Investment in Value:
A Copula Approach

Gustaf Soldan Patrikson
22107@student.hhs.se

Victor Andrée
22584@student.hhs.se

Stockholm School of Economics
Master's Thesis in Finance
Fall 2016

Abstract

We evaluate how factor equity strategies are optimally combined, focusing on the role of the value factor (HML) against the background of a recent academic discussion about its potential redundancy, and the discovery of the investment (CMA) and profitability (RMW) factors. The analysis is centered around a conditional joint return distribution from a dynamic copula model, which allows for simulation with a time-varying and non-normal dependence structure. We study portfolios of six of the most common equity factors (market (Mkt.RF), size (SMB), value (HML), investment (CMA), profitability (RMW) and momentum (Mom)) on weekly US data 1963–2016, applying two different optimization strategies: mean-variance and conditional diversification benefit, where the latter is based on expected shortfall. Our results show that HML remains an important factor that increases the Sharpe Ratio and also decreases the tail risk of portfolios. However, HML should only be combined carefully with CMA, as they overlap to some extent. In parallel, we find that RMW is fundamentally different from HML and CMA, and that the factor is significantly more impactful on the risk-reward profile of portfolios.

Supervisor: Michael Halling
Keywords: factor investing, copula, tail dependence, diversification

Acknowledgments: We would like to express our gratitude to Michael Halling for his valuable input and guidance, and to Alexios Ghalanos for his work on the rugarch R package.

Contents

1 Introduction

Fama and French (2015) find that the value factor (HML) is redundant in a five-factor model including investment (CMA) and profitability (RMW), as it has no explanatory power on monthly returns in a US 1963–2013 sample. This could mean that the classic value factor is an inferior proxy for what truly comprises value, and the paper has sparked a debate on whether the HML factor is a poor proxy for the value effect. In response to Fama and French (2015), Asness, Frazzini, Israel, and Moskowitz (2015) resurrect the explanatory power and value of HML in a portfolio context, by including momentum as a factor.

In this paper, we put ourselves in the shoes of an investor who optimizes factor exposures of a portfolio and is curious about what the new factors bring to the table. We consider the six factors most commonly discussed in the literature: market (Mkt.RF), size (SMB), value (HML), investment (CMA), profitability (RMW) and momentum (Mom), as available from Kenneth French's data library (French, 2016). Specifically, we focus on the role of value (HML) in factor investing, and its apparent similarity to investment (CMA), against the background of Fama and French (2015) and Asness et al. (2015). We also consider the impact on risk and reward of including the other new factor, profitability (RMW).

While most factor pairs exhibit relatively low correlation, the HML–CMA pair stands out with an unconditional correlation in our sample (weekly returns 1963–2016) of 0.63. Value firms have been found to invest less, and growth firms to invest more, and there is a negative empirical relation between past investment and current book-to-market ratios (Zhang, 2005; Anderson & Garcia-Feijóo, 2006). Therefore, there is reason to expect a degree of overlap between the portfolios that comprise HML and CMA.

The value premium has been explained both as a rational risk premium, that compensates the bearer for taking on some risk that materializes in bad times, and as an anomaly, that exists due to market frictions or investor irrationality. If the premia earned on these strategies are compensations for risk, the source of risk in the HML and CMA factors might in fact be the same. An investor who allocates to both these factors might unwillingly double up on exposure to the same risk source. Similarly, if the premia earned are due to market frictions and irrational investor behavior, the naïve investor would double up on exposure to the risk that the anomaly goes away. We believe that, regardless of the interpretation of the value premium, there is reason to place additional emphasis on the role of HML relative to CMA from a portfolio choice and risk perspective.

Our main research question can thus be expressed as: What role should value (HML) play in factor investing given the discovery of investment (CMA) and profitability (RMW)?

CMA was included in the five-factor model jointly with RMW. RMW is typically considered an anomaly, as it has proven to be especially hard to rationalize as a risk premium (Wang & Yu, 2013). High profitability is generally indicative of a favorable market position, a strong brand and a host of competitive advantages, which all contribute to lower risk. The reverse holds for the unprofitable firms. Furthermore, RMW has had high positive returns in times of crisis, indicating that the factor shields against risk at times when protection is needed the most. There seem to be strong benefits to including such a factor in a portfolio, and we believe that the RMW factor deserves to be analyzed separately.

Our secondary research question therefore considers: What is the impact on the risk-reward trade-off of including profitability (RMW) in factor portfolios?

To answer these questions, we optimize portfolios based on two different strategies and compare the resulting risk-reward profiles of including HML, CMA and RMW respectively.

Before delving into optimization methods, we revisit the zero-cost regressions of Fama and French (2015) and Asness et al. (2015). In these regressions, each factor is separately regressed on all the remaining factors, to determine whether there is any additional abnormal return to the left-hand side (LHS) factor after accounting for the variation explained by the right-hand side (RHS) factors. In other words, such regressions examine whether the LHS factor is subsumed by the RHS factors, in which case it has no significantly estimated intercept. In fact, the intercept is also equivalent to the Jensen's alpha of including the LHS variable to a portfolio of the RHS factors (Jensen, 1968). Fama and French (2015) find zero alpha of adding HML to a four-factor portfolio, and conjecture that this implies that a mean-variance (MV) investor would in fact not improve the tangent portfolio's Sharpe Ratio (SR) by including HML. Asness et al. (2015) challenge the notion that HML factor is subsumed by the addition of CMA and RMW. In their study, they resurrect the alpha of the value factor by adding a momentum factor to the regressions and by modifying the HML factor.

We find similar regression results in our weekly data set of factor returns 1963–2016. However, while the regressions of zero-cost portfolios in Fama and French (2015) and Asness et al. (2015) indicate which factors should be included in MV investing, neither paper actually carries out MV optimization of portfolios including CMA and RMW. This thesis fills that gap by optimizing weights of both five-factor (as in Fama and French (2015)) and six-factor portfolios (including momentum as in Asness et al. (2015)).

Thus, the first optimization method is mean-variance (MV) analysis – a conventional risk-reward perspective, where weights are chosen to maximize the Sharpe Ratio of the portfolio. However, MV analysis only considers the expected returns and covariances of the return distribution. Factor strategies have been shown to be inherently non-normal: they have high levels of skewness and kurtosis and also exhibit tail dependence, i.e. the notion that there might be significantly different dependence patterns when returns simultaneously realize in the lower or upper tail, as opposed to close to the center of the joint distribution (Christoffersen & Langlois, 2013). Means and covariances provide an incomplete description of factor return distributions, and therefore also of the risk in factor portfolios. Regardless of the result of MV analysis, the non-normal features might constitute another reason altogether to include (or exclude) either HML, CMA or RMW.

To analyze the lower tail of the distribution, the second optimization method is based on the *conditional diversification benefit* (CDB) statistic, a measure introduced in Christoffersen, Errunza, Jacobs, and Langlois (2012). CDB is based on Expected Shortfall (ES) and measures the diversification benefit in the lower tail of the distribution. The statistic studies how close a portfolio's ES is to the portfolio's Value-at-Risk (VaR), and provides an additional dimension to the risk-reward trade-off in the MV setting. In optimization, weights are chosen to maximize tail diversification (i.e. maximize the CDB statistic).

While MV analysis could be carried out using static sample estimators of means and covariances alone, CDB analysis is impossible without a model of returns, from which ES and VaR can be derived. Furthermore, a conditional model allows us to study dynamic portfolio weights.

The choice of a return model is central. While the ARMA-GARCH model family is the norm of univariate time-series modeling, multivariate modeling has proven harder as the multivariate extensions of such models are often computationally infeasible and ridden with dimensionality problems. Recently, however, copula models have attracted considerable attention in the risk management field, as they offer a numerically stable and flexible way of estimating joint probability distributions.

Following closely the method of Christoffersen and Langlois (2013), we build a copula model of the joint factor returns. The specification we use is designed to recognize time-varying correlation and tail dependence, which are two important features of factor returns (Christoffersen & Langlois, 2013). We measure time-varying correlation with rolling one-year correlation and tail dependence with threshold correlation (also known as exceedance correlation), i.e. the linear correlation when factors simultaneously realize in the upper or lower tail (Ang & Chen, 2002). In in-sample robustness tests, our copula model is shown to generate the time-varying correlation patterns in the data. It can also, to a limited extent, reproduce the tail dependence.

Based on copula estimates of means and covariances, our MV optimization shows that HML does indeed improve the tangency portfolio's Sharpe Ratio, subject to a constraint of non-negative weights. HML receives an average portfolio weight of 18% and improves the SR by 0.16 in the five-factor model. We therefore agree with the conclusions of Asness et al. (2015), and suggest that the discussion about HML's redundancy is likely to be caused by omitting the momentum factor from the zero-cost regressions in Fama and French (2015).

However, our MV analysis also highlights the risk of over-allocating to the value factor as HML is highly similar to CMA. When HML is included, it mainly cannibalizes on the weight that CMA had before, and vice versa, indicating that the variables proxy for each other to a high extent. This is in line with the theoretical and empirical support for an overlap in the stocks that comprise HML and CMA (Zhang, 2005; Anderson & Garcia-Feijóo, 2006). Investors who do not consider this similarity, for example by equal-weighting the factors, risk over-allocating to the same return premium.

We also find that RMW has a much greater impact on the tangency portfolio than do HML and CMA. When RMW is excluded, the realized Sharpe Ratio falls 0.36, compared to a drop of 0.16 and 0.10 for HML and CMA, respectively, in the five-factor model. This illustrates the unique diversifying nature of RMW, which is not captured or proxied well by the other factors.

We find that all of our MV results are qualitatively similar, albeit less pronounced, when full sample estimators of means and covariances are used instead of model inputs.

Having considered only means and variances in the MV analysis, we shift the focus to the tail of the distribution in the CDB analysis and investigate whether HML and CMA differ substantially in terms of their contribution to tail risk. The CDB is alternately higher for the exclusion of HML or CMA, but no pattern emerges.

Based on the CDB measure of tail risk, there is no reason to remove HML from factor investing. However, we find that excluding either one of HML or CMA has a very modest impact on tail risk. We interpret this as another sign of the overlap between HML and CMA. As in the mean-variance analysis, the factors proxy for each other, making an exclusion less dramatic. Still, we see no reason for investors to choose either one or the other, as both provide valuable diversification, and even better, they do so at different times.

Excluding RMW has much greater impact. The CDB drastically worsens, with substantially greater and more frequent declines. We find that the RMW factor is very efficient in reducing tail risk.

2 Literature review

We review previous literature on equity factor strategies and summarize the discussion of why there are factor premia. We also discuss practical factor investing and the work on modeling factors with copulas.

2.1 The five- and six-factor models

Fama and French (2015) introduce two additional factors to complement the Fama and French (1993) three-factor asset pricing model. In what is referred to as the five-factor model, the traditional factors (market, size and value) are complemented by an investment factor and a profitability factor. Both factors represent zero-cost portfolios: the investment factor, denoted CMA (conservative-minus-aggressive), is long firms with low investment rate and short firms with high investment rate, and the profitability factor, denoted RMW (robust-minus-weak), is long firms with high operating profitability and short firms with low operating profitability.

The five-factor model is found to be a significant improvement (in terms of explaining cross-sectional returns) to the three-factor model, and the two new factors appear to have made the value factor (HML) redundant. More specifically, the authors show that there is no significant intercept in regressions of HML on the remaining four factors, while each of the other factors have significant intercepts in similar regressions. The returns of the HML factor appear to be fully explained by the remaining four factors. In an investment context, this can be interpreted as the HML factor adding no alpha to a portfolio holding the remaining four factors.

Asness et al. (2015) challenge the notion that the value factor is subsumed by the addition of investment and profitability. In their study, they add a momentum factor, as well as an enhanced HML factor, and resurrect the alpha of the value factor in a six-factor model. This leads us to believe that momentum could play an important role in recognizing the effect of HML. The momentum factor was originally studied by Jegadeesh and Titman (1993) and has since been shown to be present in many financial return series (Asness, Moskowitz, & Pedersen, 2013).

The investment and profitability factors have only recently made their way into academic literature. Cooper, Gulen, and Schill (2008) investigate investment, measured as the percentage change in total balance sheet assets, and show that a related zero-cost-portfolio provides significant abnormal returns. Investment also has incremental predictive ability in the cross-section of stock returns, taking both value and size into account.

Novy-Marx (2013) studies the return differences between firms with high and low gross profitability and shows significant abnormal returns to a profitability factor. Profitability is also shown to have approximately the same power in predicting the cross-section of stock returns as does value (HML). Furthermore, the profitability strategy is negatively correlated with value, and can improve the investing performance of a value strategy.

While all other factor pairs exhibit correlations at or below zero, the value (HML) and investment (CMA) factors are highly positively correlated. Zhang (2005) predicts this positive relation in a model setting, and Anderson and Garcia-Feijóo (2006) confirm it on empirical data. More specifically, the empirical study shows that past investment has a significant negative relation with the book-to-market ratio. In other words, value firms (with high book-to-market) might be value firms precisely because they have invested little, and vice versa. Fama and French (2015) consider it a fact that value firms invest less than growth firms.

2.2 Factor premia – anomalies, risk premia, or both?

It is debated whether factor return premia constitute rational risk premia or whether they are the consequences of market imperfections and irrational behavior. There are some appealing rational stories for the return premium of HML, which could also explain the premium of CMA as the factors overlap. Fama and French (1993) show that the HML factor is related to systematic patterns of profitability and growth, and could proxy for a systematic risk source. This is supported by Liew and Vassalou (2000), who show that the value factor can predict real GDP growth on data in several markets. Zhang (2005) uses a neoclassical model with rational expectations and competitive equilibrium to show that value firms have more tangible assets and are burdened by industry over-capacity in downturns, leading to higher down-market betas. Petkova and Zhang (2005) find that the conditional betas of value stocks covary positively with the expected market risk premium. Despite there being a number of rational theories, they all predict effects that are fairly small and cannot fully motivate the value premium. So far, the most pervasive explanations of the value premium are based on market imperfections and irrational behavior.[1]

Lakonishok, Shleifer, and Vishny (1994) argue that the value factor is driven by investors' overreaction to changes in earnings. They show that value firms have often experienced a decline in earnings over the last three years, lowering their book-to-market ratios. When earnings have gone down, investors as a group extrapolate the trend into the future and push prices away from fundamentals, giving rise to higher average returns for value firms and vice versa. Similar to Lakonishok et al. (1994), Barberis and Huang (2001) draw on the fact that value firms have experienced decreasing earnings, but suggest that the premium is driven by investors' loss aversion bias. In this story, current value firms have had falling earnings, which has led to to lower share prices and negative returns. Then, many investors are deterred from value firms by the past performance of negative returns in itself, and are less willing to hold value stocks. This unwillingness to hold value firms create a risk-reward upside for those who do.

