Tail Risk, Core Risk, and Expected Stock Returns

Johannes Breckenfelder* Bastien Buchwalter† Roméo Tédongap‡

May 17, 2020

Abstract

Expected returns should not only include rewards for accepting the risk of a potential downside loss, but also discounts for potential upside gains. Since investors care differently about upside gains versus downside losses, they require a risk premium for bearing the relative downside risk. We validate this perception empirically as well as theoretically and show that conditional asymmetry forecasts equity market returns in the short run. The results hold not only for different return series and different data frequencies (daily and intra-daily) but also for various subsamples. Our short-term expected return predictor, the asymmetric realized volatility measure, captures more variation in equity returns than the variance risk premium, a forward-looking measure, or the price-earnings ratio, and can easily be extracted from realized return series. We formalize this intuition with a closed-form asset pricing model that incorporates disappointment aversion and time-varying macroeconomic uncertainty.

Keywords: realized variance, realized semivariance, stock market return predictability, asymmetric realized volatility, variance risk premium, high-frequency data, equilibrium asset pricing

JEL Classification: G1, G12, G11, C1, C5

*European Central Bank

†ESSEC Business School, bastien.buchwalter@essec.edu

‡ESSEC Business School

1 Introduction

The predictability of the market return is of great interest to academics and practitioners alike. The tools used to study this age-old topic crucially depend on the time horizon for which the market return is to be predicted, i.e. short-term (one to three months), medium-term (three to twelve months) and long-term (one year or above). The literature already covers extensively the predictability for medium and long-term horizons.[1] The present paper contributes to this extensive literature by developing new measures that forecast the stock market return in the short-term.

Fundamentally speaking, our approach is based on Tversky and Kahneman (1979)'s prospect theory. That is, investors care differently about upside gains as compared to downside losses, and that those who face downside risk require a relative downside risk premium (Ang, Hodrick, Xing and Zhang 2006). Hence, when studying the predictability of the market return, it proves useful to distinguish between downside and upside uncertainty. The idea to compute variation above and below a certain threshold was pioneered by Roy (1952) and Markowitz (1952). This approach was later extended to the realm of intra-daily data by Andersen et al. (2001) and Barndorff-Nielsen et al. (2008). More precisely, Barndorff-Nielsen et al. (2008) define downside (upside) realized variance (RV) as the sum of all intra-daily log returns below (above) zero. This concept of semi variances is extensively used in the literature of market return predictability. For instance, Feunou et al. (2018) and Kilic and Shaliastovich (2019) use downside and upside variance risk premium (VRP), Segal et al. (2015) use positive (negative) variations of industrial production growth as a proxy for upside (downside) uncertainty, and Jurado et al. (2015) use a vast list of time series variables including, among others, real output and income as well as price indices and stock market indexes to compute measures of upside and downside uncertainty. Implicitly, all these papers have a common approach; distinguishing between upside and downside uncertainty because gains and losses are perceived differently by investors. Harvey and Siddique (2000) show that this asymmetry of risk is linked to the skewness of return distributions. As pointed out by Jondeau and Rockinger (2003) return distributions are

[1] Fama and French (1988) show that the price-dividend ratio forecasts the stock market return very well in the long-term. To predict the stock market return in the medium-term, Bollerslev et al. (2009) and Drechsler and Yaron (2011) use the variance risk premium which is defined as the difference between the realized volatility and the implied volatility.

generally negatively skewed which means that downside returns display occasional outliers, whereas upside returns or more concentrated. In line with this, the high frequency data used in this paper (five-minute log returns of the S&P 500 Index between January 1990 and September 2016) displays negative skewness. Therefore, the predictors developed in this paper will be in the line with the recent literature and distinguish between downside and upside RV.

However, the predictors we develop in the present paper do not only consider the skewness, but also (as opposed to the existing literature mentioned above) consider the kurtosis of the return distribution. That is, we notice that return distributions generally display high excess kurtosis. In other words, a significant amount of information is captured by the tails of the return distribution. In order to investigate whether this information differs from the one at core of the return distribution we distinguish between 'extreme' events, i.e. large returns, and 'normal' events, i.e. small returns. These two types of returns correspond, respectively, to observations in the tails and the core of the return distribution. By combining our two motivations derived from the third and fourth moment of the return distribution, we obtain a total of four uncertainty measures; downside tail, downside core, upside core and upside tail RV. It is important to understand, however, that we are not interested in the decomposition of variance according to a jump-diffusion threshold as suggested by Patton and Sheppard (2015). Indeed, the authors' approach only allows to forecast volatility, but not returns. In the present paper, we compute tail and core components for a multitude of thresholds and investigate the threshold for which short-term predictability of the market return is maximized. To the best of our knowledge, we are the first to investigate such a decomposition which allows for predictability of the excess market return on a monthly horizon.

Our analysis is conducted in a monthly vector autoregression (VAR) framework. The adjusted R^2 of predicting the market excess return is 9.27%. This result is driven in particular by tail asymmetry (which is defined as the difference between upside and downside tail) and the core asymmetry (which are defined as the difference between upside and downside core). The coefficient estimates are negative and statistically significant at the 5% and 1% level for the tail and core asymmetry, respectively. The negativity of the coefficient estimates implies that higher upside uncertainty (as measured by upside tail and core

RVs) yields a lower return. Conversely, higher downside uncertainty (as measured by the downside tail and core RVs) yields a higher return. Hence, we show that expected returns include rewards for accepting the risk of a potential downside loss, but also discounts for potential upside gains. As such, these results are in line with the inter-temporal capital asset pricing model (ICAPM) as introduced by Merton (1973) who advocates a liner relation between variance and expected excess stock market return. The systematic risk-return trade-off suggests that the conditional excess market return varies with its conditional variance. While previous literature is finding mostly insignificant, and even negative, systematic risk-reward trade-offs, the present paper confirms a positive risk-return trade-off for downside uncertainty, and also establishes a negative risk-return trade-off for upside uncertainty.

We strengthen our results by comparing the performance of our tail and core measures to the one of existing predictors, i.e. the VRP and the price-dividend (PD) ratio. While our predictors yield an adjusted R^2 of 9.27% on monthly basis, the VRP and the PD ratio only yield, respectively, an adjusted R^2 of 4.90% and 3.37% for the same time horizon. To investigate the predictability for longer time horizons we use two different approaches; the *overalapping* and the *VAR* methodology. The former consists of aggregating predictors backwards and predicted variables forwards. Hence, by averaging variables over several periods this approach increases the persistence of variables. However, as pointed out by Stambaugh (1999), and Valkanov (2003), the persistence of variables can artificially inflate statistical significance and R^2. Hence, in the present paper we advocate the use of an alternative approach to study predictability over long horizons, i.e. the *VAR* methodology. Intuitively, rather than aggregating predictors and forecasted variables over multiple time horizons, we exploit the stationarity feature of the VAR to estimate the predictive regression for longer time horizons. Independent of the methodology (*overalapping* or *VAR*), our predictors outperform both the VRP and the PD ratio for short time-horizons of up to 3 months.

The remainder of the paper is structured as follows. In section 2 we formally introduce our tail and core measures. We also discuss the implementation of the *overalapping* and the *VAR* methodologies. Section 3 presents the data. Section 4 discusses the main results of the paper. Section 5 concludes.

2 Tail/Core and Downside/Upside Analysis of Stock Variance

2.1 Risk Measures

In line with risk-return trade-off suggested by Merton (1973), a large body of the financial literature that analyzes the relationship between expected returns and risk has focused on the total variance of returns as the ultimate measure of equity risk. However, we argue that this can be misleading as behavioral biases induce equity investors to have heterogeneous perceptions and unequal treatment of the different segments of the asset return distribution. In this section, we introduce our methodological analysis of stock market variance by distinguishing between its upside and downside components, each of which can further be categorized into a tail and a core outlook. Therefore, we decompose the asset returns distribution into four different segments namely the downside tail, the downside core, the upside core and the upside tail, and we measure the variance associated with each of these segments.

