of SP and GC return variance and covariance. Each day, we solve a portfolio optimisation problem in which the expected return for each asset equals its in-sample mean return and the conditional covariance matrix is estimated out-of-sample using the Realized Beta GARCH model for one strategy, and the Realised FTS GARCH model for the second strategy. In each case, we solve for the portfolio weights that minimise conditional volatility subject to a target expected return of 10%.

Chapter 4

The Impact of Equity Tail Risk on Bond Risk Premia: Evidence of FTS in the US Term Structure

Abstract

This chapter quantifies the effects of equity tail risk on the term structure of US government securities. We combine the downside jump intensity factors of international stock market indices into a single measure of equity tail risk and show that it is strongly priced in an affine term structure model for the US interest rates. Consistent with the theory of flight-to-safety, we find that the response of Treasury bond yields and future excess returns to a contemporaneous shock to the equity tail factor is negative and opposite to what happens in the stock market. The significance of these results decreases with the maturity of the bonds, suggesting that the short end of the US yield curve is more strongly affected by flight-to-safety than the long end.

JEL Codes: C52, C58, G12, E43.

Keywords: Flight to safety, bond risk premium, equity tail risk, affine term structure model.

4.1 Introduction

In times of financial distress, the disengagement from risky assets, such as stocks, and the simultaneous demand for a safe haven, such as top-tier government bonds, generate a flight-to-safety (FTS) event in the capital markets. A large body of literature examines the linkages between the stock and bond markets during crisis periods and their implications for asset pricing, see Hartmann et al. (2004), Vayanos (2004), Chordia et al. (2005), Connolly et al. (2005) and Adrian et al. (2018), among others. We add to this literature by estimating a model of the term structure of interest rates that incorporates the effect of extreme events happening in the stock market. If Treasury bonds are a major beneficiary of the FTS flows occurring when the stock market is hit by heavy losses, then we expect the downside tail risk of equity to affect bond risk premia and determine both stock and bond prices during distress periods. We investigate this conjecture by considering a Gaussian affine term structure model (ATSM) for US interest rates where the pricing factors are the principal components of the yield curve and an equity left tail factor derived from options on international stock market indices. This provides, to the best of our knowledge, the first evidence for the importance of equity tail risk in bond pricing.

Understanding the dynamics of bond yields is particularly useful for forecasting financial and macro variables, for making debt and monetary policy decisions and for derivative pricing. Most of these applications require the decomposition of yields into expectations of future short rates (averaged over the lifetime of the bond) and term premia, i.e. the additional returns required by investors for bearing the risk of long-term commitment. Gaussian affine term structure models have long been used for this purpose, see, e.g., Duffee (2002), Kim and Wright (2005) and Abrahams et al. (2016). In the setup of a Gaussian ATSM, a number of pricing factors that affect bond yields are selected and assumed to evolve according to a vector autoregressive (VAR) process of order one. The yields of different maturities are all expressed as linear functions of the factors with restrictions on the coefficients that prevent arbitrage opportunities, implying that long-term yields are merely risk-adjusted expectations of future short rates.

The selection of pricing factors typically starts by extracting from the cross-section of bond yields a given number of principal components (PCs), which are linear combinations of the rates themselves. The first three PCs are prime candidates as they generally explain over 99% of the variability in the term structure and, due to their loadings on yields, may be interpreted as the level, slope and curvature factor. However, it is well established in the literature that additional factors are needed to explain the cross-section of bond returns. For this reason, the first five principal components of the US Treasury yield curve are used as pricing factors in Adrian et al. (2013), while Malik and Meldrum (2016) adopt a four-factor specification for UK government bond yields. Furthermore, several studies suggest that a great deal of information about bond risk premia can be found in factors that are not principal components of the yield curve. Cochrane and Piazzesi (2005) discover a new linear combination of forward rates which is a strong predictor of future excess bond returns and, based on this evidence, Cochrane and Piazzesi (2008) use it in an ATSM along with the classical level, slope and curvature factors. More recently, Duffee (2011) and Joslin et al. (2014) show that valuable information about bond premia is located outside of the yield curve and contained, for example, in macro variables that have little or no impact on current yields but strong predictive power for future bond returns.

This chapter explores the use of factors, other than combinations of yields, to drive the curve of US Treasury rates and explain bond returns. In contrast to earlier work, however, we draw on the literature that deals with comovements in the equity and bond markets and we consider the possibility that pricing factors of Treasury bonds originate also in the stock market. The findings of Connolly et al. (2005) and Baele et al. (2010) suggest that measures linked to stock market uncertainty explain time variation in the stock-bond return relation and have important cross-market pricing effects.[1] Therefore, we select a risk measure which is known to predict equity returns and we examine its role in an ATSM. The existing literature suggests that the variance risk premium (VRP) forecasts stock returns at shorter horizons than other predictors like dividend yields or price-to-earning ratios, see Bollerslev et al. (2009), Bollerslev et al. (2014) and Bekaert and Hoerova (2014), among others. In view of recent studies showing that the

[1] Connolly et al. (2005) find that when the implied volatility from equity index options, measured by the VIX, increases to a considerable extent, bond returns tend to be higher than stock returns (flight-to-quality) and the correlation between the two assets over the next month is lower. Baele et al. (2010) show that the time-varying and sometimes negative stock-bond return correlations cannot be explained by macro variables but instead by liquidity factors and the variance risk premium, which represents the compensation demanded by investors for bearing variance risk and is defined as the difference between the risk-neutral and statistical expectations of the future return variation. Although the variance risk premium is a major contributor to the stock-bond return correlation dynamics, Baele et al. (2010) find significant exposures to it only for stock but not for bond returns.

predictive power of the VRP for future equity returns stems from a jump tail risk component (see, for example, Bollerslev et al. (2015) and Andersen et al. (2017a,b)), we opt for the jump intensity factor extracted from the Andersen et al. (2015b) model to assess the impact of equity tail events on US Treasury bonds. Hence, our main contribution is to integrate the downside tail risk of the stock market into the study of Treasury bond risk premia.

Caballero et al. (2017) describe US government securities as global stores of value in what they call "safe asset shortage". They illustrate how US government debt rose between 2007 and 2011 in response to an increased demand for safe assets. During the same years, however, the supply of safe assets contracted since debt from fiscally weak sovereigns, such as Italy and Spain, stopped being perceived as safe. The observed surge in cross-border purchases of safe assets indicate that Treasury bonds may react to tail events that originate in the stock market of countries other than the US. Therefore, in order to analyse the effect of the international stock market, we follow the methodology of Bollerslev et al. (2014) and define the equity tail risk measure of this study as the market capitalization weighted average of the downside jump intensity factors extracted from US, UK and Euro-zone equity-index options.

Our empirical analysis relies on monthly data for the US zero-coupon yield curve and the S&P 500, FTSE 100 and EURO STOXX 50 index options over the period 2007 to 2017. We obtain the equity tail factor from option data, which are well known to embed rich information about the pricing of extreme events, and then we estimate an ATSM as in Adrian et al. (2013) but including also the equity tail factor. Overall, the results show that equity jump tail risk is strongly priced within the term structure model. This risk premium in the Treasury market is consistent with the evidence in Krishnamurthy and Vissing-Jorgensen (2012), who document the existence of a significant price for the safety attribute of Treasuries. Further, we find that bond prices, which move inversely to yields, increase and future expected excess returns shrink in response to a contemporaneous shock to the equity tail factor. These observations confirm the role of US Treasuries as a safe haven and, when combined with the previously documented positive relationship between jump tail risk and future equity returns, indicate the presence of a common predictor across the two asset classes. This remark is in line with the findings reported by Adrian et al. (2018), who show that the same nonlinear function of the VIX can forecast both stock and bond returns, but in opposite directions as predicted by the theory of FTS. Finally, we observe that the predictive power of the equity tail factor for lower bond returns is statistically significant only at the short end of the US yield curve. Based on this evidence, we claim that short-term bonds are more sensitive to flight-to-safety than are long-term bonds.

This study is related to the work of Kaminska and Roberts-Sklar (2015), who document the importance of global market sentiment for the term structure of UK government bonds. The authors observe that future excess returns on UK bonds load positively on a VRP-based proxy of risk aversion. Their results are consistent with the findings of Bekaert et al. (2010), who show that both equity and bond premia increase with risk aversion, but contrast with the negative relationship between US bond returns and our measure of downside equity tail risk.

The remainder of the chapter is structured as follows. In Section 4.2 we review the methodology used to identify a left tail factor for the stock market. Section 4.3 outlines the term structure modelling approach. Section 4.4 covers the empirical application of equity tail risk in an ATSM. Section 4.5 concludes.

4.2 Equity Left Tail Factor

This section summarises the estimation of the equity tail risk measure whose impact on US Treasuries is discussed later in the chapter. This measure, which we denote by $\widetilde{U}^{Equity}$, is obtained as the market capitalization weighted average of downside jump intensity factors driving the returns of international stock market indices. To identify each of these index-specific factors, we rely on the Three-Factor Double Exponential Model proposed for option pricing by Andersen et al. (2015b).[2] The authors specify a parametric model for the risk-neutral dynamics of equity-index returns that includes two volatility factors, V_1 and V_2, plus a separate jump intensity factor, U, which is capable of detecting the priced downside risk in the option surface.[3] The model features two separate jump components: one captures co-jumps in the level of the

[2]The interested reader is directed to Andersen et al. (2015b) for an in-depth description of the formulation since here we limit ourselves to highlighting the distinctive features.

[3]Nonparametric and seminonparametric approaches to estimating a tail risk measure from option data are also available, see, e.g., Aït-Sahalia and Lo (2000), Bollerslev and Todorov (2011), Bollerslev et al. (2015) and Andersen et al. (2017a). For tail risk measures that are computed from the cross-section of returns without relying on option price information see, e.g., Kelly and Jiang (2014) and Almeida et al. (2017).

index, X, the first volatility factor, V_1, and the tail factor, U, and one captures jumps that affect U only. The distribution for the size of return jumps is assumed to be double exponential with two distinct parameters governing the decay of left and right tail. Although the time variation in positive and negative jumps is not the same, both intensities are affine functions of the state vector (V_1, V_2, U). This procedure allows for "cross self-exciting" jumps: a shock to one factor can increase the jump intensity, which in turns increases the probability of future jumps in that and all other factors. The Three-Factor Double Exponential Model is represented by the following equations:

$$\begin{aligned}
\frac{dX_t}{X_{t-}} &= (r_t - \delta_t)dt + \sqrt{V_{1,t}}\ dW^{\mathbb{Q}}_{1,t} + \sqrt{V_{2,t}}\ dW^{\mathbb{Q}}_{2,t} + \eta\sqrt{U_t}\ dW^{\mathbb{Q}}_{3,t} + \int_{\mathbb{R}^2} (e^x - 1)\widetilde{\mu}^{\mathbb{Q}}(dt, dx, dy), \\
dV_{1,t} &= \kappa_1(\bar{v}_1 - V_{1,t})dt + \sigma_1\sqrt{V_{1,t}}\ dB^{\mathbb{Q}}_{1,t} + \mu_1 \int_{\mathbb{R}^2} x^2 1_{\{x<0\}}\mu(dt, dx, dy)\ , \\
dV_{2,t} &= \kappa_2(\bar{v}_2 - V_{2,t})dt + \sigma_2\sqrt{V_{2,t}}\ dB^{\mathbb{Q}}_{2,t}\ , \\
dU_t &= -\kappa_u U_t dt + \mu_u \int_{\mathbb{R}^2} [(1-\rho_u)x^2 1_{\{x<0\}} + \rho_u y^2]\mu(dt, dx, dy)\ .
\end{aligned} \tag{4.1}$$

where r_t is the risk-free rate, δ_t is the dividend yield on the index, and $(W^{\mathbb{Q}}_{1,t}, W^{\mathbb{Q}}_{2,t}, W^{\mathbb{Q}}_{3,t}, B^{\mathbb{Q}}_{1,t}, B^{\mathbb{Q}}_{2,t})$ is a five-dimensional Brownian motion with $\text{corr}\left(W^{\mathbb{Q}}_{1,t}, B^{\mathbb{Q}}_{1,t}\right) = \rho_1$, $\text{corr}\left(W^{\mathbb{Q}}_{2,t}, B^{\mathbb{Q}}_{2,t}\right) = \rho_2$, and mutual independence for the remaining Brownian motions. In addition, μ is the jump counting measure with instantaneous intensity, under the risk-neutral measure, given by $dt \otimes \nu^{\mathbb{Q}}_t(dx, dy)$. The difference $\widetilde{\mu}^{\mathbb{Q}}(dt, dx, dy) = \mu(dt, dx, dy) - dt\nu^{\mathbb{Q}}_t(dx, dy)$ constitutes the associated martingale measure. The contemporaneous co-jumps in X, V_1 and, if $\rho_u < 1$, also in U are captured by x, while y represents the independent shocks to the U factor. The jump component x is distributed according to a double exponential density function with separate tail decay parameters, λ_- and λ_+, for negative and positive jumps, respectively. The jump component y is distributed identically to the negative price jumps. Moreover, $c^-(t)$ and $c^+(t)$ define the time-varying intensities of, respectively, negative and positive jumps as follows,

$$c^-(t) = c^-_0 + c^-_1 V_{1,t} + c^-_2 V_{2,t} + c^-_3 U_t\ , \qquad c^+(t) = c^+_0 + c^+_1 V_{1,t} + c^+_2 V_{2,t} + c^+_3 U_t\ . \tag{4.2}$$

Finally, the jump compensator characterises the conditional jump distribution and is given by,

$$\frac{\nu_t^{\mathbb{Q}}(dx,dy)}{dxdy} = \begin{cases} (c^-(t)\cdot 1_{\{x<0\}}\lambda_- e^{-\lambda_-|x|} + c^+(t)\cdot 1_{\{x>0\}}\lambda_+ e^{-\lambda_+ x}) \ , & \text{if } y=0 \\ c^-(t)\lambda_- e^{-\lambda_-|y|} \ , \quad \text{if } x=0 \text{ and } y<0 \end{cases}$$

Supported by the data, Andersen et al. (2015b) constrain the statistically insignificant parameters to zero and set c_3^- to unity for identification purposes.[4] The implication of this is that U becomes a left tail factor that affects the intensity of only negative jumps and does not contribute directly to the diffusive spot variance. Given these characteristics, we are motivated to formulate the equity tail risk measure of the present chapter in terms of the U factor.

The period-by-period estimates of the state variables, $(V_{1,t}\ V_{2,t}\ U_t)$, together with values for the model parameters, are obtained by using the penalised nonlinear least squares estimator developed by Andersen et al. (2015a). The Andersen et al. (2015b) model is fitted to a panel of equity-index options by minimising the weighted sum of squared deviations of the Black-Scholes implied volatilities generated by the model from the observed ones.[5] In solving this minimisation problem, the estimator also penalises for discrepancies between the model-implied spot volatilities and those estimated, in a nonparametric fashion, from high-frequency data on the underlying asset returns. Using the same notation as in Andersen et al. (2015b), we denote the parameter vector of the model by θ and the state vector at time t by $\mathbf{Z_t} = (V_{1,t}\ V_{2,t}\ U_t)$. Further, we use $\kappa(k,\tau,\mathbf{Z_t},\theta)$ and $\bar{\kappa}(t,k,\tau)$ to denote, respectively, the model-implied Black-Scholes implied volatility (IV) and the observed Black-Scholes IV corresponding to the average of bid and ask quotes of the option with tenor τ and log-forward moneyness k at time t. As for the diffusive spot variance, we denote the model-implied measure by $V(\mathbf{Z_t},\theta) = V_{1,t} + V_{2,t} + \eta^2 U_t$ and its nonparametric estimator constructed from intraday returns by $\widehat{V}_t$.[6] Finally, letting N_t denote the number of option contracts available on day t, the estimator takes the form,

$$\left(\{\hat{V}_{1,t}, \hat{V}_{2,t}, \hat{U}_t\}_{t=1,\ldots,T}, \hat{\theta}\right) = \underset{\{\mathbf{Z_t}\}_{t=1,\ldots T}, \theta\in\Theta}{\arg\min} \sum_{t=1}^{T} \left\{ \frac{\text{Option Fit}_t + \lambda \times \text{Vol Fit}_t}{V_t^{ATM}} \right\} \tag{4.3}$$

[4]The restrictions on the parameters are the same as those shown in Table 4.1, 4.2 and 4.3 of this chapter.

[5]Andersen et al. (2015b) provide in their *Supplementary Appendix* the conditional characteristic function of log-returns needed to price options according to the Three-Factor Double Exponential Model. The obtained option prices are then expressed in Black-Scholes implied volatility units for estimation purposes.

[6]For jump-robust volatility estimators based on high-frequency data see, for instance, Andersen et al. (2012).

$$\text{Option Fit}_t = \frac{1}{N_t}\sum_{j=1}^{N_t}\Big(\bar{\kappa}(t,k_j,\tau_j)-\kappa(k_j,\tau_j,\mathbf{Z_t},\theta\Big)^2 \ , \qquad \text{Vol Fit}_t = \left(\sqrt{\widehat{V}_t}-V(\mathbf{Z_t},\theta)\right)^2 \quad (4.4)$$

where λ is a calibration parameter that we set to 0.05 as in Andersen et al. (2017b), and V_t^{ATM} is the squared Black-Scholes IV obtained for the option closest to at-the-money with the shortest available maturity on day t.[7] The standardisation by V_t^{ATM} in the estimator is such that days with high market volatility are underweighted because option pricing errors tend to be larger.

Throughout the rest of the chapter, we use the residual of the regression of the U factor on the spot variance V to construct our equity left tail factor and study the effect of equity tail risk on bond risk premia. This choice is motivated by the work of Andersen et al. (2015b, 2017b) who have recently shown that the component of the left jump tail intensity factor unspanned by volatility, the so-called "pure tail" factor, has strong predictive power for future equity returns. Building on the significant stock-bond return comovements documented in times of elevated risk, we investigate the explanatory power of this equity "pure tail" factor for future bond returns. As mentioned earlier, due to the scarcity of safe asset producers other than the United States (Caballero et al., 2017), we treat US Treasury bonds as global safe haven and examine their response to the downside tail risk of the international stock market. Therefore, if we denote by $\widetilde{U}^i$ the "pure tail" factor relating to the i-th stock market index, we construct the equity left tail factor of this chapter as follows,

$$\widetilde{U}_t^{Equity} = \sum_{i=1}^{I} w_t^i \ \widetilde{U}_t^i \ , \tag{4.5}$$

where w_t^i is the time-t market capitalization of the i-th stock market index divided by the sum of the market capitalizations of the I indices at time t.[8]

[7] In practice, the joint optimisation over parameters and state vector realisations is performed by concentrating, or profiling, the state vector and optimising over the model parameters. Indeed, given a candidate vector θ, it is easy to obtain estimates of $(V_{1,t}\ V_{2,t}\ U_t)$ with local optimisation search methods. By contrast, the search of a global optimum is done for vector θ.

[8] In Appendix A, we show that the option-implied "pure tail" factor of the US stock market provides, within the model for the US term structure, qualitatively identical performance to that of $\widetilde{U}_t^{Equity}$.

4.3 Term Structure Modelling

We now introduce the term structure framework adopted in this chapter and we present its estimation procedure. To set up the model, we rely on the approach suggested by Adrian et al. (2013), which has the advantage that the pricing factors of bonds are not restricted to linear combinations of yields. Factors can indeed also be of different origin, such as the equity tail measure defined in Section 4.2 and whose use in a model for US interest rates is the main novelty of this chapter. After deriving the data generating process of log excess bond returns from a dynamic asset pricing model with an exponentially affine pricing kernel, Adrian et al. (2013) propose a new regression-based estimation technique for the model parameters. The linear regressions of this simple estimator avoid the computational burden of maximum likelihood methods, which have previously been the standard approach to the pricing of interest rates.

The formulation and estimation of the Gaussian ATSM in Adrian et al. (2013) can be summarised as follows. A $K \times 1$ vector of pricing factors, $\mathbf{X}_t$, is assumed to evolve according to a VAR process of order one:

$$\mathbf{X}_{t+1} = \boldsymbol{\mu} + \boldsymbol{\phi}\mathbf{X}_t + \mathbf{v}_{t+1} \ , \tag{4.6}$$

where the shocks $\mathbf{v}_{t+1} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ are conditionally Gaussian with zero mean and variance-covariance matrix $\boldsymbol{\Sigma}$. Letting $P_t^{(n)}$ denote the price of a zero-coupon bond with maturity n at time t, the assumption of no-arbitrage implies the existence of a pricing kernel M_{t+1} such that,

$$P_t^{(n)} = \mathrm{E}_t\Big[M_{t+1}P_{t+1}^{(n-1)}\Big] \ . \tag{4.7}$$

The pricing kernel M_{t+1} is assumed to have the following exponential form:

$$M_{t+1} = \exp\Big(- r_t - \frac{1}{2}\boldsymbol{\lambda}_t'\boldsymbol{\lambda}_t - \boldsymbol{\lambda}_t'\boldsymbol{\Sigma}^{-1/2}\mathbf{v}_{t+1}\Big) \ , \tag{4.8}$$

where $r_t = -\ln P_t^{(1)}$ is the continuously compounded one-period risk-free rate and $\boldsymbol{\lambda}_t$ is the $K \times 1$ vector of market prices of risk, which are affine in the factors as in Duffee (2002):

$$\boldsymbol{\lambda}_t = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\lambda}_0 + \boldsymbol{\lambda}_1\mathbf{X}_t) \ . \tag{4.9}$$

The log excess one-period return of a bond maturing in n periods is defined as follows,

$$rx_{t+1}^{(n-1)} = \ln P_{t+1}^{(n-1)} - \ln P_t^{(n)} - r_t \ . \tag{4.10}$$

After assuming the joint normality of $\{rx_{t+1}^{(n-1)}, \mathbf{v}_{t+1}\}$, Adrian et al. (2013) derive the return generating process for log excess returns, which takes the form[9],

$$rx_{t+1}^{(n-1)} = \boldsymbol{\beta}^{(n-1)'}(\boldsymbol{\lambda}_0 + \boldsymbol{\lambda}_1 \mathbf{X}_t) - \frac{1}{2}(\boldsymbol{\beta}^{(n-1)'}\boldsymbol{\Sigma}\boldsymbol{\beta}^{(n-1)} + \sigma^2) + \boldsymbol{\beta}^{(n-1)'}\mathbf{v}_{t+1} + e_{t+1}^{(n-1)} \ , \tag{4.11}$$

where the return pricing errors $e_{t+1}^{(n-1)} \sim$ i.i.d. $(0, \sigma^2)$ are conditionally independently and identically distributed with zero mean and variance σ^2. Letting N be the number of bond maturities available and T be the number of time periods at which bond returns are observed, Adrian et al. (2013) rewrite equation (4.11) in the stacked form,

$$\mathbf{rx} = \boldsymbol{\beta}'(\boldsymbol{\lambda}_0 \boldsymbol{\iota}_T' + \boldsymbol{\lambda}_1 \mathbf{X}_{\_}) - \frac{1}{2}(\mathbf{B}^* \text{vec}(\boldsymbol{\Sigma}) + \sigma^2 \boldsymbol{\iota}_N)\boldsymbol{\iota}_T' + \boldsymbol{\beta}'\mathbf{V} + \mathbf{E} \ , \tag{4.12}$$

where $\mathbf{rx}$ is an $N \times T$ matrix of excess bond returns, $\boldsymbol{\beta} = \left[\boldsymbol{\beta}^{(1)} \ \boldsymbol{\beta}^{(2)} \ ... \ \boldsymbol{\beta}^{(N)}\right]$ is a $K \times N$ matrix of factor loadings, $\boldsymbol{\iota}_T$ and $\boldsymbol{\iota}_N$ are a $T \times 1$ and $N \times 1$ vector of ones, $\mathbf{X}_{\_} = [\mathbf{X}_0 \ \mathbf{X}_1 \ ... \ \mathbf{X}_{T-1}]$ is a $K \times T$ matrix of lagged pricing factors, $\mathbf{B}^* = \left[\text{vec}(\boldsymbol{\beta}^{(1)}\boldsymbol{\beta}^{(1)'}) \ ... \ \text{vec}(\boldsymbol{\beta}^{(N)}\boldsymbol{\beta}^{(N)'})\right]'$ is an $N \times K^2$ matrix, $\mathbf{V}$ is a $K \times T$ matrix and $\mathbf{E}$ is an $N \times T$ matrix.

The main novelty of the approach taken by Adrian et al. (2013) to model the term structure of interest rates is the use of ordinary least squares to estimate the parameters of equation (4.12). In particular, the authors propose the following three-step procedure:

1. Estimate the coefficients of the VAR model in equation (4.6) by ordinary least squares.[10] Stack the estimates of the innovations $\hat{\mathbf{v}}_{t+1}$ into matrix $\hat{\mathbf{V}}$ and use this to construct an estimator of the variance-covariance matrix $\hat{\boldsymbol{\Sigma}} = \hat{\mathbf{V}}\hat{\mathbf{V}}'/T$.

2. From the excess return regression equation $\mathbf{rx} = \mathbf{a}\boldsymbol{\iota}_T' + \boldsymbol{\beta}'\hat{\mathbf{V}} + \mathbf{c}\mathbf{X}_{\_} + \mathbf{E}$, obtain estimates of $\hat{\mathbf{a}}$, $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{c}}$. Use $\hat{\boldsymbol{\beta}}$ to construct $\hat{\mathbf{B}}^*$. Stack the residuals of the regression into matrix $\hat{\mathbf{E}}$

[9]For the full derivation of the data generating process see Section 2.1 in Adrian et al. (2013).