For the profitability factor, RMW, risk based explanations are harder to come by (Novy-Marx, 2013). It is hard to pinpoint reasons for profitable firms to be more risky than unprofitable. Wang and Yu (2013) investigate the relationship between macro risks and the profitability factor and find risk based stories to be implausible. Instead they suggest that the RMW factor is driven mainly by systematic underreaction, causing a negative alpha in unprofitable stocks.

2.3 Factor strategy investing

Variations of the long-short factor strategies have become staple strategies of both quantitative and qualitative hedge funds, often under the "equity market neutral" or "fundamental quantitative" labels. Factor equity strategies have also become increasingly accessible for retail investors, especially with the advent of smart beta exchange traded funds (ETFs).[2] A number of large money managers including AQR, BlackRock and Robeco today provide factor investing based products, and MSCI provides indices on factor strategies. Generally, these managers advise against factor timing and use static strategies that are based on equal-weighting – a simple heuristic that has proven hard to beat out-of-sample. The managers blend the equal-weights approach with optimization routines, including mean-variance and minimum-volatility, to arrive at policy weights.[3]

The use of leverage in hedge funds can exacerbate the flow patterns in factor strategies, as highlighted by the quant crash in July-August 2007. Khandani and Lo (2011) and Khandani and Lo (2007) revisit the sudden and large losses of factor strategies (including value and size) during this period, and provide evidence for the "Unwind hypothesis": The crash started with rapid sell-offs of large blocks of factor strategy portfolios, for which there was not enough liquidity to maintain prices. The price drops, in turn, led to further liquidations due to 1) margin calls in other leveraged and long-short funds and 2) risk management policies, even in traditional long-

[1]See i.a. Ilmanen (2011) for a summary.
[2]See i.a. Pedersen (2015), AQR Capital Management, LLC (2015) and McKinsey & Company (2014).
[3]See i.a. AQR Capital Management, LLC (2016), BlackRock (2016), MSCI (2015) and Robeco (2014).

only funds. This liquidity and margin spiral is very similar to that proposed by Brunnermeier (2009) and Brunnermeier and Pedersen (2009), and provides reason to suspect systemic risks in crowded factor strategies.

2.4 Modeling of factor returns

Recently, copula models have attracted much attention in the field of risk management, as they provide a flexible way to infer a multivariate probability distribution. Furthermore, copulas are flexible in the sense that they can capture tail dependence, i.e. when the dependence structure changes in extreme times. Copula models are most often estimated taking popular univariate models such as ARMA-GARCH models as a starting point, and use a copula function to explain the multivariate dependence structure.

There are only a handful of papers that study factor strategies using copula methods. A working paper by Chollete and Ning (2012) examines dynamic correlations between a four factors (market, size, value and momentum) and aggregate US consumption, and find evidence for tail dependence across the five risk factors. Christoffersen and Langlois (2013) study the same four factors on US data 1963-2010, and show significant and asymmetric tail dependence that cannot be captured by standard linear correlation measures. A skewed *t* copula model is found to be able to generate the data fairly well, and the authors proceed with 20 years of out-of-sample analysis on investing based on conditional expectations from the copula model, leading to significant improvements for investors with a CRRA utility function.

3 Data

In the data section, we describe the source of our data and how the factor strategies are constructed. We then present summary statistics including tests for autocorrelation and volatility clustering, as well as quantile-quantile (QQ) plots. Finally, we discuss the unconditional correlations of the factor strategies.

3.1 Data description

We use US data on factor strategies 1963–2016, which we download from Kenneth French's data library (French, 2016).[4] We merge the daily Fama-French five-factor data set with the daily momentum data set. Both are available since 1963-07-01, making 1963-07-05 the first week of data. In our sample, 2016-07-01 is the last data point.

While the related papers Fama and French (2015) and Asness et al. (2015) both use monthly data, we choose the weekly frequency for two main reasons: (1) We take a portfolio perspective rather than an asset pricing perspective, and believe that a more frequent horizon than monthly

[4] Data sets *Fama/French 5 Factors (2x3) [Daily]* and *Momentum Factor (Mom) [Daily]*.

data is relevant for both rebalancing and risk management objectives, and (2) Due to computational limitations, the copula methodology discourages us from going to the daily frequency, as optimizations become significantly more time-consuming.

We proceed with descriptions of how factors are constructed, from French (2016). The Mkt.RF factor is long the value-weighted return of CRSP firms on NYSE, AMEX or NASDAQ with CRSP share codes 10 or 11 and short the one-month Treasury bill rate. The remaining return series are based on zero-cost portfolios that are long certain equities and short other equities, according to a 2 x 3 sort: First, firms are sorted into one of two size groups, small and big, depending on whether the market cap is above or below the median. In the small and big firm groups, each factor then sorts into one of three groups depending on whether the variable of interest falls below the 30$^{\text{th}}$ percentile, between the 30$^{\text{th}}$ and the 70$^{\text{th}}$ or above the 70$^{\text{th}}$. For the six-factor data set, the remaining five factors are:

- High-minus-low (HML), is long firms above the 70$^{\text{th}}$ percentile book-to-market and short stocks below the 30$^{\text{th}}$ percentile, in the small and big firm group respectively.
- Conservative-minus-aggressive (CMA), is long firms above the 70$^{\text{th}}$ percentile total asset growth and short firms below the 30$^{\text{th}}$ percentile.
- Robust-minus-weak (RMW), is long firms above the 70$^{\text{th}}$ percentile operating profitability and short firms below the 30$^{\text{th}}$ percentile.
- Small-minus-big (SMB), is long firms below the 50$^{\text{th}}$ percentile market cap and short firms above the 50$^{\text{th}}$ percentile, in each of the three groups HML, CMA and RMW.
- Momentum (Mom), is long firms above the 70$^{\text{th}}$ percentile prior 2-12 month return (i.e. excluding the last month) and short stocks below the 30$^{\text{th}}$ percentile, in the small and big firm group respectively.

The sort ensures that SMB includes firms small and big firms equally from the remaining factors, and that the other factors include equal amounts of small and big firms. Note that momentum originates from a different data set and does not affect the SMB composition. French's financial statement data originates from Compustat, stock return data is from CRSP and Treasury return data is from Ibbotson Associates.

3.2 Summary statistics

For all factors, there are 2766 consecutive data points and no missing data. Mkt.RF is the most volatile and extreme of the series, with weekly returns between -18.0% and 13.5% and a weekly volatility of 2.2%. CMA and RMW seem to be less extreme than HML, with less negative minimums and smaller volatilities. The factor strategies have excess kurtosis, or fat tails, which is typical for financial returns. The excess kurtoses of the Mom, HML and RMW factors are even

Table 1: Summary statistics of weekly factor returns

Kurtosis is excess kurtosis. *LB* is the weighted Ljung-Box test up to 5 and 10 lags respectively, where the null hypothesis is no serial correlation (Fisher & Gallagher, 2012). Based on weekly data 1963–2016.

	Mkt.RF	SMB	HML	CMA	RMW	Mom
Mean (%)	0.12	0.04	0.08	0.07	0.06	0.16
SD (%)	2.19	1.21	1.23	0.88	0.88	1.84
Maximum (%)	13.46	6.13	12.46	5.53	9.87	12.79
Minimum (%)	−18.00	−9.34	−7.93	−4.28	−5.99	−16.04
Median (%)	0.27	0.06	0.05	0.04	0.05	0.23
Skewness	−0.44	−0.38	0.50	0.37	0.92	−1.07
Kurtosis	5.27	4.50	7.95	3.24	14.18	10.19
p-values of Ljung-Box (LB) tests on returns and squared returns						
Returns LB(5)	0.34	0.00	0.00	0.00	0.00	0.00
Returns LB(10)	0.01	0.00	0.00	0.00	0.00	0.00
Squared returns LB(5)	0.00	0.00	0.00	0.00	0.00	0.00
Squared returns LB(10)	0.00	0.00	0.00	0.00	0.00	0.00
Correlations	Mkt.RF	SMB	HML	CMA	RMW	Mom
Mkt.RF	1.00					
SMB	0.08	1.00				
HML	−0.28	−0.03	1.00			
CMA	−0.42	−0.05	0.63	1.00		
RMW	−0.15	−0.34	−0.05	−0.06	1.00	
Mom	−0.12	−0.01	−0.22	0.07	0.08	1.00

higher than the kurtosis of the Mkt.RF factor, where RMW is exceptionally high at 14.8. However, while market returns are negatively skewed, HML, CMA and RMW instead exhibit positive skewness. QQ-plots versus normal theoretical quantiles in Figure 1 graphically show the non-normality.

We conduct Ljung-Box tests of the factor returns to control for weekly autocorrelation.[5] The p-values of these tests are given in Table 1 and are very low for all factors except Mkt.RF, leading to a strong rejection of the zero autocorrelation null hypothesis. For Mkt.RF, the p-value is not small enough for a rejection of zero autocorrelation at the 5 week maximum lag length, but strongly rejected at the 10 week maximum lag length. We also conduct Ljung-Box tests of the squared factor returns to control for volatility clustering (ARCH effects). Here, the null hypothesis is that there are no ARCH effects, and p-values given in Table 1 strongly reject the null for all factors, both at max lag length of 5 and 10 weeks.[6]

We conclude that factor return series are non-normal and that returns are not independently distributed over time – more specifically, past returns have predictive power on future returns, and past volatility has predictive power on future volatility, i.e. the series exhibit both autocorrela-

[5] For a detailed description of the test, see Appendix C.

[6] The lag lengths were chosen after visual inspection of autocorrelation function plots.

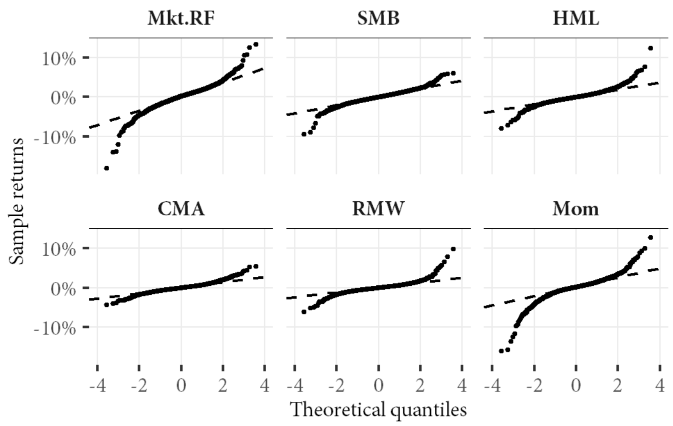


Figure 1: QQ-plots of return series

Data from a normal distribution should line up on the dashed line. Based on weekly returns 1963–2016.

tion as well as autoregressive heteroscedasticity. These predictable phenomena in financial return data are typically captured by models that incorporate autoregressive components for both the conditional mean and variance equations, such as the family of ARMA-GARCH models, which is further discussed in section 5.

A plot of cumulative gross returns (Figure 2) clearly show the high returns to the momentum strategy throughout the sample period. Taking out momentum, Figure 3 shows that the Mkt.RF factor has the second highest cumulative gross return, but also that it is the most volatile of the remaining strategies. We normalize the series to 10% annual volatility in Figure 4, which gives a more nuanced picture of risk-adjusted performance. Since 1963, each of the strategies except for SMB has outperformed the market factor. Furthermore, factor strategies seem to crash at different times and diversify each other (e.g. Mom performed well during the bubble of 1999–2000 and RMW performed well during the recession of 2007–2009).

Looking at the correlation matrix of factor returns, the generally low or even negative correlation coefficients indicate the diversification benefits of factor strategies. The HML–CMA pair does stand out, however, with an unconditional correlation of 0.63, which could be related to a partial overlap of the factor components, as discussed in section 2 – past investment is shown to be negatively empirically related to the current book-to-market ratio. The substantially higher correlation in this asset pair indicates smaller diversification benefits. We also note that the new factor RMW has an interesting pattern of low of negative correlations to all other factors, which indicates diversification benefits.

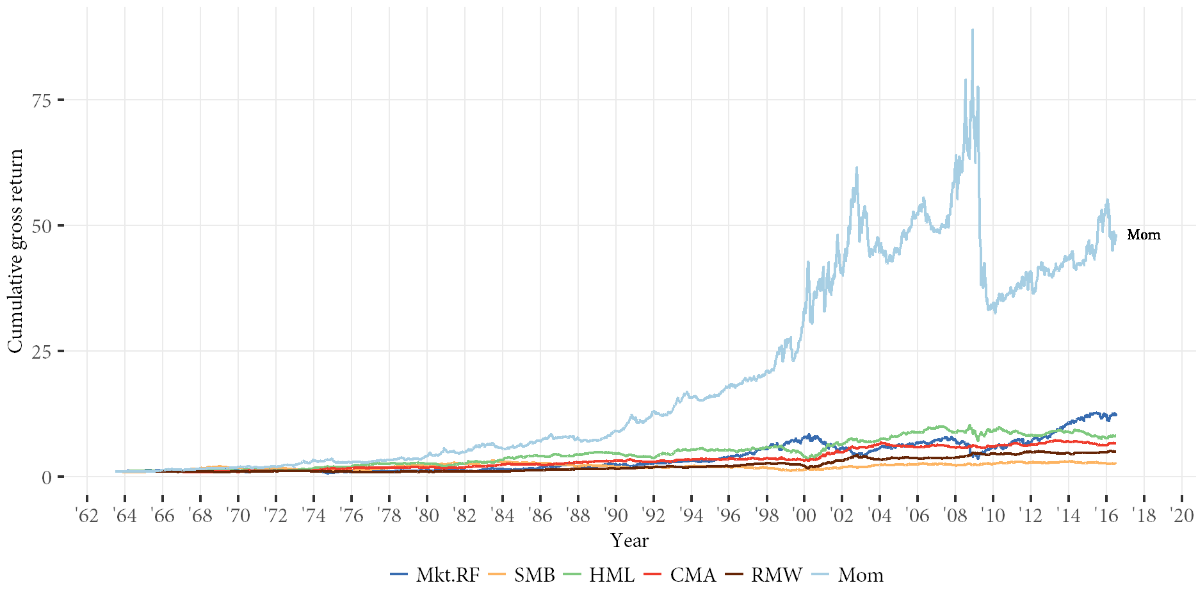


Figure 2: Cumulative returns to factor strategies

Cumulative returns to investing one dollar in each factor strategy, beginning 1963-07-05. Based on weekly returns 1963–2016.