We use high-frequency equity logarithmic (log) returns[2] to compute monthly measures of tail and core variances of the log returns distribution. In terms of notation, let $1/\Delta$ be the number of high-frequency returns per month, e.g., $\Delta \approx 1/176$ for a sixty-minute return series and $\Delta \approx 1/1,870$ for a five-minute return series, and let $r_{t-1+j\Delta}$ denote the jth intra-monthly return of the current month, t, where $j = 1, 2, \ldots, 1/\Delta$. Let $r_{t-1,t}$ be the current month's return. We have

$$r_{t-1,t} = \sum_{j=1}^{1/\Delta} r_{t-1+j\Delta}. \tag{1}$$

For $0 < \alpha \leq 1/2$, let $q_t(\alpha)$ be the α-quantile of the intra-monthly returns series $\{r_{t-1+j\Delta},\ j = 1, 2, \ldots, 1/\Delta\}$. Also, let $n_t^{T-}(\alpha)$ be the number of intra-monthly returns smaller than $q_t(\alpha)$, and let $n_t^{T+}(\alpha)$ be the number of intra-monthly returns higher than $q_t(1-\alpha)$. That is,

$$n_t^{T-}(\alpha) = \sum_{j=1}^{1/\Delta} I\left(r_{t-1+j\Delta} < q_t(\alpha)\right) \quad \text{and} \quad n_t^{T+}(\alpha) = \sum_{j=1}^{1/\Delta} I\left(r_{t-1+j\Delta} \geq q_t(1-\alpha)\right), \tag{2}$$

[2]In empirical studies, to ensure that our analysis is not driven by a few outliers, we winsorize the log returns at the 0.1% level, i.e. 0.05% from the top and 0.05% from the bottom. This is commonly done in related literature; for example, Ang, Chen and Xing (2006) winsorize in their cross-sectional study on downside risk, and Drechsler and Yaron (2011) winsorize their exogenous variables at the 1% level.

where $I(\cdot)$ is the indicator function. Likewise, let $\mu_t^{T-}(\alpha)$ be the average of intra-monthly returns smaller than $q_t(\alpha)$, and let $\mu_t^{T+}(\alpha)$ be the average of intra-monthly returns higher than $q_t(1-\alpha)$. That is,

$$\begin{aligned}
\mu_t^{T-}(\alpha) &= \frac{1}{n_t^{T-}(\alpha)} \sum_{j=1}^{1/\Delta} r_{t-1+j\Delta} I\left(r_{t-1+j\Delta} < q_t(\alpha)\right) \\
\mu_t^{T+}(\alpha) &= \frac{1}{n_t^{T+}(\alpha)} \sum_{j=1}^{1/\Delta} r_{t-1+j\Delta} I\left(r_{t-1+j\Delta} \geq q_t(1-\alpha)\right).
\end{aligned} \tag{3}$$

We define the monthly downside and upside tail RVs, respectively, as

$$\begin{aligned}
RV_{t-1,t}^{T-}(\alpha) &= \frac{1/\Delta}{n_t^{T-}(\alpha)} \sum_{j=1}^{1/\Delta} \left(r_{t-1+j\Delta} - \mu_t^{T-}(\alpha)\right)^2 I\left(r_{t-1+j\Delta} < q_t(\alpha)\right) \\
RV_{t-1,t}^{T+}(\alpha) &= \frac{1/\Delta}{n_t^{T+}(\alpha)} \sum_{j=1}^{1/\Delta} \left(r_{t-1+j\Delta} - \mu_t^{T+}(\alpha)\right)^2 I\left(r_{t-1+j\Delta} \geq q_t(1-\alpha)\right).
\end{aligned} \tag{4}$$

Similarly, the monthly downside and upside core RVs are, respectively,

$$\begin{aligned}
RV_{t-1,t}^{C-}(\alpha) &= \frac{1/\Delta}{n_t^{C-}(\alpha)} \sum_{j=1}^{1/\Delta} \left(r_{t-1+j\Delta} - \mu_t^{C-}(\alpha)\right)^2 I\left(q_t(\alpha) \leq r_{t-1+j\Delta} < q_t(1/2)\right) \\
RV_{t-1,t}^{C+}(\alpha) &= \frac{1/\Delta}{n_t^{C+}(\alpha)} \sum_{j=1}^{1/\Delta} \left(r_{t-1+j\Delta} - \mu_t^{C+}(\alpha)\right)^2 I\left(q_t(1-\alpha) > r_{t-1+j\Delta} \geq q_t(1/2)\right)
\end{aligned} \tag{5}$$

where

$$\begin{aligned}
\mu_t^{C-}(\alpha) &= \frac{1}{n_t^{C-}(\alpha)} \sum_{j=1}^{1/\Delta} r_{t-1+j\Delta} I\left(q_t(\alpha) \leq r_{t-1+j\Delta} < q_t(1/2)\right) \\
\mu_t^{C+}(\alpha) &= \frac{1}{n_t^{C+}(\alpha)} \sum_{j=1}^{1/\Delta} r_{t-1+j\Delta} I\left(q_t(1-\alpha) > r_{t-1+j\Delta} \geq q_t(1/2)\right)
\end{aligned} \tag{6}$$

are the average of intra-monthly returns smaller than the median and higher than $q_t(\alpha)$, and the average of intra-monthly returns higher than the median and smaller than $q_t(1-\alpha)$, respectively. Likewise,

$$
\begin{aligned}
n_t^{C-}(\alpha) &= \sum_{j=1}^{1/\Delta} I\left(q_t(\alpha) \leq r_{t-1+j\Delta} < q_t(1/2)\right) \\
n_t^{C+}(\alpha) &= \sum_{j=1}^{1/\Delta} I\left(q_t(1-\alpha) > r_{t-1+j\Delta} \geq q_t(1/2)\right)
\end{aligned}
\tag{7}
$$

are the number of intra-monthly returns smaller than the median and higher than $q_t(\alpha)$, and the number of intra-monthly returns higher than the median and smaller than $q_t(1-\alpha)$, respectively.

2.2 Time Series Analysis

Each of the four variances defined in equations (4) and (5) is valued differently by equity investors and this disaggregation of equity risk can be important for enhancing our understanding of the equity risk-return tradeoff. To analyze the risk-return tradeoff, we use a VAR framework and define the state vector

$$
z_t^{\top} = [\ r_{t-1,t} \quad TV_{t-1,t}(\alpha) \quad CV_{t-1,t}(\alpha) \quad TA_{t-1,t}(\alpha) \quad CA_{t-1,t}(\alpha)\ ] \tag{8}
$$

where $TV_{t-1,t}(\alpha) = RV_{t-1,t}^{T+}(\alpha) + RV_{t-1,t}^{T-}(\alpha)$ and $CV_{t-1,t}(\alpha) = RV_{t-1,t}^{C+}(\alpha) + RV_{t-1,t}^{C-}(\alpha)$ are the total tail realized variance and the total core realized variance, respectively, while $TA_{t-1,t}(\alpha) = RV_{t-1,t}^{T+}(\alpha) - RV_{t-1,t}^{T-}(\alpha)$ and $CA_{t-1,t}(\alpha) = RV_{t-1,t}^{C+}(\alpha) - RV_{t-1,t}^{C-}(\alpha)$ are the tail realized asymmetry and the core realized asymmetry, respectively.

The dependent vector z_t in equation (8) may be augmented to include other potential return predictors such as the VRP, PD ratio, etc. We assume that the vector z_t has the following VAR dynamics:

$$
z_t = \theta_0 + \Theta z_{t-1} + u_t \quad \text{where} \quad \mathbb{E}[u_t] = 0 \quad \text{and} \quad \mathbb{E}\left[u_t u_t^{\top}\right] = \Sigma. \tag{9}
$$

Let v be a vector with the same dimension as z_t, and let V be a matrix with number of columns equal to the dimension of z_t and number of rows less than or equal to the dimension of z_t. We define the variables $y_t = v^{\top} z_t$ and $x_t = V z_t$, and we are interested in the predictive linear regression

$$y_{t+1:t+h} = a_{mh} + b_{mh}^{\top} x_{t-m+1:t} + u_{t+h}^{(m)} \tag{10}$$

where

$$y_{t+1:t+h} = \frac{1}{h}\sum_{j=1}^{h} y_{t+j} \quad \text{and} \quad x_{t-m+1:t} = \frac{1}{m}\sum_{i=1}^{h} x_{t-i+1}. \tag{11}$$

The approach to estimate a_{mh} and b_{mh} in equation (11) is referred to as *overlapping* methodology. The gist consists of aggregating the predictors, $x_{t-m+1:t}$, backwards over m periods and the forecasted variables, $y_{t+1:t+h}$, forwards over h periods. It is important to notice that such aggregation increases the persistence of variables as it averages variables over several periods. However, as pointed out by Stambaugh (1999), and Valkanov (2003), the persistence of variables can artificially inflate statistical significance and R^2. Hence, in the present paper we advocate the use of an alternative approach to study predictability over long horizons, i.e. the *VAR* methodology. Intuitively, rather than aggregating predictors and forecasted variables over multiple time horizons, we exploit the stationarity feature of the VAR to estimate the drift, a_{mh}, and the slope coefficient, b_{mh}. In the appendix we show that the drift and slope coefficients of the predictive regression (10) can be expressed in terms of the VAR parameters through the formulas

$$a_{mh} = \left(v - V^{\top} b_{mh}\right)^{\top} (\mathbf{I} - \Theta)^{-1} \theta_0 \quad \text{and} \quad b_{mh} = \left(\frac{1}{m}\Sigma_{mm}^{xx}\right)^{-1} \left(\frac{1}{h}\sigma_{mh}^{xy}\right) \tag{12}$$

where $\mathbf{I}$ denotes the appropriate identity matrix,

$$\begin{aligned} \sigma_{mh}^{xy} &= V\Sigma^{zz} \left((\mathbf{I} - \Theta)^{-1} (\mathbf{I} - \Theta^{m})\right)^{\top} \left((\mathbf{I} - \Theta)^{-1} \left(\Theta - \Theta^{h+1}\right)\right)^{\top} v \\ \Sigma_{mm}^{xx} &= V\Sigma_{mm}^{zz} V^{\top}, \end{aligned} \tag{13}$$

and where

$$\begin{aligned} \Sigma_{mm}^{zz} = m\Sigma^{zz} &+ \Sigma^{zz} \left(\left((\mathbf{I} - \Theta)^{2}\right)^{-1} \left(\Theta^{m+1} - m\Theta^{2} + (m-1)\,\Theta\right)\right)^{\top} \\ &+ \left(\left((\mathbf{I} - \Theta)^{2}\right)^{-1} \left(\Theta^{m+1} - m\Theta^{2} + (m-1)\,\Theta\right)\right) \Sigma^{zz}, \end{aligned} \tag{14}$$

with $\mathrm{vec}\left(\Sigma^{zz}\right) = (\mathbf{I} - \Theta \otimes \Theta)^{-1} \mathrm{vec}\left(\Sigma\right)$.