[10]For estimation purposes, Adrian et al. (2013) advise to set $\boldsymbol{\mu} = 0$ in case of zero-mean pricing factors.

and use this to construct an estimator of the variance $\hat{\sigma}^2 = \text{tr}(\hat{\mathbf{E}}\hat{\mathbf{E}}^{'})/NT$.

3. Noting from equation (4.12) that $\mathbf{a} = \boldsymbol{\beta}^{'}\boldsymbol{\lambda}_0 - \frac{1}{2}(\mathbf{B}^{*}\text{vec}(\boldsymbol{\Sigma}) + \sigma^2\boldsymbol{\iota}_N)$ and $\mathbf{c} = \boldsymbol{\beta}^{'}\boldsymbol{\lambda}_1$, estimate the price of risk parameters $\boldsymbol{\lambda}_0$ and $\boldsymbol{\lambda}_1$ via cross-sectional regressions,

$$\hat{\boldsymbol{\lambda}}_0 = (\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}^{'})^{-1}\hat{\boldsymbol{\beta}}\Big(\hat{\mathbf{a}} + \frac{1}{2}(\hat{\mathbf{B}}^{*}\text{vec}(\hat{\boldsymbol{\Sigma}}) + \hat{\sigma}^2\boldsymbol{\iota}_N)\Big) \ , \tag{4.13}$$

$$\hat{\boldsymbol{\lambda}}_1 = (\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}^{'})^{-1}\hat{\boldsymbol{\beta}}\hat{\mathbf{c}} \ . \tag{4.14}$$

The analytical expressions of the asymptotic variance and covariance of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\Lambda}} = [\hat{\boldsymbol{\lambda}}_0 \ \hat{\boldsymbol{\lambda}}_1]$, which we do not report here to save space, are provided in Appendix A.1 of Adrian et al. (2013). From the estimated model parameters, Adrian et al. (2013) show how to generate a yield curve. Indeed, within the proposed framework, bond prices are exponentially affine in the pricing factors. Consequently, the yield of a zero-coupon bond with maturity n at time t, $y_t^{(n)}$, can be expressed as follows,

$$y_t^{(n)} = -\frac{1}{n}[a_n + \mathbf{b}_n^{'}\mathbf{X}_t] + u_t^{(n)} \ , \tag{4.15}$$

where the coefficients a_n and $\mathbf{b}_n$ are obtained from the following no-arbitrage recursions,

$$a_n = a_{n-1} + \mathbf{b}_{n-1}^{'}(\boldsymbol{\mu} - \boldsymbol{\lambda}_0) + \frac{1}{2}(\mathbf{b}_{n-1}^{'}\boldsymbol{\Sigma}\mathbf{b}_{n-1} + \sigma^2) - \delta_0 \ , \tag{4.16}$$

$$\mathbf{b}_n^{'} = \mathbf{b}_{n-1}^{'}(\boldsymbol{\phi} - \boldsymbol{\lambda}_1) - \boldsymbol{\delta}_1^{'} \ , \tag{4.17}$$

subject to the initial conditions $a_0 = 0$, $\mathbf{b}_n = \mathbf{0}$, $a_1 = -\delta_0$ and $\mathbf{b}_1 = -\boldsymbol{\delta}_1$. The parameters δ_0 and $\boldsymbol{\delta}_1$ are estimated by regressing the short rate, $r_t = -\ln P_t^{(1)}$, on a constant and contemporaneous pricing factors according to,

$$r_t = \delta_0 + \boldsymbol{\delta}_1\mathbf{X}_t + \epsilon_t \ , \quad \epsilon_t \sim i.i.d.\ (0, \sigma_\epsilon^2) \ . \tag{4.18}$$

By setting the price of risk parameters $\boldsymbol{\lambda}_0$ and $\boldsymbol{\lambda}_1$ to zero in equation (4.16) and (4.17), Adrian et al. (2013) obtain a_n^{RN} and $\mathbf{b}_n^{\text{RN}}$, which they use to generate the risk-neutral yields, $y_t^{(n)\ \text{RN}}$. These yields reflect the average expected short rate over the current and the subsequent $(n-1)$

periods and are computed as follows,

$$y_t^{(n)\ \mathrm{RN}} = \frac{1}{n}\sum_{i=0}^{n-1}\mathrm{E}_t[r_{t+i}] = -\frac{1}{n}[a_n^{\mathrm{RN}} + \mathbf{b}_n^{\mathrm{RN}\prime}\mathbf{X}_t]\ . \tag{4.19}$$

Given equation (4.15) and (4.19), the term premium $TP_t^{(n)}$, which is the additional compensation required for investing in long-term bonds relative to rolling over a series of short-term bonds, can be calculated as follows,

$$TP_t^{(n)} = y_t^{(n)} - y_t^{(n)\ \mathrm{RN}}\ . \tag{4.20}$$

Starting from the expressions for the zero-coupon bond yields, it is possible to show that also forward rates are affine functions of the pricing factors. In particular, we calculate $f_t^{m,n}$, which denotes the forward rate at time t for an investment that starts m periods after time t and terminates n periods after time t, as follows,

$$f_t^{(m,n)} = \frac{1}{n-m}\Big[(a_m - a_n) + (\mathbf{b}_m^{\prime} - \mathbf{b}_n^{\prime})\mathbf{X}_t\Big]\ . \tag{4.21}$$

By replacing a_m, a_n, $\mathbf{b}_m$ and $\mathbf{b}_n$ in equation (4.21) with their risk-neutral counterparts a_m^{RN}, a_n^{RN}, $\mathbf{b}_m^{\mathrm{RN}}$ and $\mathbf{b}_n^{\mathrm{RN}}$, we obtain the risk-neutral forward rates $f_t^{(m,n)\ \mathrm{RN}}$ which we use to calculate the forward term premium $FTP_t^{(m,n)}$ according to,

$$FTP_t^{(m,n)} = f_t^{(m,n)} - f_t^{(m,n)\ \mathrm{RN}}\ . \tag{4.22}$$

In the next section we specify and estimate a term structure model for US interest rates following the procedure outlined above. The main difference between the Gaussian ATSM in Adrian et al. (2013) and ours is that we use a different set of pricing factors. Indeed, we include in $\mathbf{X}_t$ not only factors of bond-market origin (principal components of the yield curve) but also the left jump tail risk measure extracted from equity-index options and described in Section 4.2.

4.4 Empirical Application

We provide in this section an application to data of a bond pricing model featuring equity tail risk. We present empirical results using government bond data from the US market and equity option data from the US, UK and Euro-zone markets. We start by examining the role of the equity left tail factor in predicting bond returns for horizons up to one year. We then estimate a Gaussian ATSM that uses the equity left tail factor, along with the first five principal components of Treasury yields, to explain the cross-section of one-month excess bond returns. We report the estimation results for the full sample and we claim that equity jump tail risk is strongly priced within the model and is a significant predictor of lower expected returns on short-term bonds. We further assess the out-of-sample performance of the proposed model relative to that of a benchmark ATSM that does not include equity tail risk. Finally, we discuss how equity tail risk has influenced the Treasury term structure over time.

4.4.1 Data

All data considered here are sampled at the end of each month, or the previous trading day if the month-end value is missing, for the period from January 2007 through December 2017. In this study on bond premia, the start date of the sample is chosen in accordance with Andersen et al. (2017b), who analyse the impact of market tail risk on the equity risk premium instead.[11]

To construct the equity left tail factor, we use the closing bid and ask prices reported by OptionMetrics IvyDB US for the European style S&P 500 equity-index (SPX) options, and the last prices reported by OptionMetrics IvyDB Europe for the European style FTSE 100 (FTSE) and EURO STOXX 50 (ESTOXX) equity-index options. We apply the following standard filters to our dataset. We discard options with a tenor of less than seven days or more than one year. We discard options with zero bid prices and options with non-positive open interest. We only use options with non-negative bid-ask spread and options with an ask-to-bid ratio smaller than five. We retain only options whose prices are at least threefold the minimum tick size. For each

[11]Since 2007, the available maturity and strike coverage of the option data of this study is broad enough to estimate the Three-Factor Double Exponential Model discussed in Section 4.2 instead of the simplified two-factor model used by Andersen et al. (2017b).

day in the sample, we retain only option tenors for which we have at least five pairs of call and put contracts with the same strike price. We exploit these cross sections to derive, via put-call parity, the risk-free rate and the underlying asset price adjusted for the dividend yield that apply to a given option tenor on a given day. Finally, we discard all in-the-money options and we use only out-of-the-money options whose volatility-adjusted log-forward moneyness is between -15 and 5. The option data so obtained are supplemented by the time series of the three indices' Bipower Variation (5-min) provided by the OxfordMan Institute's "realised library". This is the nonparametric estimator of variance, constructed from high-frequency returns, that we use in equation (4.4) to compute Vol Fit_t. The estimator belongs to the class of jump-robust measures of volatility and was introduced by Barndorff-Nielsen and Shephard (2004).

The term structure model of this chapter is estimated using the Gürkaynak et al. (2007) zero-coupon bond yields derived from US Treasuries.[12] In line with the range of maturities in Adrian et al. (2013) and Abrahams et al. (2016), for our analysis we consider bonds maturing in less than or equal to ten years. More specifically, we extract the principal components, which we then use as pricing factors in the ATSM, from yields of maturities $n = 3, 6, ..., 120$ months. Furthermore, setting the risk-free short rate equal to the $n = 1$ month yield, we calculate the one-month excess returns for Treasury bonds with maturities $n = 6, 12, ..., 120$ months.

4.4.2 Option-Implied Equity Factors

The estimated parameters of the Three-Factor Double Exponential Model applied to S&P 500, FTSE 100 and EURO STOXX 50 equity-index option data are listed, respectively, in Table 4.1, 4.2 and 4.3. We note that our estimates for the S&P 500 index are very close to those reported by Andersen et al. (2015b) in their Table 4.

[Insert Table 4.1 here]

[Insert Table 4.2 here]

[12] These yield data are available at a daily frequency for annually spaced maturities ranging from 1 to 30 years from the Federal Reserve website https://www.federalreserve.gov/pubs/feds/2006/200628/200628abs.html. The parameters used to calculate the yields of any desired maturity are also available.

[Insert Table 4.3 here]

Using the index-specific estimated parameter vector $\hat{\theta}$, we recover the month-by-month realisations of the state variables for the S&P 500, FTSE 100 and EURO STOXX 50 equity-index returns, which are displayed in Figure 4.1. The top panel shows the model-implied diffusive spot variance, which we denote by V and is given by the sum of the two volatility factors, V_1 and V_2. The middle panel displays the downside jump intensity factor, U. The bottom panel presents the pure tail factor, $\widetilde{U}$, which corresponds to the residual obtained from the linear regression of U on V, and then normalised to have mean zero and unit variance. High values of $\widetilde{U}$ reflect a high perception, or fear, of future negative return jumps in the stock market. The equity left tail risk factor that we use in the term structure model of this chapter is given by the market capitalization weighted average of the $\widetilde{U}$ factor of the three stock market indices.[13]

[Insert Figure 4.1 here]

Inspection of Figure 4.1 immediately reveals that, as documented in Andersen et al. (2015b, 2017b), the negative jump intensity factor is far more persistent than diffusive volatility in the years following a crisis. In those previous studies, the component of the left jump intensity factor unspanned by volatility, i.e. $\widetilde{U}$, is shown to be a strong predictor of future excess equity-index returns, with predictive power superior to that of the VRP. Starting from this result and inspired by the theory of flight-to-safety, we explore the relationship of $\widetilde{U}$ with the US government bond market.

4.4.3 Bond Return Predictability

We start by considering the role of S&P 500 option-implied measures in explaining Treasury risk premia. To this end, we regress the future excess returns of one-, two-, three-, four-, five-,

[13] All the code has been written in Python to benefit from its computational speed. Also, since the estimation of state variables in the option pricing exercise is inherently independent from one day to another, we relied on BlueCrystal, the High Performance Computing (HPC) machine provided by the Advanced Computing Research Centre at University of Bristol, to exploit the power of multiple CPUs at the same time. Finally, it is a pleasure to acknowledge the invaluable help and advice received from Nicola Fusari in support of this part of the work.

seven- and ten-year Treasury bonds ($n = 12, 24, 36, 48, 60, 84, 120$ months, respectively) on the left jump intensity factor (orthogonal to spot variance) of S&P 500 equity-index returns, $\widetilde{U}^{\text{SPX}}$, and the variance risk premium orthogonal to $\widetilde{U}^{\text{SPX}}$, which we denote by $VRP^{\perp}$. We calculate the variance risk premium (on a monthly basis) as the difference between the square of the VIX index (obtained from the Chicago Board Options Exchange (CBOE)) and the summation of current and previous 20 trading days' daily realised variances and overnight squared returns of the S&P 500 (obtained from the OxfordMan Institute's "realised library").[14] As is pointed out in Andersen et al. (2017b), there is a strong correlation between the tail factor and the variance risk premium since the former is part of the risk-neutral measure of expected return variation used to calculate the latter. For this reason, the auxiliary explanatory power of the VRP is assessed in the following bivariate regressions,

$$rx_{t+h}^{(n-h)} = c_{0,h} + c_{u,h} \cdot \widetilde{U}_t^{\text{SPX}} + c_{p,h} \cdot VRP_t^{\perp} + \xi_{t+h} \ , \tag{4.23}$$

where h is the holding period (in months) and $rx_{t+h}^{(n-h)} = \ln P_{t+h}^{(n-h)} - \ln P_t^{(n)} - r_t$ is the h-month excess log-return on a bond with maturity n (in months) at time t. The risk-free rate used in the calculation of the excess returns is the yield of a zero-coupon bond with maturity h at time t. We run the predictive regressions using the full sample of monthly data over forecast horizons $h = 1, 2, ..., 12$ months. We compute the robust Newey-West standard errors using a window of twice as many lags as the number of months within the holding period horizon.[15] The significance of the regression coefficients and the degree of explanatory power provided by $\widetilde{U}^{\text{SPX}}$ and $VRP^{\perp}$ are presented in Figure 4.2.

[Insert Figure 4.2 here]

The immediate point that stands out is that Treasury risk premia load negatively on the

[14]This is the same methodology used by Bollerslev et al. (2009) and Bollerslev et al. (2014) to approximate the variance risk premium. This proxy of VRP has the advantage of being directly observable and completely model-free as the risk-neutral expectation of the future return variation is measured by the CBOE VIX squared, while the statistical expectation is quantified by the total realised variation over the previous month.

[15]The regression standard errors so constructed should also account for the estimation error in the projection generating the pure tail factor, see Andersen et al. (2015b, 2017b).

S&P 500 option-implied tail factor, $\widetilde{U}^{\mathrm{SPX}}$. Hence, as reflecting flight-to-safety, a higher fear of abrupt negative return shocks to the US equity market forces a contraction in the risk premia that investors require for holding US government bonds. For bonds with maturity less than five years, we find that the excess returns are significantly related to $\widetilde{U}^{\mathrm{SPX}}$, for most of the forecast horizons considered. For longer-term bonds, the explanatory power of $\widetilde{U}^{\mathrm{SPX}}$ is never statistically significant at the 10% level. Examining the coefficient of $VRP^{\perp}$, we observe that the component of the variance risk premium orthogonal to $\widetilde{U}^{\mathrm{SPX}}$ is highly significant for the longer return horizons of short-term bonds and for the shorter horizons of long-term bonds. The sign of the coefficient is positive in the former case and negative in the latter. Therefore, since the variance risk premium captures both jump and diffusive spot volatility, the risk premium associated with the diffusive part tends to predict lower bond returns at short horizons and higher returns at long horizons. Thus we can conclude that the component of the variance risk premium unrelated to the tail factor may have some predictive power for the Treasury risk premia. However, since the relevant return horizon in this chapter is one month and $VRP^{\perp}$ does not seem to have (at this horizon) predictive power over-and-above $\widetilde{U}^{\mathrm{SPX}}$ for most of the excess bond returns, it seems reasonable to us to consider the equity tail index, rather than the variance risk premium, as a potentially relevant pricing factor for the bond market.

Motivated by the global safe haven status of US government bonds (Caballero et al., 2017) and the commonalities across countries in stock return predictability (Bollerslev et al., 2014; Andersen et al., 2017b), we now extend our analysis to the explanatory power of the UK and Euro-zone equity tail risk measures.[16] We explore predictive regressions for Treasury risk premia including the option-implied left jump intensity factor (orthogonal to spot variance) of FTSE 100 equity-index returns, $\widetilde{U}^{\mathrm{FTSE}}$, and the option-implied left jump intensity factor (orthogonal to spot variance) of EURO STOXX 50 equity-index returns, $\widetilde{U}^{\mathrm{ESTOXX}}$, as single explanatory variables. Hoping to uncover stronger predictability than the one provided by the individual country measures, we also use as predictor $\widetilde{U}^{Equity}$, which is the market capitalization weighted

[16]In their multi-country study, Bollerslev et al. (2014) and Andersen et al. (2017b) find, respectively, that a country's equity variance risk premium and "pure tail" factor forecast the future equity market returns of that country. The predictability pattern that holds true for a number of countries in addition to the United States leads Bollerslev et al. (2014) to define a "global" variance risk premium, which, they find, is a highly significant predictor for all of the different country returns.

average of the option-implied pure tail factor of S&P 500, FTSE 100 and EURO STOXX 50 equity-index returns, calculated from equation (4.5). The univariate regressions take the form,

$$rx_{t+h}^{(n-h)} = d_{0,h} + d_{y,h} \cdot Y_t + \xi_{t+h}^{Y} \ , \tag{4.24}$$

where Y_t denotes one of the possible predictors, and the quantities on the left-hand side are the same log-returns as those used for equation (4.23). The results are presented in Figure 4.3 and 4.4.[17]

[Insert Figure 4.3 here]

[Insert Figure 4.4 here]

Inspection of Figure 4.3 reveals that excess Treasury returns are linked to the left jump tail intensity factor driving stock index returns in the UK and Euro-area markets, for horizons beyond four-five months and with a peak at around six months. Importantly, the same holds true for bonds with maturity of five years or longer, for which we failed to find significant exposures to $\widetilde{U}^{\text{SPX}}$ in Figure 4.2. Furthermore, examining the short return horizons, we note that the explanatory power of $Y_t = \widetilde{U}_t^{\text{ESTOXX}}$ is statistically insignificant. On the contrary, the option-implied tail factor of the UK stock market, $Y_t = \widetilde{U}_t^{\text{FTSE}}$, contains predictive power for the one-month excess returns on short-term bonds. A potential explanation for the observed inverse relationship between risk premia in the US government bond market and downside tail risk in the UK equity market can be the role of safe asset producer that the United States play for many economies, including the United Kingdom. For instance, as reported by the Economist (2015), the world aggregate demand for safe assets shifted away from the UK supply toward the US one in 1920-45, when Britain ceased to be the world's pre-eminent power and passed the safe asset baton to the United States.[18] Figure 4.4 shows qualitative features that are similar to those of Figure 4.2 and 4.3. That is, Treasury risk premia, computed over horizons up to one

[17]The standard errors are computed in the same way as those for the regression coefficients in equation (4.23).

[18]See Chiţu et al. (2014) for a thorough examination of how and why the dollar overtook sterling as the dominant currency of denomination for international bonds.

year, load negatively on a measure of downside equity tail risk, this time associated with the international stock market, $Y_t = \widetilde{U}_t^{Equity}$. Despite the negative sign of the coefficient, which is in agreement with the theory of flight-to-safety, we note, however, that the results tend to be statistically significant at the 10% level only for short-maturity bonds.

On the basis of the significant interactions observed between excess Treasury returns and the $\widetilde{U}$ factor of all three equity market indices considered, we deem it best to evaluate the overall effect of the international stock market on US government securities. Hence, the equity left tail factor that we use throughout the rest of the chapter is $\widetilde{U}^{Equity}$, which we interpret as a measure of international fear of future abrupt negative return shocks to the equity market.

4.4.4 Equity Tail Risk and Bond Pricing

The limited statistical power of the preliminary analysis presented in Section 4.4.3 may be justified by the misspecification of equation (4.23) and (4.24) that should logically include additional explanatory variables of bond risk premia. To go further in the analysis, we now estimate a Gaussian ATSM for US interest rates that includes as pricing factors the first five principal components of Treasury yields and the equity left tail factor, $\widetilde{U}^{Equity}$. This is a richer framework that allows us to explore in detail the effect of equity tail risk on contemporaneous bond yields and future excess bond returns. The first five principal components of the US yield curve have proven to be remarkably effective in fitting the cross-section of bond yields and returns in Adrian et al. (2013). Based on this evidence, we let these PCs drive the interest rates of our model as well, but with a slight modification of the methodology. Indeed, in order to have pricing factors that are uncorrelated with each other, we follow Cochrane and Piazzesi (2008) and extract the principal components not from the conventional yields, but instead from the yields orthogonalized to the extra factor, which in our study is $\widetilde{U}^{Equity}$. By doing so, we obtain yield curve factors that are unrelated to the pricing of tail risk in the stock market, which is entirely ascribed to the $\widetilde{U}^{Equity}$ factor. The choice of those state variables for our model is supported by the following observations. First, we note that the equity left tail factor is poorly spanned by the first five PCs extracted from the non-orthogonalized yields. Indeed, a regression of $\widetilde{U}^{Equity}$ on the traditional level, slope and curvature factors augmented with the fourth and fifth principal components results in an R^2 of only 28%. We find no significant relationships between the equity left tail factor and these PCs, as the largest correlation coefficient is -0.34 with the level factor. Therefore, if we want to capture the effect of equity tail risk on bond risk premia we must include $\widetilde{U}^{Equity}$ separately in the vector of pricing factors, $\mathbf{X}_t$, and orthogonalize for convenience the remaining factors. The second observation that we make about the choice of the state variables in $\mathbf{X}_t$ is that we cannot exclude the fourth and fifth principal components of the yield curve. The regressions of $PC4$ and $PC5$ on the equity left tail factor yield an R^2 of, respectively, 6% and 1%. These results imply that $\widetilde{U}^{Equity}$ does not subsume the predictive ability of the fourth and fifth principal components of the yield curve, which are, therefore, needed to explain the cross-section of bond returns as well as the model of Adrian et al. (2013) does. In view of these considerations, we employ the following set of pricing factors in our Gaussian ATSM,

$$\mathbf{X}_t = \left[\widetilde{U}_t^{Equity},\ PC1_t,\ PC2_t,\ PC3_t,\ PC4_t,\ PC5_t\right]', \tag{4.25}$$

where $\widetilde{U}^{Equity}$ is the equity left tail factor from equation (4.5) and $PC1$–$PC5$ are the first five principal components estimated from an eigenvalue decomposition of the covariance matrix of zero-coupon bond yields of maturities $n = 3, 6, ..., 120$ months, orthogonal to $\widetilde{U}^{Equity}$. All factors have mean zero and unit variance, and they are plotted in Figure 4.5. The panels of $PC1$–$PC5$ also present the principal components of the conventional non-orthogonalized bond yields. We find that estimates of the factors extracted using the two yield curves track each other quite closely, with the largest differences occurring for $PC1$ at the onset of the financial crisis. Therefore, the orthogonalization of the rates with respect to $\widetilde{U}^{Equity}$ does not appear to significantly alter the interpretation and role of the principal components in describing the characteristics of the US Treasury yield curve.

[Insert Figure 4.5 here]

Given the vector of state variables in (4.25), we estimate our Gaussian ATSM using the method put forward by Adrian et al. (2013) and discussed in Section 4.3. In particular, we use a total of $N = 20$ one-month excess returns for Treasury bonds with maturities $n = 6, 12, ..., 120$ months to fit the cross-section of yields. The estimation approach by Adrian et al. (2013) allows for direct testing of the presence of unspanned factors, i.e. factors that do not help explain variation in Treasury returns. The specification test is implemented as a Wald test of the null hypothesis that bond return exposures to a given factor are jointly equal to zero. Letting $\boldsymbol{\beta}_i$ be the i-th column of $\boldsymbol{\beta}'$, the Wald statistic, under the null $H_0 : \boldsymbol{\beta}_i = \mathbf{0}_{N\times 1}$, is defined as follows,

$$W_{\beta_i} = \hat{\boldsymbol{\beta}}_i' \hat{\mathcal{V}}_{\beta_i}^{-1} \hat{\boldsymbol{\beta}}_i \overset{\alpha}{\sim} \chi^2(N) \ , \tag{4.26}$$

where $\hat{\mathcal{V}}_{\beta_i}$ is an $N \times N$ diagonal matrix that contains the estimated variances of the $\hat{\boldsymbol{\beta}}_i$ coefficient estimates.[19] The results of the Wald test on the pricing factors of both the proposed ATSM with equity tail risk and a benchmark model based on only the first five PCs of the yield curve are shown in Table 4.4. As we can see, we strongly reject the hypothesis of unspanned factor for each of our state variables. This means that the data support the use of the equity left tail factor $\widetilde{U}^{Equity}$, together with the yield curve factors indicated by Adrian et al. (2013), for pricing government bonds in the US market over the period 2007 – 2017.