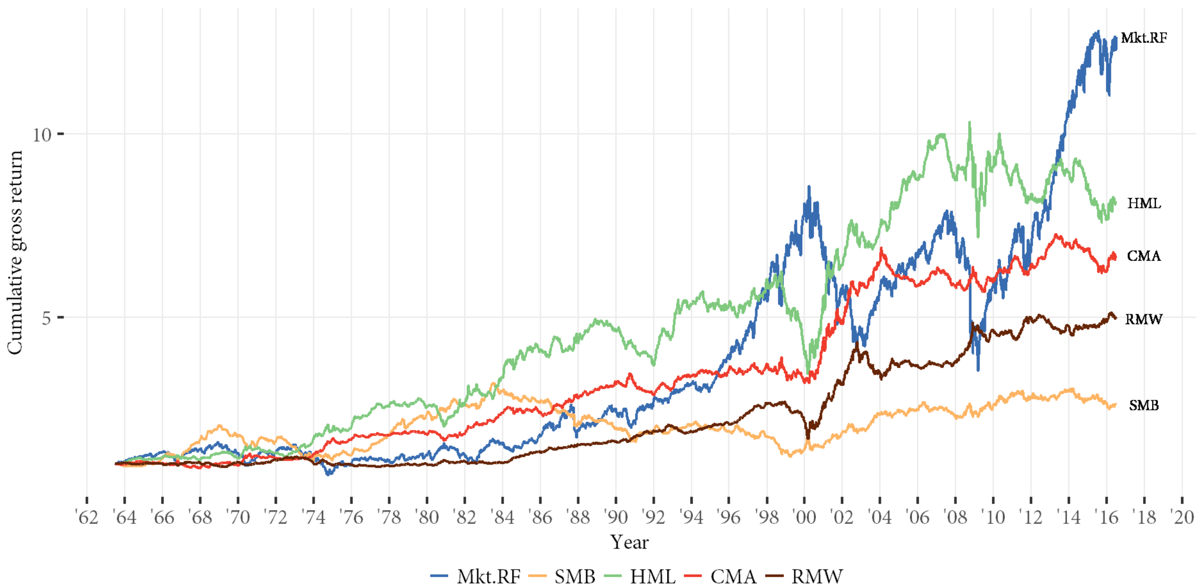


Figure 3: Cumulative returns to factor strategies, excl. momentum

Cumulative returns to investing one dollar in each factor strategy, beginning 1963-07-05. Based on weekly returns 1963–2016.

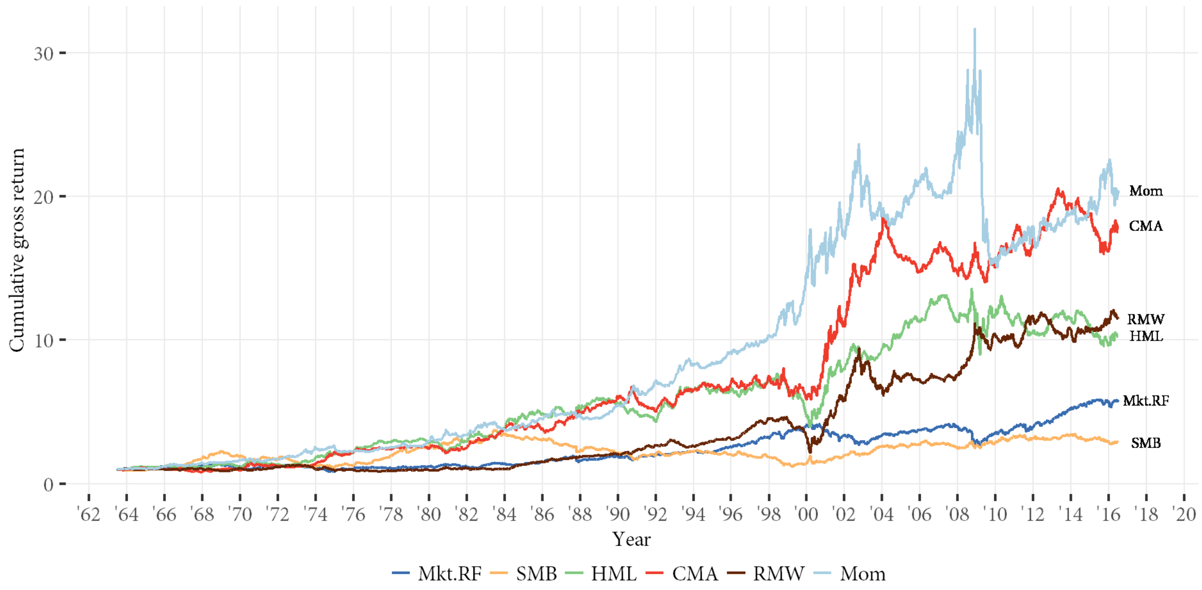


Figure 4: Standardized cumulative returns to factor strategies

Cumulative returns to investing one dollar in each factor strategy, standardized to 10% annual volatility, beginning 1963-07-05. Based on weekly returns 1963–2016.

4 Zero-cost portfolio regressions

Fama and French (2015) and Asness et al. (2015) run factor regressions where both the LHS variable and the RHS variables are zero-cost factor portfolios. The intercept in this type of regression is to be interpreted as the abnormal return, or Jensen's alpha, of adding the LHS factor to a portfolio already consisting of the RHS factors (Jensen, 1968). In this section, we replicate and discuss the regressions where HML, CMA and RMW are the LHS variables, and find that previous results hold up in our weekly data set.

4.1 Model specification

As a specific example, we begin by considering the regression that has caused the discussion on whether HML is a redundant factor. Fama and French (2015) run the regression

$$r_t^{HML} = \alpha + \beta_1 r_t^{Mkt.RF} + \beta_2 r_t^{SMB} + \beta_3 r_t^{RMW} + \beta_4 r_t^{CMA} + \varepsilon_t \tag{4.1}$$

where r_t^i denote monthly returns. The central finding is that HML is completely subsumed by the four factors Mkt.RF, SMB, RMW and CMA – i.e. the alpha of the regression is very small and not statistically significant. In other words, adding HML to a portfolio of the other four factors should give no abnormal return.

Our regression analysis deviates from that in Fama and French (2015) as we consider weekly return data and approximately two more years of recent data. However, the main regression specifications are the same. We also consider a six-factor model, as done by Asness et al. (2015), who show that there is in fact added value of HML when momentum is included. We use standard errors that are adjusted for serial correlation found in the return data, following Newey and West (1987).

4.2 Regression results

In Table 2, regressions for the five-factor (excluding momentum) and six-factor (including momentum) models are presented. Each column represents one unique regression, with one of the factors as the LHS variable and with the remaining four (or five) factors as RHS variables (in rows).

First, we examine regression (1) in a five-factor model where HML is the LHS variable. We note that the alpha of HML is not significant, indicating that the factor is completely subsumed by the remaining four factors and does not create additional value in a portfolio setting, in line with Fama and French (2015). More specifically, the only factor that explains HML is CMA, with a high coefficient of 0.85, with all other factor loadings insignificant and close to zero. Put differently, this suggests that for a factor portfolio already loaded on Mkt.RF, SMB, CMA and RMW, adding HML will load up additionally on CMA risk plus the idiosyncratic risk of HML, without adding any additional return.

Second, we turn to regression (2) in a five-factor model where CMA is the LHS variable. Here, the alpha is significant, indicating that the factor does provide an additional 0.06% weekly beyond the existing four factors. While the CMA factor loads positively 0.39 on HML, this is substantially lower than HML's loading on CMA of 0.85. With CMA as the LHS variable, there is a significant negative loading on Mkt.RF and a significant negative loading on RMW. The CMA portfolio loads relatively less on the market and relatively more on unprofitable stocks than does HML.

Third, we study regression (3) where RMW is the LHS variable. The alpha is significant, indicating that the RMW factor adds abnormal return of 0.09% weekly to the four existing factors. In terms of factor loadings, profitability loads zero or negatively on all four remaining factors in the five-factor model. This is evidence of the diversification that RMW provides. The low explanatory power of the other factors is summarized by a low 15% R-squared.

Now, we move to the six-factor regression results, where we include the momentum factor Mom. First, in regression (4) with HML as the LHS variable, we note that the addition of HML makes the alpha of HML positive and significant – in line with Asness et al. (2015). As momentum is correlated with both the LHS and RHS factors, it constitutes an omitted variable bias on the beta factor loadings in the five-factor model. HML has a substantial negative loading on the Mom factor of -0.18, while the CMA regression instead has a positive Mom loading of 0.09, suggesting that the seemingly similar factors HML and CMA are quite different in terms of momentum properties. The momentum factor explains an additional 8% of the variance in the HML factor.

Table 2: Zero-cost portfolio regressions (1963–2016)

Six regressions of zero-cost equity factor portfolios on 2766 weekly returns 1963–2016, following the analysis of Fama and French (2015) and Asness, Frazzini, Israel, and Moskowitz (2015). Alpha and Beta (factor loadings) of the column's portfolio on other factors. Heteroskedacity and autocorrelation robust standard errors in parentheses, following Newey and West (1987). Significance given by $^{*}p < 10\%$; $^{**}p < 5\%$; $^{***}p < 1\%$

	Five factor universe			Six factor universe		
	(1) HML	(2) CMA	(3) RMW	(4) HML	(5) CMA	(6) RMW
Alpha (%)	0.02	0.06^{***}	0.09^{***}	0.05^{**}	0.04^{***}	0.09^{***}
	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)
Mkt.RF	−0.02	-0.11^{***}	-0.08^{***}	−0.03	-0.10^{***}	-0.07^{***}
	(0.03)	(0.02)	(0.01)	(0.03)	(0.01)	(0.01)
SMB	0.00	−0.03	-0.24^{***}	0.01	-0.03^{*}	-0.24^{***}
	(0.03)	(0.02)	(0.04)	(0.03)	(0.02)	(0.05)
HML		0.39^{***}	−0.01		0.42^{***}	0.01
		(0.04)	(0.06)		(0.03)	(0.06)
CMA	0.85^{***}		-0.15^{**}	0.87^{***}		-0.17^{***}
	(0.04)		(0.07)	(0.04)		(0.06)
RMW	−0.02	-0.09^{**}		0.01	-0.10^{**}	
	(0.09)	(0.04)		(0.07)	(0.04)	
Mom				-0.18^{***}	0.09^{***}	0.03
				(0.04)	(0.02)	(0.03)
R^2	0.39	0.46	0.15	0.47	0.49	0.15

CMA and RMW are to a lesser extent than HML correlated with Mom, and the factor loadings in regressions (5) and (6) therefore change less as we go to the six-factor model.

Although we employ weekly data, our results are qualitatively similar to the results in Fama and French (2015) as well as in Asness et al. (2015). The alpha of HML is only recognized in a model including momentum. This indicates that the insignificant alpha of HML in the five-factor model might be due to the omission of an important control variable, momentum, that is included in the six-factor-model.

5 Modeling of factor returns

This section presents our model for the joint behavior of returns. A multivariate model of returns allows us to make conditional forecasts of the distribution of returns one week ahead, which take into account the dependence between factors. In the mean-variance analysis, the model is used to provide dynamic inputs that give us optimal weights over time. The model is also used in the analysis of diversification benefits, where we shift the focus to the tail risk of factor portfolios. First, we describe why we choose the copula model among different multivariate models. Second, we define the model and interpret the parameterization. Third, we select and estimate univariate models that are building blocks of the copula. Fourth, we analyze residuals from univariate models, to determine what type of multivariate dependence the copula should capture. Fifth, we estimate the copula and choose the best specification. Last, we conduct a robustness check of whether the copula captures the dependence structure.

5.1 Choosing a multivariate model

The ARMA-GARCH family of models has become the norm of modeling univariate financial return series, beginning with Bollerslev (1986). The straightforward extension of univariate GARCH models to multiple return series has, however, proven difficult. Unrestricted multivariate GARCH (MGARCH) models that directly model the conditional covariance matrix become impossible to estimate, as the number of covariances grows exponentially with the number of series. It thus becomes necessary to restrict the model's parameter space, of which the BEKK model is a common example (Engle & Kroner, 1995).

A more parsimonious solution to the dimensionality problem is to separate the modeling of return and volatility dynamics from the modeling of conditional correlations. The separation allows for consistent (albeit inefficient) two-step estimation, and makes large-scale estimation feasible. One such approach is the *dynamic conditional correlation* (DCC) model, originally proposed by (Engle, 2002). In the DCC model, univariate GARCH models are first estimated on each series. Then, an autoregressive process for the correlation matrix is fitted to the standardized residuals z_t of those models.

DCC is a useful and tractable model for estimating time-varying correlations between return series. However, it is a model of correlations only; it is not flexible enough to model the univariate components differently. More specifically, it is not constructed to generate tail dependence, which is the notion that correlation dynamics can be very different in extreme realizations.

Copula models have recently attracted much attention in the field of risk management, as they provide a flexible way to infer a multivariate probability distribution. Copula models are, just like DCC models, based on two-step estimation and work well in large scale applications. Furthermore, copulas are flexible enough to generate tail dependence, which is shown to be an important feature of factor return data (Christoffersen & Langlois, 2013).

Copula models are most often constructed by estimating univariate models from the ARMA-GARCH family in the first step. The residuals from the ARMA-GARCH models are then used in the copula function, which explains the multivariate dependence, including dynamic correlations and tail dependence.

Among copula models, there are three main routes of interest: (1) Archimedean copulas, (2) multivariate normal and Student's *t* copulas, and (3) vine copulas. While Archimedean copulas, such as the Gumbel and Clayton specifications, are attractive in many settings, they fail to generate both high threshold correlations and simultaneously low unconditional correlations, and are hard to generalize beyond the bivariate case – making them less attractive for factor return series (Christoffersen & Langlois, 2013). Vine copulas, or pairwise copula constructions, are made up of a combination of bivariate copulas in a tree structure (hence the name vine copulas), and pose an interesting alternative to multivariate normal and Student's *t* copulas. However, the vine method is far less parsimonious as both the number of bivariate combinations and the number of different tree structures increases exponentially with the number of assets modeled (Aas, Czado, Frigessi, & Bakken, 2009).