Likewise, the R^2 of the predictive regression (10) is given by

$$R^2_{mh} = \frac{\sigma^{xy\top}_{mh}\left(\Sigma^{xx}_{mm}\right)^{-1}\sigma^{xy}_{mh}}{\left(\sigma^y_h\right)^2} \tag{15}$$

where σ^{xy}_{mh} and Σ^{xx}_{mm} are given by equation (13), and where

$$\left(\sigma^y_h\right)^2 = v^\top \Sigma^{zz}_{hh} v. \tag{16}$$

In the empirical part of the paper we present the results for both the *overlapping* and the *VAR* methodology.

3 Data

To compute the measures of market uncertainty we rely on high-frequency data of the S&P 500 Index provided by Olsen financial services. The sample period starts in January 1990 and ends in September 2016, i.e. 321 months. We eliminate obvious outliers and exclude short trading days.[3] Next, we construct a five-minute grid according to two criteria; (i) we include the first and last observed level of each day and (ii) we select all intra day levels of the S&P 500 Index at 00, 05, 10, 15 minute of every hour and compute log returns. We eliminate overnight returns as the focus lies on high frequency data.[4] The cleaned data set contains 527530 five-minute log returns, i.e. roughly 1643 observations per month. The annualized return and standard deviation are 4.89% and 14.38%, respectively. As mentioned above, the return distribution exhibits negative skewness (-0.18) and high excess kurtosis (42.88) which motivate our decomposition of the return distribution. We compute measures for market uncertainty on a monthly frequency as depicted in equation (8). The summary statistics are presented in Table 1.

Note that we only present and discuss the summary statistics for the combined measures as depicted in equation (8) and not the individual measures as introduced in equations (4) and (5). Throughout the analysis we will focus on the combined measures rather than individual measures as they present a higher correlation (see Figure 1).

The correlations of the individual measures is always larger than 0.5. Conversely the

[3] We note five short trading days, namely 1997.01.21, 1997.01.22, 2000.04.21, 2001.09.11 and 2015.03.30.

[4] This is in contradiction to Drechsler and Yaron (2011) or Andersen et al. (2010)) who treat the overnight returns and the returns over the weekend as one high-frequency return.

Table 1: **Summary Statistics of Market Uncertainty Measures**

The table reports the summary statistics of our predictors. The threshold to distinguish between tail and core measures is $\alpha = 0.28$, i.e. we set α such that R^2 is maximized. Further $\mathbb{E}[\,\cdot\,]$, $\mathbb{V}[\,\cdot\,]$, $\mathbb{S}[\,\cdot\,]$, and $\mathbb{K}[\,\cdot\,]$ correspond to the first, second, third and fourth moment, respectively and $\alpha_1[\,\cdot\,]$ captures the autocorrelation of order 1. The sample period is from beginning of January 1990 to end of September 2016.

	$\mathbb{E}[\,\cdot\,]$	$\mathbb{V}[\,\cdot\,]$	$\mathbb{S}[\,\cdot\,]$	$\mathbb{K}[\,\cdot\,]$	$\alpha_1[\,\cdot\,]$
Tail Asymmetry (TA)	-0.12	5.93	-8.07	129.44	0.08
Core Asymmetry (CA)	-0.01	0.08	-3.91	30.52	0.30
Tail Variance (TV)	17.95	28.95	7.66	85.60	0.62
Core Variance (CV)	0.57	1.11	7.73	82.38	0.69

correlation for combined measures is almost always smaller than 0.5 in absolute terms. It is preferable from an econometric perspective to rely on the regressors with lower correlations (see e.g., Stambaugh, 1999 and Valkanov, 2003).

4 Results

In this section we discuss our results. We start by highlighting the performance of our predictors in a one-month horizon framework. We then compare the performance of our measures in forecasting the one-month excess return to the one of existing predictors. Lastly, we compare the performance of all predictors for longer time horizons using both the *VAR* and the *overlapping* methodology.

4.1 Monthly VAR

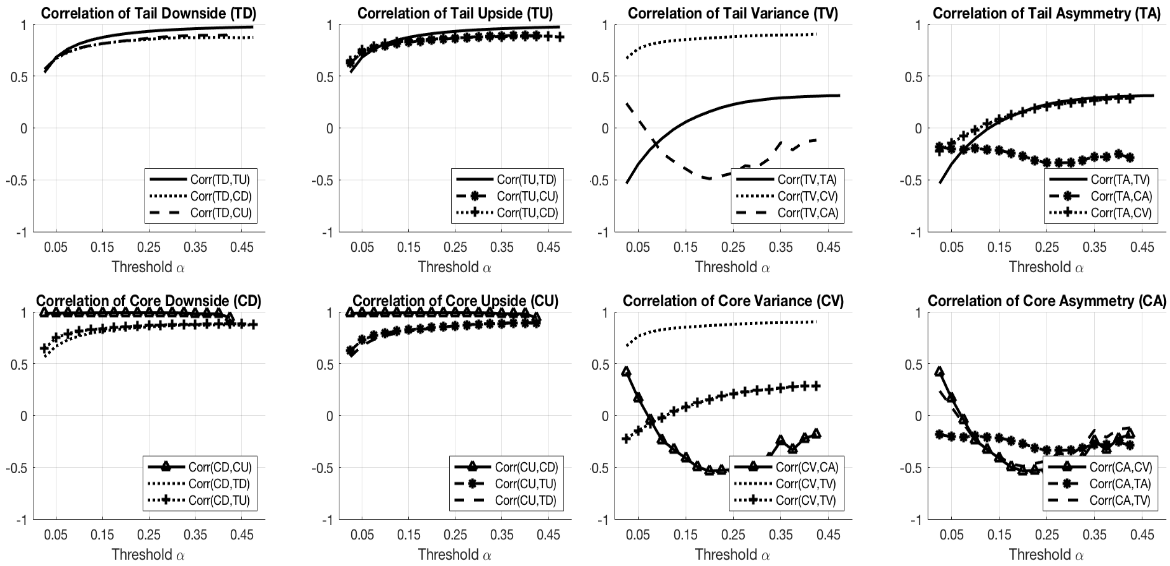
We run the regression depicted in equation (9) for several values of α. Our analysis reveals that for $\alpha = \alpha_{max} = 0.28$ the adjusted R^2 in the prediction of the one-month excess market return is highest, i.e. R^2=9.27%. In the remainder of the paper we focus on this threshold to discuss our findings.[5] The results are presented in Table 2

The analysis yields two relevant predictors to forecast the market return on the one month horizon; the tail asymmetry and the core asymmetry. We notice that both coefficients are negative. This means the higher the upside tail (core) RV, i.e. the higher the

[5] In Table 5 in the appendix we present our results for alternatives values of α.

Figure 1: **Correlations**

This figure reports the pairwise correlations of our predictors. The plot on the left hand side depicts the pairwise correlation for the individual measures (downside tail/core and upside tail/core RV), whereas the plot on the right hand side depicts the pairwise correlation of our combined measures (tail/core asymmetries and tail/core variances).



upside uncertainty, the lower the expected return. Conversely, the higher the downside tail (core) RV, i.e. the higher the downside uncertainty, the higher the expected return. These opposite effects of downside and upside uncertainty highlight why previous literature finds mostly insignificant, and even negative, systematic risk-return trade-offs.

In Table 4 in the appendix we conduct a similar analysis which also includes an additional variable, $r^- = \sum_i max(-r_{i,t}; 0)$ that captures all the negative returns within a given month, i.e. the total losses of the month. Taking a closer look at the second row we find that higher returns yield lower losses. We also show that losses forecast themselves well with an adjusted R^2=59.64% and a statistical significance at the 1% level. In unreported result we also show that losses forecast gains, denoted r^+, where $r^+ = \sum_i max(r_{i,t}; 0)$ well with an adjusted R^2=63.04% and a statistical significance also at the 1% level. It is interesting to see that losses (r^-) are able to predict both gains (r^+) and losses (r^-) but not difference of the two ($r^+ - r^-$), which represent the overall market return. As mentioned above the market return is only forecastable with the asymmetry of upside and downside variances.