[Insert Table 4.4 here]

The summary statistics of the pricing errors implied by our term structure model, which accounts for equity tail risk, and the benchmark PC-only specification are provided in Table 4.5. Overall the results indicate a good fit between the data and the proposed model with equity tail risk. Indeed, both the mean and the standard deviation of our yield pricing errors remain well below half of a basis point for all maturities and they never exceed, in absolute value, those of the benchmark. As for the return pricing errors, we can see that including our equity tail risk measure explicitly in the Gaussian ATSM improves the fit especially to the short end of the US yield curve. Moreover, consistent with the way Adrian et al. (2013) construct their framework for the term structure of interest rates, we observe a strong autocorrelation in the yield pricing

[19]See Appendix A.1 in Adrian et al. (2013) for the analytical expressions of the asymptotic variance of the estimators.

errors and a negligible one in the return pricing errors. The success of our model in fitting the yield curve is shown graphically in the left panels of Figure 4.6. In these plots, the solid black lines of observed yields are visually indistinguishable from the dashed grey lines of model-implied yields. Similarly, the right panels of Figure 4.6 display the tight fit between actual and fitted excess Treasury returns. The dashed red lines plot the model-implied dynamics of bond term premia in the left panels and of the expected component of excess returns in the right panels.

[Insert Table 4.5 here]

[Insert Figure 4.6 here]

We now examine whether the risk factors that we use in our Gaussian ATSM are priced in the cross-section of Treasury returns. To this end, we follow Adrian et al. (2013) and perform a Wald test of the null hypothesis that the market price of risk parameters associated with a given model factor are jointly equal to zero. Letting $\boldsymbol{\lambda}_i^{'}$ be the i-th row of $\boldsymbol{\Lambda} = [\boldsymbol{\lambda}_0 \ \boldsymbol{\lambda}_1]$, the Wald statistic, under the null $H_0 : \boldsymbol{\lambda}_i^{'} = \mathbf{0}_{1\times(K+1)}$, is defined as follows,

$$W_{\Lambda_i} = \hat{\boldsymbol{\lambda}}_i^{'} \hat{\mathcal{V}}_{\lambda_i}^{-1} \hat{\boldsymbol{\lambda}}_i \overset{\alpha}{\sim} \chi^2(K+1) \ , \tag{4.27}$$

where $\hat{\mathcal{V}}_{\lambda_i}$ is a square matrix of order $(K+1)$ that contains the estimated variances of the $\hat{\boldsymbol{\lambda}}_i$ coefficient estimates.[20] In addition, in order to test whether the market prices of risk are time-varying, Adrian et al. (2013) propose the following Wald test which focuses on $\boldsymbol{\lambda}_1$ and excludes the contribution of $\boldsymbol{\lambda}_0$. Letting $\boldsymbol{\lambda}_{1_i}^{'}$ be the i-th row of $\boldsymbol{\lambda}_1$, the Wald statistic of this second test, under the null $H_0 : \boldsymbol{\lambda}_{1_i}^{'} = \mathbf{0}_{1\times(K)}$, is defined as follows,

$$W_{\lambda_{1_i}} = \hat{\boldsymbol{\lambda}}_{1_i}^{'} \hat{\mathcal{V}}_{\lambda_{1_i}}^{-1} \hat{\boldsymbol{\lambda}}_{1_i} \overset{\alpha}{\sim} \chi^2(K) \ . \tag{4.28}$$

In Table 4.6, we report the estimates and t-statistics for the market price of risk parameters in the proposed Gaussian ATSM, together with the Wald statistics and p-values for the two tests

[20] See Appendix A.1 in Adrian et al. (2013) for the analytical expressions of the asymptotic variance of the estimators.

just described. Examining the first row of the table, we note that equity tail risk, as measured by exposure to $\widetilde{U}^{Equity}$, is strongly priced in our term structure model with a p-value of 6.2%. We detect statistically significant time variations in the market price of equity tail risk, which are mostly explained by the equity left tail factor itself. Furthermore, we find that nearly all the coefficients in the second column of the table are statistically significant at the 1% level. These results suggest that $\widetilde{U}^{Equity}$ is an important driver of the market price of risk related to the factors that explain the yield curve movements. The only exception is in the risk associated with $PC4$, which, however, does not show significant time variation in its market price. Finally, we observe that $PC2$ carries a significant price of risk in our term structure model. This result, together with the fact that Adrian et al. (2013) find a significant market price of slope risk only after adding an unspanned real activity factor to their framework, corroborates the hypothesis that valuable information about bond premia is located outside of the yield curve.

[Insert Table 4.6 here]

We now discuss the impact of the state variables of our Gaussian ATSM on the pricing of US Treasury bonds. The loadings of the yields on all model factors are reported in Figure 4.7, whereas the loadings of the expected one-month excess returns are displayed in Figure 4.8. From an examination of the state variables that are in common with the work of Adrian et al. (2013), we can see that our results are broadly consistent with the well-established role of these factors. Indeed, given the sign of the yield loadings on $PC1$, $PC2$ and $PC3$, we can argue that the first three principal components of yields preserve in our study the interpretation of, respectively, level, slope and curvature of the term structure. Moreover, the yield loadings on $PC4$ and $PC5$ are both quite small, reflecting the modest variability of bond rates explained by these factors. As can be seen from Figure 4.8, however, all the principal components, including the higher order ones, are important to explain variation in Treasury returns. Specifically, in line with previous findings concerning the predictability of bond returns with yield spreads, our evidence suggests that an increase in the slope factor forecasts higher expected excess returns on bonds of all maturities. Now turning to the new pricing factor that we propose in this chapter, we observe from the top left panel of Figure 4.7 that the yield loadings on $\widetilde{U}^{Equity}$ are negative across all maturities. These results imply that bond prices, which move inversely to yields, rise in response to a contemporaneous shock to the equity left tail factor. And since, by construction, $\widetilde{U}^{Equity}$ is associated with a downturn in the international stock market, we confirm the hypothesis that US Treasury bonds benefit from flight-to-safety flows during periods of turmoil. Further, it is worth noting that, according to the size of the loadings, the contemporaneous effect of the equity left tail factor on the yield curve is not negligible compared to that of the first three principal components. Additional evidence of flight-to-safety is provided in the top left panel of Figure 4.8 where the expected excess return loadings on $\widetilde{U}^{Equity}$ are displayed. The coefficients are negative and tend to decrease with the maturity of the bond. Therefore, calculation of the second term in equation (4.11) suggests that the risk premium required by investors for holding US Treasury securities for one month shrinks in response to a contemporaneous shock to the equity left tail factor. In particular, we find that a one standard deviation increase in the $\widetilde{U}^{Equity}$ factor reduces the annualised expected excess return by up to about 2% for short-maturity and medium-maturity bonds. These observations about Treasury returns, combined with the previously documented positive relationship between the "pure tail" factor and future equity returns (Andersen et al., 2015b, 2017b), indicate a common predictor across the two asset classes, whose existence can be justified by the safe haven potential of US Treasuries.

[Insert Figure 4.7 here]

[Insert Figure 4.8 here]

In order to assess the significance of our results, we report in Table 4.7 the estimates and t-statistics of the expected excess return loadings associated with the $N = 20$ Treasury maturities used to fit the cross-section of yields.[21] To ease visual interpretation of the results, Figure 4.9 plots the absolute value of the t-statistics against the critical value of 1.64 for the 10%

[21] We use a delta method approach to estimate the standard errors of the expected excess return loadings on the pricing factors of the model. The coefficients are calculated as $\boldsymbol{\beta}^{(n)\prime}\boldsymbol{\lambda}_1^{(i)}$ and represent the response of the expected one-month excess return on the n-month bond to a contemporaneous shock to the i-th factor. The use of $\boldsymbol{\beta}^{(n)\prime}$ in place of $\mathbf{b}_n^{\prime}$ is allowed since the derivation of log bond prices in the ATSM is exact conditional on the equivalence of the two measures. Standard errors are calculated using the analytical expressions for the asymptotic variance and covariance of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\Lambda}}$ provided in Appendix A.1 of Adrian et al. (2013).

significance level. From an examination of the loadings on $\widetilde{U}^{Equity}$, it seems that, although the equity left tail factor predicts lower future returns across the whole yield curve, the significance of the results decreases with the maturity of the bonds. Indeed, we find that the $\widetilde{U}^{Equity}$ factor has highly significant explanatory power for future returns only on Treasuries with maturities ranging from one to four years. Based on this evidence, we argue that when the equity market tumbles, the short end of the US yield curve is more strongly affected by flight-to-safety than the long end. When looking at the return loadings on the remaining pricing factors of the Gaussian ATSM, we note a remarkably strong predictive ability of $PC1$ and $PC2$ over a wide range of maturities. By contrast, the higher order principal components have significant forecast power for future returns on Treasuries with either only short maturity or only long maturity.

[Insert Table 4.7 here]

[Insert Figure 4.9 here]

The analysis presented thus far can be related to the work of Kaminska and Roberts-Sklar (2015), who assess the importance of global market sentiment for the term structure of UK government bonds. The authors use the variance risk premium of US, UK and Euro-area equity markets to construct a proxy of global risk aversion, which then they introduce explicitly as a pricing factor into a Gaussian ATSM.[22] However, by studying the impact of risk aversion on UK bond data, Kaminska and Roberts-Sklar (2015) reach a different conclusion from ours: future excess bond returns for all maturities load positively on the equity market factor. We can interpret this as evidence of "weak" FTS affecting the UK term structure if we believe that their VRP-based measure and our equity left tail factor capture similar attributes of the stock market. Alternatively or concomitantly, the different conclusions drawn can be traced to differences in the equity factor used in the Gaussian ATSM of the two studies.

We now turn to the question of how equity tail risk has affected bond term premia over the course of time. To conduct the analysis, we use forward rates because, as suggested by Abrahams

[22]Kaminska and Roberts-Sklar (2015) obtain the global measure of risk aversion either as the market capitalization weighted average or as the first principal component of the individual VRPs. As the authors claim, the results are not sensitive to the choice of the aggregation method.

et al. (2016), their variation may be ascribed more to changes in risk premia than to changes in the expected future short rate. The left panels of Figure 4.10 show the dynamics of the 2-3y, 2-5y and 5-10y forward Treasury rates and their components, whereas the right panels illustrate the effect of the equity left tail factor on the term premia of those forward rates. We determine the contribution of $\widetilde{U}^{Equity}$ to FTP in equation (4.22) as the difference between the component of fitted forward rates and the component of their risk-neutral counterparts that the model attributes to the equity left tail factor. The following remarks can be made by observing Figure 4.10. As anticipated, we confirm that the expectations of future short spot rates embedded in forward yields (and represented by the risk-neutral forward rates) remain stable throughout time, especially in the case of far in the future forwards. Therefore it follows that oscillations in forward rates reflect, in large part, adjustments in the required term premia. We note that the outburst of the 2008-09 financial crisis marks the beginning of a long period of declining rates which was interrupted only briefly by the Federal Reserve's "taper tantrum" in 2013. Although the same pattern is observed for all yields presented in Figure 4.10, it is interesting to see how the $\widetilde{U}^{Equity}$ factor influenced the downward trend of term premia differently depending on the maturity. Indeed, from the right panels of Figure 4.10, it appears that the term premium of short-maturity forward rates was strongly affected by equity tail risk, whereas the response of far in the future forward rates was consistently negligible. This further corroborates our previous conclusion that short-term bonds provide a more effective shelter against equity market losses than long-term bonds do. For the 2-3y and 2-5y forward Treasury rates, we measure the impact of $\widetilde{U}^{Equity}$ on FTP to be as large as -80 and -70 basis points, respectively, at the peak of the crisis. The forward term premia show strong downward oscillations also in the first half of 2010 and second half of 2011, when the equity left tail factor increased in response to the intensification of the European sovereign debt crisis. In both these instances, the extent of the reduction in bond term premia that can be credited to equity tail risk is approximately 50 basis points. In conclusion, we can state that equity jump tail risk has played a central role in shaping the short end of the US Treasury yield curve since the outburst of the recent financial crisis.

[Insert Figure 4.10 here]

4.4.5 Out-of-Sample Evidence

Up to this point, we have only discussed results based on the full sample of data from 2007 to 2017. We conclude this section by considering two out-of-sample exercises as in Adrian et al. (2013). We assess the out-of-sample performance, both in the cross-section and in the time series, of the ATSM that uses the equity left tail factor, along with the first five PCs of Treasury yields, relative to that of the benchmark PC-only specification. We first evaluate the time series fit by examining the predictive ability of the models for future short-term interest rates. To this end, we begin by estimating the models using the first four years of data and holding back all subsequent observations. The holdout sample thus starts on January 31, 2011. By adopting an increasing forecasting scheme in which models are re-estimated monthly by recursively adding one observation to the end of the sample and keeping fixed the initial estimation date, we generate forecasts for the average short rate up to four years ahead.[23] The top panel of Figure 4.11 reports the relative root mean squared forecast error of the ATSM that includes equity tail risk, over forecast horizons from one month to four years. The mean forecast losses of the Equity Tail Risk ATSM have been divided by those of the benchmark model. Overall, the results suggest that the Equity Tail Risk ATSM outperforms the benchmark across a wide range of forecast horizons. The gains in forecast accuracy are over 10% at horizons between seven months and three years. However, a slightly worse performance than the benchmark is observed at very short forecasting horizons. Now turning to the out-of-sample performance in the cross-section, we consider the fit to maturities up to twenty years produced by models estimated using excess returns only on bonds with maturities up to ten years.[24] The bottom panel of Figure 4.11 shows the observed average yields for maturities up twenty years, along with the average yields implied by the Equity Tail Risk ATSM and the benchmark PC-only specification. The only remark needed here is that both models seem to produce fairly accurate long-maturity yields, although slightly higher than the realised mean values. Furthermore, we note that, while the gains in forecasting the future short rates were standing out clearly, the improvements in the

[23]These are the $t+1$ forecasts made at time t for the risk-neutral yields on bonds with maturities up to four years. See equation (4.19) for definition of risk-neutral yields.

[24]For estimation purposes, we continue to use the same $N = 20$ bond returns as before, i.e. one-month excess returns for Treasury bonds with maturities $n = 6, 12, ..., 120$ months.

out-of-sample cross-sectional fit offered by the ATSM that includes equity tail risk are limited and barely visible from the chart.

[Insert Figure 4.11 here]

To sum up, compared to the traditional PC-only specification, an ATSM that accounts for equity tail risk not only produces a good in-sample fit but also may yield superior out-of-sample forecasts of future short rates and be less sensitive to the choice of maturities used in estimation.

4.5 Conclusion

In this chapter, we study the response of US Treasury bonds to extreme events happening in the international stock market. We propose an affine term structure model in which the main drivers of interest rates are the principal components of the zero-coupon yield curve and a downside jump intensity factor extracted from S&P 500, FTSE 100 and EURO STOXX 50 equity-index options. While earlier approaches to pricing bonds with factors other than combinations of yields have proven useful when macro variables are considered, we focus here on the safe haven potential of US Treasuries and use a factor that originates in the equity option markets of developed countries.

The results of our main application to US bond market and international stock market data are summarised as follows. First, equity jump tail risk is strongly priced and exhibits significant time variation within the term structure model. Second, consistent with the theory of flight-to-safety, bond prices increase and future expected excess returns shrink in response to a contemporaneous shock to the equity left tail factor. Third, the equity left tail factor has significant explanatory power for future returns on Treasuries with maturities ranging from one to four years. Fourth, large drops in term premia at the short end of the US yield curve are attributable to equity tail risk since the outburst of the recent financial crisis. Finally, the inclusion of the equity left tail factor in the Gaussian ATSM leads to improved out-of-sample forecasts of future short rates.

A natural direction for future research is to apply the methodology outlined in this chapter to the yield curve of a wide range of developed countries. In particular, the model would allow to explore in detail the effect of equity tail risk and, therefore, uncover evidence of FTS in the government bond market of countries other than the United States.

Given our findings with a downside jump intensity factor related to the international stock market, it would also be worth assessing the impact on the yield curve of a tail factor implied by Treasury options. For instance, it would be interesting to see whether the downside tail risk of the bond market receives compensation in a term structure model and how its pricing differs from that of equity jump tail risk. This would contribute to the recent literature on the auxiliary role of Treasury variance risk premium in predicting higher expected bond returns (Mueller et al., 2016). We leave investigation of such possibilities to future research.

4.6 Tables of Chapter 4

Table 4.1 – SPX – Three-Factor Double Exponential Model - Estimation Results

Parameter	Estimate	Constrained	Parameter	Estimate	Constrained
ρ_1	-0.961	-	ρ_u	0.513	-
$\bar{v}_1$	0.003	-	c_0^-	0.000	✓
κ_1	10.521	-	c_0^+	0.377	-
σ_1	0.256	-	c_1^-	110.294	-
μ_1	12.504	-	c_1^+	25.988	-
ρ_2	-0.979	-	c_2^-	0.000	✓
$\bar{v}_2$	0.001	-	c_2^+	82.612	-
κ_2	1.755	-	c_3^-	1.000	✓
σ_2	0.172	-	c_3^+	0.000	✓
η	0.000	✓	λ_-	25.413	-
μ_u	7.147	-	λ_+	36.880	-
κ_u	0.087	-			

Notes: This table provides the in-sample estimates of parameter vector θ of the Three-Factor Double Exponential Model discussed in Section 4.2 and applied to S&P 500 equity-index options. All parameters are expressed in annualised terms. A ✓ in the "Constrained" column means that the corresponding parameter is not freely estimated, but instead is set to the value reported in the "Estimate" column. Model is estimated using data sampled at the end of each month over the period from January 2007 through December 2017.

Table 4.2 – FTSE – Three-Factor Double Exponential Model - Estimation Results

Parameter	Estimate	Constrained	Parameter	Estimate	Constrained
ρ_1	-0.946	-	ρ_u	0.473	-
$\bar{v}_1$	0.003	-	c_0^-	0.000	✓
κ_1	12.393	-	c_0^+	0.173	-
σ_1	0.521	-	c_1^-	148.335	-
μ_1	13.193	-	c_1^+	35.754	-
ρ_2	-0.992	-	c_2^-	0.000	✓
$\bar{v}_2$	0.010	-	c_2^+	73.134	-
κ_2	0.314	-	c_3^-	1.000	✓
σ_2	0.158	-	c_3^+	0.000	✓
η	0.000	✓	λ_-	25.711	-
μ_u	6.836	-	λ_+	64.987	-
κ_u	0.030	-			

Notes: This table provides the in-sample estimates of parameter vector θ of the Three-Factor Double Exponential Model discussed in Section 4.2 and applied to FTSE 100 equity-index options. All parameters are expressed in annualised terms. A ✓ in the "Constrained" column means that the corresponding parameter is not freely estimated, but instead is set to the value reported in the "Estimate" column. Model is estimated using data sampled at the end of each month over the period from January 2007 through December 2017.

Table 4.3 – ESTOXX – Three-Factor Double Exponential Model - Estimation Results

Parameter	Estimate	Constrained	Parameter	Estimate	Constrained
ρ_1	-0.927	-	ρ_u	0.943	-
$\bar{v}_1$	0.007	-	c_0^-	0.000	✓
κ_1	8.961	-	c_0^+	0.023	-
σ_1	0.311	-	c_1^-	116.125	-
μ_1	11.725	-	c_1^+	0.500	-
ρ_2	-1.000	-	c_2^-	0.000	✓
$\bar{v}_2$	0.000	-	c_2^+	120.000	-
κ_2	1.925	-	c_3^-	1.000	✓
σ_2	0.000	-	c_3^+	0.000	✓
η	0.000	✓	λ_-	26.658	-
μ_u	1.003	-	λ_+	14.314	-
κ_u	0.000	-			

Notes: This table provides the in-sample estimates of parameter vector θ of the Three-Factor Double Exponential Model discussed in Section 4.2 and applied to EURO STOXX 50 equity-index options. All parameters are expressed in annualised terms. A ✓ in the "Constrained" column means that the corresponding parameter is not freely estimated, but instead is set to the value reported in the "Estimate" column. Model is estimated using data sampled at the end of each month over the period from January 2007 through December 2017.

Table 4.4 – Gaussian ATSM - Factor Risk Exposures

	Equity Tail Risk ATSM		PC-only ATSM	
Factor	W_{β_i}	p-value	W_{β_i}	p-value
$\widetilde{U}^{Equity}$	24348047.824	0.000	-	-
$PC1$	93189771.518	0.000	81008679.455	0.000
$PC2$	22249278.004	0.000	22570673.342	0.000
$PC3$	3157441.193	0.000	2845956.394	0.000
$PC4$	412034.906	0.000	402369.064	0.000
$PC5$	36850.820	0.000	34767.623	0.000

Notes: This table provides the Wald statistics and corresponding p-values for the Wald test of whether the exposures of bond returns to a given model factor are jointly zero. Under the null $H_0 : \boldsymbol{\beta}_i = \mathbf{0}_{N\times 1}$ the i-th pricing factor is unspanned, i.e. Treasury returns are not exposed to that factor. The p-values of the statistics are obtained from a chi-squared distribution with $N = 20$ degrees of freedom. The test is conducted on the pricing factors of both the proposed ATSM specified with equity tail risk and a benchmark PC-only model specification.

Table 4.5 – Gaussian ATSM - Fit Diagnostics

Panel A: Equity Tail Risk ATSM

	$n = 12$	$n = 24$	$n = 36$	$n = 60$	$n = 84$	$n = 120$
Panel A1: Yield Pricing Errors						
Mean	−0.001	0.000	0.000	0.000	0.000	0.000
Standard Deviation	0.002	0.001	0.001	0.001	0.001	0.002
Skewness	−1.057	1.947	1.767	−0.958	1.223	−1.140
Kurtosis	5.946	9.999	7.480	5.917	6.708	6.517
$\rho(1)$	0.756	0.754	0.830	0.762	0.777	0.778
$\rho(6)$	0.201	0.221	0.295	0.161	0.188	0.057
Panel A2: Return Pricing Errors						
Mean	−0.001	0.001	0.001	−0.001	0.000	0.000
Standard Deviation	0.022	0.018	0.023	0.056	0.034	0.184
Skewness	−0.310	−0.933	−1.324	0.149	−0.534	0.054
Kurtosis	5.884	15.014	10.564	5.490	12.067	4.905
$\rho(1)$	−0.095	−0.199	−0.034	−0.129	−0.154	−0.036
$\rho(6)$	0.028	0.204	0.177	0.035	0.199	0.028

Panel B: PC-only ATSM

	$n = 12$	$n = 24$	$n = 36$	$n = 60$	$n = 84$	$n = 120$
Panel B1: Yield Pricing Errors						
Mean	0.001	0.000	0.001	0.000	−0.001	0.000
Standard Deviation	0.004	0.002	0.001	0.002	0.001	0.002
Skewness	−1.200	2.343	2.205	−1.085	0.390	−1.402
Kurtosis	3.913	10.582	9.608	3.348	3.106	6.193
$\rho(1)$	0.908	0.826	0.872	0.866	0.882	0.787
$\rho(6)$	0.571	0.341	0.376	0.471	0.454	0.122
Panel B2: Return Pricing Errors						
Mean	0.000	−0.003	0.000	0.000	−0.006	0.006
Standard Deviation	0.025	0.020	0.026	0.052	0.040	0.171
Skewness	−0.371	−0.129	−2.076	−0.256	−0.148	−0.229
Kurtosis	14.188	15.218	17.183	11.281	8.139	6.232
$\rho(1)$	−0.046	−0.186	−0.076	−0.135	−0.126	−0.063
$\rho(6)$	0.117	0.272	0.249	0.108	0.188	0.066

Notes: This table contains the summary statistics of the pricing errors implied by the Gaussian ATSM that includes equity tail risk (Panel A) and by the benchmark model that only uses the first five PCs of the yield curve (Panel B). Models are estimated over the period 2007 to 2017. Reported are the sample mean, standard deviation, skewness, kurtosis and the autocorrelation coefficients of order one and six. Panels A1 and B1: properties of the yield pricing errors $\hat{u}$. Panels A2 and B2: properties of the return pricing errors $\hat{e}$. n denotes the maturity of the bonds in months.

Table 4.6 – Gaussian ATSM - Market Prices of Risk

Factor	λ_0	$\lambda_{1,1}$	$\lambda_{1,2}$	$\lambda_{1,3}$	$\lambda_{1,4}$	$\lambda_{1,5}$	$\lambda_{1,6}$	W_{Λ_i}	$W_{\lambda_{1_i}}$
$\widetilde{U}^{Equity}$	0.076	0.809	0.022	−0.120	0.199	−0.037	−0.231	13.435	13.308
	(0.475)	(3.543)	(0.140)	(−0.747)	(1.207)	(−0.223)	(−1.414)	(0.062)	(0.038)
$PC1$	−0.008	0.325	−0.042	−0.100	0.061	0.023	−0.053	16.283	16.252
	(−0.136)	(3.778)	(−0.684)	(−1.619)	(0.966)	(0.374)	(−0.849)	(0.023)	(0.012)
$PC2$	−0.051	−0.293	0.012	−0.006	−0.063	0.074	0.095	15.431	14.913
	(−0.816)	(−3.531)	(0.188)	(−0.102)	(−0.989)	(1.159)	(1.482)	(0.031)	(0.021)
$PC3$	0.106	−0.175	0.042	0.075	−0.121	0.039	0.030	26.028	21.076
	(2.224)	(−3.413)	(0.879)	(1.565)	(−2.522)	(0.813)	(0.635)	(0.000)	(0.002)
$PC4$	−0.159	−0.059	0.029	−0.057	0.050	−0.182	0.012	12.682	8.019
	(−2.185)	(−0.672)	(0.406)	(−0.779)	(0.676)	(−2.476)	(0.161)	(0.080)	(0.237)
$PC5$	0.025	0.186	−0.026	−0.012	0.111	−0.159	−0.116	38.089	37.813
	(0.525)	(3.867)	(−0.550)	(−0.245)	(2.361)	(−3.394)	(−2.453)	(0.000)	(0.000)

Notes: This table provides the estimates of the market price of risk parameters $\boldsymbol{\lambda}_0$ and $\boldsymbol{\lambda}_1$ in equation (4.9) for the Gaussian ATSM specified with equity tail risk. Estimated t-statistics are reported in parentheses. Wald statistics for tests of the rows of $\boldsymbol{\Lambda}$ and of $\boldsymbol{\lambda}_1$ being different from zero are reported along each row, with the corresponding p-values in parentheses below. The null hypothesis underlying W_{Λ_i} is that the risk related to a given factor is not priced in the term structure model. The null hypothesis underlying $W_{\lambda_{1_i}}$ is that the price of risk associated with a given factor does not vary over time.