We choose to work with multivariate Student's *t* copula models, as they can (1) estimate the joint distribution function in large scale applications, (2) model different univariate models for the different factors, and (3) incorporate both tail dependence and dynamic correlations. Next, we define and describe the copula model.

5.2 Definition of copula model

Each week t, the conditional joint density of returns on N assets $R_{t+1} = \{r_{i,t+1}, \dots, r_{N,t+1}\}$ is described by a joint density function $f_t(R_{t+1})$. Following Christoffersen et al. (2012), who build on Patton (2006) and Sklar (1959), we decompose the joint density function into the product of a joint copula function $c_t(U_{t+1})$ of uniformly distributed variables $U_{t+1} \sim U(0,1)$ and marginal densities $f_{i,t}(r_{i,t+1})$:

$$f_t(R_{t+1}) = c_t(U_{t+1}) \prod_{i=1}^{N} f_{i,t}(r_{i,t+1}) \tag{5.1}$$

The elements of $U_{t+1} = \{u_{i,t+1}, \ldots, u_{N,t+1}\}$ are related to the original returns by the probability integral transform, i.e the cumulative distribution of $r_{i,t+1}$:

$$u_{i,t+1} = F_{i,t}(r_{i,t+1}) = \int_{-\infty}^{r_{i,t+1}} f_{i,t}(r)dr \tag{5.2}$$

The copula function $c_t(U_{t+1})$ is a multivariate skewed t distribution. This distribution is parameterized by a single degrees of freedom parameter ν_c, controlling the degree of dependence, a vector of N skewness parameters γ_c, controlling the asymmetry in dependence, and a potentially time-varying correlation matrix Ψ_t.[7] The skewed t distribution nests the standard (hereafter symmetric) t distribution when all $\gamma_{i,c} = 0$ and the standard normal distribution when additionally $\nu_c = \infty$.

The log-likelihood of the model is constructed from Equation 5.1:

$$L = \underbrace{\sum_{t=1}^{T} \log(c_t(U_{t+1}))}_{\text{Copula}} + \underbrace{\sum_{t=1}^{T}\sum_{i=1}^{N} \log(f_{i,t}(r_{i,t+1}))}_{\text{Marginals}} \tag{5.3}$$

At this point, it is worth noting that the joint density $c_t(U_{t+1})$ need not be of the same family as the marginal densities $f_{i,t}(r_{i,t+1})$ – nor are we restricted to modeling $f_{i,t}(r_{i,t+1})$ jointly for all factors. In fact, we take advantage of this flexibility and choose to model the marginal densities independently as ARMA-GARCH processes, which allows us to capture a number of predictable features in the univariate series – serial correlation, volatility clustering and leverage effects. The marginal models are estimated independently by maximizing the likelihood(s) of the second term, and then the copula is estimated by maximizing the first term – using the residuals of the marginal models as given.

This procedure is called multi-stage maximum log-likelihood or inference functions for margins and greatly simplifies the estimation procedure, while yielding relatively efficient estimates (Patton, 2006; Joe, 1997). The modeling and estimation of our ARMA-GARCH models is detailed in the upcoming subsection, whereas the remainder of this subsection describes how we make the correlation matrix Ψ_t, and thus the dependence between factors, dynamic.

The copula is made dynamic by fitting a dynamic conditional correlation (DCC) process for Ψ_t to copula residuals z^*_{t+1} (Engle, 2002). Using the notation from Christoffersen and Langlois (2013):

$$Q_t = (1 - \alpha - \beta)Q + \beta Q_{t-1} + \alpha \bar{z}^*_{t-1}\bar{z}^{*\top}_{t-1} \tag{5.4}$$

[7] We describe the details of the skewed t distribution, including the expanded form of c_t, in Appendix A.

where Q_t is normalized to the correlation matrix Ψ_t:

$$\Psi_t = Q_t^{-\frac{1}{2}} Q_t Q_t^{-\frac{1}{2}} \tag{5.5}$$

The Q_t process is comprised of three components that are weighted according to α, β: (1) a time-invariant component: Q, (2) an innovation component from copula shocks: $\bar{z}^*_{t-1}\bar{z}^{*\top}_{t-1}$,[8] and (3) an autoregressive component of order one: Q_{t-1}. In order for the the correlation matrix Ψ_t to be positive definite, Q_t has to be positive definite, which is ascertained by requiring that $\alpha \geq 0$, $\beta \geq 0$ and $(\alpha + \beta) < 1$. The model nests a constant copula when $\alpha = \beta = 0$.

The model for $c_t(U_{t+1})$ is comprised of $1 + N$ distribution parameters $\{\nu_c, \gamma_c\}$ and $2 + \frac{N(N-1)}{2}$ dynamics parameters $\{\alpha, \beta, Q\}$, where the elements of Q are estimated using moment matching, and the remaining parameters $\{\alpha, \beta, \nu_c, \gamma_c\}$ are estimated using maximum likelihood.[9]

ARMA-GARCH modeling allows us to filter time-varying effects, leaving independent *standardized residuals* $z_{i,t}$, which are assumed to follow a constant distribution $f_i(z_{i,t})$. These residuals are first transformed into uniform variables $u_{i,t+1}$ by the probability integral transform of the densities above, and then made to follow the *copula* distribution by the *inverse* probability integral transform of the *copula*:

$$z^*_{i,t+1} = F^{-1}_{\nu_c,\gamma_{i,c}}(F_i(z_{i,t+1})) \tag{5.6}$$

The interpretation of the copula parameterization is closely associated to the structure of multivariate dependence. By different restrictions on the parameters in the DAC model, we are able to activate or deactivate certain features of the copula: First, the degree of freedom parameter ν_c is to be interpreted as the measure of tail dependency. When $\nu \neq 0$, the lower and upper tails of the joint distribution are fatter than in the normal case, which is coherent with earlier evidence of threshold correlations (Christoffersen & Langlois, 2013). Second, the skewness parameters $\gamma_{c,i}$ are to be interpreted as the extent of asymmetry in the correlation structure. When $\gamma \neq 0$, there is asymmetry in correlations. Third, the α and β parameters determine whether the copula generates time-varying correlations. If $\alpha \neq 0$ and $\beta \neq 0$, the copula is dynamic. An overview of the six copula models is given in Table 3.

5.3 Univariate modeling of returns

We proceed by estimating models of each factor's return series, which attempt to capture predictable autocorrelation, volatility clustering and leverage effects. By fitting ARMA-GARCH models, we can filter these effects and reduce the time-varying densities $f_{i,t}(r_{i,t+1})$ to constant densities of *standardized residuals* $f_i(z_{i,t+1})$.

[8]Where $\bar{z}^*_{i,t+1} = z^*_{i,t+1}\sqrt{q_{ii,t}}$ is due to a correction by Aielli (2013), that improves the reliability of the estimation procedure.

[9]A detailed description of the copula estimation procedure can be found in Appendix B.

Table 3: Conceptual matrix of copula parameterizations

		Normal	Symmetric t	Skewed t
		$\nu_c = \infty$ $\gamma_{i,c} = 0$	$\nu_c < \infty$ $\gamma_{i,c} = 0$	$\nu_c < \infty$ $\gamma_{i,c} \neq 0$
Constant	$\alpha = 0$ $\beta = 0$	Constant normal	Constant symmetric t	Constant skewed t
Dynamic	$\alpha > 0$ $\beta > 0$	Dynamic normal	Dynamic symmetric t	Dynamic skewed t

5.3.1 General univariate model: ARMA-GJR-GARCH

The ARMA-GARCH is a broad model family designed to model predictable components of financial return series, and was originally introduced by Bollerslev (1986). The models use autoregressive and moving average lags to capture serial correlation in return data (ARMA), as well as autoregressive and moving average lags to capture ARCH effects in residuals from the mean equation (GARCH). We evaluate the GJR-GARCH model of Glosten, Jagannathan, and Runkle (1993), which is a parsimonious extension of the standard GARCH(1, 1). The GJR-GARCH is designed to also capture leverage effects (Glosten et al., 1993), i.e. when positive and negative return shocks have different impact on future volatility (Black, 1976).

We estimate conditional mean equations for each factor *up to* ARMA(3, 3):

$$r_{i,t} = \varphi_{i,0} + \sum^{p} \varphi_{i,p} r_{i,t-p} + \sum^{q} \theta_{i,q} \varepsilon_{i,t-q} + \varepsilon_{i,t} \tag{5.7}$$

where $r_{i,t}$ are weekly returns of each factor. The conditional volatility evolves according to the GJR-GARCH specification:

$$\varepsilon_{i,t} = \sigma_{i,t} z_{i,t} \tag{5.8}$$

$$\sigma^2_{i,t} = \omega_i + (\alpha_i + \eta_i I_{\varepsilon_{i,t-1} \leq 0})\varepsilon^2_{i,t-1} + \beta_i \sigma^2_{i,t-1} \tag{5.9}$$

where I is an indicator function that is equal to one when $\varepsilon_{i,t-1} \leq 0$.

A positive η_i captures the leverage effect by increasing the current period's volatility if the previous period's residual $\varepsilon_{i,t-1}$ was below zero. A significant η_i thus introduces asymmetric volatility in the model. For the market factor, it is expected that η_i is positive, reflecting the leverage effect in the market itself, and no impact from the short risk-free component. However, for the other factors, which are constructed as all-equity zero-cost long-short portfolios, the direction of η_i is less obvious (Christoffersen & Langlois, 2013). If there are leverage effects for stocks in general, negative shocks will lead to more volatility than positive shocks in a portfolio of stocks. But in a zero-cost portfolio, the leverage effects of the long positions in stocks could be offset by the short positions in other firms. The level of the leverage effect in a zero-cost portfolio therefore depends on the relative strength of leverage effects in the long and short components.

The ARMA-GARCH models are estimated independently on each series using maximum likelihood estimation, with assumed distributions of standardized residuals $z_{i,t}$. Similar to the multivariate copula, we evaluate models where the standardized residuals are assumed to follow univariate skewed t distributions with v_i degrees of freedom and skewness γ_i, nesting the symmetric t when $\gamma_i = 0$ and the standard normal when $v_i = \infty$. A skewed t distribution allows for additional asymmetry beyond the GJR-GARCH leverage effect (Christoffersen et al., 2012).

5.3.2 Factor specific model selection process

Our selection process is as follows.

(i) For each factor strategy, we estimate GJR-GARCH models on the full dataset ($T = 2,766$) up to ARMA(3, 3) and GARCH(1, 1) under normal, symmetric t and skewed t residuals, with and without η_i fixed to zero (in which case we obtain the basic GARCH(1, 1) model).

(ii) We then compute the Bayesian Information Criterion (Schwarz, 1978, BIC) for each factor strategy and specification and select the ARMA order with the lowest BIC as our primary candidates.

For the candidate models

(i) We check for remaining serial correlation and ARCH effects using weighted portmanteau tests.

(ii) We examine whether a sign bias test concludes that there are significant leverage effects that warrant the use of a GJR-GARCH instead of a standard GARCH.

(iii) We use QQ-plots to control for misspecification in the residual process, and to find a suitable distribution for the standardized residuals z_t.

In a well-specified model, we expect there to be no significant serial correlation, ARCH effects or leverage effects in the residuals. We employ weighted Ljung-Box, ARCH LM and sign bias tests that are detailed in Appendix C. Furthermore, the QQ-plots of the standardized residuals should show that their empirical distribution is comparable to the assumed theoretical distribution (i.e. be distributed around the 45 degree line).

5.3.3 Model selection and estimation results

The result of our selection and estimation procedure are presented in Table 4. The Mkt.RF factor is the only model that requires a GJR-GARCH ($\eta_i \neq 0$), while the remaining models are all standard GARCH(1, 1). The minimization of BIC leads to ARMA(0, 0) for Mkt.RF, ARMA(1, 0) for CMA and Mom, and ARMA(1, 1) for the remaining factors SMB, HML and RMW.

Table 4: ARMA-GARCH parameter estimates (1963–2016)

Models from Equation 5.7 and Equation 5.9 on 2766 weekly returns 1963–2016, with skewed t innovations. Robust standard errors in parentheses, following White (1982). ω is set using variance targeting, following Engle and Mezrich (1995). Ljung-Box and ARCH-LM tests are the weighted portmanteau tests from Fisher and Gallagher (2012) and the sign bias test is from Engle and Ng (1993) (see Appendix C). Significance given by $^{*}p < 10\%$; $^{**}p < 5\%$; $^{***}p < 1\%$

	Mkt.RF	SMB	HML	CMA	RMW	Mom
μ (%)	0.12^{***}	0.03	0.06^{**}	0.04^{***}	0.05^{***}	0.14^{***}
	(0.03)	(0.03)	(0.03)	(0.01)	(0.02)	(0.05)
φ_1		0.77^{***}	0.72^{***}	0.11^{***}	0.59^{***}	0.13^{**}
		(0.05)	(0.08)	(0.02)	(0.19)	(0.05)
θ_1		-0.65^{***}	-0.61^{***}		-0.47^{**}	
		(0.06)	(0.09)		(0.21)	
α	0.03^{**}	0.11^{***}	0.11^{***}	0.09^{***}	0.08^{***}	0.18^{***}
	(0.01)	(0.02)	(0.00)	(0.00)	(0.00)	(0.01)
β	0.85^{***}	0.84^{***}	0.87^{***}	0.90^{***}	0.92^{***}	0.80^{***}
	(0.01)	(0.04)	(0.00)	(0.00)	(0.00)	(0.01)
η	0.19^{***}					
	(0.02)					
ν	13.25^{***}	11.56	10.20^{***}	11.09^{***}	10.95^{*}	13.36
	(3.51)	(12.08)	(2.67)	(3.48)	(5.65)	(40.04)
γ	-2.36^{***}	-0.65	0.63	0.48	0.25	-2.12
	(0.84)	(1.04)	(0.41)	(0.49)	(0.42)	(9.90)
ω (ppm)	15.52	6.25	2.67	0.99	0.61	6.94
Log-likelihood (LLH), unconditional volatility (UV) and variance persistence (VP)						
LLH	7,051	8,567	8,788	9,574	9,872	7,941
UV (%)	2.19	1.20	1.22	0.88	0.87	1.83
VP (%)	96.75	95.66	98.20	98.71	99.18	97.93
p-values of Ljung-Box (LB), ARCH-LM (ARCH) and sign bias tests						
LB(5)	0.15	0.27	1.00	0.18	1.00	0.83
LB(10)	0.10	0.72	0.98	0.07	0.06	0.72
ARCH(5)	0.81	0.63	0.12	0.84	0.72	0.05
ARCH(10)	0.93	0.88	0.39	0.95	0.91	0.13
Sign bias(-)	0.88	0.33	0.09	0.47	0.20	0.84
Sign bias(+)	0.16	0.09	0.40	0.07	0.65	0.38

Based on these ARMA-GARCH specifications, the Ljung-Box and LM tests indicate no remaining serial correlation or ARCH effects at a 5% significance level.