Table 2: **VAR Results**

This table present the results for a VAR(1). The threshold α which defines the differences between tail and core measures is defined as $\alpha_{max} = 0.28$. The t-statistics in brackets are calculated using HAC standard errors. The statistical significance at the 1%, 5% and 10% level is indicated by, $*$, ‡, and †, respectively.

	r_t	TA_t	CA_t	TV_t	CV_t	R^2
r_{t+1}	0.04	-0.09‡	-14.60*	-0.02	-0.13	9.27
	(0.59)	(-2.19)	(-3.86)	(-0.48)	(-0.15)	
TA_{t+1}	-0.18	0.02	5.94	0.01	1.18	8.96
	(-1.50)	(0.60)	(1.14)	(0.35)	(1.24)	
CA_{t+1}	0.00‡	0.00	-0.01	0.00	-0.06‡	36.00
	(2.51)	(1.21)	(-0.08)	(0.91)	(-2.40)	
TV_{t+1}	-0.96‡	0.26	66.40	0.47	4.38	40.89
	(-1.99)	(0.65)	(1.31)	(1.24)	(0.58)	
CV_{t+1}	-0.04‡	0.00	2.62	0.01	0.46†	54.04
	(-2.20)	(0.33)	(1.35)	(0.83)	(1.73)	

4.2 Performance Compared to Existing Measures

In order to better capture the performance of our predictors we now compare them to established predictors in the finance literature. More precisely, we compare our measures to the VRP and the PD ratio. The results are presented in Table 3.

Table 3: **Comparing the Performance of Different Predictors**
This table compares the performance of our predictors to the ones of the academic literature. The threshold α which defines the differences between tail and core measures is defined as $\alpha_{max} = 0.28$. The t-statistics in brackets are calculated using HAC standard errors. The statistical significance at the 1%, 5% and 10% level is indicated by, $*$, ‡, and †, respectively.

	r_t	TA_t	CA_t	TV_t	CV_t	VRP_t	PD_t	R^2
I	0.04	-0.09‡	-14.60*	-0.02	-0.13			9.27
	(0.59)	(-2.19)	(-3.86)	(-0.48)	(-0.15)			
II	0.04					0.04*		4.90
	(0.55)					(3.60)		
III	0.03						-2.00‡	2.98
	(0.48)						(-2.25)	
IV	0.05	-0.08†	-12.89*	-0.01	-0.01	0.02†		10.08
	(0.88)	(-1.87)	(-3.12)	(-0.34)	(-0.01)	(1.71)		
V	0.04	-0.10‡	-12.95*	-0.03	0.26		-1.75†	10.47
	(0.58)	(-2.42)	(-3.29)	(-0.85)	(0.28)		(-1.94)	
VI	0.06	-0.08‡	-10.93‡	-0.02	0.42	0.02†	-1.89‡	11.54
	(0.91)	(-2.07)	(-2.44)	(-0.73)	(0.44)	(1.81)	(-2.05)	
VII	0.04					0.04*	-2.17‡	7.04
	(0.59)					(3.80)	(-2.44)	

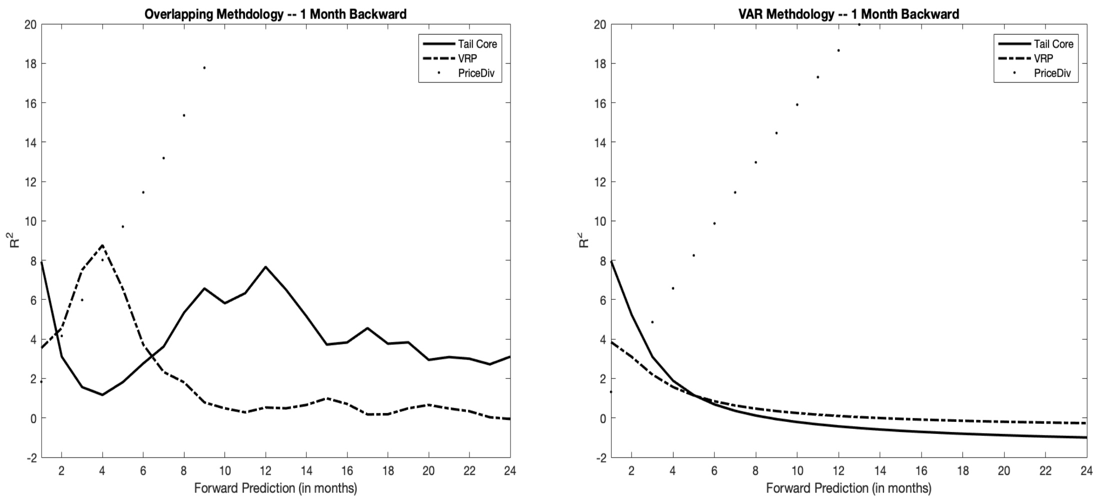
Model I corresponds to our base model as highlighted in the VAR analysis. Model II (III) captures the ability of the VRP (PD ratio) to forecast the excess market return on a one-month horizon. A quick look at the adjusted R^2 clearly verifies the better performance of our predictors compared to the existing ones. Taking a closer look at Models III, V, VI and VII, we notice that the statistical and economical significance of the PD ratio is generally not affected by the presence/absence of other predictors. Conversely, comparing Models II and VII to Models IV and VI, we notice that the statistical and economical significance of the VRP as well as the asymmetry measures decreases. Hence, to some extend our predictors capture information that is related to the VRP. This common information content stems from the physical variation of the market returns. More precisely, our measures are solely based on the historical return distribution whereas the VRP is defined as the difference between the risk-neutral and the psychical variance of market returns. Despite the fact that VRP contains more information, our predictors (i) outperform the VRP significantly when predicting the excess monthly market return, and (ii) are easier to compute as they can be extracted from the realize return series and, hence, do not require any assumption regarding the risk neutral variation.

4.3 Multi-horizon Predictability

So far we have only discussed the predictability on a one-month horizon. We now turn our attention to the performance of our predictors for longer time horizons. As mentioned above, when investigating the multi-horizon predictability we distinguish between the *overlapping* and the *VAR* methodology. While the former approach is straight forward to implement, it raises the issue of persistence of variables, which in turn, artificially inflates statistical significance and R^2. The latter approach is able to circumvent this problem by exploiting the stationarity feature of the VAR. Figure 2 represents the R^2 for both approach for a multitude of horizons.

Figure 2: **Predictability for Longer Time Horizons**

This figure contains the predictability of the market return as measured with the adjusted R^2 using different predictors. The plot on the right (left)-hand side of Figure 2 contains the R^2 using the *overlapping* (VAR) methodology.



The plot on the right-hand side of Figure 2 contains the R^2 using the *overlapping* methodology. It confirms the findings of Bollerslev et al. (2009) and Drechsler and Yaron

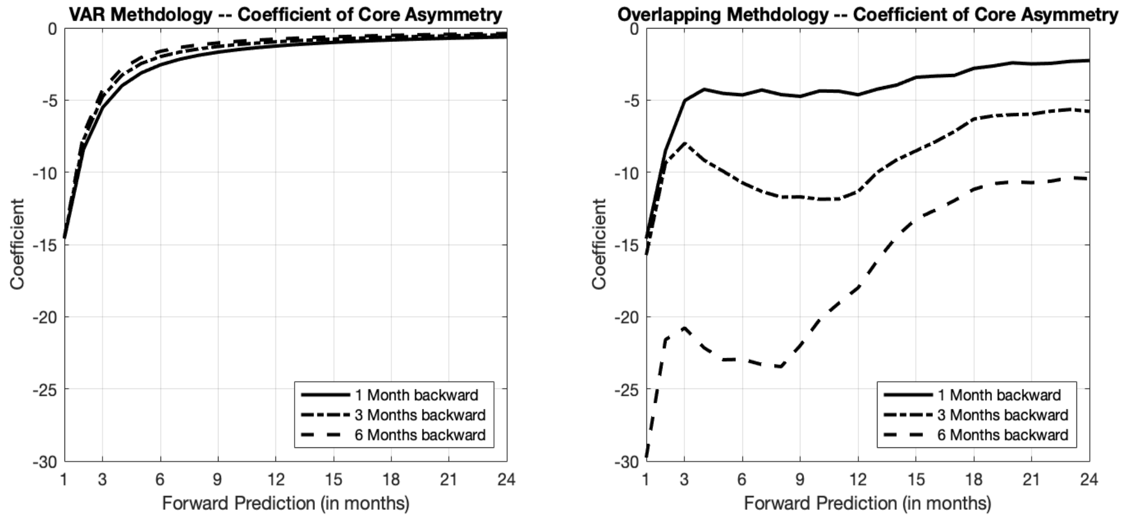
(2011) who show that the VRP is able to best predict the market return on a quarterly horizon. However, when using the *VAR* methodology we find that our measures are able to outperform the VRP on a quarterly horizon. In other words, the predictably performance of the VRP simply stems from the persistence which in turn originates form the aggregation of the predictors. Independent of the methodology (*overalapping* or *VAR*), the PD ratio performs well for long time-horizons predictions.

In Figure 4 in the appendix we also report result for the *VAR* and the *overlapping* methodology for alternative lengths m of backward aggregation. With the *VAR* methodology, we notice that for $m > 1$ the predictability of the VRP and our measures vanishes. The PD ratio is unaffected. However, using the overlapping methodology we find that our measures yield better predications than the VRP for $m \leq 6$. Similarly, our measures also outperform the PD ratio for forward's prediction of up to 9 months. This means, that using the *overlapping* methodology (which is used most commonly in the literature), our measures outperform both the VRP and the PD ratio for a broad range of forward and backward aggregation specifications.