Table 4.7 – Gaussian ATSM - Expected Excess Return Loadings

Maturity	$\widetilde{U}^{Equity}$	$PC1$	$PC2$	$PC3$	$PC4$	$PC5$
$6m$	−0.004	0.023	0.006	0.017	−0.009	−0.011
	(−0.793)	(4.100)	(1.026)	(3.031)	(−1.669)	(−2.064)
$12m$	−0.024	0.049	0.023	0.025	−0.011	−0.023
	(−1.963)	(3.973)	(1.834)	(1.991)	(−0.863)	(−1.827)
$18m$	−0.050	0.077	0.047	0.025	−0.010	−0.033
	(−2.382)	(3.660)	(2.207)	(1.170)	(−0.477)	(−1.575)
$24m$	−0.075	0.104	0.074	0.020	−0.013	−0.044
	(−2.434)	(3.396)	(2.384)	(0.660)	(−0.415)	(−1.425)
$30m$	−0.097	0.132	0.103	0.015	−0.022	−0.057
	(−2.323)	(3.186)	(2.478)	(0.368)	(−0.530)	(−1.358)
$36m$	−0.113	0.158	0.134	0.011	−0.039	−0.071
	(−2.142)	(3.011)	(2.535)	(0.215)	(−0.734)	(−1.344)
$42m$	−0.125	0.183	0.166	0.010	−0.062	−0.088
	(−1.932)	(2.854)	(2.574)	(0.154)	(−0.974)	(−1.360)
$48m$	−0.131	0.206	0.199	0.012	−0.093	−0.106
	(−1.718)	(2.708)	(2.603)	(0.151)	(−1.220)	(−1.393)
$54m$	−0.133	0.226	0.232	0.016	−0.128	−0.126
	(−1.511)	(2.569)	(2.626)	(0.185)	(−1.455)	(−1.431)
$60m$	−0.132	0.243	0.266	0.024	−0.166	−0.147
	(−1.317)	(2.433)	(2.644)	(0.241)	(−1.666)	(−1.468)
$66m$	−0.128	0.258	0.300	0.035	−0.207	−0.169
	(−1.142)	(2.301)	(2.657)	(0.308)	(−1.850)	(−1.501)
$72m$	−0.123	0.270	0.334	0.047	−0.249	−0.190
	(−0.986)	(2.173)	(2.667)	(0.379)	(−2.002)	(−1.525)
$78m$	−0.117	0.280	0.368	0.061	−0.290	−0.212
	(−0.852)	(2.049)	(2.673)	(0.448)	(−2.125)	(−1.541)
$84m$	−0.110	0.288	0.401	0.077	−0.330	−0.232
	(−0.738)	(1.930)	(2.676)	(0.513)	(−2.218)	(−1.548)
$90m$	−0.105	0.293	0.435	0.093	−0.369	−0.251
	(−0.646)	(1.817)	(2.676)	(0.571)	(−2.284)	(−1.545)
$96m$	−0.100	0.298	0.469	0.108	−0.405	−0.268
	(−0.572)	(1.711)	(2.675)	(0.621)	(−2.326)	(−1.535)
$102m$	−0.097	0.301	0.502	0.124	−0.438	−0.284
	(−0.517)	(1.611)	(2.672)	(0.662)	(−2.348)	(−1.517)
$108m$	−0.096	0.302	0.535	0.139	−0.468	−0.299
	(−0.479)	(1.519)	(2.668)	(0.694)	(−2.352)	(−1.493)
$114m$	−0.097	0.303	0.568	0.153	−0.496	−0.311
	(−0.455)	(1.433)	(2.664)	(0.719)	(−2.342)	(−1.463)
$120m$	−0.100	0.303	0.600	0.165	−0.519	−0.322
	(−0.445)	(1.355)	(2.659)	(0.735)	(−2.320)	(−1.429)

Notes: This table provides the estimates and t-statistics (in parentheses) of the expected excess return loadings on the factors of the proposed ATSM with equity tail risk. These coefficients are calculated as $\boldsymbol{\beta}^{(n)'}\boldsymbol{\lambda}_1^{(i)}$ and can be interpreted as the response of the expected one-month excess return on the n-month bond to a contemporaneous shock to the i-th pricing factor. Results are provided for the $N = 20$ Treasury returns used for model estimation.

4.7 Figures of Chapter 4

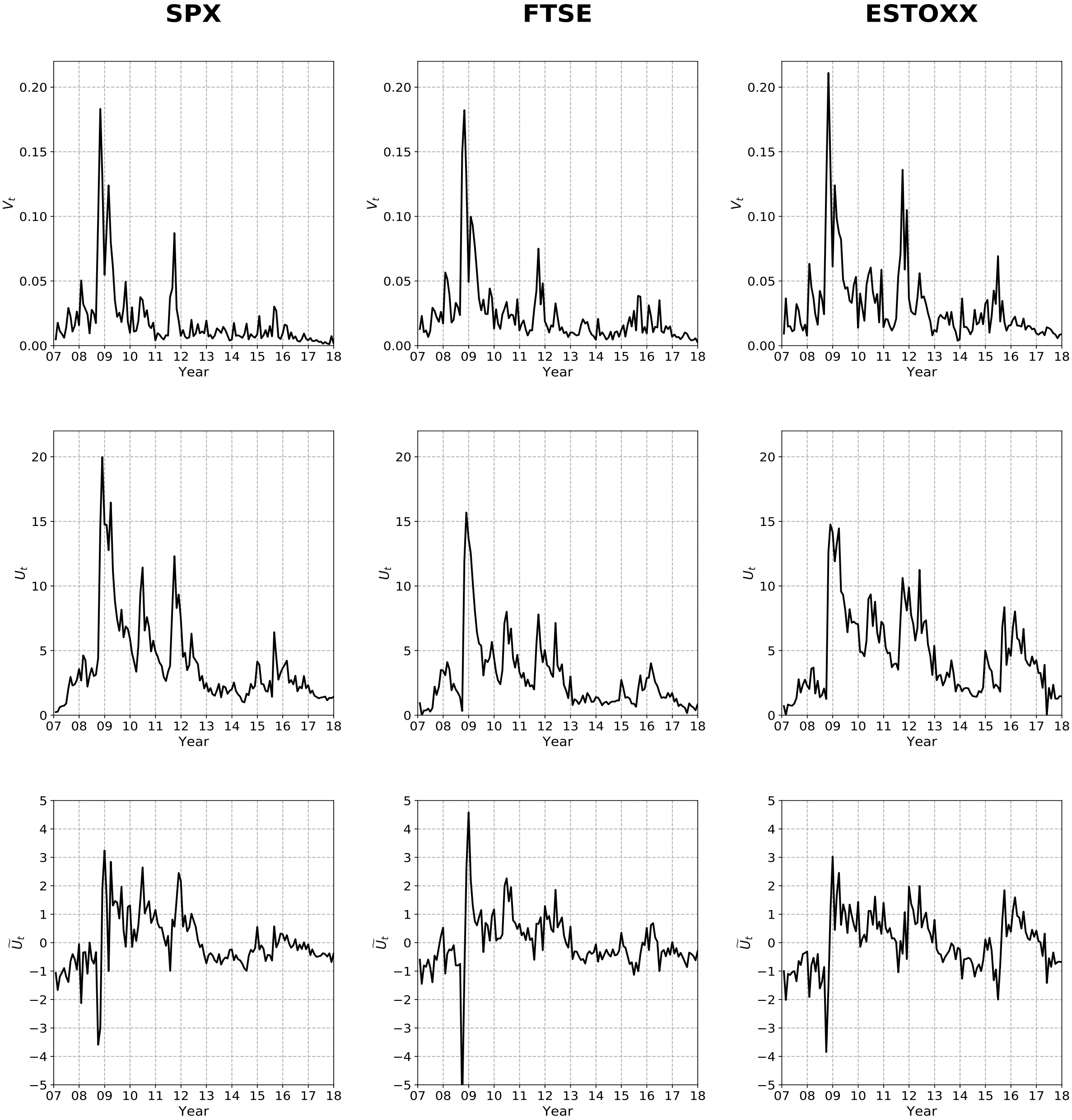


Figure 4.1 – Monthly option-implied state variables for the S&P 500, FTSE 100 and EURO STOXX 50 equity-index returns. Estimates are obtained using the model parameter values from Table 4.1, 4.2 and 4.3. Top panel: annualised spot variance. Middle panel: annualised negative jump intensity factor. Bottom panel: component of the negative jump intensity factor orthogonal to spot variance and normalised to have mean zero and unit variance. The equity left tail risk factor that we use in the Gaussian ATSM for US interest rates is obtained as the market capitalization weighted average of the $\widetilde{U}$ factor of the three stock market indices.

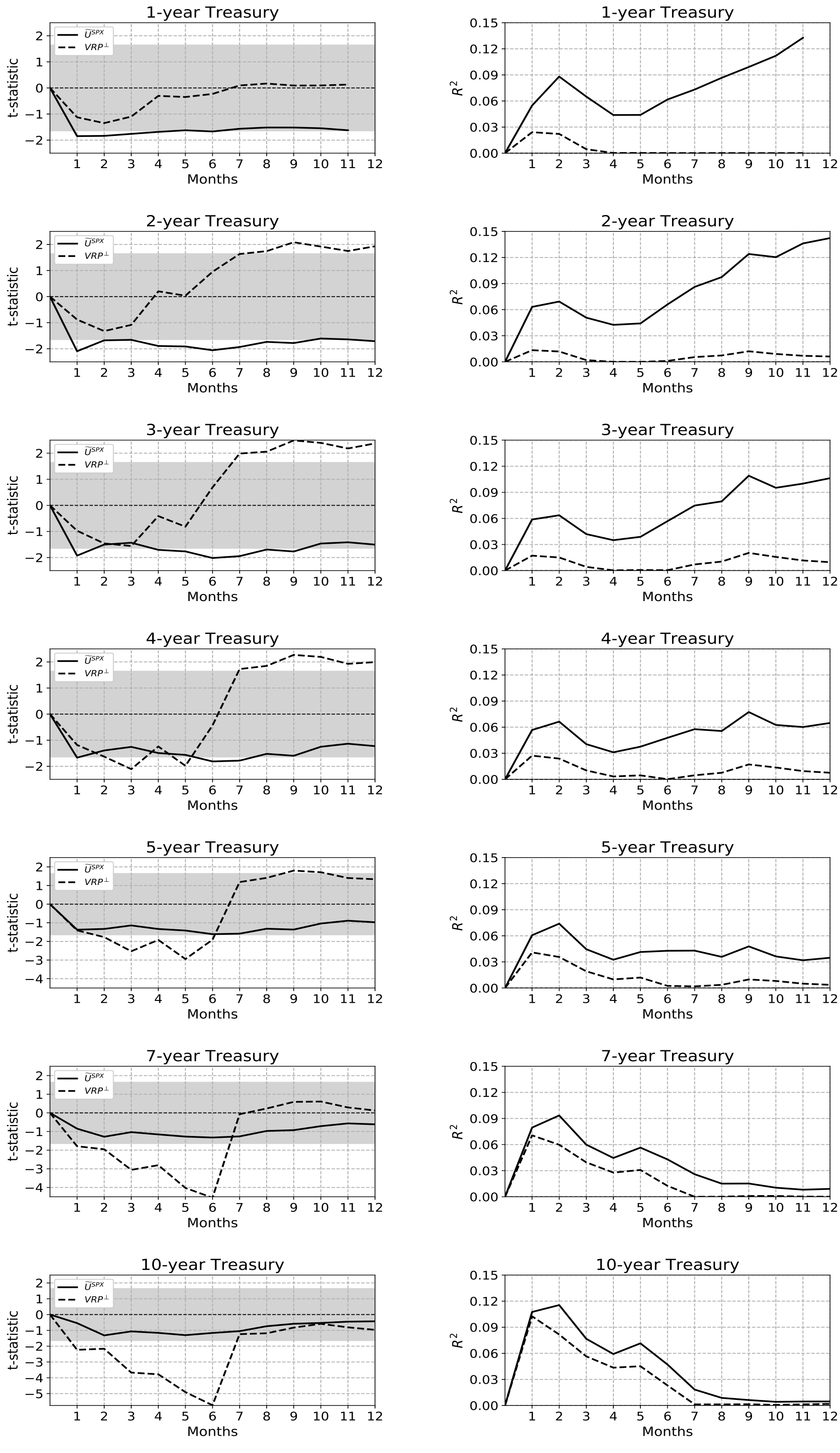


Figure 4.2 – Results of the regressions of excess returns of US Treasury bonds with 1-, 2-, 3-, 4-, 5-, 7- and 10-year maturities on the option-implied left jump intensity factor (orthogonal to spot variance) of S&P 500 equity-index returns, $\widetilde{U}^{\text{SPX}}$, and the variance risk premium orthogonal to $\widetilde{U}^{\text{SPX}}$, $VRP^{\perp}$. Regressions are run for holding periods from 1 to 12 months using the full sample of data from 2007 to 2017. Left panels: Newey-West t-statistics for the regression slopes, along with regions of statistical insignificance at the 10% level (shaded area). Right panels: regression R^2 obtained using just $VRP^{\perp}$ (dashed line) and both $\widetilde{U}^{\text{SPX}}$ and $VRP^{\perp}$ (solid line).

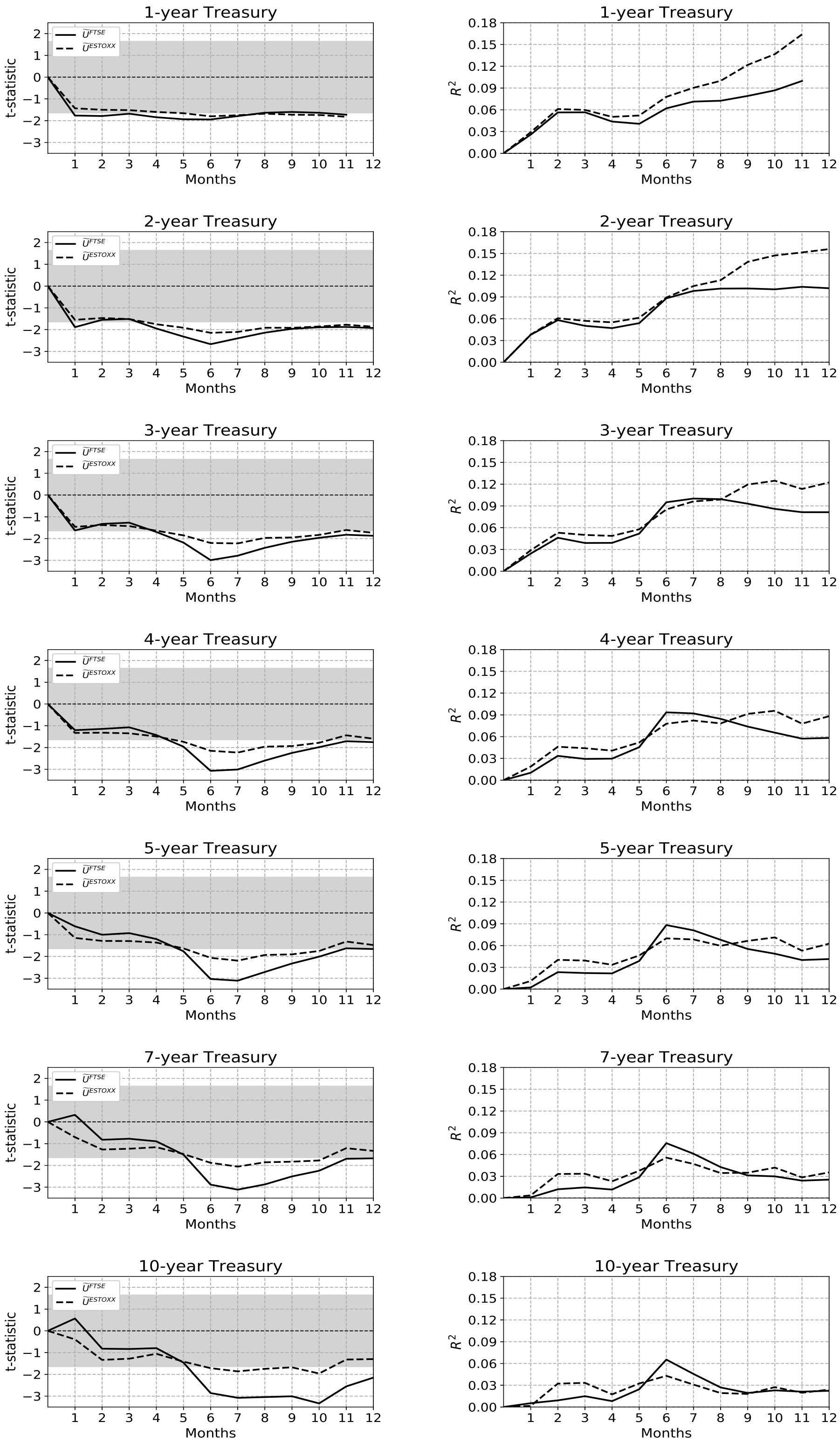


Figure 4.3 – Results of the regressions of excess returns of US Treasury bonds with 1-, 2-, 3-, 4-, 5-, 7- and 10-year maturities on the option-implied left jump intensity factor (orthogonal to spot variance) of FTSE 100, $\widetilde{U}^{\mathrm{FTSE}}$, and of EURO STOXX 50 equity-index returns, $\widetilde{U}^{\mathrm{ESTOXX}}$. Regressions are run separately on each intensity factor, for holding periods from 1 to 12 months using the full sample of data from 2007 to 2017. Left panels: Newey-West t-statistics for the regression slopes, along with regions of statistical insignificance at the 10% level (shaded area). Right panels: R^2 of the regressions on $\widetilde{U}^{\mathrm{FTSE}}$ (solid line) and on $\widetilde{U}^{\mathrm{ESTOXX}}$ (dashed line).

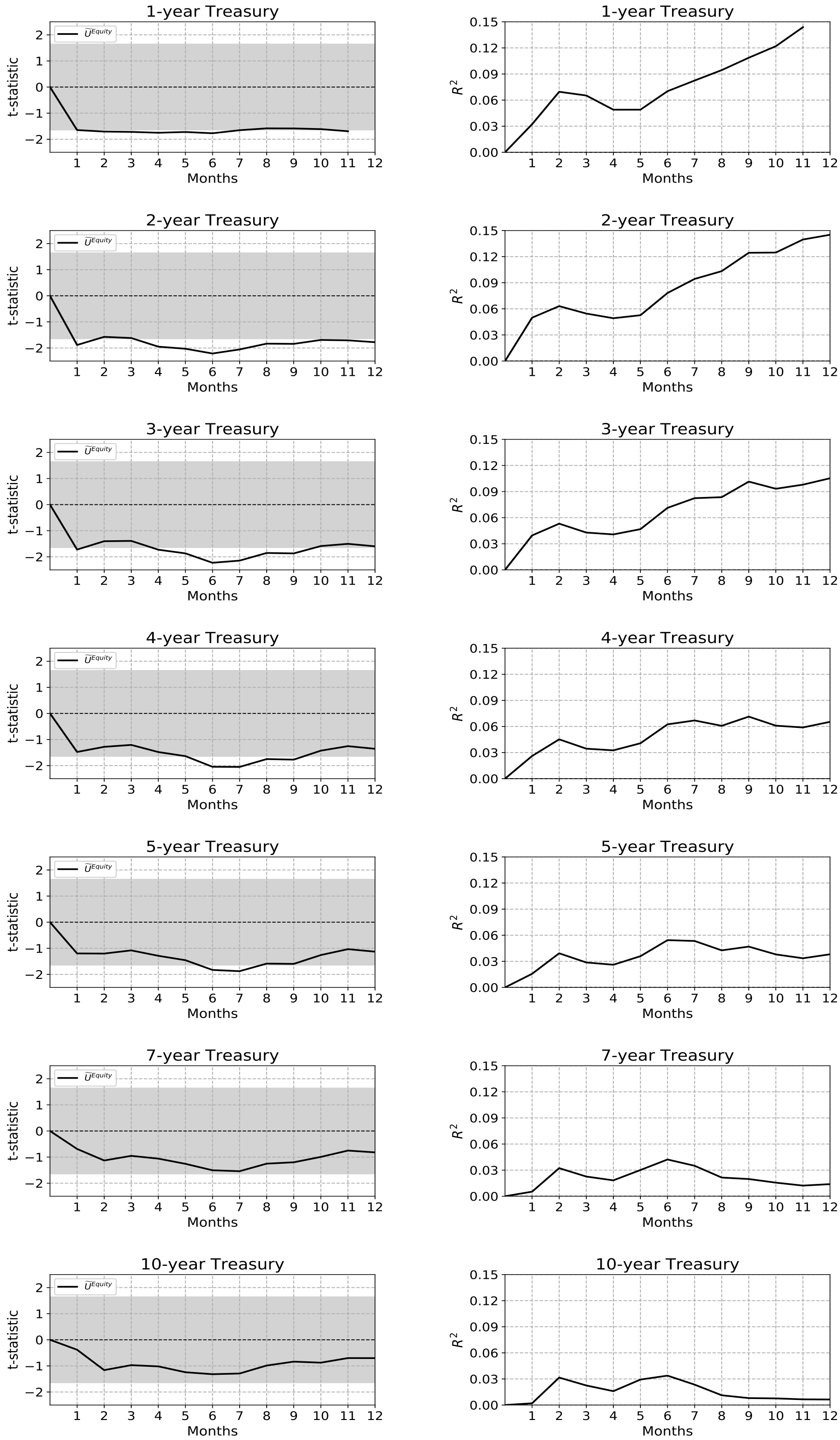


Figure 4.4 – Results of the regressions of excess returns of US Treasury bonds with 1-, 2-, 3-, 4-, 5-, 7- and 10-year maturities on a constant and $\widetilde{U}^{Equity}$, which is the market capitalization weighted average of the option-implied left jump intensity factor (orthogonal to spot variance) of S&P 500, FTSE 100 and EURO STOXX 50 equity-index returns. Regressions are run for holding periods from 1 to 12 months using the full sample of data from 2007 to 2017. Left panels: Newey-West t-statistics for the coefficient of $\widetilde{U}^{Equity}$, along with regions of statistical insignificance at the 10% level (shaded area). Right panels: R^2 of the regressions.