The lack of significant sign bias in the GARCH specifications for all models except Mkt.RF is interesting, and in line with the argument that any leverage effects could cancel out in a zero-cost long-short equity portfolio; the Mkt.RF is the only factor that is net-long equities and also exhibited leverage effects, with a negative sign bias as a GARCH model. We note that the sign bias of Mkt.RF has been eliminated in the GJR-GARCH model.

The candidate specifications under normal and symmetric t distributed innovations all display misaligned QQ-plots (see Figure 5). The empirical distributions deviate from the 45 degree theoretical lines, especially in the more extreme quantiles. This indicates asymmetry in the residual series. In unreported results, we have controlled that the misspecification of normal and symmetric t residuals is present even if GJR-GARCH models are fitted for all factors – i.e. leverage effects cannot explain the misspecification. By comparison, the QQ-plots with skewed t innovations seem to fit the data well. We proceed with skewed t residual distributions.

Many of the estimates of γ_i, the skewness of the skewed t GARCH innovation process, are statistically insignificant. This is also the case for the degree of freedom estimates, ν for the SMB and Mom models. Although these parameters are not significantly estimated, we believe that including them is essential, as QQ-plots indicate misspecification for the models with normal and symmetric t innovations.

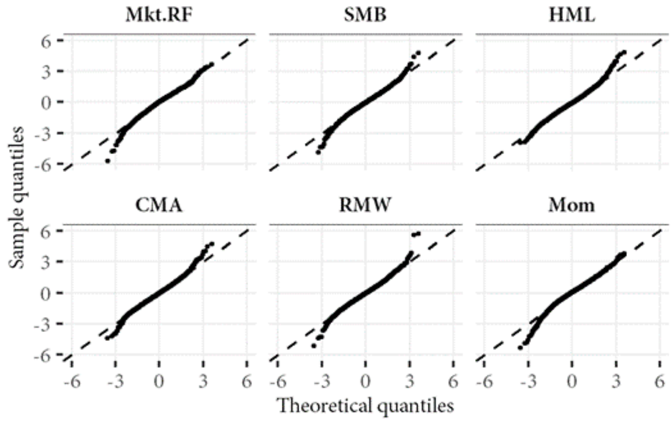
5.4 Multivariate dependence

In this subsection, we demonstrate that the standardized residuals $z_{i,t}$ in our chosen ARMA-GARCH models display both asymmetric and time-varying dependence, shown by threshold and rolling correlations. ARMA-GARCH filtering has little effect on the (time-varying) correlations between factors. However, the use of the skewed t distribution does remove a degree of asymmetry in the dependence. These patterns in multivariate dependence are the motivating reasons for a copula model, as they suggest that factor returns are not independent of each other after filtering for univariate effects, and that this dependence is not well-approximated by a normal model.[10]

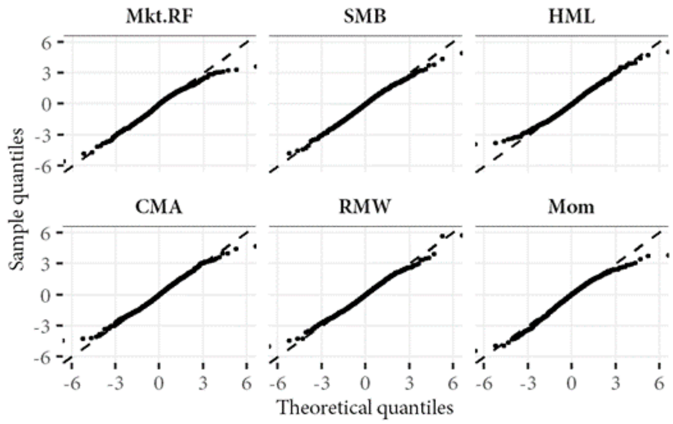
5.4.1 Threshold correlations

Threshold (or exceedance) correlations have previously been used to highlight the asymmetric dependence structure of i.a. country equity indices (Longin & Solnik, 2001), portfolios by industry, size, value and momentum (Ang & Chen, 2002) and factor strategies (Christoffersen & Langlois, 2013). The following analysis is still new as it adds the investment (CMA) and profitability (RMW) factors. We follow Christoffersen and Langlois (2013) definition of threshold

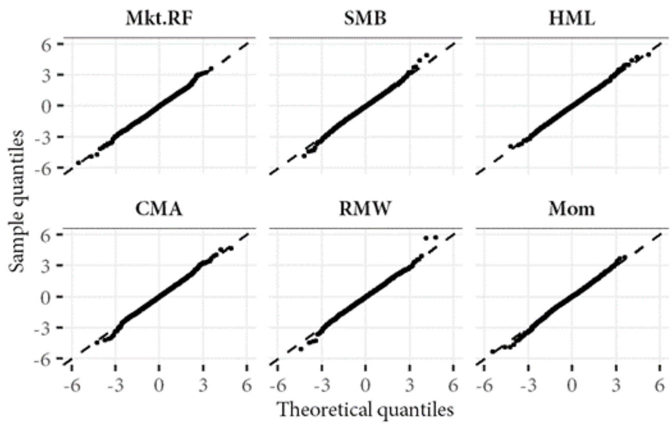
[10] A visual comparison of dependence measures on returns compared to standardized residuals can be found in Appendix E.



(a) Normal



(b) Symmetric *t*



(c) Skewed *t*

Figure 5: QQ plots of standardized residuals

Standardized residuals from the best (lowest BIC) ARMA-GARCH model specifications, with normal, symmetric t and skewed t innovations. Data from the theoretical distribution should line up on the dashed line. Based on weekly data 1963–2016.

correlation:

$$ThCorr(r_i, r_j) = \begin{cases} Corr\Big(r_i, r_j \,|\, r_i < F_i^{-1}(p), r_j < F_j^{-1}(p)\Big) & \text{for } p < 0.5 \\ Corr\Big(r_i, r_j \,|\, r_i \geq F_i^{-1}(p), r_j \geq F_j^{-1}(p)\Big) & \text{for } p \geq 0.5 \end{cases} \tag{5.10}$$

where $F_i^{-1}(p)$ is the empirical quantile of r_i at percentile p. Threshold correlations thus reflect how series correlate when both are simultaneously realizing in their respective tails. This subsetting of data is illustrated in Figure 6. In the left hand plots, we see the scatter of ARMA-GARCH residuals of Mkt.RF and HML respectively, and how the threshold p, found on the x-axis of the right hand plot, determines the subset of data that is included in the correlation calculation. We note that the unconditional (standard) correlation, given by the dashed line in the right hand plot, is clearly negative, while threshold correlations in the first and third quadrants are significantly more positive, which shows that not taking threshold correlations into account provides a vaguer picture of the dependence structure when both factor series realize in the tails.

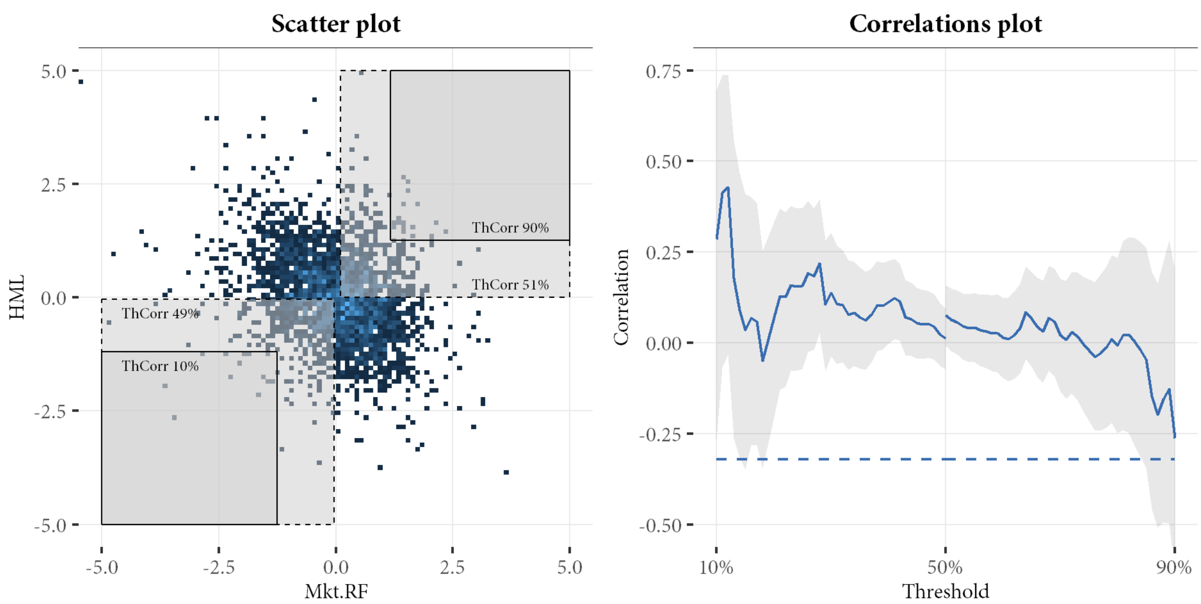


Figure 6: Illustration of threshold correlations

ARMA-GARCH residuals from the Mkt.RF–HML asset pair. 95% shaded confidence bounds. The unconditional correlation is given by the dashed line. Based on weekly data 1963–2016.

We now plot threshold correlations without the adjacent scatter graph. Figure 7 displays threshold correlations for HML, CMA and RMW against each other as well as against the the other factors Mkt.RF, SMB and Mom. We note that for most asset pairs, the threshold correlation is significantly different from the unconditional correlation coefficient given by the dashed line.

We also note that there is asymmetry around the median for some factor pairs, including the Mom–CMA, RMW–HML, RMW–CMA, and to a lesser extent Mkt.RF–RMW, asset pairs. For example, in the Mom-CMA asset pair, the threshold correlation jumps up for the first percentile below the median, indicating that the correlation is higher when both realize below the median than when both realize above the median. This type of asymmetric property, where downside (below the median) correlation is higher than upside correlation is unwanted, as it reflects a poorer diversification in bad times. The opposite type of asymmetry can be seen for the HML–RMW and CMA–RMW asset pairs; When these factors simultaneously realize above the median, they are significantly more correlated. This pattern presents no diversification problem.

Although estimated with substantial uncertainty, the threshold correlations do not seem to be constant as the threshold p approaches either zero or one. For example, the Mkt.RF–HML asset pair seems to have a downward pattern, where correlations are the most positive in the lowest percentiles of residuals and the most negative in the highest percentiles of residuals. In fact, this pattern is unwanted from a diversification perspective, as series tend to coincide more in extreme negative events.

The CMA–HML pair stands out from the other factors. The pair exhibits an unusually high correlation in excess of 0.60 with a virtually flat threshold pattern. HML and CMA are also generally similar to each other in their respective threshold patterns to other factors – most notably in the HML/CMA–RMW pairs. They differ in the presence of a break around the median in Mom–CMA not present in Mom–HML.

RMW is the only factor to be virtually uncorrelated with Mkt.RF in the lower tails, suggesting that it is a good diversifier in market downturns. This is very different from the pattern of higher threshold correlations as p approaches zero for e.g. Mkt.RF–HML. For Mkt.RF–HML, the higher lower tail correlation could be related to the industry over-capacity hypothesis discussed in section 2, i.e. that value firms are particularly sensitive to market downturns due to unproductive capital.

While the patterns in threshold correlations are interesting, we are careful not to draw conclusions regarding diversification benefit based on solitary threshold correlation graphs – what is interesting is the total pattern, and our key point is that there seems to be tail dependence that should not be ignored in the copula specification.

5.4.2 Rolling correlations

We compute rolling 52-week correlations between the factors on standardized residuals of our ARMA-GARCH models, according to the formula:

$$RCorr(r_{i,t}, r_{j,t})_t^{52} = \frac{\sum_{t-51}^{t}(r_{i,t} - \bar{r}_i)(r_{j,t} - \bar{r}_j)}{\sqrt{\sum_{t-51}^{t}(r_{i,t} - \bar{r}_i)^2}\sqrt{\sum_{t-51}^{t}(r_{j,t} - \bar{r}_j)^2}} \tag{5.11}$$

where r_i, r_j are the different pairs of the factor strategies' ARMA-GARCH residuals.[11] Results are presented in Figure 8. First, we note that for most factor pairs, the rolling 52-week correlations

[11] Rolling correlations for the returns themselves are available in Figure 17 (Appendix E).