To further compare the *VAR* and *methodology*, Figure 3 depicts the coefficients of the core asymmetry for different lengths of forward and backward aggregation. With the *VAR* methodology, we notice that the economic significance of the coefficients gradually diminishes with increasing forward and backward aggregation. Using the *overlapping* methodology, we obtain different insights. Regarding the backward aggregation, we notice that the longer the period of backward aggregation, the higher the economic significance of the core asymmetry. This is in line with Stambaugh (1999) and Valkanov (2003) who show that coefficients are inflates with increasing persistency of the variables. Regarding the forward aggregation we notice that overall the coefficients are decreasing with time (similar as the VAR methodolgy). However, we notice that in a range of three to nine months, the economic significance either stabilizes of increases in case or longer backward aggregation. Such stylized facts (together with the high adjusted R^2 in the same time range) explain why a large body of the literature focuses on three to nine months horizons when predicting the market return.

Figure 3: **Coefficients**

This figure depicts the coefficients of the core asymmetry for different lengths of forward and backward aggregation. The two traps corresponds to coefficients using the *VAR* and the *overlapping* methodology.



5 Conclusion

In the present paper we develop a novel decomposition of the realized variance into four components; downside tail, downside core, upside core and upside tail RV. To compute the measures of uncertainty we follow two motivations. *First*, given that returns are negatively skewed and that upside gains and downside loses are perceived differently by investor, it is desirable to distinguish between of upside and downside uncertainty. *Second*, given that the average monthly return distribution presents a high excess kurtosis, i.e. a significant amount of information is captured by the tails, it also proves relevant to distinguish between 'extreme' events, i.e. large returns and 'average' events, i.e. small returns. These two types of returns correspond respectively to observations in the tails and the core of the return distribution. Consequently, by combining our two motivations, we obtain a total of four MU measures: downside tail, downside core, upside core and upside tail RV.

Our analysis is conducted in a monthly vector autoregression (VAR) framework. The adjusted R^2 of predicting the market excess return is 9.27%. This result is driven in particular by tail asymmetry (which is defined as the difference between upside and downside tail) and the core asymmetry (which are defined as the difference between upside and downside core). The coefficient estimates are negative and statistically significant at the

5% and 1% level for the tail and core asymmetry, respectively. The negativity of the coefficient estimates implies that higher upside uncertainty (as measured by upside tail and core RVs) yields a lower return. Conversely, higher downside uncertainty (as measured by the downside tail and core RVs) yields a higher return. Hence, we show that expected returns include rewards for accepting the risk of a potential downside loss, but also discounts for potential upside gains. As such, these results are in line with the inter-temporal capital asset pricing model (ICAPM) as introduced by Merton (1973) who advocates a liner relation between variance and expected excess stock market return. The systematic risk-return trade-off suggests that the conditional excess market return varies with its conditional variance. While previous literature is finding mostly insignificant, and even negative, systematic risk-reward trade-offs, the present paper confirms a positive risk-return trade-off for downside uncertainty, and also establishes a negative risk-return trade-off for upside uncertainty.

Appendix

6 Additional Results

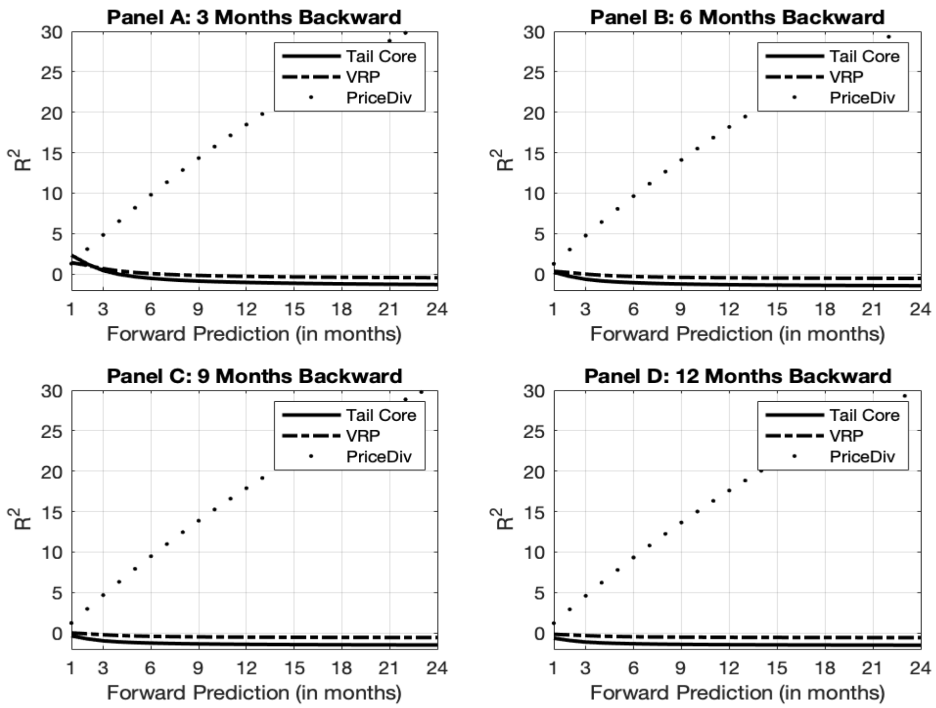
Table 4: **Results for Alternative Dependent Vector in VAR**

The present table presents result for an alternative VAR specification. More precisely, we also include the losses, denoted by r_t^-.

	r_t	r_t^-	TA_t	CA_t	TV_t	CV_t	R^2
r_{t+1}	0.04	0.00	-0.09‡	-14.61*	-0.02	-0.17	8.98
	(0.58)	(0.11)	(-2.15)	(-3.85)	(-0.49)	(-0.18)	
r_{t+1}^-	-0.62†	0.88*	-0.03	39.24	-0.03	-2.38	59.64
	(-1.67)	(7.69)	(-0.12)	(1.06)	(-0.11)	(-0.34)	
TA_{t+1}	-0.19	-0.02	0.02	6.06	0.01	1.54	8.77
	(-1.52)	(-0.69)	(0.52)	(1.16)	(0.40)	(1.32)	
CA_{t+1}	0.00*	0.00	0.00	-0.02	0.00	-0.07*	36.78
	(2.83)	(1.43)	(1.34)	(-0.11)	(0.81)	(-2.95)	
TV_{t+1}	-0.80†	0.31‡	0.33	63.95	0.44	-2.81	42.46
	(-1.82)	(2.52)	(0.82)	(1.31)	(1.19)	(-0.29)	
CV_{t+1}	-0.03‡	0.01*	0.01	2.52	0.01	0.16	55.96
	(-2.03)	(2.95)	(0.53)	(1.37)	(0.77)	(0.48)	