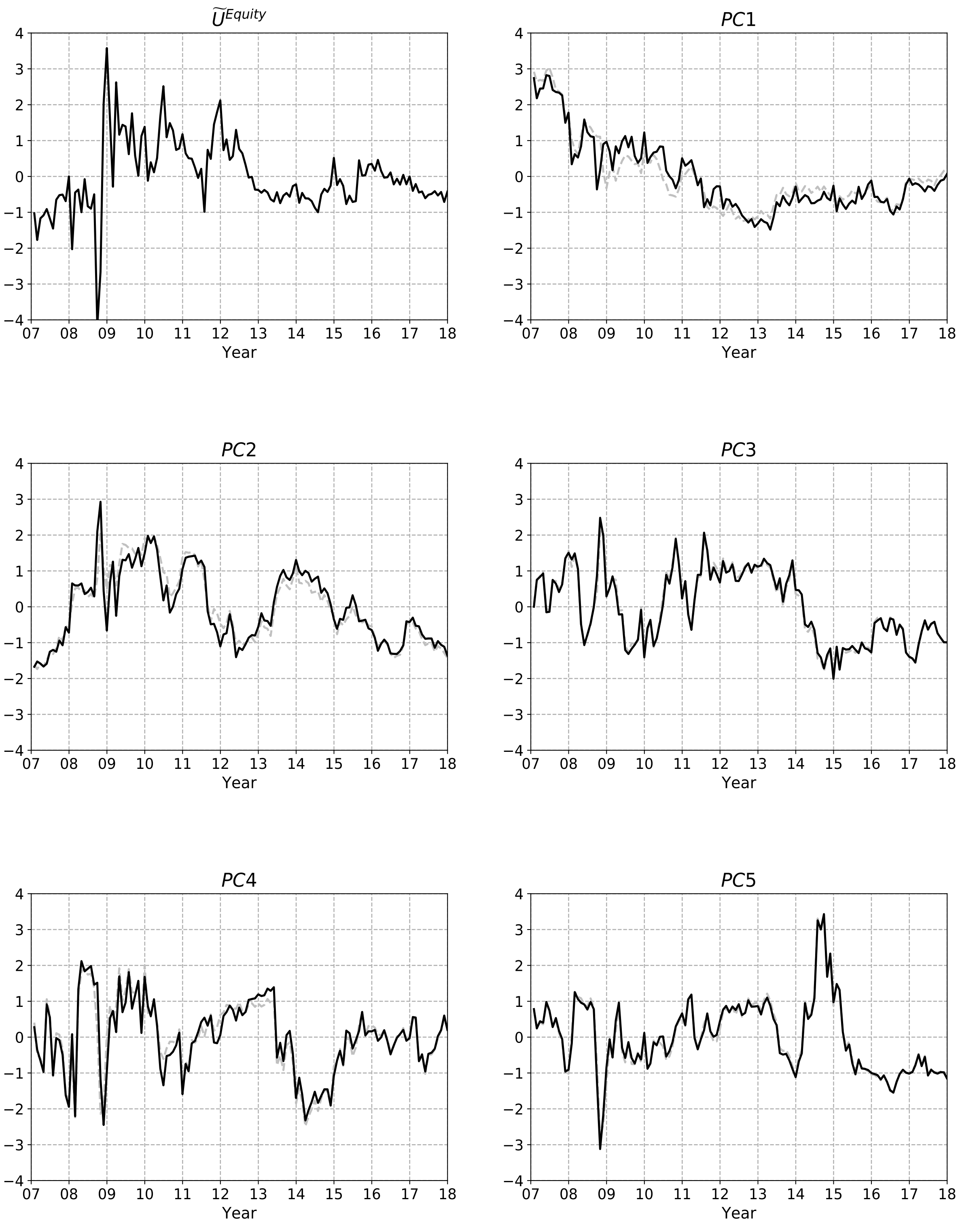


Figure 4.5 – Monthly time series of the pricing factors of our Gaussian ATSM. The top-left panel shows the equity left tail factor associated with the S&P 500, FTSE 100 and EURO STOXX 50 index returns, calculated from equation (4.5) and then normalised to have mean zero and unit variance. The remaining panels show the first five standardised principal components extracted from the US Treasury yields of maturities $n = 3, 6, ..., 120$ months, orthogonal to the $\widetilde{U}^{Equity}$ factor. The light-colored dashed lines show the principal components extracted from non-orthogonalized yields, which, however, are not used as pricing factors in our model.

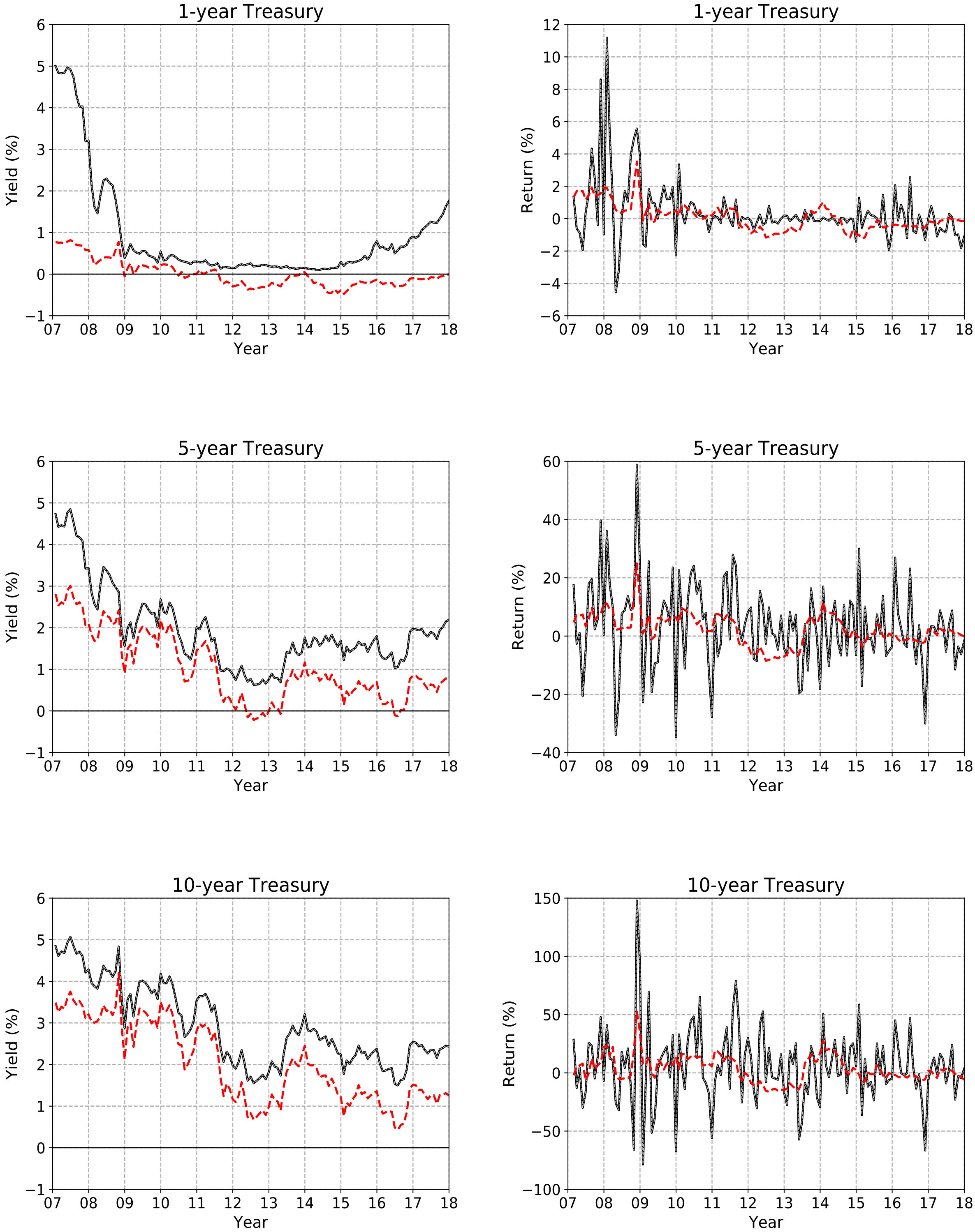


Figure 4.6 – Observed and model-implied time series of yields and one-month excess returns on US Treasury bonds with 1-, 5- and 10-year maturities. In the left panels, the solid black lines show the observed yields, the dashed grey lines plot the model-implied yields, while the dashed red lines indicate the model-implied term premia. In the right panels, the solid black lines show the observed excess returns, the dashed grey lines plot the model-implied excess returns, while the dashed red lines indicate the model-implied expected excess returns.

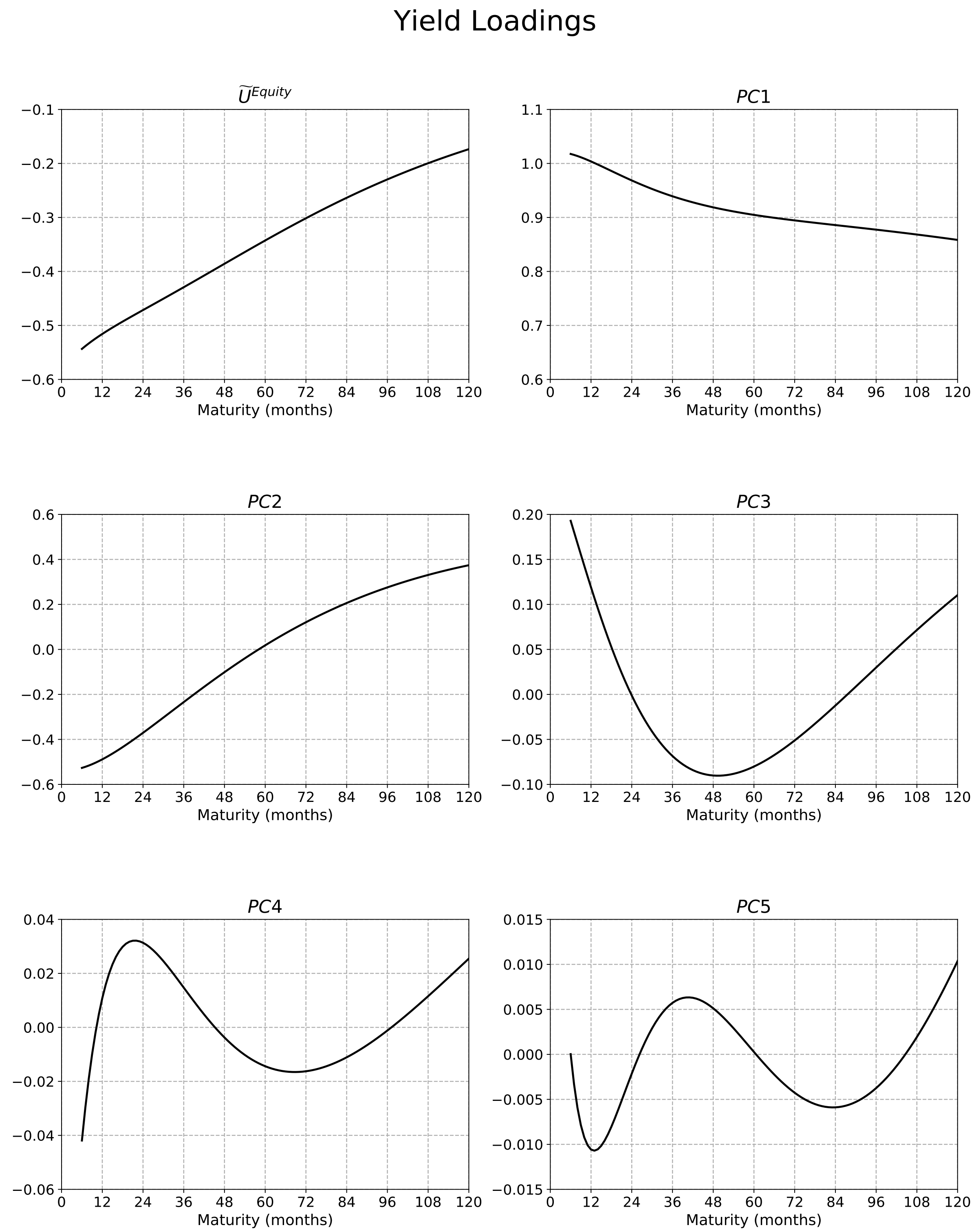


Figure 4.7 – Model-implied yield loadings on the pricing factors of the proposed ATSM with equity tail risk. These coefficients are calculated as $-(1/n)\mathbf{b}_n$ and can be interpreted as the response of the n-month yield to a contemporaneous shock to the respective factor. $\widetilde{U}^{Equity}$ represents the equity left tail factor associated with the S&P 500, FTSE 100 and EURO STOXX 50 index returns, calculated from equation (4.5) and then standardised. $PC1$ – $PC5$ denote the first five standardised principal components extracted from the US Treasury yields orthogonal with respect to the $\widetilde{U}^{Equity}$ factor.

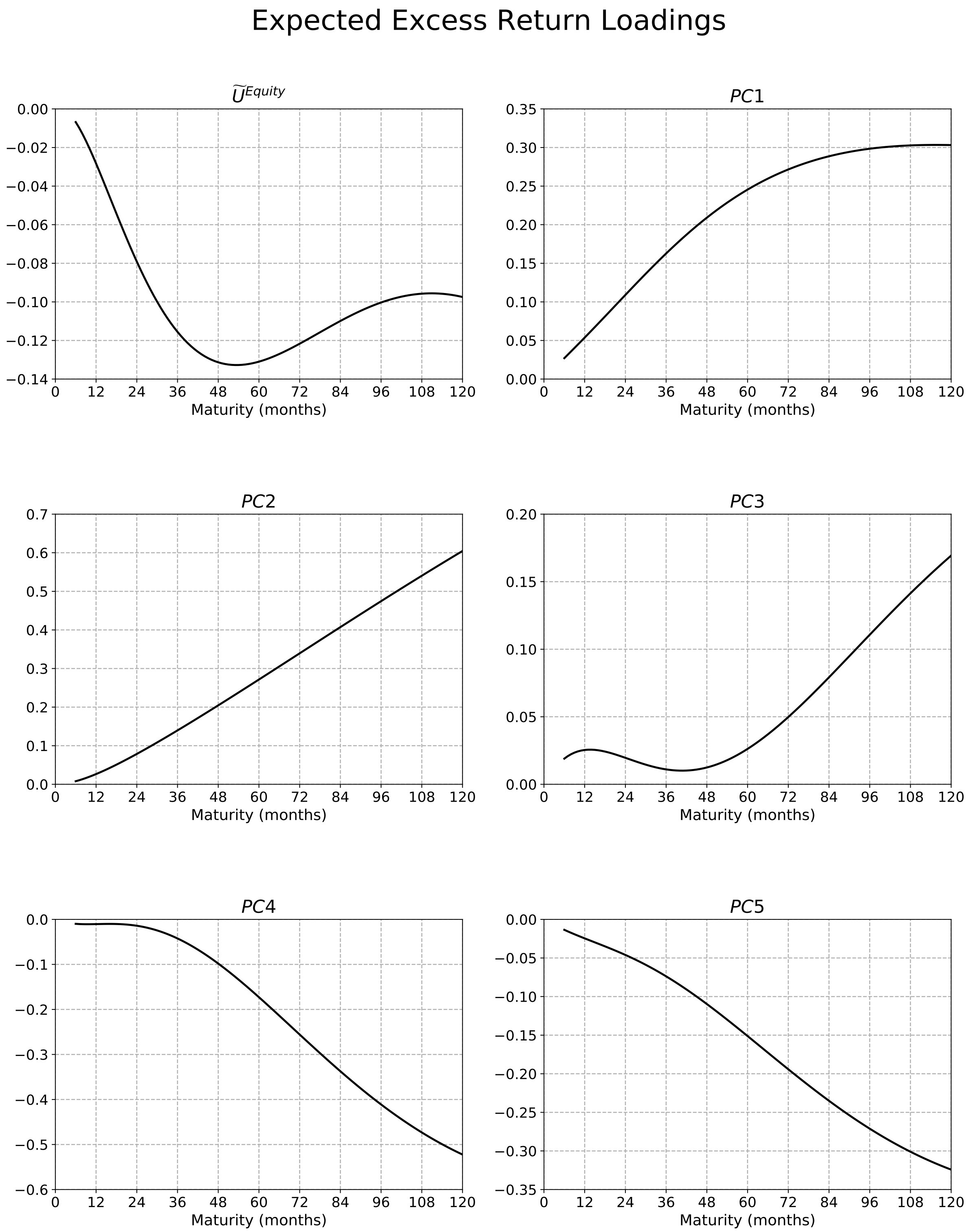


Figure 4.8 – Model-implied expected excess return loadings on the pricing factors of the proposed ATSM with equity tail risk. These coefficients are calculated as $\mathbf{b}_n^{'}\boldsymbol{\lambda}_1$ and can be interpreted as the response of the expected one-month excess return on the n-month bond to a contemporaneous shock to the respective factor. $\widetilde{U}^{Equity}$ represents the equity left tail factor associated with the S&P 500, FTSE 100 and EURO STOXX 50 index returns, calculated from equation (4.5) and then standardised. $PC1$ – $PC5$ denote the first five standardised principal components extracted from the US Treasury yields orthogonal with respect to the $\widetilde{U}^{Equity}$ factor.

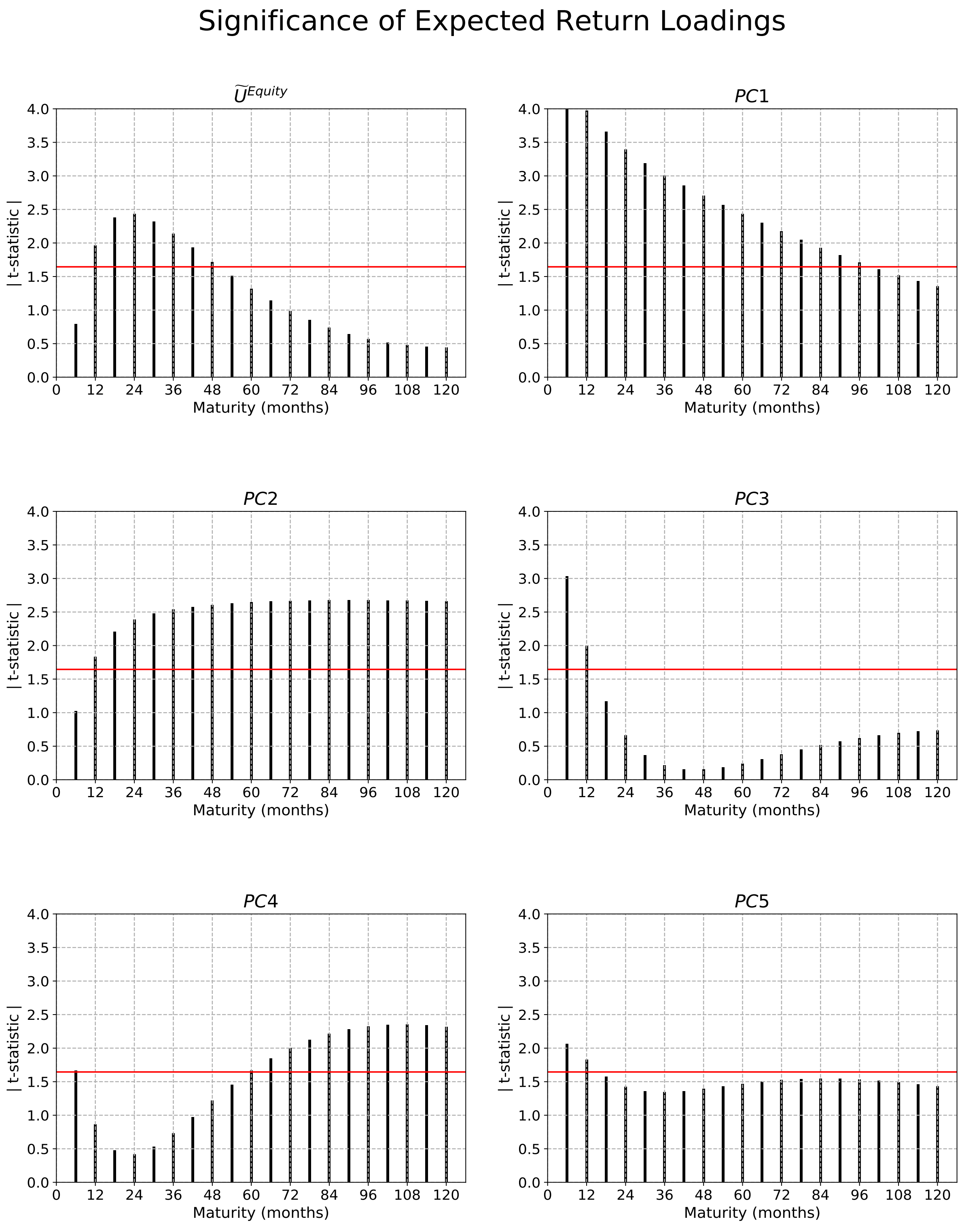


Figure 4.9 – Significance of expected return loadings on the pricing factors of the proposed ATSM with equity tail risk. The absolute value of the t-statistic is reported for the $N = 20$ one-month excess Treasury returns used to fit the cross-section of yields. The solid red lines depict the critical value of the statistics for the significance level of 10%. $\widetilde{U}^{Equity}$ represents the equity left tail factor associated with the S&P 500, FTSE 100 and EURO STOXX 50 index returns, calculated from equation (4.5) and then standardised. $PC1$ – $PC5$ denote the first five standardised principal components extracted from the US Treasury yields orthogonal with respect to $\widetilde{U}^{Equity}$.

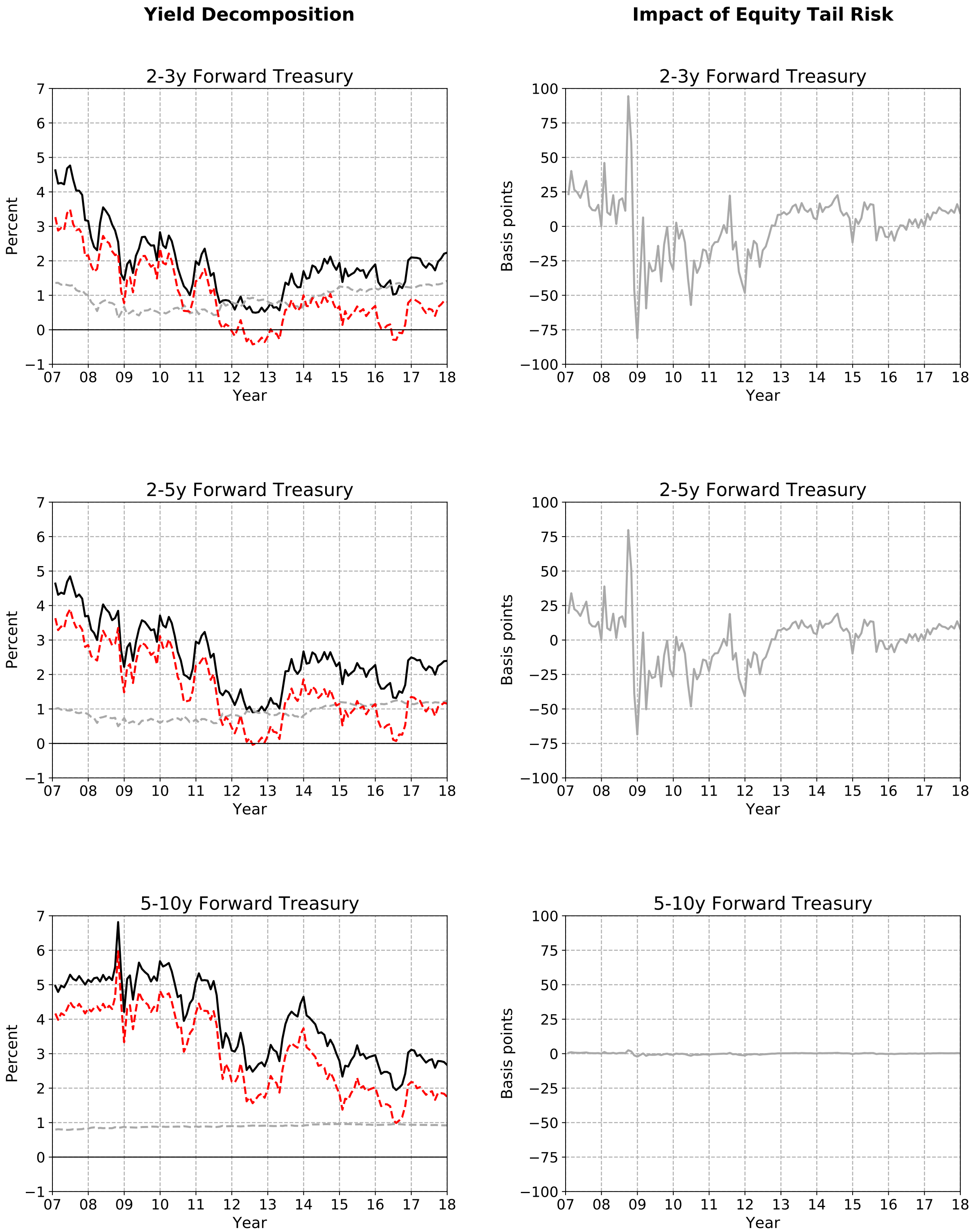
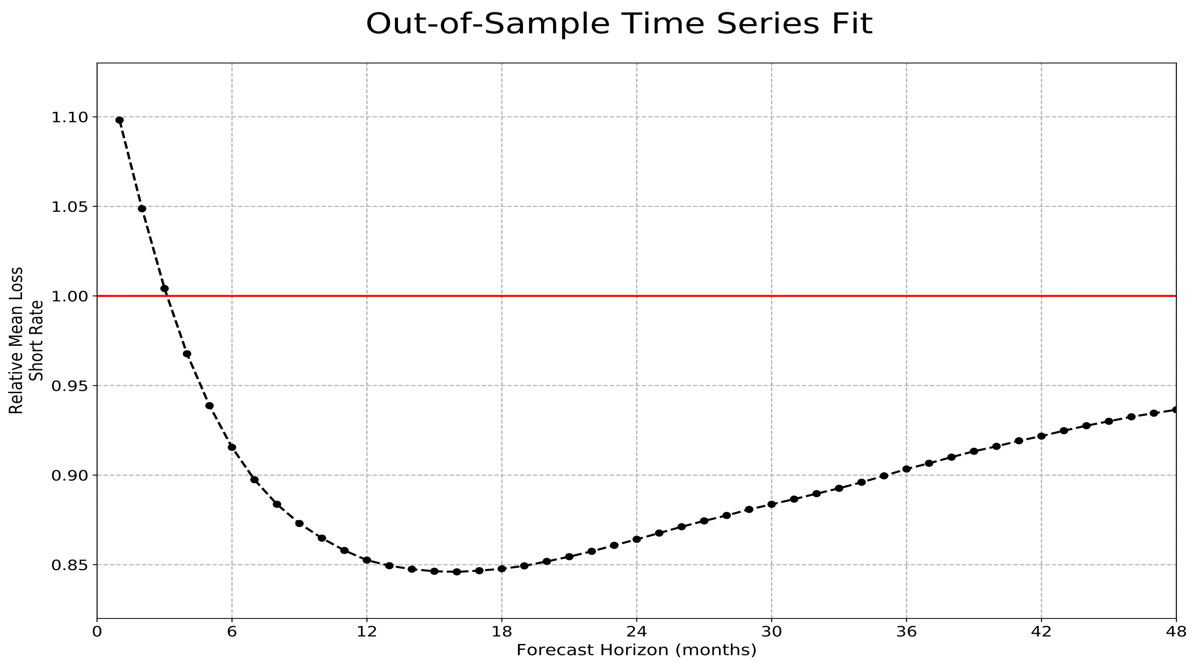


Figure 4.10 – Contribution of changes in equity tail risk to the term premia embedded in the 2-3y, 2-5y and 5-10y forward Treasury rates. In the left panels, the solid black lines show the model-implied m-n forward rates, the dashed grey lines plot the risk-neutral m-n forward rates (the average expectation of the short rates over the next m to n periods), while the dashed red lines indicate the model-implied term premia embedded in the m-n forward rates. In the right panels, the solid dark grey lines show the impact over time of the equity left tail factor $\widetilde{U}^{Equity}$ on the model-implied term premium of the m-n forward Treasury rates.