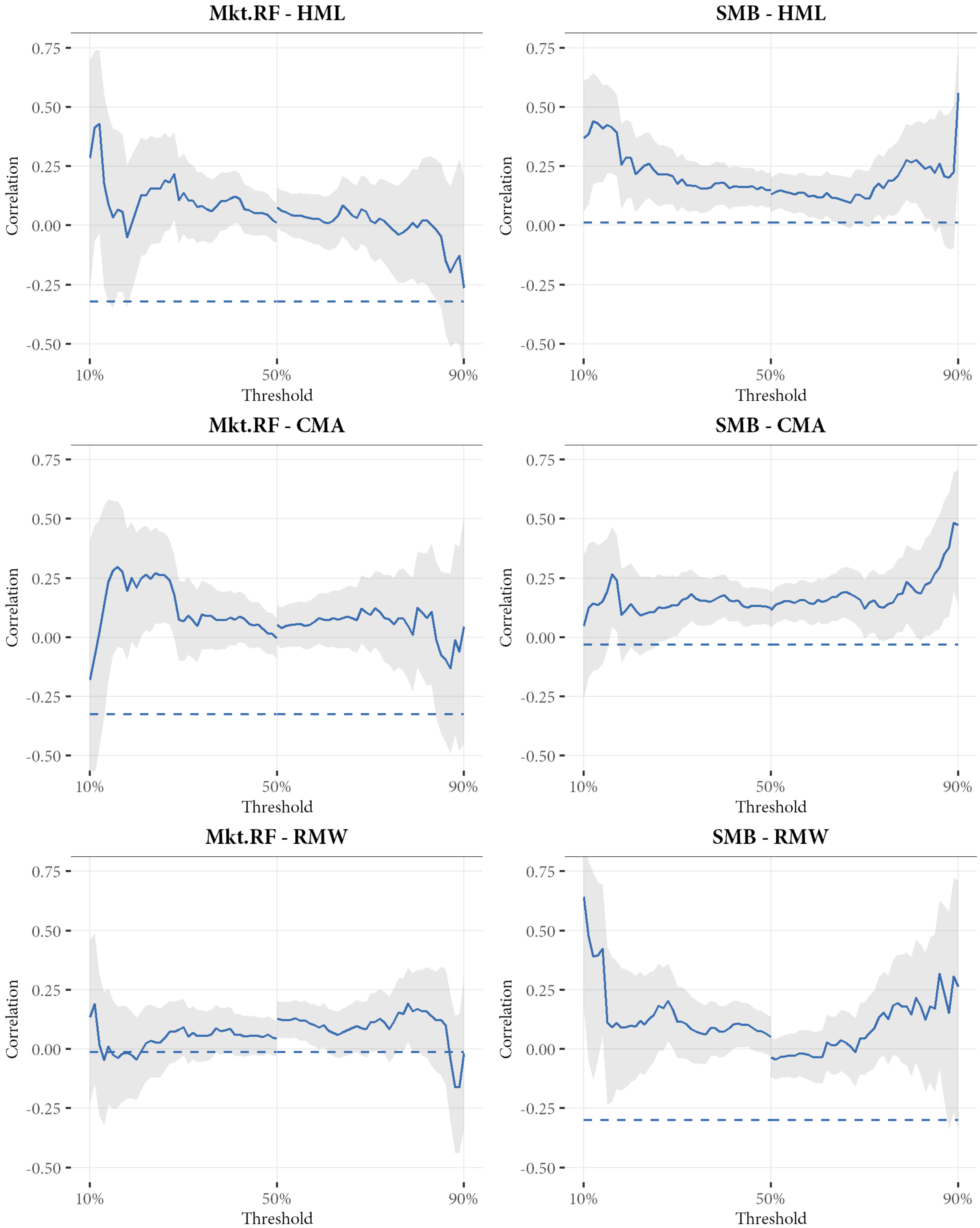


Figure 7: Threshold correlations of ARMA-GARCH standardized residuals

The formula for threshold correlations for a threshold p is given in Equation 5.10. 95% shaded confidence bounds, taking the model as given. The unconditional correlation is given by the dashed line. Based on weekly data 1963–2016.

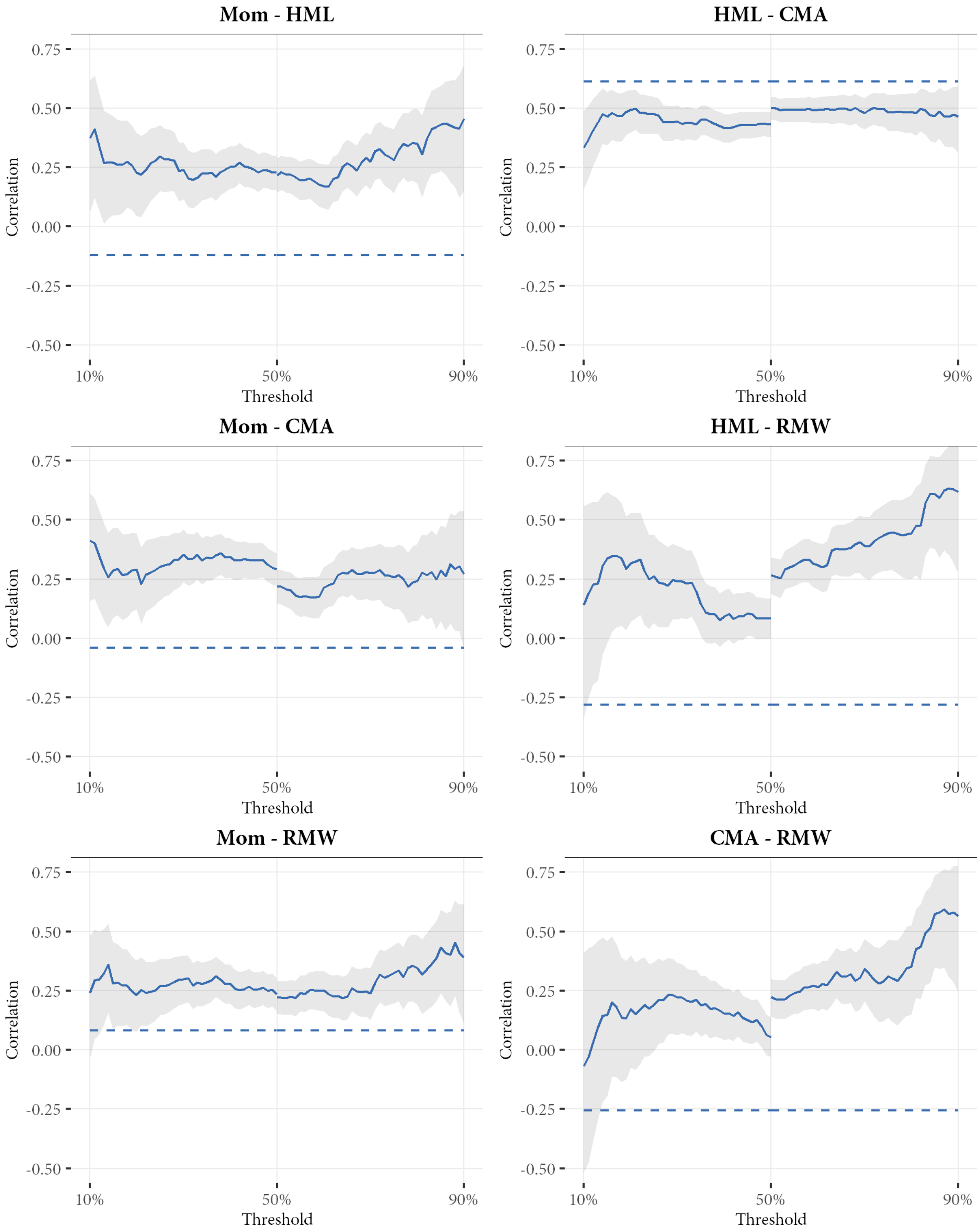


Figure 7: Threshold correlations of ARMA-GARCH standardized residuals (cont.)

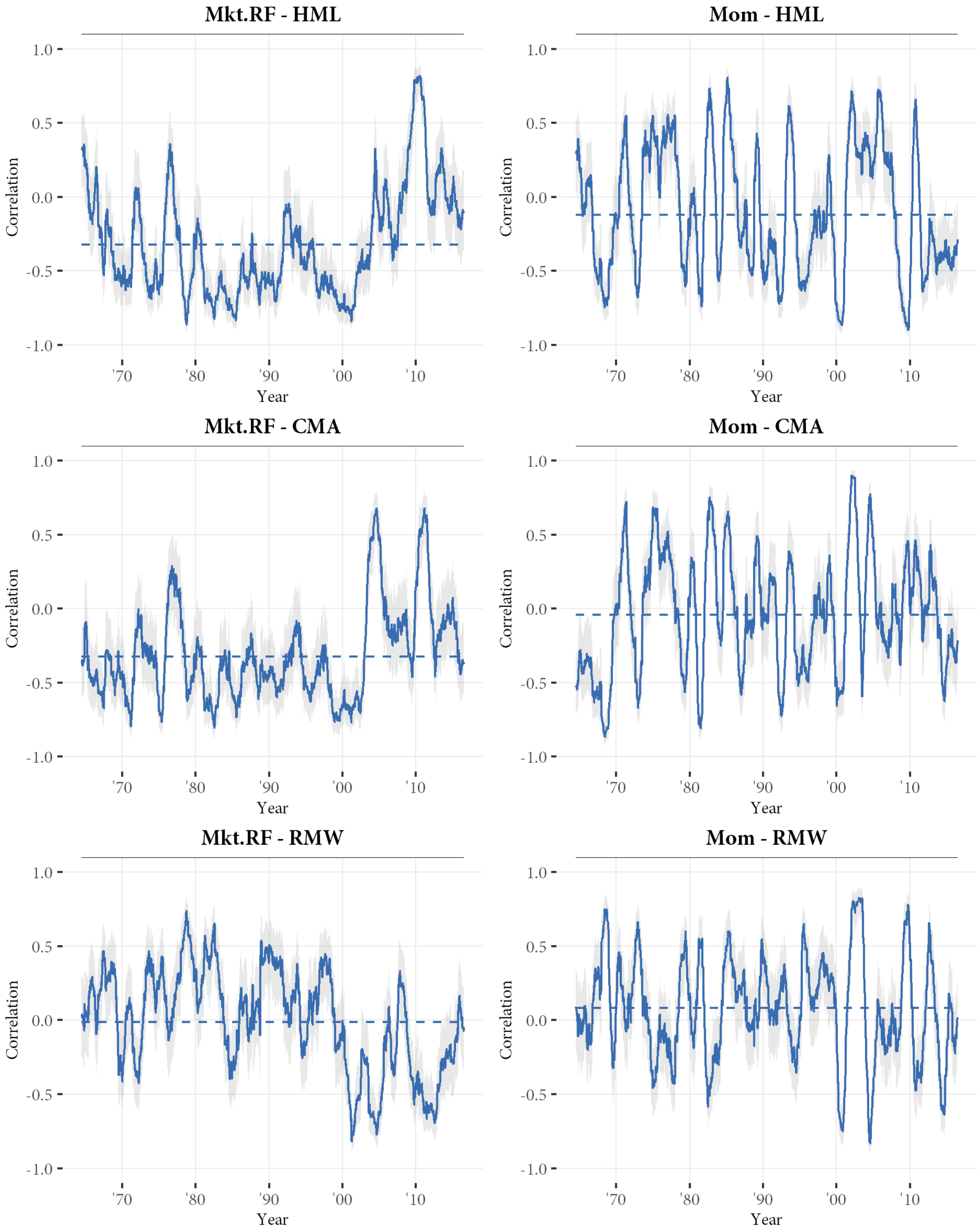


Figure 8: Rolling correlations of ARMA-GARCH standardized residuals

95% shaded confidence bounds, taking the model as given. The unconditional correlation is given by the dashed line. Based on weekly data 1963–2016.

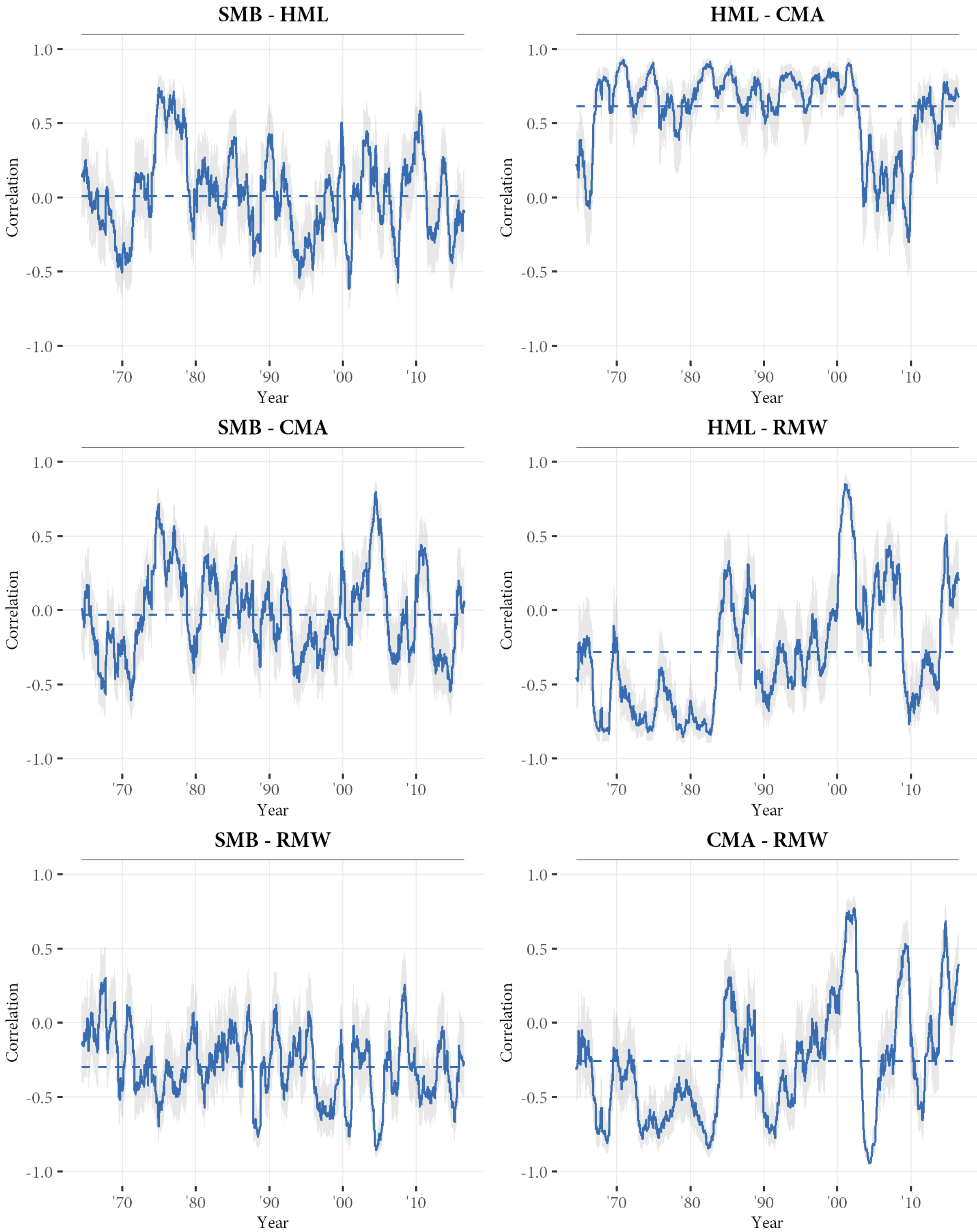


Figure 8: Rolling correlations of ARMA-GARCH standardized residuals (cont.)

are time-varying, and indeed appear to swing wildly. The unconditional correlation of Mkt–HML is negative in the studied time period, but rolling correlations range between -0.75 and 0.75. Also of note is the momentum factor's rapid shifts between positive and negative correlations to the other factors.