Figure 4: **Predictability for Alternative Lengths of Backward Aggregation**

VAR methodology



Overlapping methodology

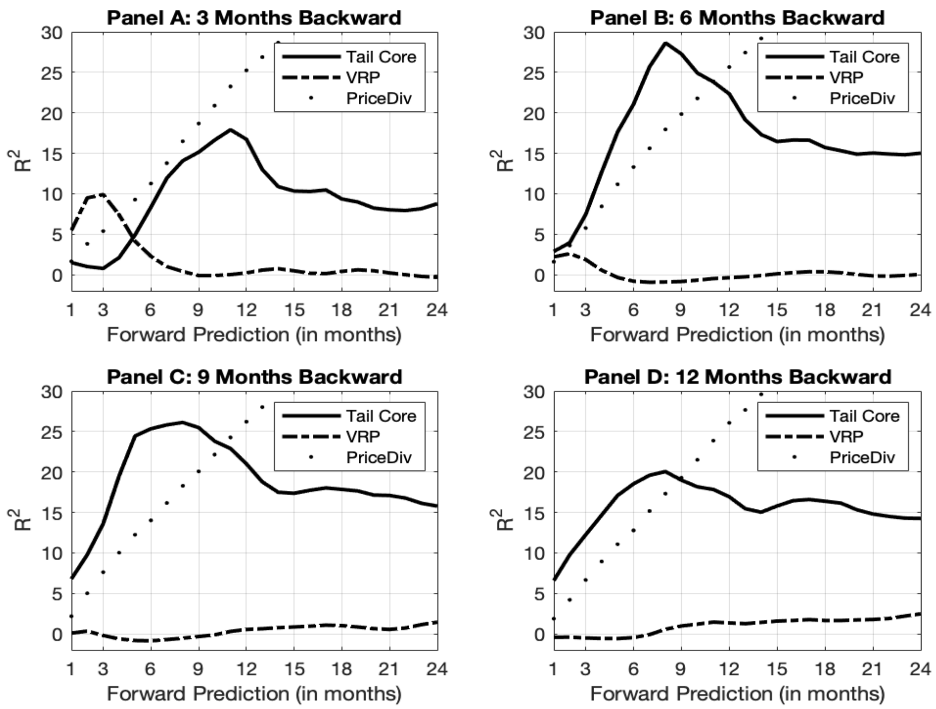


Table 5: **Results for Alternative Tail-core Thresholds**

	$\alpha = 0.10$						$\alpha = 0.15$						$\alpha = 0.20$					
	r_t	TA_t	CA_t	TV_t	CV_t	R^2	r_t	TA_t	CA_t	TV_t	CV_t	R^2	r_t	TA_t	CA_t	TV_t	CV_t	R^2
r_{t+1}	0.02	-0.06^{*}	0.43	$-0.04^{\ddagger}$	$0.22^{\ddagger}$	4.48	0.02	-0.08^{*}	0.68	$-0.05^{\ddagger}$	$0.44^{\ddagger}$	4.50	0.03	-0.09^{*}	-2.39	$-0.05^{\ddagger}$	$0.58^{\dagger}$	5.08
	(0.28)	(-3.11)	(0.41)	(-2.32)	(2.01)		(0.31)	(-3.24)	(0.43)	(-2.39)	(2.30)		(0.50)	(-3.16)	(-1.05)	(-2.15)	(1.70)	
TA_{t+1}	-0.33	-0.02	1.72	-0.01	0.33	0.69	-0.25	-0.00	1.28	0.01	0.39	3.01	-0.21	0.01	2.28	0.01	0.59	5.46
	(-1.08)	(-0.72)	(0.62)	(-0.39)	(1.16)		(-1.17)	(-0.09)	(0.41)	(0.24)	(0.96)		(-1.26)	(0.25)	(0.58)	(0.35)	(0.99)	
CA_{t+1}	-0.00	-0.00	0.12	-0.00	0.00	1.02	0.00	-0.00	-0.04	-0.00	-0.02	6.84	$0.01^{\dagger}$	0.00	-0.12	0.00	$-0.04^{\ddagger}$	19.65
	(-0.36)	(-0.44)	(0.81)	(-0.32)	(0.07)		(0.45)	(-0.48)	(-0.22)	(-0.06)	(-0.81)		(1.72)	(0.18)	(-0.73)	(0.53)	(-1.97)	
TV_{t+1}	$-1.46^{\dagger}$	0.37	24.11	0.45	0.33	27.12	$-1.24^{\dagger}$	0.32	33.99	0.46	1.13	34.09	$-1.06^{\ddagger}$	0.27	46.10	0.46	2.54	37.76
	(-1.67)	(1.05)	(1.29)	(1.22)	(0.19)		(-1.84)	(0.92)	(1.29)	(1.29)	(0.50)		(-1.96)	(0.79)	(1.25)	(1.31)	(0.85)	
CV_{t+1}	$-0.22^{\dagger}$	0.04	2.04	0.05	0.39	55.99	$-0.13^{\dagger}$	0.02	1.96	0.04	0.40	55.12	$-0.08^{\ddagger}$	0.01	2.24	0.02	$0.44^{\dagger}$	54.67
	(-1.89)	(0.84)	(0.76)	(0.98)	(1.52)		(-1.92)	(0.67)	(0.69)	(0.98)	(1.50)		(-2.14)	(0.49)	(0.82)	(0.96)	(1.73)	

	$\alpha = 0.25$						$\alpha = 0.30$						$\alpha = 0.35$					
	r_t	TA_t	CA_t	TV_t	CV_t	R^2	r_t	TA_t	CA_t	TV_t	CV_t	R^2	r_t	TA_t	CA_t	TV_t	CV_t	R^2
r_{t+1}	0.04	-0.09^{*}	-9.48^{*}	-0.03	0.28	7.96	0.04	$-0.09^{\ddagger}$	-16.85^{*}	-0.01	-0.18	8.74	0.04	$-0.11^{\dagger}$	$-12.45^{\ddagger}$	-0.04	1.76	4.81
	(0.62)	(-2.59)	(-2.99)	(-1.17)	(0.49)		(0.58)	(-2.00)	(-3.97)	(-0.40)	(-0.16)		(0.56)	(-1.78)	(-2.27)	(-0.73)	(0.61)	
TA_{t+1}	-0.19	0.02	4.58	0.01	0.95	7.77	-0.18	0.03	5.78	0.02	1.14	9.48	$-0.17^{\dagger}$	0.05	$11.37^{\ddagger}$	0.03	1.24	11.56
	(-1.42)	(0.46)	(0.94)	(0.32)	(1.20)		(-1.56)	(0.73)	(1.08)	(0.53)	(0.98)		(-1.74)	(1.03)	(2.11)	(0.57)	(0.46)	
CA_{t+1}	$0.00^{\ddagger}$	0.00	0.01	0.00	$-0.05^{\dagger}$	30.95	$0.00^{\ddagger}$	0.00	-0.02	0.00	-0.06^{*}	40.41	0.00	$0.00^{\dagger}$	-0.35^{*}	$0.00^{\dagger}$	-0.14^{*}	45.74
	(2.45)	(0.82)	(0.08)	(0.56)	(-1.79)		(2.33)	(1.21)	(-0.11)	(1.05)	(-2.98)		(1.01)	(1.88)	(-3.29)	(1.72)	(-3.08)	
TV_{t+1}	$-0.98^{\ddagger}$	0.22	58.22	0.43	4.27	39.79	$-0.95^{\ddagger}$	0.37	48.50	0.62	-0.54	40.71	$-0.92^{\ddagger}$	0.64	-40.38	1.03	-27.88	43.14
	(-1.99)	(0.64)	(1.27)	(1.28)	(0.95)		(-1.97)	(0.72)	(1.20)	(1.23)	(-0.04)		(-1.98)	(0.87)	(-0.65)	(1.34)	(-0.65)	
CV_{t+1}	$-0.05^{\ddagger}$	0.01	2.61	0.01	$0.49^{\ddagger}$	54.93	$-0.03^{\ddagger}$	0.01	1.78	0.01	0.27	52.14	$-0.02^{\ddagger}$	0.01	0.02	0.01	-0.17	50.30
	(-2.24)	(0.32)	(1.17)	(0.90)	(2.20)		(-2.13)	(0.47)	(1.30)	(0.92)	(0.65)		(-2.03)	(0.67)	(0.02)	(1.04)	(-0.23)	

7 VAR Methodology

Before we derive the results of our parametric model it proves usefull to be familiar with some results of the VAR framework. Let $z_t = [r_t, RV_t^+, RV_t^-]^\top$ follow a VAR model

$$\underset{q\times 1}{z_t} = \underset{q\times 1}{\Psi} + \underset{q\times q}{\Phi}\underset{q\times 1}{z_{t-1}} + \underset{q\times 1}{\epsilon_t} \tag{17}$$

where $\mathbb{E}[\epsilon_t] = 0$ and $\mathbb{V}(\epsilon_t)=\underset{q\times q}{\Sigma}$.

Further, the stationarity of the VAR process let's solve for closed form solution of $\mathbb{E}[z_t]$. Rewriting equation (17) yields

$$\begin{aligned}\mathbb{E}[z_t] =& \Psi + \Phi\mathbb{E}[z_{t-1}]\\ =& \Psi + \Phi\mathbb{E}[z_t]\end{aligned}$$

$$\begin{aligned}\Rightarrow \qquad (I-\Phi)\mathbb{E}[z_t] =& \Psi\\ \mathbb{E}[z_t] =& (I-\Phi)^{-1}\Psi\end{aligned}$$

Similarily we can derive a closed form solution for expectation about future outcomes.

$$\begin{aligned}\mathbb{E}_t[z_{t+j}] =& \Psi + \Phi\mathbb{E}_t[z_{t+j-1}] + \underbrace{\mathbb{E}[\epsilon_{t+j}]}_{=0}\\ =& \Psi + \Phi\mathbb{E}[\Psi + \Phi\mathbb{E}_t[z_{t+j-2}]]\\ =& \Psi + \Phi\Psi_{j-1} + \Phi\cdot\Phi_{j-1}\mathbb{E}_t[z_{t+j-2}]]\end{aligned} \tag{18}$$

for t large enough, we can state that $\mathbb{E}_t[z_{t+j}]] = \mathbb{E}_t[z_{t+j-2}]]$. We can hence derive a system of equation for equation (18) such that we can postulate

$$\mathbb{E}[z_{t+j}] = \Psi_j + \Phi_j z_t \qquad , j \geq 0$$

where

$$\begin{cases} \Psi_j = \Psi + \Phi\Psi_{j-1} \\ \Phi_j = \Phi\Phi_{j-1} \end{cases}$$
$$\Leftrightarrow \begin{cases} \Psi_j - (I-\Phi)^{-1}\Psi = \Phi\big(\Psi_{j-1} - (I-\Phi)^{-1}\Psi\big) \\ \Phi_j = \Phi^j \end{cases}$$

We can solve for Ψ_j explicitely

$$\begin{aligned}
\Psi_j =& (I-\Phi)^{-1}\Psi + \Phi^{j-1}\big(\Psi_1 - (I-\Phi)^{-1}\Psi\big) \\
=& (I-\Phi)^{-1}\Psi + \Phi^{j-1}\big(\Psi - (I-\Phi)^{-1}\Psi\big) \\
=& (I-\Phi)^{-1}\Psi + \Phi^{j-1}(I-\Phi)^{-1}\big((I-\Phi)\Psi - \Psi\big) \\
=& (I-\Phi)^{-1}\Psi + \Phi^{j-1}(I-\Phi)^{-1}\Phi\Psi \\
=& (I-\Phi)^{-1}\Psi + \Phi^{j}(I-\Phi)^{-1}\Psi \\
=& (I-\Phi^j)(I-\Phi)^{-1}\Psi
\end{aligned}$$