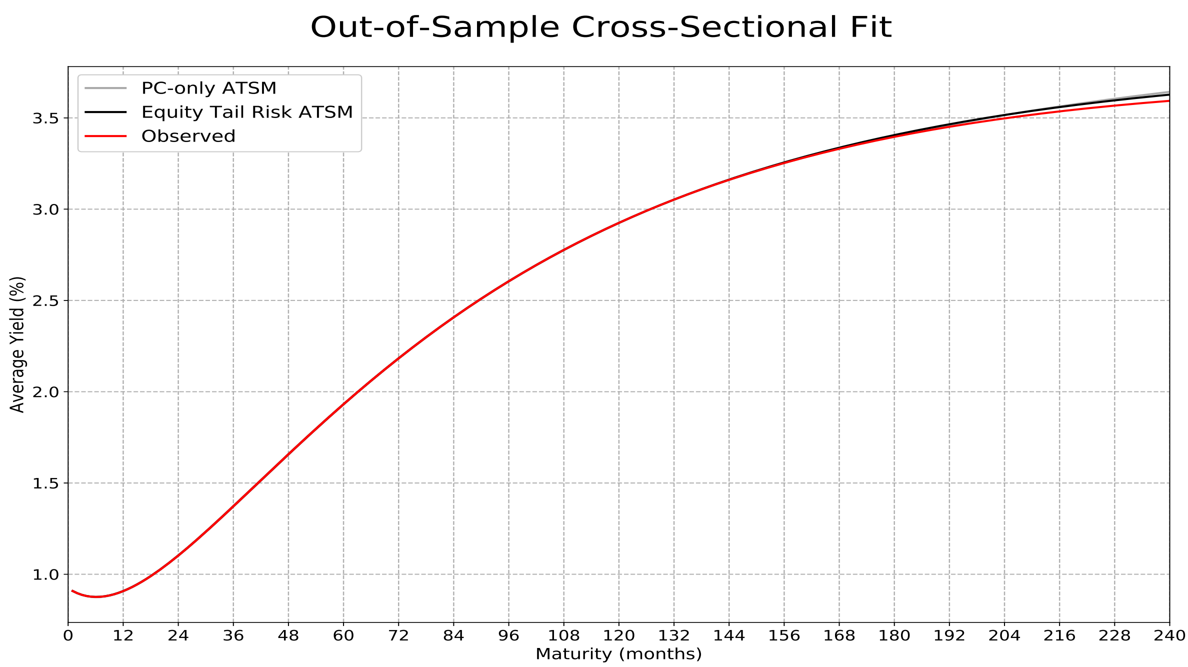


Figure 4.11 – Out-of-sample performance of the proposed ATSM with equity tail risk. Top panel: root mean squared forecast error for the Equity Tail Risk ATSM relative to the benchmark PC-only specification (red line), plotted against the forecast horizon. Results are based on the out-of-sample forecasts of the average short rate over horizons from one month to four years. Bottom panel: observed average yields for maturities up twenty years (red line), along with the average yields implied by the Equity Tail Risk ATSM (black line) and the benchmark PC-only specification (grey line) estimated using only maturities up to ten years.

4.8 Appendix A: Robustness

In this appendix we provide tables and figures illustrating the performance of a Gaussian ATSM in which the pricing factors are the first five principal components of Treasury yields and the left jump intensity factor (orthogonal to spot variance) of S&P 500 equity-index returns, $\widetilde{U}^{SPX}$. As we did in Section 4.4.4, we estimate the principal components from an eigenvalue decomposition of the covariance matrix of zero-coupon bond yields of maturities $n = 3, 6, ..., 120$ months, orthogonal to the extra factor, $\widetilde{U}^{SPX}$. By using the option-implied pure tail factor of the US stock market to drive the curve of US Treasury rates and explain bond returns, we find qualitatively similar results to those obtained in Section 4.4.4 using $\widetilde{U}^{Equity}$, i.e. the market capitalization weighted average of the S&P 500, FTSE 100 and EURO STOXX 50 option-implied tail factors. The following points summarise the main findings. First, with respect to the S&P 500 option-implied tail factor we strongly reject the hypothesis of unspanned factor in the ATSM for US interest rates. Second, we find a good in-sample fit between the data and the proposed model with US-only equity tail risk. As we have seen in Table 4.5, the return pricing errors on short-term bonds are smaller than in the PC-only model specification. Third, equity tail risk, as measured by exposure to $\widetilde{U}^{SPX}$, is significantly priced (p-value equal to 8.3%) and time-varying (p-value equal to 5.5%) within the term structure model for US interest rates. Fourth, contemporaneous bond yields and future expcted bond returns load negatively on the S&P 500 option-implied tail factor. This result provides evidence of flights from the riskiness of the US stock market to the safety of US Treasury bonds. However, as it was the case for $\widetilde{U}^{Equity}$, the predictive power of $\widetilde{U}^{SPX}$ for lower bond returns is statistically significant only at the short end of the US yield curve. Finally, the inclusion of the pure tail factor of the US stock market in the Gaussian ATSM leads to out-of-sample forecasts of future short rates improved with respect to those produced by a benchmark PC-only specification. Nonetheless, the gains in forecast accuracy are smaller than those reported in Section 4.4.5, where we considered the downside tail risk of the international stock market.

Table 4.8 – US-only equity tail risk ATSM - Factor Risk Exposures

	US-only Equity Tail Risk ATSM		PC-only ATSM	
Factor	W_{β_i}	p-value	W_{β_i}	p-value
$\widetilde{U}^{SPX}$	22274116.085	0.000	-	-
$PC1$	92988864.429	0.000	81008679.455	0.000
$PC2$	22353915.728	0.000	22570673.342	0.000
$PC3$	3140929.332	0.000	2845956.394	0.000
$PC4$	407193.929	0.000	402369.064	0.000
$PC5$	37184.076	0.000	34767.623	0.000

Notes: This table provides the Wald statistics and corresponding p-values for the Wald test of whether the exposures of bond returns to a given model factor are jointly zero. Under the null $H_0 : \boldsymbol{\beta}_i = \mathbf{0}_{N\times 1}$ the i-th pricing factor is unspanned, i.e. Treasury returns are not exposed to that factor. The p-values of the statistics are obtained from a chi-squared distribution with $N = 20$ degrees of freedom. The test is conducted on the pricing factors of both the proposed ATSM specified with US-only equity tail risk and a benchmark PC-only model specification.

Table 4.9 – US-only equity tail risk ATSM - Fit Diagnostics

Panel A: US-only Equity Tail Risk ATSM

	$n = 12$	$n = 24$	$n = 36$	$n = 60$	$n = 84$	$n = 120$
Panel A1: Yield Pricing Errors						
Mean	−0.001	0.000	0.000	0.000	0.000	0.000
Standard Deviation	0.002	0.001	0.001	0.001	0.001	0.002
Skewness	−0.916	1.828	1.713	−0.868	1.151	−1.066
Kurtosis	5.383	9.591	7.159	5.507	6.525	6.223
$\rho(1)$	0.749	0.749	0.829	0.757	0.774	0.777
$\rho(6)$	0.212	0.213	0.288	0.161	0.185	0.063
Panel A2: Return Pricing Errors						
Mean	−0.001	0.001	0.001	−0.001	0.000	−0.001
Standard Deviation	0.022	0.017	0.023	0.055	0.034	0.183
Skewness	−0.343	−0.925	−1.340	0.112	−0.508	0.029
Kurtosis	6.142	14.835	10.349	5.650	11.793	4.942
$\rho(1)$	−0.106	−0.200	−0.024	−0.141	−0.153	−0.044
$\rho(6)$	0.039	0.202	0.169	0.047	0.197	0.040

Panel B: PC-only ATSM

	$n = 12$	$n = 24$	$n = 36$	$n = 60$	$n = 84$	$n = 120$
Panel B1: Yield Pricing Errors						
Mean	0.001	0.000	0.001	0.000	−0.001	0.000
Standard Deviation	0.004	0.002	0.001	0.002	0.001	0.002
Skewness	−1.200	2.343	2.205	−1.085	0.390	−1.402
Kurtosis	3.913	10.582	9.608	3.348	3.106	6.193
$\rho(1)$	0.908	0.826	0.872	0.866	0.882	0.787
$\rho(6)$	0.571	0.341	0.376	0.471	0.454	0.122
Panel B2: Return Pricing Errors						
Mean	0.000	−0.003	0.000	0.000	−0.006	0.006
Standard Deviation	0.025	0.020	0.026	0.052	0.040	0.171
Skewness	−0.371	−0.129	−2.076	−0.256	−0.148	−0.229
Kurtosis	14.188	15.218	17.183	11.281	8.139	6.232
$\rho(1)$	−0.046	−0.186	−0.076	−0.135	−0.126	−0.063
$\rho(6)$	0.117	0.272	0.249	0.108	0.188	0.066

Notes: This table contains the summary statistics of the pricing errors implied by the Gaussian ATSM that includes US-only equity tail risk (Panel A) and by the benchmark model that only uses the first five PCs of the yield curve (Panel B). Models are estimated over the period 2007 to 2017. Reported are the sample mean, standard deviation, skewness, kurtosis and the autocorrelation coefficients of order one and six. Panels A1 and B1: properties of the yield pricing errors $\hat{u}$. Panels A2 and B2: properties of the return pricing errors $\hat{e}$. n denotes the maturity of the bonds in months.

Table 4.10 – US-only equity tail risk ATSM - Market Prices of Risk

Factor	λ_0	$\lambda_{1,1}$	$\lambda_{1,2}$	$\lambda_{1,3}$	$\lambda_{1,4}$	$\lambda_{1,5}$	$\lambda_{1,6}$	W_{Λ_i}	$W_{\lambda_{1_i}}$
$\widetilde{U}^{SPX}$	0.079	0.575	−0.045	−0.056	0.204	0.023	−0.233	12.579	12.352
	(0.532)	(3.230)	(−0.301)	(−0.380)	(1.338)	(0.153)	(−1.542)	(0.083)	(0.055)
$PC1$	−0.008	0.231	−0.064	−0.076	0.059	0.043	−0.049	16.260	16.244
	(−0.155)	(3.582)	(−1.180)	(−1.407)	(1.062)	(0.779)	(−0.893)	(0.023)	(0.013)
$PC2$	−0.053	−0.210	0.038	−0.031	−0.067	0.051	0.098	13.455	12.770
	(−0.870)	(−3.022)	(0.623)	(−0.505)	(−1.088)	(0.829)	(1.596)	(0.062)	(0.047)
$PC3$	0.108	−0.124	0.050	0.065	−0.116	0.032	0.025	23.403	17.772
	(2.386)	(−2.709)	(1.101)	(1.428)	(−2.574)	(0.720)	(0.549)	(0.001)	(0.007)
$PC4$	−0.161	0.009	0.044	−0.077	0.056	−0.200	0.006	15.268	10.130
	(−2.295)	(0.116)	(0.629)	(−1.101)	(0.782)	(−2.821)	(0.080)	(0.033)	(0.119)
$PC5$	0.022	0.124	−0.039	0.003	0.110	−0.149	−0.114	29.795	29.555
	(0.482)	(2.650)	(−0.850)	(0.065)	(2.362)	(−3.204)	(−2.450)	(0.000)	(0.000)

Notes: This table provides the estimates of the market price of risk parameters $\boldsymbol{\lambda}_0$ and $\boldsymbol{\lambda}_1$ in equation (4.9) for the Gaussian ATSM specified with US-only equity tail risk. Estimated t-statistics are reported in parentheses. Wald statistics for tests of the rows of $\boldsymbol{\Lambda}$ and of $\boldsymbol{\lambda}_1$ being different from zero are reported along each row, with the corresponding p-values in parentheses below. The null hypothesis underlying W_{Λ_i} is that the risk related to a given factor is not priced in the term structure model. The null hypothesis underlying $W_{\lambda_{1_i}}$ is that the price of risk associated with a given factor does not vary over time.

Table 4.11 – US-only equity tail risk ATSM - Expected Excess Return Loadings

Maturity	$\widetilde{U}^{SPX}$	$PC1$	$PC2$	$PC3$	$PC4$	$PC5$
$6m$	−0.004 (−0.811)	0.022 (4.084)	0.006 (1.053)	0.017 (3.038)	−0.009 (−1.665)	−0.011 (−2.075)
$12m$	−0.024 (−1.960)	0.049 (3.983)	0.023 (1.864)	0.024 (1.961)	−0.011 (−0.853)	−0.022 (−1.802)
$18m$	−0.050 (−2.377)	0.077 (3.680)	0.047 (2.240)	0.024 (1.125)	−0.010 (−0.462)	−0.032 (−1.537)
$24m$	−0.075 (−2.442)	0.105 (3.416)	0.075 (2.421)	0.019 (0.611)	−0.012 (−0.391)	−0.043 (−1.386)
$30m$	−0.098 (−2.352)	0.132 (3.200)	0.105 (2.522)	0.013 (0.321)	−0.021 (−0.499)	−0.055 (−1.324)
$36m$	−0.116 (−2.196)	0.158 (3.015)	0.137 (2.586)	0.009 (0.173)	−0.036 (−0.695)	−0.070 (−1.317)
$42m$	−0.130 (−2.015)	0.182 (2.846)	0.170 (2.633)	0.008 (0.117)	−0.059 (−0.927)	−0.086 (−1.341)
$48m$	−0.139 (−1.830)	0.204 (2.687)	0.204 (2.671)	0.009 (0.120)	−0.088 (−1.165)	−0.105 (−1.381)
$54m$	−0.145 (−1.652)	0.222 (2.535)	0.239 (2.701)	0.014 (0.160)	−0.122 (−1.392)	−0.126 (−1.427)
$60m$	−0.149 (−1.486)	0.238 (2.387)	0.274 (2.727)	0.022 (0.221)	−0.159 (−1.596)	−0.147 (−1.471)
$66m$	−0.150 (−1.337)	0.251 (2.243)	0.309 (2.748)	0.033 (0.293)	−0.198 (−1.772)	−0.170 (−1.510)
$72m$	−0.150 (−1.205)	0.261 (2.104)	0.345 (2.764)	0.046 (0.369)	−0.238 (−1.919)	−0.192 (−1.539)
$78m$	−0.149 (−1.091)	0.269 (1.971)	0.381 (2.775)	0.060 (0.442)	−0.278 (−2.035)	−0.214 (−1.559)
$84m$	−0.149 (−0.996)	0.274 (1.844)	0.417 (2.783)	0.076 (0.509)	−0.316 (−2.124)	−0.235 (−1.569)
$90m$	−0.149 (−0.918)	0.278 (1.725)	0.453 (2.787)	0.092 (0.570)	−0.353 (−2.186)	−0.254 (−1.569)
$96m$	−0.149 (−0.856)	0.280 (1.613)	0.488 (2.788)	0.108 (0.621)	−0.387 (−2.226)	−0.273 (−1.560)
$102m$	−0.151 (−0.809)	0.281 (1.510)	0.523 (2.787)	0.124 (0.663)	−0.419 (−2.246)	−0.289 (−1.544)
$108m$	−0.155 (−0.777)	0.281 (1.415)	0.558 (2.785)	0.139 (0.696)	−0.447 (−2.249)	−0.304 (−1.520)
$114m$	−0.161 (−0.757)	0.281 (1.329)	0.592 (2.782)	0.153 (0.720)	−0.473 (−2.238)	−0.317 (−1.490)
$120m$	−0.168 (−0.747)	0.280 (1.251)	0.626 (2.778)	0.165 (0.737)	−0.496 (−2.217)	−0.327 (−1.456)

Notes: This table provides the estimates and t-statistics (in parentheses) of the expected excess return loadings on the factors of the proposed US-only equity tail risk ATSM. These coefficients are calculated as $\boldsymbol{\beta}^{(n)\prime}\boldsymbol{\lambda}_1^{(i)}$ and can be interpreted as the response of the expected one-month excess return on the n-month bond to a contemporaneous shock to the i-th pricing factor. Results are provided for the $N = 20$ Treasury returns used for model estimation.

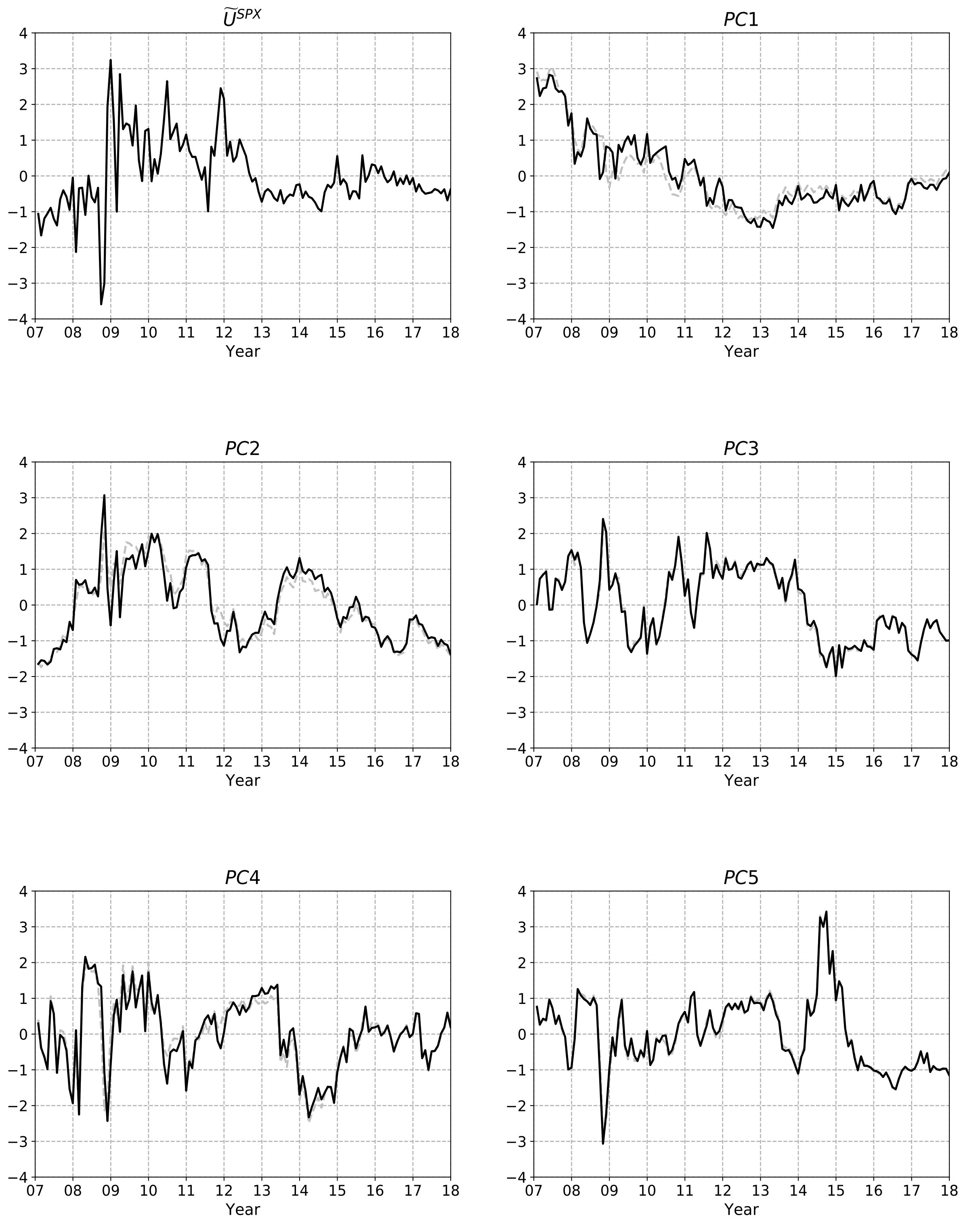


Figure 4.12 – US-only equity tail risk ATSM: Monthly time series of the pricing factors of the proposed US-only equity tail risk ATSM. The top-left panel shows the equity left tail factor associated with the S&P 500, normalised to have mean zero and unit variance. The remaining panels show the first five standardised principal components extracted from the US Treasury yields of maturities $n = 3, 6, ..., 120$ months, orthogonal to the $\widetilde{U}^{SPX}$ factor. The light-colored dashed lines show the principal components extracted from non-orthogonalized yields, which, however, are not used as pricing factors in our model.

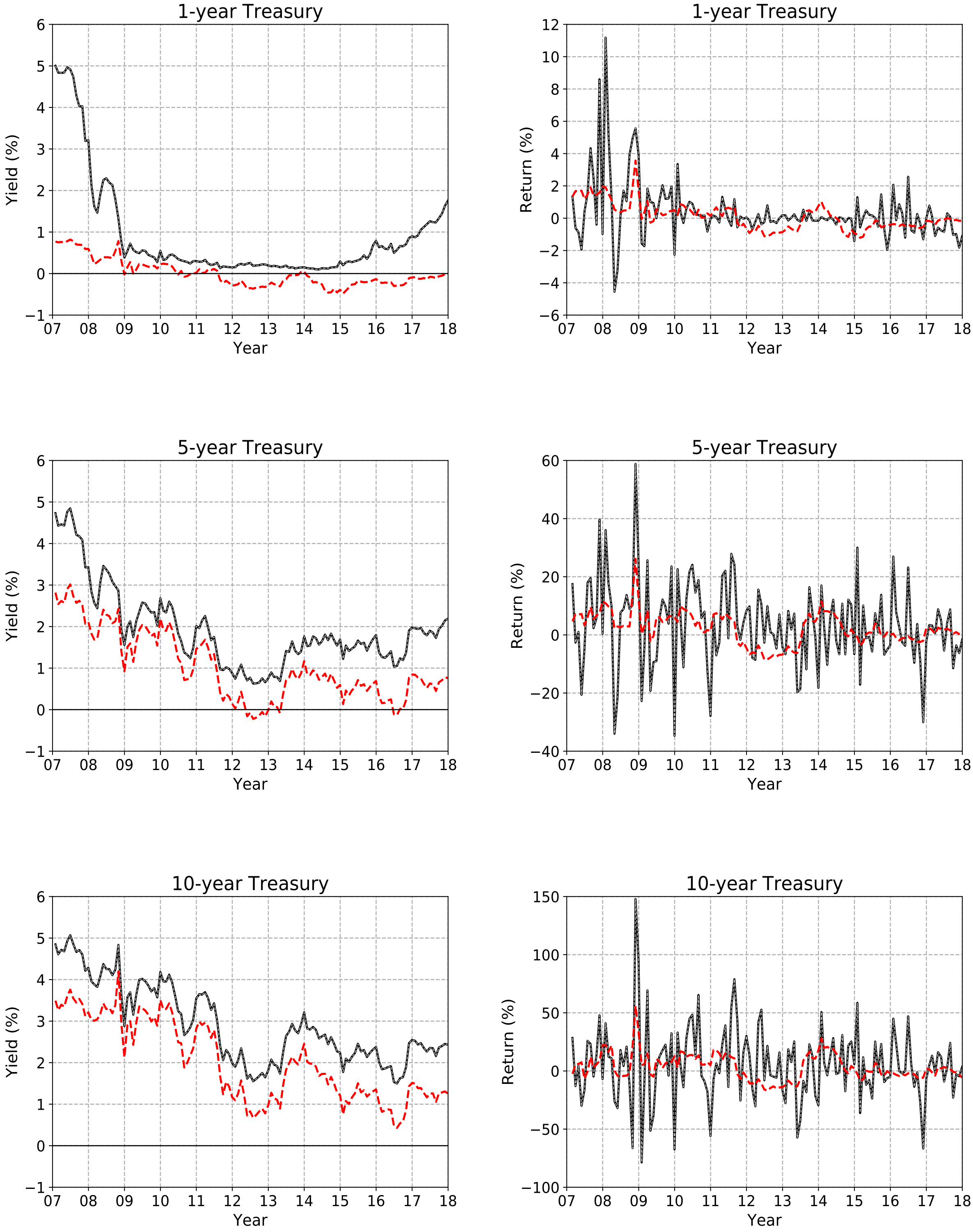


Figure 4.13 – US-only equity tail risk ATSM: Observed and model-implied time series of yields and one-month excess returns on US Treasury bonds with 1-, 5- and 10-year maturities. In the left panels, the solid black lines show the observed yields, the dashed grey lines plot the model-implied yields, while the dashed red lines indicate the model-implied term premia. In the right panels, the solid black lines show the observed excess returns, the dashed grey lines plot the model-implied excess returns, while the dashed red lines indicate the model-implied expected excess returns.