Second, by visual inspection, we see no obvious trend in the correlations between factor pairs. There are, however, notable patterns around the 2000–2001 bubble period – here, the correlations of HML–RMW, CMA–RMW and Mkt–CMA appear to jump. Another interesting pattern is that the correlations of Mkt–RMW went down sharply around this period – in line with the idea that profitable firms are stronger and better at weathering crises than the average firm (Novy-Marx, 2013). The 2000–2001 period may represent a structural break in the dependency patterns between factors, with the appearance of persistent differences before and after – however, there is not enough post-2000 data to support such a conclusion, yet.

Third, the HML–CMA factor pair again stands out as different from other factor pairs. The unconditional correlation is much closer to the rolling estimates than for other factor pairs, with a dip in the 2000–2010 period that appears to have gone away. Clearly, the HML–CMA pair is the most strongly correlated factor pair, even when considering subperiods of the data.

Our key takeaway from the rolling correlations is that there seems to be persistence in the time-variation in correlations. This could be incorporated in the copula specification, which then needs to have a time-varying correlation matrix, Ψ_t.

5.5 Copula specification and estimation results

Given the results of the dependence structure of residuals, we now discuss the best choice of copula model and present estimation results of the six competing copula specifications.

We have estimated constant and dynamic normal, symmetric t and skewed t copula models on the full dataset of uniform GARCH residuals. Results are presented in Table 5.

First, we examine the choice between a normal, symmetric t or skewed t copula. We note that ν_c is clearly significant and suggests one of the Student's t models with tail dependence over the normal model. Second, we examine the asymmetric specification and find that few of the γ_c estimates appear significant. This indicates that the asymmetry is hard to capture, or that it is not well represented by this type of model. This is supported by the relatively small improvement in log-likelihood in going from a symmetric t to skewed t copula, and also by the fact that the BIC criterion prefers the symmetric t model in the dynamic case.

Second, we examine the choice between a constant and dynamic copula correlation matrix. There is a significant improvement in log-likelihood and BIC when moving from a constant to a dynamic copula, which suggests that time-varying dependence shown by rolling correlation is captured and improves the model's fit. We also find a high persistence of the correlation process, as $\alpha + \beta$, is close to a unit root.

In summary, we find that the dynamic symmetric t copula is the best specification, as it has the lowest BIC, well defined parameters, and is strongly supported by the dependence pattern showcased by threshold and rolling correlation analyses. While the skewed t copula is an interesting model, we believe that the asymmetry patterns in data are too irregular to be well captured by a copula model with only one asymmetry parameter for each series (this is further discussed in the subsequent robustness discussion, see subsection 5.6).

5.6 Copula robustness check

This subsection provides an in-sample robustness check of how well the copula models can reproduce the threshold correlations and rolling correlations found in the dependence analysis of ARMA-GARCH residuals. By comparing simulated data from the copulas to ARMA-GARCH residual data, we find that the main features are captured. However, we highlight that tail dependence is only reproduced to a certain extent.

5.6.1 Threshold correlations in constant copulas

By simulating 250,000 weeks of shocks in the copula, and then transforming these shocks into standardized residuals for each of the factors, we can test the constant t copulas' abilities to generate the threshold correlations in the ARMA-GARCH residuals.[12] If a Student's t copula specification reasonably well captures tail dependence, the threshold correlations from the empirical and the copula specification should align. The results are presented in Figure 9.

First, we note that for most factors, the normal copula is the far away from generating threshold correlations that correspond to the ARMA-GARCH residuals around the median. This is highly expected, as the normal copula does not generate tail dependence, and hence the need for the Student's t based copula models. The symmetric t and skewed t copulas better capture the threshold correlations, as the fatter tails of the Student's t distribution allows for tail dependence. For example, note how the normal copula generates negative threshold correlations for both the Mom–HML and RMW–HML asset pairs, while the Student's t based copulas are much closer to the higher values in the data. On the other hand, the Student's t based copulas sometimes seem to overshoot the empirical threshold correlation, as in the Mkt.RF–RMW asset pair.

Second, we find that the skewed t generates some asymmetry around the median, which can be seen most clearly for the Mom–RMW and RMW–HML asset pairs. The generated asymmetry does, however, appear to be too small to capture the features of the data.

In conclusion, comparing threshold correlations from ARMA-GARCH residual data and simulated data shows that the constant copula specifications capture some of the tail dependence.[13] Although the symmetric t and skewed t results do not align perfectly with the data, they constitute clear improvements to the normal copula in modeling tail dependence.

[12] This robustness check is inspired by Christoffersen and Langlois (2013).

[13] Note that, in order to make the threshold correlation comparison valid, we use the constant copula specifications. The dynamic version is still the workhorse for all continued analysis in the mean-variance and diversification benefit sections.

Table 5: Copula parameter estimates (1963–2016)

Models from Equation 5.4 on 2766 weekly standardized residuals from the ARMA-GARCH models in Table 4. Stationary bootstrap standard errors in parentheses, following (Politis & Romano, 1994) (see Appendix D). nu_c is the degree of freedom parameter and $gamma_i$ are the copula skewness parameters. α, β control the correlation dynamics, and Persistence is $\alpha + \beta$. Elements of $\hat{Q}$ are estimated using moment matching (see Appendix B). The constant copula is achieved by forcing $\alpha = \beta = 0$. For each model, there are 15 parameters in the Q time-invariant correlation matrix, which are not reported in the table. Significance given by $^{*}p < 10\%$; $^{**}p < 5\%$; $^{***}p < 1\%$

	Constant Copula			Dymamic Copula		
	Normal	Symmetric t	Skewed t	Normal	Symmetric t	Skewed t
ν_c		6.61^{***}	6.64^{***}		11.77^{***}	11.63^{***}
		(0.89)	(0.10)		(0.89)	(1.00)
γ_{Mkt}			-0.06			-0.05
			(0.05)			(0.05)
γ_{SMB}			-0.11^{*}			-0.14^{**}
			(0.06)			(0.06)
γ_{Mom}			-0.20^{**}			-0.12
			(0.07)			(0.07)
γ_{HML}			0.10			-0.02
			(0.06)			(0.06)
γ_{CMA}			0.08			-0.05
			(0.06)			(0.07)
γ_{RMW}			0.02			0.18^{**}
			(0.07)			(0.07)
α				0.07^{***}	0.07^{***}	0.07^{***}
				(0.01)	(0.01)	(0.01)
β				0.91^{***}	0.91^{***}	0.91^{***}
				(0.01)	(0.01)	(0.01)
Log-likelihood (LLH), Number of parameters (# params.), BIC and Correlation Persistence (CP)						
LLH	1,169	1,555	1,573	2,790	2,983	2,995
# params.	15	16	22	17	18	24
BIC	$-2,337$	$-3,103$	$-3,090$	$-5,564$	$-5,943$	$-5,919$
CP (%)				97.73	98.01	97.98

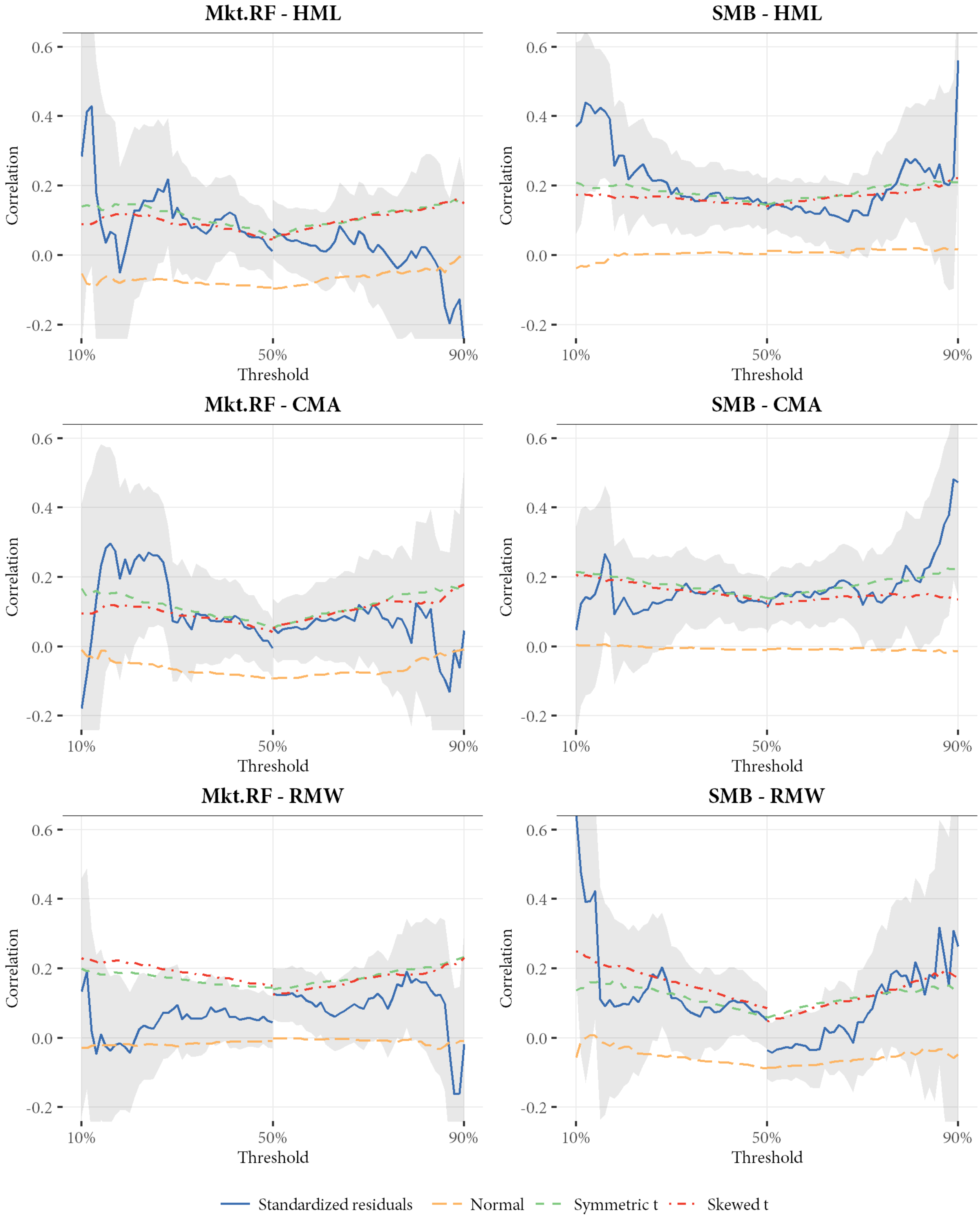


Figure 9: Threshold correlations of standardized residuals from the constant copulas

Threshold correlations of simulated constant copulas, compared to ARMA-GARCH standardized residuals (95% confidence bounds taking the ARMA-GARCH models as given). The simulated threshold correlations are based on 250,000 simulated returns each. ARMA-GARCH models based on empirical weekly data 1963–2016.

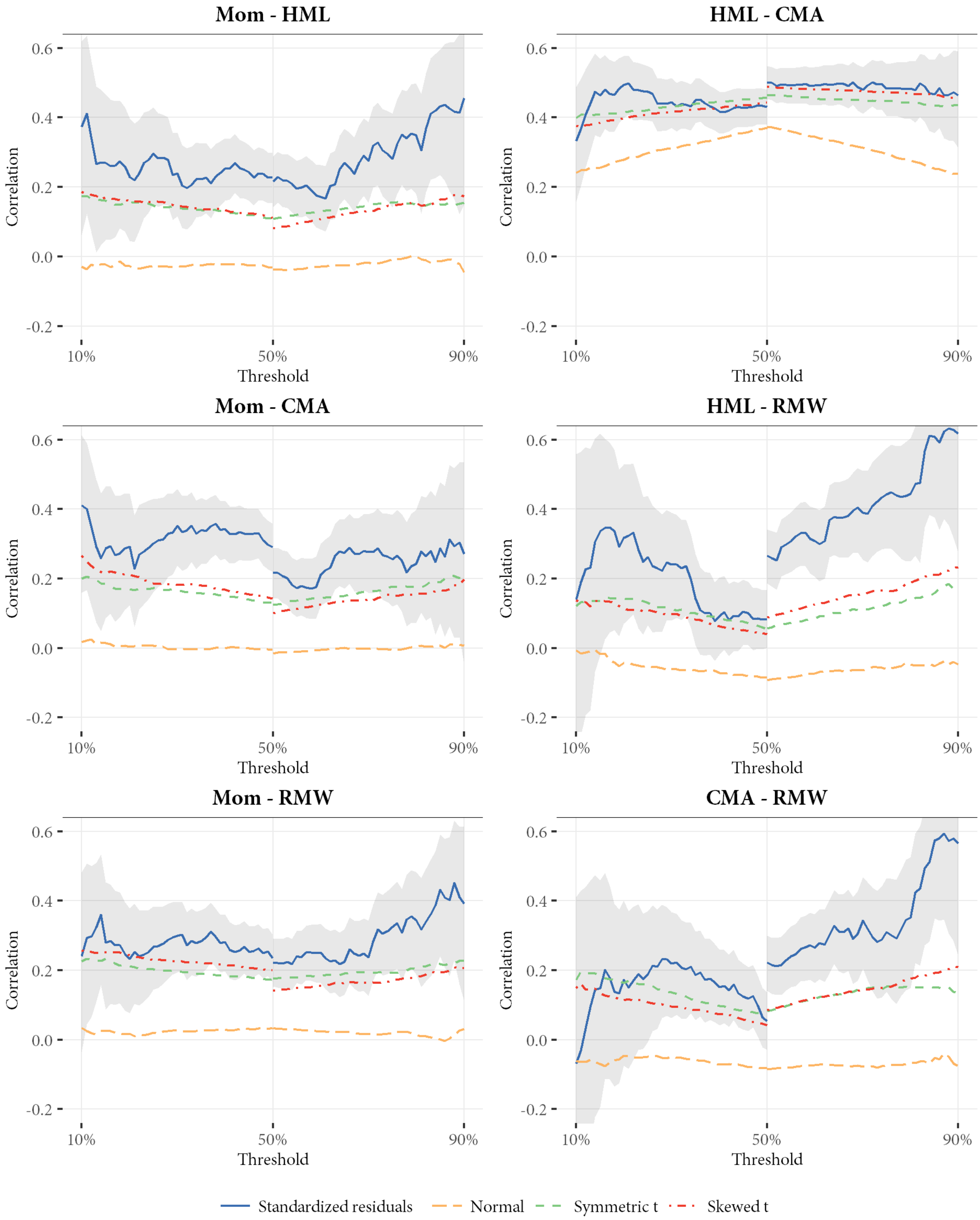


Figure 9: Threshold correlations of standardized residuals from the constant copulas (cont.)