We can hence write

$$\begin{aligned}
\mathbb{E}_t[z_{t+j}] =& \Psi_j + \Phi_j z_t \\
=& (I-\Phi^j)(I-\Phi)^{-1}\Psi + \Phi^j z_t
\end{aligned} \tag{19}$$

8 Parametric estimation as an alternative to aggregation

We define forward and backwards aggregated variables respectively as

$$z_{t+1:t+h} = \sum_{j=1}^{k} z_{t+j} = z_{t+1} + z_{t+2} + ... + z_{t+h}$$
$$z_{t-m+1:t} = \sum_{i=1}^{m} z_{t-i+1} = z_{t-m+1} + z_{t-m+2} + ... + z_t$$

We further define a selection mechanism

$$\begin{aligned}
y_t =& \iota^\top z_t \qquad \text{where } \iota \text{ is a selection vector} \\
x_t =& \Lambda z_t \qquad \text{where } \Lambda \text{ is a selection matrix}
\end{aligned}$$

So we can rewrite the regression for the selected variables as follows:

$$y_{t+1:t+h} = \psi_{m,h} + \Phi_{m,h}^{\top} x_{t-m+1:t} + u_{t+h}^{(m)} \tag{20}$$

and solve for the estimates

$$\begin{aligned}
\psi_{m,h} &= \mathbb{E}[y_{t+1:t+h}] - \Phi_{m,h}^{\top}\mathbb{E}[x_{t-m+1:t}] \\
\Phi_{m,h} &= \left(\Sigma_{mm}^{xx}\right)^{-1}\Sigma_{m,h}^{xy}
\end{aligned}$$

We now want to explicitly express these parameters as functions of the VAR parameters as well as the corresponding periods of forward and backward aggregation, denoted by h and m respectively. The gist lies in the stationarity of the underlying process.

Let's start with $\psi_{m,h}$:

$$\begin{aligned}
\mathbb{E}[y_{t+1:t+h}] =& \sum_{j=1}^{h} \mathbb{E}[y_{t+j}] = \sum_{j=1}^{h} \mathbb{E}[y_t] \\
=& h\mathbb{E}[y_t] = h\mathbb{E}[\iota^{\top} z_t] = h\iota^{\top}\mathbb{E}[z_t] \\
=& h\iota^{\top}(I-\Phi)^{-1}\Psi
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[x_{t-m+1:t}] =& \sum_{i=1}^{m} \mathbb{E}[x_{t-i+1}] = \sum_{i=1}^{m} \mathbb{E}[x_t] \\
=& m\mathbb{E}[x_t] = m\mathbb{E}[\Lambda z_t] = m\Lambda\mathbb{E}[z_t] \\
=& m\Lambda(I-\Phi)^{-1}\Psi
\end{aligned}$$

Hence, we can rewrite $\psi_{m,h}$ as

$$\begin{aligned}
\psi_{m,h} &= \mathbb{E}[y_{t+1:t+h}] - \Phi_{m,h}^{\top}\mathbb{E}[x_{t-m+1:t}] \\
&= h\iota^{\top}(I-\Phi)^{-1}\Psi - \Phi_{m,h}^{\top} m\Lambda(I-\Phi)^{-1}\Psi \qquad (21)
\end{aligned}$$

We now want to express $\Phi_{m,h}$ in a similiar fashion. To do so we use the same approach to solve for $\Sigma_{m,h}^{xy}$ and $\Sigma_{m,m}^{xx}$. It proves useful to introduce the following theorem

Theorem

$$\mathbb{C}(X,Y) = \mathbb{C}\big[\mathbb{E}_c[X], \mathbb{E}_c[Y]\big] + \mathbb{E}\big[\mathbb{C}_c(X,Y)\big] \tag{22}$$

$$
\begin{aligned}
\Sigma^{xy}_{m,h} =& \mathbb{C}[x_{t-m+1:t}, y_{t+1:t+h}] \\
=& \mathbb{C}\big[\mathbb{E}_t[x_{t-m+1:t}], \mathbb{E}_t[y_{t+1:t+h}]\big] + \underbrace{\mathbb{E}\big[\mathbb{C}_t[x_{t-m+1:t}], y_{t+1:t+h}]\big]}_{=0} \\
=& \mathbb{C}\big[x_{t-m+1:t}, \mathbb{E}_t[y_{t+1:t+h}]\big] \\
=& \mathbb{C}\left[\sum_{i=1}^{m} x_{t-i+1}, \sum_{j=1}^{h} \mathbb{E}_t[y_{t+j}]\right] \\
=& \mathbb{C}\left[\sum_{i=1}^{m} \Lambda z_{t-i+1}, \sum_{j=1}^{h} \mathbb{E}_t[\iota^\top z_{t+j}]\right] = \mathbb{C}\left[\Lambda \sum_{i=1}^{m} z_{t-i+1}, \iota^\top \sum_{j=1}^{h} \big(\Psi_j + \Phi_j z_t\big)\right] \\
=& \Lambda \sum_{i=1}^{m} \mathbb{C}\left[z_{t-i+1}, , \iota^\top \sum_{j=1}^{h} \big(\Phi^j\big) z_t\right] \\
=& \Lambda \sum_{i=1}^{m} \mathbb{C}[z_{t-i+1}, z_t] \left(\sum_{j=1}^{h} \Phi^j\right)^\top \iota \\
=& \Lambda \left\{\mathbb{C}[z_t, z_t] + \sum_{i=2}^{m} \mathbb{C}[z_{t-i+1}, , z_t]\right\} \left(\sum_{j=1}^{h} \Phi^j\right)^\top \iota \\
=& \Lambda \left\{\Sigma^{zz} + \sum_{i=2}^{m} \mathbb{C}\big[z_{t-i+1}, \mathbb{E}_{t-i+1}[z_t]\big]\right\} \left(\sum_{j=1}^{h} \Phi^j\right)^\top \iota \\
=& \Lambda \left\{\Sigma^{zz} + \sum_{i=2}^{m} \mathbb{C}\big[z_{t-i+1}, \Psi_{i-1} + \Phi_{i-1} z_{t-i+1}\big]\right\} \left(\sum_{j=1}^{h} \Phi^j\right)^\top \iota \\
=& \Lambda \left\{\Sigma^{zz} + \sum_{i=2}^{m} \mathbb{C}\big[z_{t-i+1}, \Phi^{i-1} z_{t-i+1}\big]\right\} \left(\sum_{j=1}^{h} \Phi^j\right)^\top \iota \\
=& \Lambda \left\{\Sigma^{zz} + \sum_{i=2}^{m} \Big(\mathbb{C}\big[z_{t-i+1}, z_{t-i+1}\big] \big(\Phi^{i-1}\big)^\top\Big)\right\} \left(\sum_{j=1}^{h} \Phi^j\right)^\top \iota \\
=& \Lambda \left\{\Sigma^{zz} + \sum_{i=2}^{m} \Big(\Sigma^{zz} \big(\Phi^{i-1}\big)^\top\Big)\right\} \left(\sum_{j=1}^{h} \Phi^j\right)^\top \iota \\
=& \Lambda \sum_{i=1}^{m} \left(\Sigma^{zz} \Big(\Phi^{i-1}\Big)^\top\right) \left(\sum_{j=1}^{h} \Phi^j\right)^\top \iota \\
\Sigma^{xy}_{m,h} =& \Lambda \Sigma^{zz} \left(\sum_{i=1}^{m} \Phi^{i-1}\right)^\top \left(\sum_{j=1}^{h} \Phi^j\right)^\top \iota && (23) \\
\Sigma^{xy}_{m,h} =& \Lambda \Sigma^{zz} \Big((I-\Phi)^{-1}(I-\Phi^m)\Big)^\top \Big((I-\Phi)^{-1}(\Phi-\Phi^{h+1})\Big)^\top \iota && (24)
\end{aligned}
$$

where from equation (17)

$$
\begin{aligned}
\Sigma^{zz} =& \mathbb{V}[z_t] = \Phi \mathbb{V}[z_{t-1}]\Phi^\top + Var[\epsilon_t] \\
=& \Phi \Sigma^{zz} \Phi^\top + \Sigma \\
\Rightarrow vec(\Sigma^{zz}) =& vec(\Phi \Sigma^{zz} \Phi^\top) + vec(\Sigma) \\
=& (\Phi \otimes \Phi) vec(\Sigma^{zz}) + vec(\Sigma) \\
=& \ [I - \Phi \otimes \Phi]^{-1} vec(\Sigma)
\end{aligned} \tag{25}
$$

and, more generally for variance covariance between z_{t-i} and z_t is given by

$$
\begin{aligned}
\Sigma_i^{zz} =& \mathbb{C}[z_{t-i}, z_t] = \mathbb{C}\big[z_{t-i}, \mathbb{E}[z_t]\big] \\
=& \mathbb{C}[z_{t-i}, \Psi_i + \Phi_i z_{t-i}] = \Sigma^{zz}\big(\Phi^i\big)^\top
\end{aligned} \tag{26}
$$