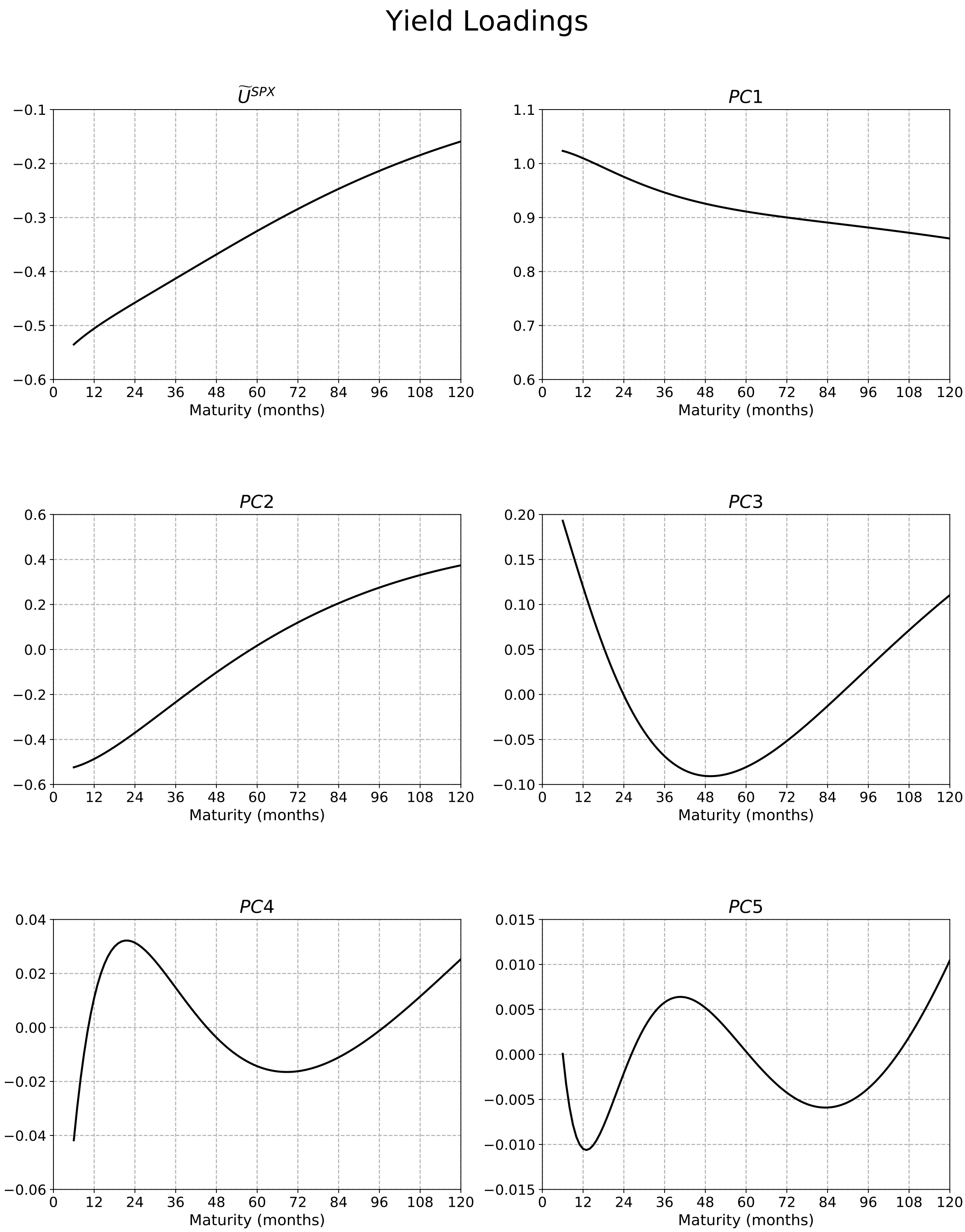


Figure 4.14 – US-only equity tail risk ATSM: Model-implied yield loadings on the pricing factors of the proposed US-only equity tail risk ATSM. These coefficients are calculated as $-(1/n)\mathbf{b}_n$ and can be interpreted as the response of the n-month yield to a contemporaneous shock to the respective factor. $\widetilde{U}^{SPX}$ represents the standardised equity left tail factor associated with the S&P 500. $PC1$ – $PC5$ denote the first five standardised principal components extracted from the US Treasury yields orthogonal with respect to the $\widetilde{U}^{SPX}$ factor.

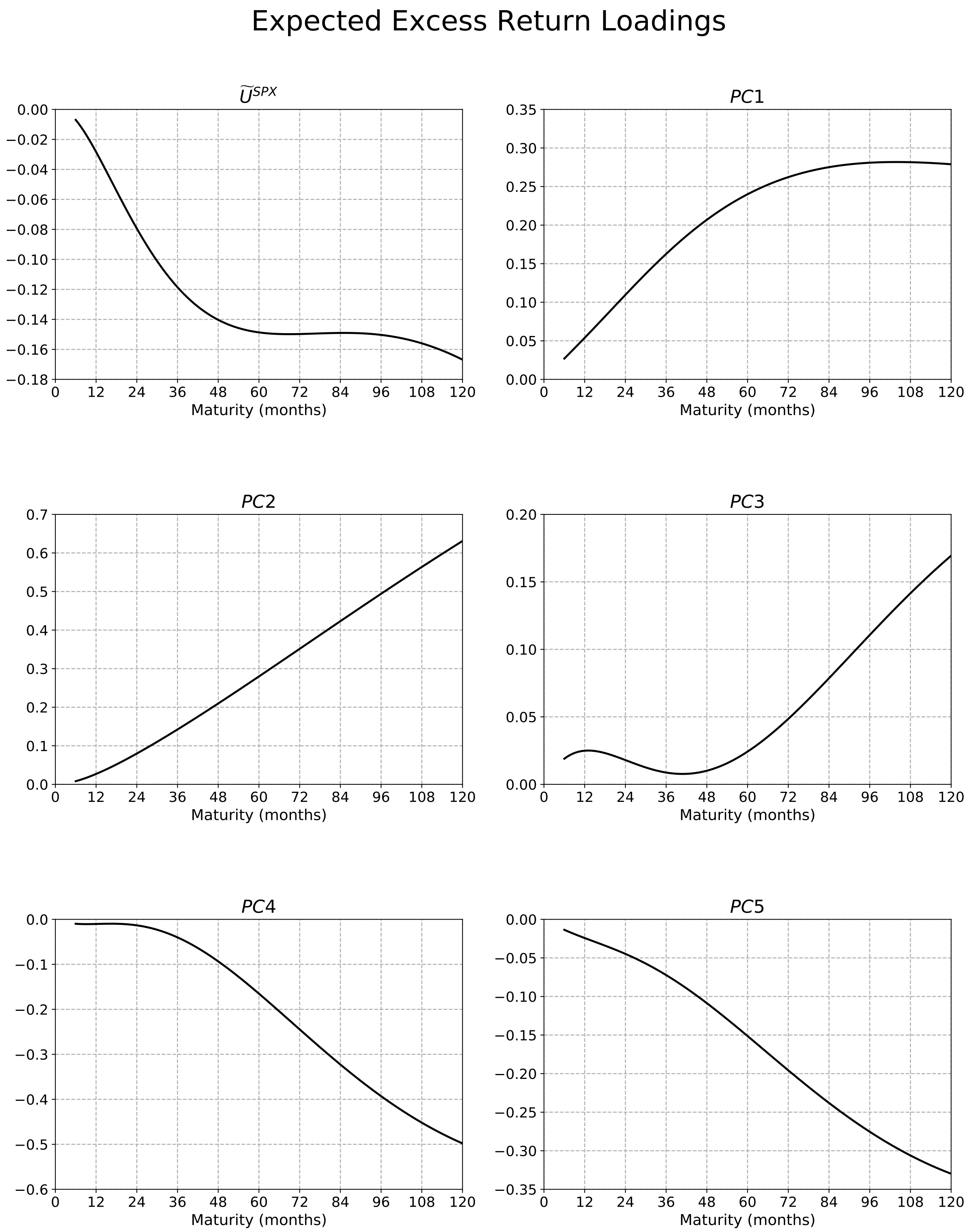


Figure 4.15 – US-only equity tail risk ATSM: Model-implied expected excess return loadings on the pricing factors of the proposed US-only equity tail risk ATSM. These coefficients are calculated as $\mathbf{b}_n^{'}\boldsymbol{\lambda}_1$ and can be interpreted as the response of the expected one-month excess return on the n-month bond to a contemporaneous shock to the respective factor. $\widetilde{U}^{SPX}$ represents the standardised equity left tail factor associated with the S&P 500. $PC1$ – $PC5$ denote the first five standardised principal components extracted from the US Treasury yields orthogonal with respect to the $\widetilde{U}^{SPX}$ factor.

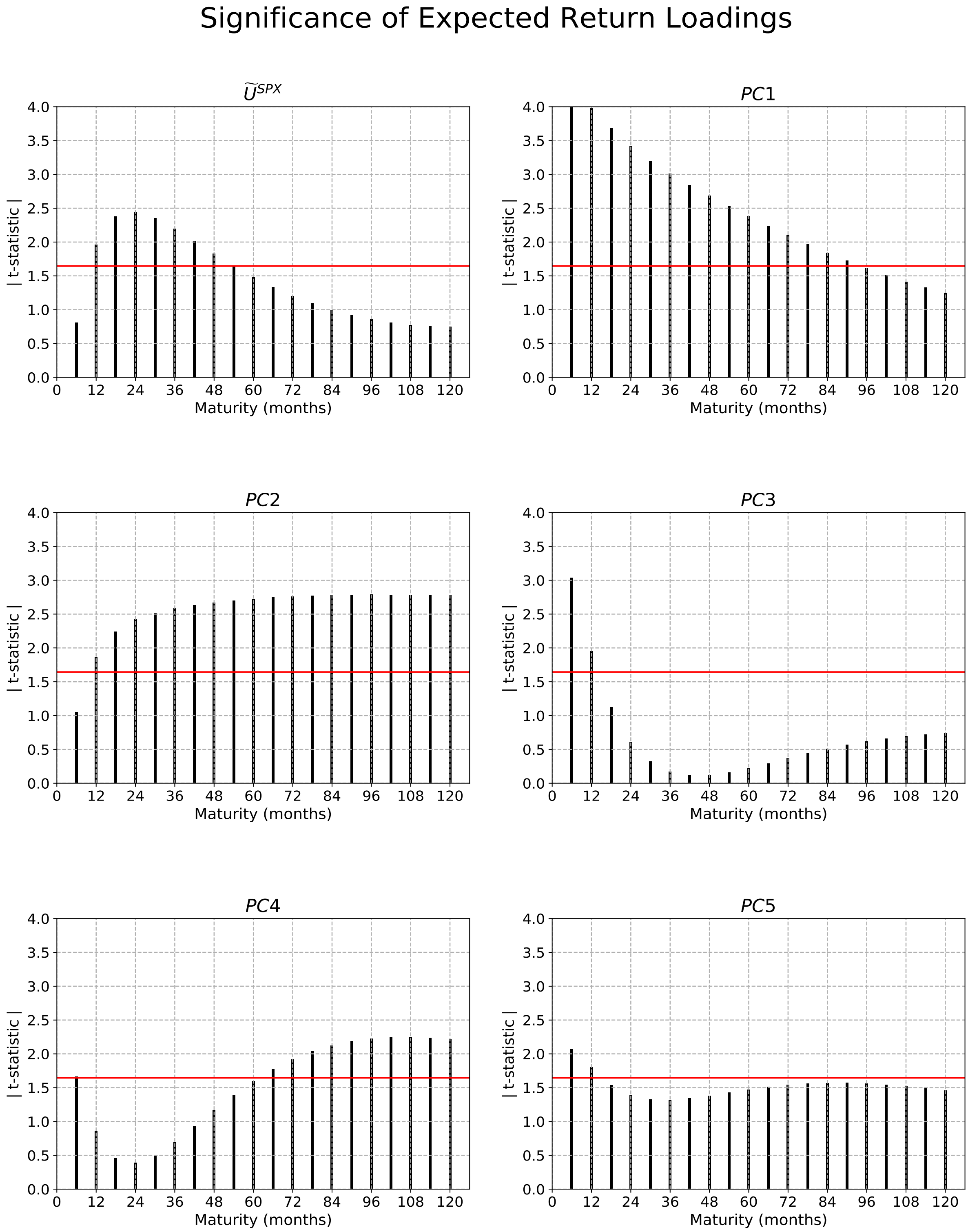


Figure 4.16 – US-only equity tail risk ATSM: Significance of expected return loadings on the pricing factors of the proposed US-only equity tail risk ATSM. The absolute value of the t-statistic is reported for the $N = 20$ one-month excess Treasury returns used to fit the cross-section of yields. The solid red lines depict the critical value of the statistics for the significance level of 10%. $\widetilde{U}^{SPX}$ represents the standardised equity left tail factor associated with the S&P 500. $PC1 - PC5$ denote the first five standardised principal components extracted from the US Treasury yields orthogonal with respect to $\widetilde{U}^{SPX}$.

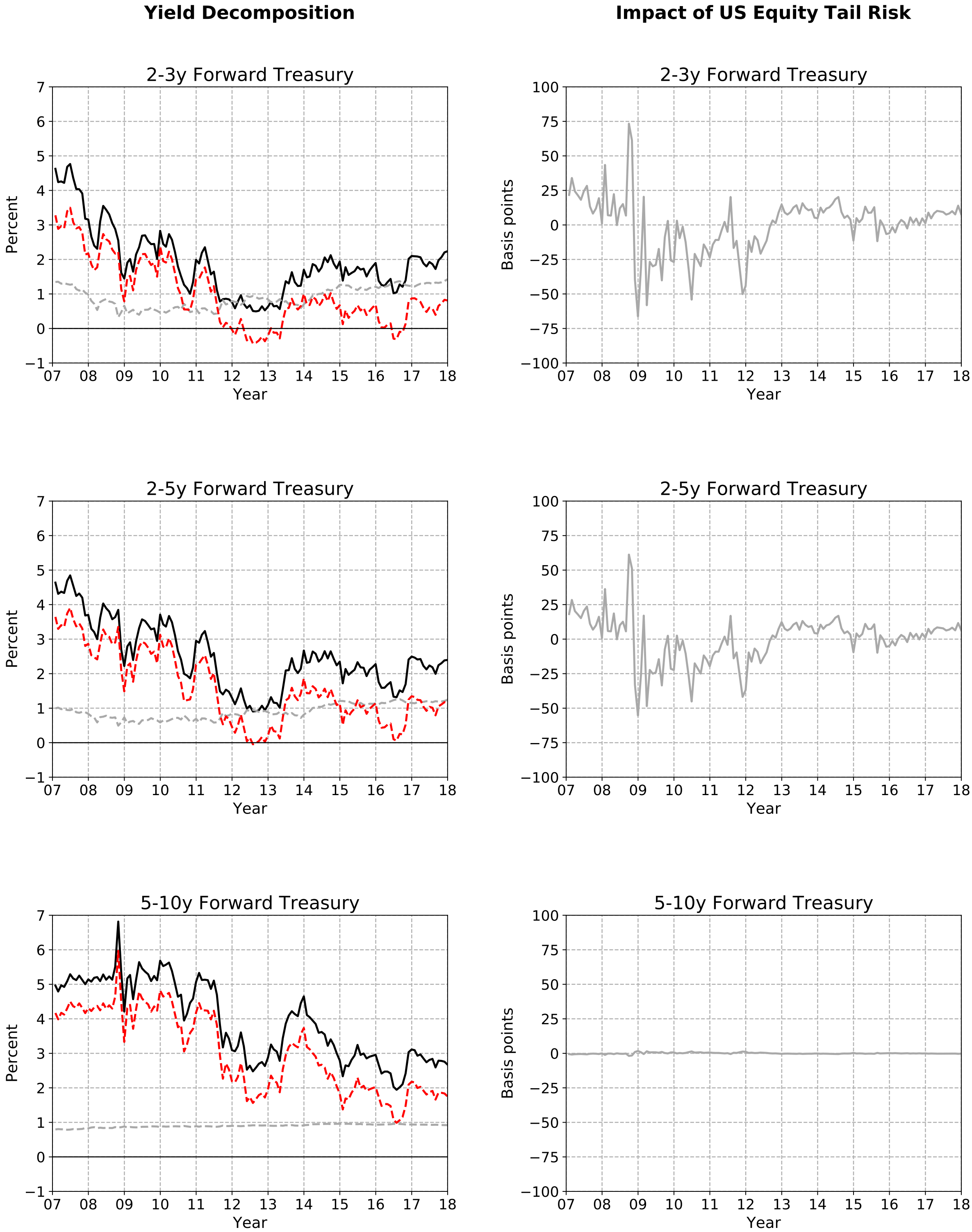


Figure 4.17 – US-only equity tail risk ATSM: Contribution of changes in US-only equity tail risk to the term premia embedded in the 2-3y, 2-5y and 5-10y forward Treasury rates. In the left panels, the solid black lines show the model-implied m-n forward rates, the dashed grey lines plot the risk-neutral m-n forward rates (the average expectation of the short rates over the next m to n periods), the dashed red lines indicate the model-implied term premia embedded in the m-n forward rates. In the right panels, the solid dark grey lines show the impact over time of the equity left tail factor $\widetilde{U}^{SPX}$ on the model-implied term premium of the m-n forward Treasury rates.

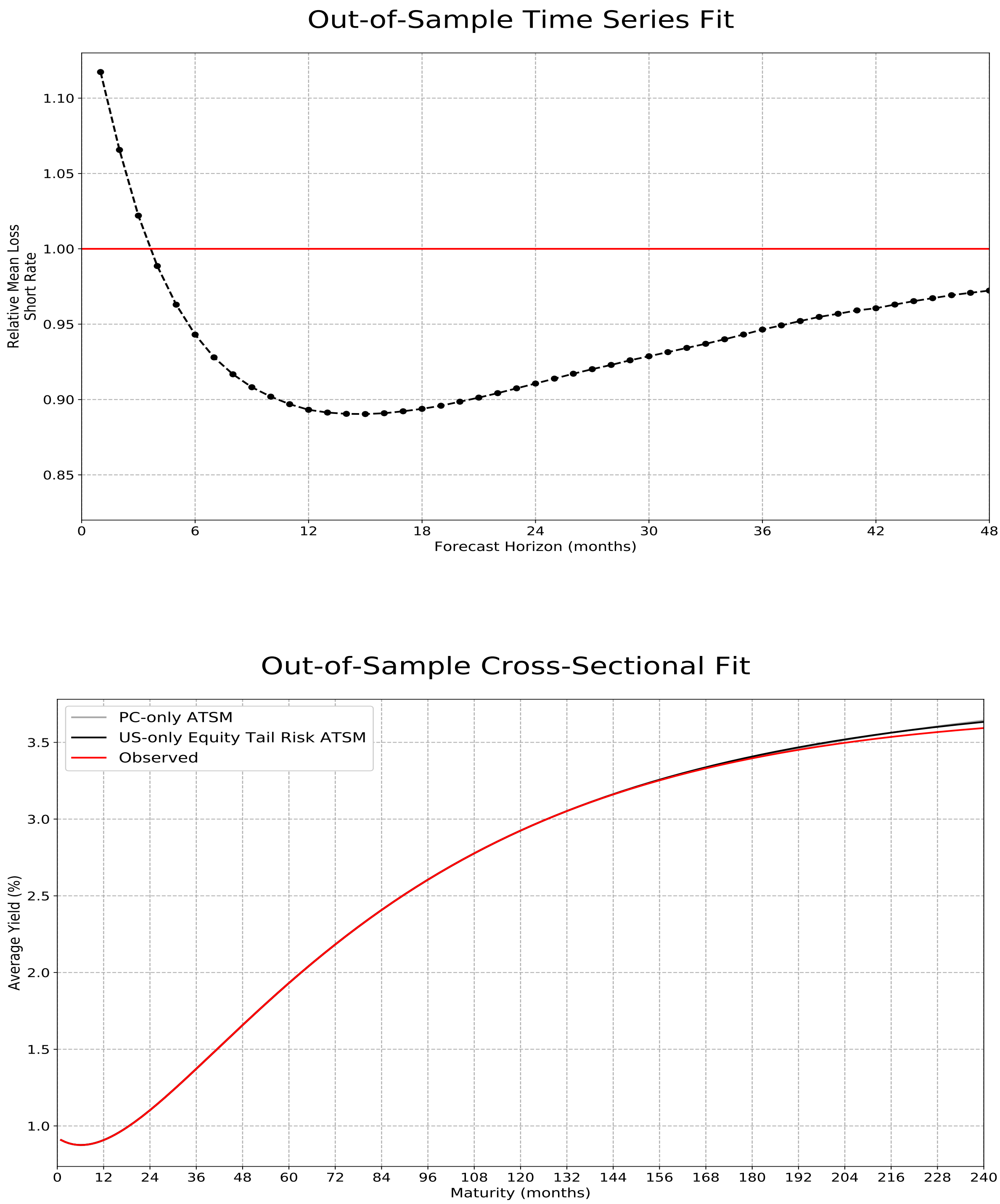


Figure 4.18 – US-only equity tail risk ATSM: Out-of-sample performance of the proposed US-only equity tail risk ATSM. Top panel: root mean squared forecast error for the US-only Equity Tail Risk ATSM relative to the benchmark PC-only specification (red line), plotted against the forecast horizon. Results are based on the out-of-sample forecasts of the average short rate over horizons from one month to four years. Bottom panel: observed average yields for maturities up twenty years (red line), along with the average yields implied by the US-only Equity Tail Risk ATSM (black line) and the benchmark PC-only specification (grey line) estimated using only maturities up to ten years.

Chapter 5

Conclusions and Future Research

5.1 Summary and Conclusions

The thesis starts by noting that the relation between stocks and government bonds significantly changed in 2000 and it is only since then that bonds have become an important hedge against stock market fluctuations. Chapter 2 then documents the relevance of a second safe haven - gold - in the empirical characterisation of flight-to-safety, especially in the period before the late 1990s. The measures that identify the occurrence of flight-to-safety days after observing abnormal movements in the US equity, Treasury bond and gold markets are potent indicators of higher next-day volatility in the stock market. Building on these results, Chapter 3 extends the research scope by examining how flight-to-safety can improve the modelling of both equity and safe haven return volatility in a multi-period setting. A proxy for flight-to-safety based on the realised semi-covariance between falling equity and rising safe haven returns is found to significantly improve the forecasts of both equity and safe haven volatility up to one month ahead. The incremental value of employing flight-to-safety for volatility forecasting is not only statistically but also economically significant. Although both Treasury bonds and gold are granted the safe haven status against stock market losses, it is the former that, when used to integrate flight-to-safety in volatility forecasting, allow to substantially enhance the asset allocation performance with a volatility-based rebalancing process. Lastly, Chapter 4 turns the attention to the role of equity jump tail risk in pricing government bonds. The investors' fear of abrupt negative return shocks to the international stock market is shown to command a risk premium in the US government bond market. Consistent with the theory of flight-to-safety, Treasury bond prices increase and future expected returns shrink when investors' fear is higher. The evidence of flight-to-safety is stronger at the short end of the US yield curve where equity tail risk has significant explanatory power for Treasury risk premia and is responsible for large term premium drops since the recent financial crisis.

In summary, this thesis adds to the literature on flight-to-safety by showing that the switching from equity to safe haven assets in times of stress has a significant role in volatility forecasting. Specifically, accounting for flight-to-safety allows for improved predictions of return volatility and better performance of volatility-timing strategies. Furthermore, this thesis provides guidance on how the safety attribute of Treasuries varies with the maturity of the bonds. Lastly, it sheds light on the existence of a common predictor between equity and bond markets, whose existence can be justified by the safe haven potential of US Treasuries.

The results of this thesis are also attractive from a practical standpoint as they have important implications for asset managers, practitioners and policy-making institutions. In particular, our findings offer practical insights into how flight-to-safety phenomena can be used to support portfolio allocation decisions and determine the response of the yield curve to changes in equity tail risk. For instance, asset managers and practitioners that form their portfolios with a volatility-based rebalancing process can achieve better risk-adjusted returns when the estimated conditional volatilities account for the flight-to-safety effect. The dominance of such volatility-timing strategies persists even in the presence of high transaction costs. Furthermore, comparison of our findings in the Treasury market with prior empirical studies in equity markets provides clear indication of a common predictor across the two asset classes. Expected equity returns rise and expected government bond returns fall in periods of high equity jump tail risk. This result is of obvious importance since it raises the possibility of superior portfolio performance for investors that use the forecasting information in the two markets to buy stocks and sell Treasuries when the perceived risk of negative return shocks to the stock market is higher. Finally, the empirical findings of this thesis can help policy makers gain a deeper understanding of the powerful forces that shape the term structure of interest rates. The flight-to-safety flows that occur when the stock market is hit by heavy losses are a dominant factor for the evolution of the short end of the yield curve. Indeed, while the unconventional monetary policies introduced by central banks to mitigate the severity of the financial crisis have been a major force in lowering longer-term yields, the reduction in shorter-term yields can be associated with the investors' increased appetite for safe assets.