Now, let's solve for the backward aggregated variable.

$$
\begin{aligned}
\Sigma_{m,m}^{xx} =& \mathbb{V}[x_{t-m+1:t}] = \mathbb{V}\left[\sum_{i=1}^{m} x_{t-i+1}\right] \\
=& \mathbb{V}\left[\Lambda \sum_{i=1}^{m} z_{t-i+1}\right] = \Lambda \mathbb{V}\left[\sum_{i=1}^{m} z_{t-i+1}\right]\Lambda^\top \\
=& \Lambda\left[m\Sigma^{zz} + \sum_{i=1}^{m-1}(m-i)\Big(\Sigma_i^{zz} + (\Sigma_i^{zz})^\top\Big)\right]\Lambda^\top \\
=& \Lambda\left[m\Sigma^{zz} + \sum_{i=1}^{m-1}(m-i)\Big(\Sigma^{zz}\big(\Phi^i\big)^\top + \big(\Phi^i\big)\Sigma^{zz}\Big)\right]\Lambda^\top \\
=& \Lambda\left[m\Sigma^{zz} + \Sigma^{zz}\left(\sum_{i=1}^{m-1}(m-i)\big(\Phi^i\big)\right)^\top + \left(\sum_{i=1}^{m-1}(m-i)\big(\Phi^i\big)\Sigma^{zz}\right)\right]\Lambda^\top \\
=& \Lambda\Bigg[m\Sigma^{zz} + \Sigma^{zz}\bigg(\Big[(I-\Phi)^2\Big]^{-1}\Big[\Phi^{m+1} - m\Phi^2 + (m-1)\Phi\Big]\bigg)^\top \\
& + \Big[(I-\Phi)^2\Big]^{-1}\Big[\Phi^{m+1} - m\Phi^2 + (m-1)\Phi\Big]\Sigma^{zz}\bigg)\Bigg]\Lambda^\top
\end{aligned} \tag{27}
$$

Proof

We define the following function

$$
\begin{aligned}
f(x) = \sum_{i=1}^{m-1}(m-i)x^i = \sum_{i=1}^{m-1} i x^{m-i} = x^{m-1}\sum_{i=1}^{m-1} x^{1-i} = x^{m-1}\left(\sum_{i=1}^{m-1} y^i\right)'\Bigg|_{y=x^{-1}} \\
=& x^{m-1}\left(\frac{y-y^m}{1-y}\right)'\Bigg|_{y=x^{-1}} \\
=& x^{m-1}\frac{(1-my^{m-1})(1-y)+(y-y^m)}{(1-y)^2}\Bigg|_{y=x^{-1}} \\
=& x^{m-1}\frac{1-my^{m-1}+(m-1)y^m}{(1-y)^2}\Bigg|_{y=x^{-1}} \\
=& x^{m-1}\frac{1-mx^{1-m}+(m-1)x^{-m}}{(1-x^{-1})^2} = \frac{x^{m-1}-m+(m-1)x^{-1}}{x^{-2}(1-x)^2} \\
=& \frac{x^{m+1}-mx^2+(m-1)x}{(1-x)^2}
\end{aligned} \tag{28}
$$

Equivalently, if x is matrix, we can rewrite equation (28) as

$$
\sum_{i=1}^{m-1}(m-1)\Phi^i = \left[(I-\Phi)^2\right]^{-1}\left[\Phi^{m+1}-m\Phi^2+(m-1)\Phi\right] \qquad Q.E.D.
$$

Finally we can rewrite $\Phi_{m,h}$ as

$$
\begin{aligned}
\Phi_{m,h} =& \left(\Sigma^{xx}_{m,m}\right)^{-1}\Sigma^{xy}_{m,h} \\
=& \Bigg(\Lambda\Bigg[m\Sigma^{zz} + \Sigma^{zz}\Bigg(\left[(I-\Phi)^2\right]^{-1}\left[\Phi^{m+1}-m\Phi^2+(m-1)\Phi\right]\Bigg)^\top \\
& + \left[(I-\Phi)^2\right]^{-1}\left[\Phi^{m+1}-m\Phi^2+(m-1)\Phi\right]\Sigma^{zz}\Bigg)\Bigg]\Lambda^\top\Bigg)^{-1} \times \\
& \Lambda\Sigma^{zz}\Big((I-\Phi)^{-1}(I-\Phi^m)\Big)^\top\Big((I-\Phi)^{-1}(\Phi-\Phi^{h+1})\Big)^\top\iota
\end{aligned}
$$

with $vec(\Sigma^{zz}) = [I-\Phi\otimes\Phi]^{-1}vec(\Sigma)$.

And recall $\psi_{m,h}$

$$\psi_{m,h} = h\iota^\top (I-\Phi)^{-1}\Psi - \Phi_{m,h}^\top m\Lambda (I-\Phi)^{-1}\Psi$$

Recall the regression

$$y_{t+1:t+h} = \psi_{m,h} + \Phi_{m,h}^\top x_{t-m+1:t} + u_{t+h}^{(m)}$$

The $R^2_{m,h}$ is given by

$$\begin{aligned} R^2_{m,h} =& \frac{\Phi_{m,h}^\top \Sigma_{m,m}^{xx} \Phi_{m,h}}{\Sigma_{h,h}^{yy}} = \frac{\Sigma_{m,h}^{xy} \left(\Sigma_{m,m}^{xx}\right)^{-1} \Sigma_{m,m}^{xx} \left(\Sigma_{m,m}^{xx}\right)^{-1} \Sigma_{m,h}^{xy}}{\Sigma_{h,h}^{yy}} \\ =& \frac{\Sigma_{m,h}^{xy} \left(\Sigma_{m,m}^{xx}\right)^{-1} \Sigma_{m,h}^{xy}}{\Sigma_{h,h}^{yy}} \end{aligned} \tag{29}$$

References

Andersen, T. G., Bollerslev, T. and Diebold, F. X. (2010). Parametric and nonparametric volatility measurement, *Handbook of Financial Econometrics: Tools and Techniques*, Elsevier, pp. 67–137.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Ebens, H. (2001). The distribution of realized stock return volatility, *Journal of financial economics* **61**(1): 43–76.

Ang, A., Chen, J. and Xing, Y. (2006). Downside risk, *The review of financial studies* **19**(4): 1191–1239.

Ang, A., Hodrick, R. J., Xing, Y. and Zhang, X. (2006). The cross-section of volatility and expected returns, *The Journal of Finance* **61**(1): 259–299.

Barndorff-Nielsen, O. E., Kinnebrock, S. and Shephard, N. (2008). Measuring downside risk-realised semivariance.

Bollerslev, T., Tauchen, G. and Zhou, H. (2009). Expected stock returns and variance risk premia, *The Review of Financial Studies* **22**(11): 4463–4492.

Drechsler, I. and Yaron, A. (2011). What's vol got to do with it, *The Review of Financial Studies* **24**(1): 1–45.

Fama, E. F. and French, K. R. (1988). Dividend yields and expected stock returns, *Journal of financial economics* **22**(1): 3–25.

Feunou, B., Jahan-Parvar, M. R. and Okou, C. (2018). Downside variance risk premium, *Journal of Financial Econometrics* **16**(3): 341–383.

Harvey, C. R. and Siddique, A. (2000). Conditional skewness in asset pricing tests, *The Journal of Finance* **55**(3): 1263–1295.

Jondeau, E. and Rockinger, M. (2003). Conditional volatility, skewness, and kurtosis: existence, persistence, and comovements, *Journal of Economic dynamics and Control* **27**(10): 1699–1737.

Jurado, K., Ludvigson, S. C. and Ng, S. (2015). Measuring uncertainty, *American Economic Review* **105**(3): 1177–1216.

Kilic, M. and Shaliastovich, I. (2019). Good and bad variance premia and expected returns, *Management Science* **65**(6): 2522–2544.

Markowitz, H. (1952). Portfolio selection, *The journal of finance* **7**(1): 77–91.

Merton, R. C. (1973). Theory of rational option pricing, *The Bell Journal of economics and management science* pp. 141–183.

Patton, A. J. and Sheppard, K. (2015). Good volatility, bad volatility: Signed jumps and the persistence of volatility, *Review of Economics and Statistics* **97**(3): 683–697.

Roy, A. D. (1952). Safety first and the holding of assets, *Econometrica: Journal of the Econometric Society* pp. 431–449.

Segal, G., Shaliastovich, I. and Yaron, A. (2015). Good and bad uncertainty: Macroeconomic and financial market implications, *Journal of Financial Economics* **117**(2): 369–397.

Stambaugh, R. F. (1999). Predictive regressions, *Journal of Financial Economics* **54**(3): 375–421.

Tversky, A. and Kahneman, D. (1979). Prospect theory: An analysis of decision under risk, *Econometrica* **47**(2): 263–291.

Valkanov, R. (2003). Long-horizon regressions: theoretical results and applications, *Journal of Financial Economics* **68**(2): 201–232.