5.2 Limitations and Recommendations for Future Research

Whereas this thesis presents some interesting results in the area of financial modelling by taking into account the impact of flight-to-safety, there are several limitations that should be considered when designing future research.

We begin by discussing some of the limiting factors of our work on the incremental value of flight-to-safety for volatility forecasting. First, Chapter 2 and Chapter 3 rely on prior empirical findings to establish a link between flight-to-safety and the future dynamics of financial return volatility. Fully understanding the theory behind it is a useful direction for future research but is beyond the scope of this thesis. It would be interesting, for example, to investigate the various channels of the transmission mechanism. The switching from equity to safe haven assets was suggested to be the actual cause for higher future volatility in the introduction of this thesis. Alternatively, the act of switching could be a sign that rational investors are anticipating high volatility in the future. Should this be the case, then option data could be a useful resource for learning more about the market dynamics and the flight-to-safety effect. Second, the finding that US government bonds, and not gold, yield the most significant improvements in forecasting and volatility-timing performance is a potentially important result and therefore, it requires additional consideration. One can imagine a number of reasons (e.g., liquidity) why Treasury bonds are more representative of a safe haven asset. A detailed investigation of this subject would be an interesting avenue to follow. Also, it would be useful to know whether our finding is true in general, or whether it is an accidental property or distortion of the Treasury market that occurs as a consequence of many years of aggressive central bank easing. Third, this thesis evaluates the economic benefits of using flight-to-safety in volatility forecasting only from the perspective of market investors who make portfolio selection decisions. However, future studies could explore the same research question by considering other relevant applications. An example would be to show the importance of our findings from the point of view of risk managers who decide the level of capital to hold based on the predicted future trends in the markets.

Now turning to the findings in Chapter 4, these suggest that negative jump tail risk predicts future returns in both equity and government bond markets. Based on this premise, a variety of future research directions can be pursued from both a theoretical and an empirical perspective.

Empirically, a first step for future research would be to apply the approach outlined in Chapter 4 to the yield curve of countries other than the United States. This would provide international evidence of flight-to-safety and help shed light on what countries supply government bonds that are an effective safe haven against stock market losses. Another interesting avenue for empirical research would be to expand on the approach of Chapter 4 and allow large panels of equity and bond returns to be examined at once. The major limitation of our study is the short time span for the availability of option data used to construct the equity left tail factor. In a recent work, Almeida et al. (2017) obtain a risk-neutralised measure of tail risk from a broad cross-section of equity returns available for a period of nearly 100 years. By exploiting the availability of long historical data for equity and bond portfolios, it would be interesting to capture a common tail risk to these assets and evaluate its relevance for the asset risk premia.

From a theoretical perspective, we feel that the most promising direction for future research in this area lies in the extension of the *approximate factor model* (see the seminal studies of Bai and Ng (2002), Bai (2003) and Stock and Watson (2002a,b)) in order to incorporate the effect of extreme market events on more than one group of assets. More specifically, it would be interesting to develop a large dimensional factor model that explains comovement between panels of equity and government bond returns in different market conditions. This can be achieved, for instance, by combining together the recent extensions of the classical approximate linear factor model that allow for either (i) time-varying loadings with threshold-type regime switches, as in Massacci (2017); or (ii) specific factors affecting only a finite number of (large) groups as in Andreou et al. (2019). This would not only represent a valuable addition to the literature on large scale latent factor models but would also provide a new approach to quantify the change in correlation between stocks and government bonds across different regimes.

Bibliography

ABRAHAMS, M., T. ADRIAN, R. K. CRUMP, E. MOENCH, AND R. YU (2016): "Decomposing real and nominal yield curves," *Journal of Monetary Economics*, 84, 182–200.

ADRIAN, T., R. K. CRUMP, AND E. MOENCH (2013): "Pricing the term structure with linear regressions," *Journal of Financial Economics*, 110, 110–138.

ADRIAN, T., R. K. CRUMP, AND E. VOGT (2018): "Nonlinearity and Flight to Safety in the Risk-Return Trade-Off for Stocks and Bonds," *Journal of Finance*, forthcoming.

ADRIAN, T., D. STACKMAN, AND E. VOGT (2019): "Global Price of Risk and Stabilization Policies," *London, Centre for Economic Policy Research.*

AÏT-SAHALIA, Y., J. FAN, AND D. XIU (2010): "High-Frequency Covariance Estimates With Noisy and Asynchronous Financial Data," *Journal of the American Statistical Association*, 105, 1504–1517.

AÏT-SAHALIA, Y. AND A. LO (2000): "Nonparametric risk management and implied risk aversion," *Journal of Econometrics*, 94, 9–51.

AÏT-SAHALIA, Y., P. A. MYKLAND, AND L. ZHANG (2005): "How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise," *The Review of Financial Studies*, 18, 351–416.

——— (2011): "Ultra high frequency volatility estimation with dependent microstructure noise," *Journal of Econometrics*, 160, 160–175.

ALMEIDA, C., K. ARDISON, R. GARCIA, AND J. VICENTE (2017): "Nonparametric Tail Risk, Stock Returns, and the Macroeconomy," *Journal of Financial Econometrics*, 15, 333–376.

Anand, A., P. Irvine, A. Puckett, and K. Venkataraman (2013): "Institutional trading and stock resiliency: Evidence from the 2007-2009 financial crisis," *Journal of Financial Economics*, 108, 773–797.

Andersen, T. G., T. Bollerslev, and F. X. Diebold (2007): "Roughing It Up: Including Jump Components in the Measurement, Modeling and Forecasting of Return Volatility," *Review of Economics and Statistics*, 89, 701–720.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens (2001): "The distribution of realized stock return volatility," *Journal of Financial Economics*, 61, 43–76.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003): "Modeling and forecasting realized volatility," *Econometrica*, 71, 579–625.

Andersen, T. G., D. Dobrev, and E. Schaumburg (2012): "Jump-robust volatility estimation using nearest neighbor truncation," *Journal of Econometrics*, 169, 75–93.

Andersen, T. G., N. Fusari, and V. Todorov (2015a): "Parametric Inference and Dynamic State Recovery from Option Panels," *Econometrica*, 83, 1081–1145.

——— (2015b): "The risk premia embedded in index options," *Journal of Financial Economics*, 117, 558–584.

——— (2017a): "Short-Term Market Risks Implied by Weekly Options," *Journal of Finance*, 72, 1335–1386.

——— (2017b): "The Pricing of Tail Risk and the Equity Premium: Evidence from International Option Markets," *Journal of Business & Economic Statistics*, forthcoming.

Andreou, E., P. Gagliardini, E. Ghysels, and M. Rubin (2019): "Inference in Group Factor Models with an Application to Mixed Frequency Data," *Econometrica*, forthcoming.

Baele, L., G. Bekaert, and K. Inghelbrecht (2010): "The determinants of stock and bond return comovements," *Review of Financial Studies*, 23, 2374–2428.

Baele, L., G. Bekaert, K. Inghelbrecht, and M. Wei (2015): "Flights to Safety," *National Bank of Belgium Working Paper*, 230.

Bai, J. (2003): "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71, 135–171.

Bai, J. and S. Ng (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221.

Bansal, N., R. Connolly, and C. Stivers (2010): "Regime-switching in stock index and Treasury futures returns and measures of stock market stress," *Journal of Futures Markets*, 30, 753–779.

Barber, B., R. Lehavy, M. McNichols, and B. Trueman (2001): "Can Investors Profit from the Prophets? Security Analyst Recommendations and Stock Returns," *Journal of Finance*, 56, 531–563.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008): "Designing Realized Kernels to Measure the ex-post Variation of Equity Prices in the Presence of Noise," *Econometrica*, 76, 1481–1536.

——— (2009): "Realized kernels in practice: trades and quotes," *Econometrics Journal*, 12, 1–33.

——— (2011): "Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading," *Journal of Econometrics*, 162, 149–169.

Barndorff-Nielsen, O. E. and N. Shephard (2004): "Power and Bipower Variation with Stochastic Volatility and Jumps," *Journal of Financial Econometrics*, 2, 1–37.

Baur, D. G. and B. M. Lucey (2009): "Flights and contagion – An empirical analysis of stock-bond correlations," *Journal of Financial Stability*, 5, 339–352.

——— (2010): "Is Gold a Hedge or a Safe Haven? An Analysis of Stocks, Bonds and Gold," *The Financial Review*, 45, 217–229.

Baur, D. G. and T. K. McDermott (2010): "Is gold a safe haven? International evidence," *Journal of Banking and Finance*, 34, 1886–1898.

——— (2012): "Safe Haven Assets and Investor Behaviour under Uncertainty," *The Institute for International Integration Studies Discussion Paper*, 392.

Beber, A., M. W. Brandt, and J. Chen (2014): "Switching Risk Off: FX Correlations and Risk Premia," *Working Paper*.

Beber, A., M. W. Brandt, and K. A. Kavajecz (2009): "Flight-to-Quality or Flight-to-Liquidity? Evidence from the Euro-Area Bond Market," *Review of Financial Studies*, 22, 925–957.

Bekaert, G. and E. Engstrom (2017): "Asset Return Dynamics under Habits and Bad Environment - Good Environment Fundamentals," *Journal of Political Economy*, 125, 713–760.

Bekaert, G., E. Engstrom, and S. R. Grenadier (2010): "Stock and bond returns with Moody Investors," *Journal of Empirical Finance*, 17, 867–894.

Bekaert, G., E. Engstrom, and Y. Xing (2009): "Risk, uncertainty, and asset prices," *Journal of Financial Economics*, 59–82.

Bekaert, G. and M. Hoerova (2014): "The VIX, the variance premium and stock market volatility," *Journal of Econometrics*, 183, 181–192.

Bekaert, G., M. Hoerova, and M. Lo Duca (2013): "Risk, uncertainty and monetary policy," *Journal of Monetary Economics*, 60, 771–788.

Bollerslev, T. (1986): "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

Bollerslev, T., J. Marrone, L. Xu, and H. Zhou (2014): "Stock Return Predictability and Variance Risk Premia: Statistical Inference and International Evidence," *Journal of Financial and Quantitative Analysis*, 49, 633–661.

Bollerslev, T., A. J. Patton, and R. Quaedvlieg (2016): "Exploiting the errors: A simple approach for improved volatility forecasting," *Journal of Econometrics*, 192, 1–18.

——— (2017): "Realized Semicovariances: Looking for Signs of Direction Inside the Covariance Matrix," *Working Paper*.

Bollerslev, T., G. Tauchen, and H. Zhou (2009): "Expected stock returns and variance risk premia," *Review of Financial Studies*, 22, 4463–4492.

Bollerslev, T. and V. Todorov (2011): "Tails, Fears, and Risk Premia," *Journal of Finance*, 66, 2165–2211.

Bollerslev, T., V. Todorov, and L. Xu (2015): "Tail risk premia and return predictability," *Journal of Financial Economics*, 118, 113–134.

Boudry, W., R. Connolly, and E. Steiner (2018): "What Really Happens During Flight to Safety: Evidence from Real Estate Markets," *Working Paper*.

Brooks, C. (2008): *Introductory Econometrics for Finance*, Cambridge University Press, Cambridge.

Brunnermeier, M. K. and L. H. Pedersen (2009): "Market Liquidity and Funding Liquidity," *Review of Financial Studies*, 22, 2201–2238.

Buraschi, A. and P. Whelan (2016): "Bond Markets and Monetary Policy," in *Handbook of Fixed–Income Securities*, ed. by P. Veronesi, Wiley, 77–92.

Busetti, F. and M. Caivano (2018): "Low frequency drivers of the real interest rate: Empirical evidence for advanced economies," *International Finance - Wiley Online Library*.

Caballero, R. and A. Krishnamurthy (2008): "Collective risk management in a flight to quality episode," *Journal of Finance*, 63, 2195–2230.

Caballero, R. J., E. Farhi, and P. O. Gourinchas (2017): "The Safe Assets Shortage Conundrum," *Journal of Economic Perspectives*, 31, 29–46.

Capie, F., T. C. Mills, and G. Wood (2005): "Gold as a hedge against the dollar," *Journal of International Financial Markets, Institutions and Money*, 15, 343–352.

Chiţu, L., B. Eichengreen, and A. Mehl (2014): "When did the dollar overtake sterling as the leading international currency? Evidence from the bond markets," *Journal of Development Economics*, 111, 225–245.

Chordia, T., A. Sarkar, and A. Subrahmanyam (2005): "An empirical analysis of stock and bond market liquidity," *Review of Financial Studies*, 18, 85–129.

Christoffersen, P., J. Hahn, and A. Inoue (2001): "Testing and comparing Value-at-Risk measures," *Journal of Empirical Finance*, 8, 325–342.

Ciner, C., C. Gurdgiev, and B. M. Lucey (2013): "Hedges and safe havens: An examination of stocks, bonds, gold, oil and exchange rates," *International Review of Financial Analysis*, 29, 202–211.

Cochrane, J. H. and M. Piazzesi (2005): "Bond Risk Premia," *The American Economic Review*, 95, 138–160.

——— (2008): "Decomposing the Yield Curve," *Working Paper*.

Connolly, R., C. Stivers, and L. Sun (2005): "Stock Market Uncertainty and the Stock-Bond Return Relation," *Journal of Financial and Quantitative Analysis*, 40, 161–194.

Cooper, I. and R. Priestley (2008): "Time-Varying Risk Premiums and the Output Gap," *The Review of Financial Studies*, 22, 2801–2833.

Corsi, F. (2009): "A simple approximate long-memory model of realized volatility," *Journal of Financial Econometrics*, 7, 174–196.

Corsi, F. and R. Renó (2012): "Discrete-Time Volatility Forecasting With Persistent Leverage Effect and the Link With Continuous-Time Volatility Modeling," *Journal of Business & Economic Statistics*, 30, 368–380.

David, A. and P. Veronesi (2013): "What Ties Return Volatilities to Price Valuations and Fundamentals?" *Journal of Political Economy*, 121, 682–746.

——— (2016): "The Economics of the Comovement of Stocks and Bonds," in *Handbook of Fixed–Income Securities*, ed. by P. Veronesi, Wiley, 313–325.

de Pooter, M., M. Martens, and D. van Dijk (2008): "Predicting the Daily Covariance Matrix for S&P 100 Stocks Using Intraday Data – But Which Frequency to Use?" *Econometric Reviews*, 27, 199–229.

Duffee, G. R. (2002): "Term premia and interest rate forecasts in affine models," *Journal of Finance*, 57, 405–443.

——— (2011): "Information in (and not in) the Term Structure," *The Review of Financial Studies*, 24, 2895–2934.

Economist (2015): "Dominant and dangerous," Available at `https://www.economist.com/leaders/2015/10/03/dominant-and-dangerous`.

Eichengreen, B., A. K. Rose, and C. Wyplosz (1996): "Contagious Currency Crises," *Scandinavian Journal of Economics*, 98, 463–484.

Engle, R. (2004): "Risk and Volatility: Econometric Models and Financial Practice," *American Economic Review*, 94, 405–420.

Engle, R. and G. M. Gallo (2006): "A multiple indicators model for volatility using intra-daily data," *Journal of Econometrics*, 131, 3–27.

Engle, R. and A. Patton (2001): "What good is a volatility model?" *Quantitative Finance*, 1, 237–245.

Epps, T. W. (1979): "Comovements in stock prices in the very short run," *Journal of the American Statistical Association*, 74, 291–298.

Farago, A. and R. Tédongap (2018): "Downside risks and the cross-section of asset returns," *Journal of Financial Economics*, 129, 69–86.

Fleming, J., C. Kirby, and B. Ostdiek (2001): "The economic value of volatility timing," *Journal of Finance*, 56, 329–352.

——— (2003): "The economic value of volatility timing using "realized" volatility," *Journal of Financial Economics*, 67, 473–509.

Forbes, K. and R. Rigobon (2002): "No Contagion, Only Interdependence: Measuring Stock Market Comovements," *Journal of Finance*, 57, 2223–2261.

Ghysels, E., A. Plazzi, and R. Valkanov (2016): "The Risk-Return Relationship and Financial Crises," *Working Paper.*

Ghysels, E., P. Santa-Clara, and R. Valkanov (2006): "Predicting volatility: getting the most out of return data sampled at different frequencies," *Journal of Econometrics*, 131, 59–95.

Gordon, D. B. and E. M. Leeper (1994): "The Dynamic Impacts of Monetary Policy: An Exercise in Tentative Identification," *Journal of Political Economy*, 102, 1228–1247.

Goyenko, R. and S. Sarkissian (2014): "Treasury Bond Illiquidity and Global Equity Returns," *Journal of Financial and Quantitative Analysis*, 49, 1227–1253.

Granger, C. W. J. (1969): "Investigating Causal Relations by Econometric Models and Cross-spectral Methods," *Econometrica*, 37, 424–438.

Greenwood, R. and D. Thesmar (2011): "Stock price fragility," *Journal of Financial Economics*, 102, 471–490.

Gürkaynak, R. S., B. Sack, and J. H. Wright (2007): "The U.S. Treasury yield curve: 1961 to the present," *Journal of Monetary Economics*, 54, 2291–2304.

Hamao, Y., R. W. Masulis, and V. Ng (1990): "Correlations in Price Changes and Volatility across International Stock Markets," *Review of Financial Studies*, 3, 281–307.

Hameed, A., W. Kang, and S. Viswanathan (2010): "Stock Market Declines and Liquidity," *Journal of Finance*, 65, 257–293.

Hamilton, J. D. (1994): *Time Series Analysis*, Princeton University Press, New York.

Hamilton, J. D. and G. Lin (1996): "Stock market volatility and the business cycle," *Journal of Applied Econometrics*, 11, 573–593.

Hansen, P. R. and Z. Huang (2016): "Exponential GARCH Modeling With Realized Measures of Volatility," *Journal of Business & Economic Statistics*, 34, 269–287.

Hansen, P. R., Z. A. Huang, and H. H. Shek (2012): "Realized GARCH: A Joint Model for Returns and Realized Measures of Volatility," *Journal of Applied Econometrics*, 27, 877–906.

Hansen, P. R., A. Lunde, and V. Voev (2014): "Realized Beta GARCH: A Multivariate GARCH Model with Realized Measures of Volatility," *Journal of Applied Econometrics*, 29, 774–799.

Hartmann, P., S. Straetmans, and C. G. de Vries (2004): "Asset Market Linkages in Crisis Periods," *Review of Economics and Statistics*, 86, 313–326.

Hayashi, T. and N. Yoshida (2005): "On covariance estimation of non-synchronously observed diffusion processes," *Bernoulli*, 11, 359–379.

Hollo, D., M. Kremer, and M. Lo Duca (2012): "CISS - A Composite Indicator of Systemic Stress in the Financial System," *ECB Working Paper*, 1426.

Ilmanen, A. (2003): "Stock–bond correlations," *The Journal of Fixed Income*, 13, 55–75.

Jaffe, J. (1989): "Gold and gold stocks as investments for institutional portfolios," *Financial Analysts Journal*, 45, 53–59.

Joslin, S., M. Priebsch, and K. J. Singleton (2014): "Risk premiums in dynamic term structure models with unspanned macro risks," *Journal of Finance*, 69, 1197–1233.

Kaminska, I. and M. Roberts-Sklar (2015): "A global factor in variance risk premia and local bond pricing," *Bank of England Staff Working Paper.*

Kaminska, I. and G. Zinna (2018): "Official Demand for U.S. Debt: Implications for U.S. Real Interest Rates," *IMF Working Paper No. 14/66.*

Kaminsky, G. L. and C. M. Reinhart (2000): "On crises, contagion, and confusion," *Journal of International Economics*, 51, 145–168.

Kelly, B. and H. Jiang (2014): "Tail Risk and Asset Prices," *Review of Financial Studies*, 27, 2841–2871.

Kim, C. J. (1994): "Dynamic linear models with Markov-switching," *Journal of Econometrics*, 60, 1–22.

Kim, D. H. and J. H. Wright (2005): "An Arbitrage-Free Three-Factor Term Structure Model and the Recent Behavior of Long-Term Yields and Distant-Horizon Forward Rates," *Federal Reserve Board Finance and Economics Discussion Series.*

King, M. A. and S. Wadhwani (1990): "Transmission of Volatility between Stock Markets," *Review of Financial Studies*, 3, 5–33.

Knight, F. H. (1921): *Risk, uncertainty and profit*, Harper & Row, New York.

Krishnamurthy, A. and A. Vissing-Jorgensen (2012): "The Aggregate Demand for Treasury Debt," *Journal of Political Economy*, 120, 233–267.

Lawrence, C. (2003): "Why is Gold different from other Assets? An Empirical Investigation," *World Gold Council, London.*

Li, J. and G. Zinna (2018): "The Variance Risk Premium: Components, Term Structures, and Stock Return Predictability," *Journal of Business & Economic Statistics*, 36, 411–425.

Li, S. and B. M. Lucey (2017): "Reassessing the role of precious metals as safe havens – What colour is your haven and why?" *Journal of Commodity Markets*, 7, 1–14.

Lin, W. L., R. Engle, and T. Ito (1994): "Do bulls and bears move across borders? International transmission of stock returns and volatility," *Review of Financial Studies*, 7, 507–538.

Longstaff, F. (2004): "The Flight-to-Liquidity Premium in U.S. Treasury Bond Prices," *Journal of Business*, 77, 511–526.

Ludvigson, S. C. and S. Ng (2009): "Macro Factors in Bond Risk Premia," *The Review of Financial Studies*, 22, 5027–5067.

Malik, S. and A. Meldrum (2016): "Evaluating the robustness of UK term structure decompositions using linear regression methods," *Journal of Banking and Finance*, 67, 85–102.

Massacci, D. (2017): "Least squares estimation of large dimensional threshold factor models," *Journal of Econometrics*, 197, 101–129.

McCown, J. R. and J. R. Zimmerman (2006): "Is Gold a Zero-Beta Asset? Analysis of the Investment Potential of Precious Metals," *Working Paper*, available at SSRN: https://ssrn.com/abstract=920496.

Mills, T. C. (2004): "Statistical analysis of daily gold price data," *Physica A: Statistical Mechanics and its Applications*, 338, 559–566.

Mueller, P., P. Sabtchevsky, A. Vedolin, and P. Whelan (2016): "Variance risk premia on stocks and bonds," *Working Paper*.

Müller, U., M. Dacorogna, R. Davé, R. Olsen, O. Pictet, and J. E. von Weizsäcker (1997): "Volatilities of different time resolutions - Analysing the dynamics of market components," *Journal of Empirical Finance*, 4, 213–239.

Odean, T. (1998): "Are Investors Reluctant to Realize Their Losses?" *Journal of Finance*, 53, 1775–1798.

Patton, A. J. (2011): "Volatility forecast comparison using imperfect volatility proxies," *Journal of Econometrics*, 160, 246–256.

Politis, D. and J. Romano (1994): "The Stationary Bootstrap," *Journal of the American Statistical Association*, 89, 1303–1313.

Sachs, J., A. Tornell, and A. Velasco (1995): "Financial Crises in Emerging Markets: The Lessons from 1995," *Brookings Papers on Economic Activity*, 1, 147–198.

Schwert, G. (1989): "Why Does Stock Market Volatility Change Over Time?" *Journal of Finance*, 44, 1115–1153.

Shefrin, H. and M. Statman (1985): "The Disposition to Sell Winners Too Early and Ride Losers Too Long: Theory and Evidence," *Journal of Finance*, 40, 777–790.

Shephard, N. and K. Sheppard (2010): “Realising the future: forecasting with high frequency based volatility (HEAVY) models,” *Journal of Applied Econometrics*, 25, 197–231.

Shiller, R. J. and A. Beltratti (1992): “Stock prices and bond yields: Can their comovements be explained in terms of present value models?” *Journal of Monetary Economics*, 30, 25–46.

Stock, J. H. and M. W. Watson (2002a): “Forecasting Using Principal Components from a Large Number of Predictors,” *Journal of the American Statistical Association*, 97, 1167 – 1179.

——— (2002b): “Macroeconomic Forecasting Using Diffusion Indexes,” *Journal of Business and Economic Statistics*, 20, 147–162.

Susmel, R. and R. Engle (1994): “Hourly volatility spillovers between international equity markets,” *Journal of International Money and Finance*, 13, 3–25.

Taylor, N. (2017): “Realised variance forecasting under Box-Cox transformations,” *International Journal of Forecasting*, 33, 770–785.

Teugels, J. L. (1990): “Some representations of the multivariate Bernoulli and binomial distributions,” *Journal of Multivariate Analysis*, 32, 256–268.

Tully, E. and B. M. Lucey (2007): “A power GARCH examination of the gold market,” *Research in International Business and Finance*, 21, 316–325.

Upper, C. (2000): “How Safe Was the ”Safe Haven”? Financial Market Liquidity during the 1998 Turbulences,” *Deutsche Bundesbank Discussion Paper*, 1/00.

Vayanos, D. (2004): “Flight to quality, flight to liquidity, and the pricing of risk,” *NBER Working Paper 10327.*

White, H. (2000): “A Reality Check for Data Snooping,” *Econometrica*, 68, 1097–1126.

Zhang, L. (2011): “Estimating covariation: Epps effect, microstructure noise,” *Journal of Econometrics*, 160, 33–47.

Zhang, L., P. A. Mykland, and Y. Aït-Sahalia (2005): "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data," *Journal of the American Statistical Association*, 100, 1394–1411.