We observe that the copula seems to lack flexibility to simultaneously generate all the asymmetries in tail dependence. This is quite expected, as the symmetric t copula only has one degree of freedom parameter that controls the fatness of tails, and the skewed t copula only has one skewness parameter for each series. This imposes limits on how strongly the model can express fat tails or asymmetries between factors A and B and simultaneously express other fat tails or asymmetries (or lack thereof) between factors A and C. For a collection of six factors with heterogeneous dependence, this is even harder. This is a clear limitation of our quite parsimonious multivariate distribution copula approach. In this regard, vine copulas that allow for unique bivariate copula specifications, as discussed in subsection 5.1, could be the solution.

Although imperfect, the multivariate copula modeling of tail dependence could constitute a significant improvement to alternatives, especially in the field of risk management, where understanding of tail events is paramount.

5.6.2 Rolling correlations in the dynamic copula

In-sample, we simulate 10,000 standardized residuals for each week from the estimated dynamic symmetric t copula model, and compute the rolling 52-week correlations. This is done to ascertain ourselves that the chosen copula specification does in fact capture the time-variation in correlations. The comparison is made between standardized residuals from the simulated copula model and standardized residuals from the ARMA-GARCH models of univariate series. If satisfactory, the rolling correlations of the simulated copula model and the ARMA-GARCH models will be roughly the same.

While we do not get a perfect overlap, the copula model generates similar time-varying correlations. When there are large swings, however, the model does not always have enough amplitude. We conclude that the model captures the dynamic features well.

6 Optimizing factor allocation

We now turn to the issue of optimizing factor allocations using our estimated copula model. We optimize portfolio weights according to two techniques: optimal Mean-Variance (MV) investing and optimal Conditional Diversification Benefit (CDB) investing, where the latter is a measure introduced by Christoffersen et al. (2012). In both optimizations, we experiment with the inclusion and exclusion of the HML, CMA and RMW factors, to discuss their marginal impact on portfolio performance measures.

The goal of the first optimization exercise is to determine the conventional risk-reward profile of including HML, CMA and RMW respectively. MV results will also test the conjecture in Fama and French (2015), postulating that the inclusion of HML does not improve the Sharpe Ratio of the tangency portfolio.

The goal of the second optimization is to consider risk beyond variance, by examining the impact on tail risk of including HML, CMA and RMW respectively. CDB analysis studies whether

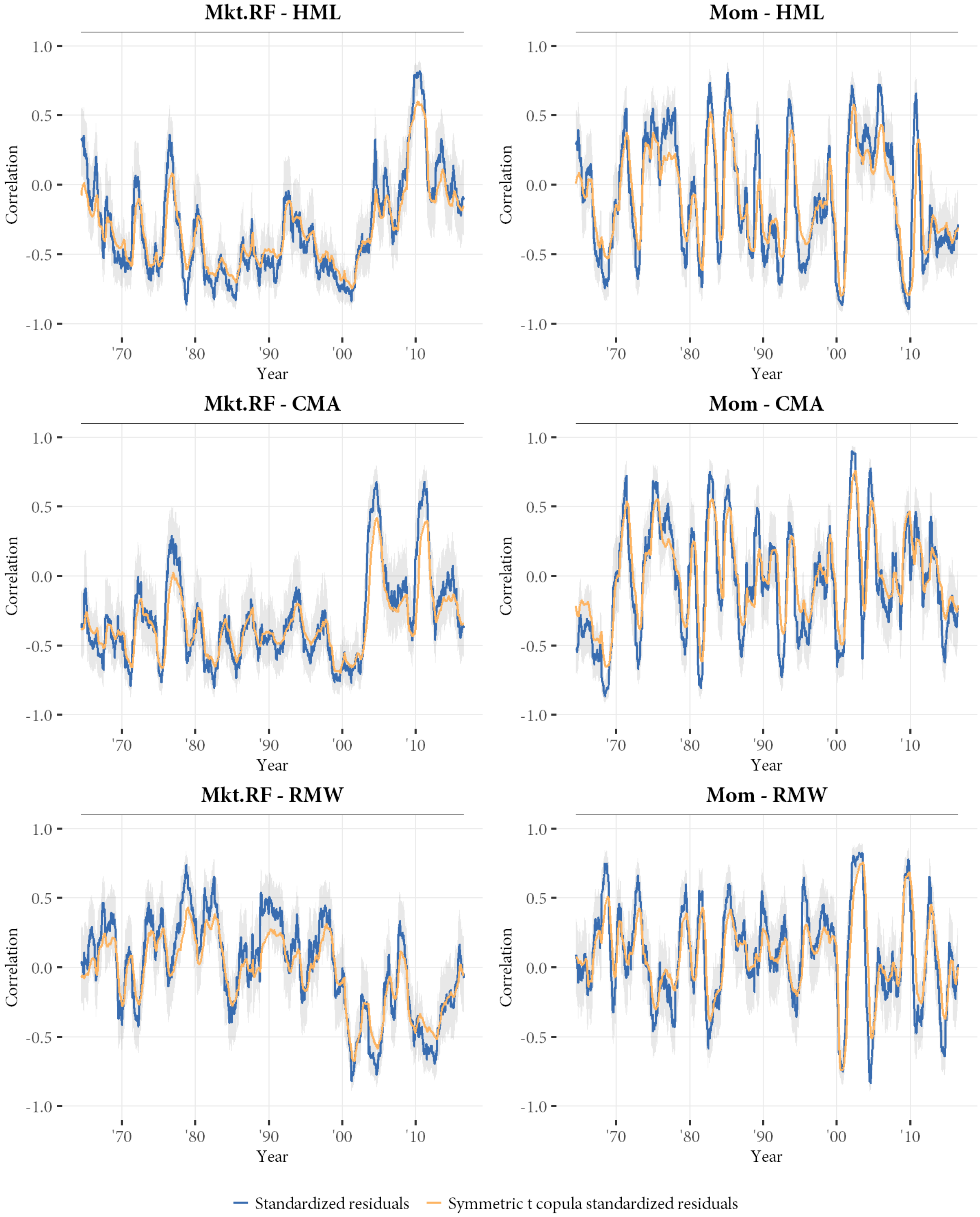


Figure 10: Rolling correlations of standardized residuals from the dynamic copula

Rolling correlations of the simulated dynamic symmetric *t* copula compared to rolling correlations on ARMA-GARCH residuals. 95% confidence bounds taking the ARMA-GARCH models as given. The simulated rolling correlations are based on 10,000 simulations each week. ARMA-GARCH models based on empirical weekly data 1963–2016.

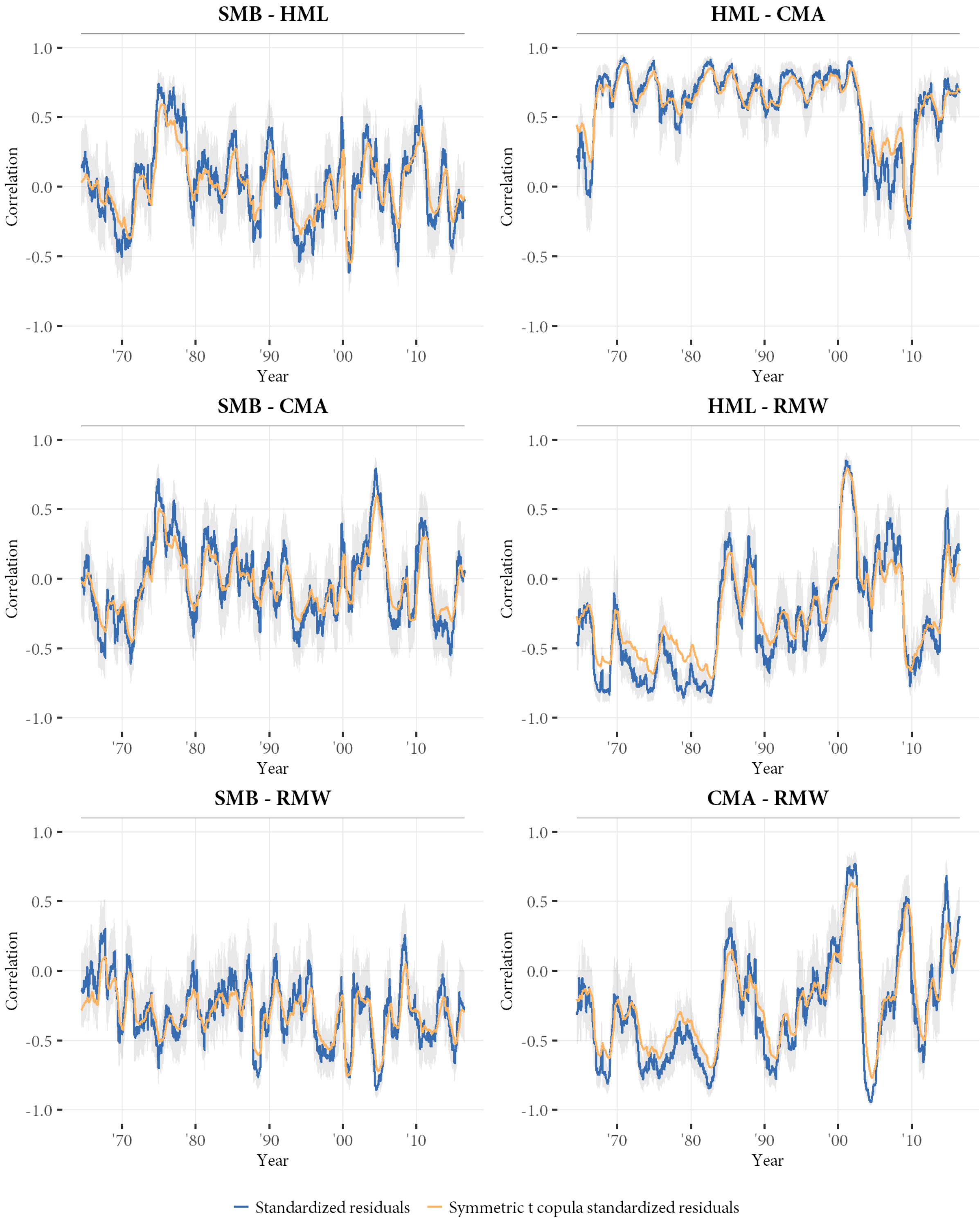


Figure 10: Rolling correlations of standardized residuals from the dynamic copula (cont.)

the non-normal features in the data (i.e. tail dependence) give any additional reason, beyond the MV results, to include or exclude the factors HML, CMA and RMW, respectively.

CDB is based on the portfolio expected shortfall (ES), i.e. the expected loss in case the return realizes below its Value-at-Risk (VaR), and therefore concerns the properties of the lower tail of the portfolio distribution. Naturally, such features are not captured by means and covariances in MV analysis.

The remainder of this section presents the construction of the CDB measure and the general optimization problem for MV and CDB. Results of the optimizations are given in the subsequent sections.

6.1 Conditional diversification benefit (CDB)

This description of CDB follows Christoffersen et al. (2012). Define ES as the expected loss in some bottom percentile q:

$$\mathrm{ES}^{q}_{i,t}(r_{i,t}) = -\mathbb{E}[r_{i,t}|r_{i,t} \leq F^{-1}_{i,t}(q)] \tag{6.1}$$

where $F^{-1}_{i,t}(q)$ is the inverse CDF of simple returns $r_{i,t}$ at q (equivalent to the $q\%$ Value-at-Risk).

The Expected Shortfall represents the expected loss when returns realize below the Value-at-Risk of the portfolio. Depending on the shape of the distribution at hand, the ES can be closer to or further away from the Value-at-Risk. Intuitively, if assets offer little diversification, then no combination of assets will reduce total portfolio risk; and ES will be higher.

For a portfolio of assets with weights w_t, the portfolio ES, $\mathrm{ES}^{q}_{t}(w_t)$, has an upper bound equal to the weighted average of each asset's ES, corresponding to the case of no diversification (Artzner, Delbaen, Eber, & Heath, 1999):

$$\overline{\mathrm{ES}}^{q}_{t}(w_t) = \sum_{i=1}^{N} w_{i,t}\mathrm{ES}^{q}_{i,t}(r_{i,t}) \tag{6.2}$$

A lower bound on portfolio ES is given by the portfolio's Value-at-Risk ($-F^{-1}_{t}(w_t, q)$), corresponding to the case of perfect diversification:

$$\underline{\mathrm{ES}}^{q}_{t}(w_t) = -F^{-1}_{t}(w_t, q) \tag{6.3}$$

CDB is defined as the portfolio's ES scaled by its lower and upper bounds:

$$\mathrm{CDB}^{q}_{t}(w_t) = \frac{\overline{\mathrm{ES}}^{q}_{t}(w_t) - \mathrm{ES}^{q}_{t}(w_t)}{\overline{\mathrm{ES}}^{q}_{t}(w_t) - \underline{\mathrm{ES}}^{q}_{t}(w_t)} \tag{6.4}$$

CDB is a number between 0 and 1, which we report scaled to 0–100. Note that the level of expected return does not enter into the measure – CDB only measures how powerful a group of assets are at achieving low tail risk.

6.2 Optimization problem

Each week, we choose portfolio weights to maximize the Sharpe Ratio in the mean-variance case, and CDB in the CDB case. We impose two restrictions on the optimization problem. First, all factor weights must be positive. This reduces the problem with extreme weights, as seen in the unconstrained optimization problem, and reflects a view that an investor will not bet against factors that have generated a history of positive premia. Second, factor weights must sum to unity, i.e. the portfolio is fully invested across factors. In light of Asness et al. (2015), we consider portfolios with and without momentum (five- and six-factor portfolios, respectively).

Together, these restrictions make the maximized Sharpe Ratio reflect tangency portfolio weights subject to a constraint of no negative weights. Due to the restrictions, the standard analytical solution to the mean-variance problem is not equal to our optimal tangency portfolio. Similarly, no analytical solution exists for optimizing CDB. Hence, for a portfolio with N factors we perform a numerical optimization where we choose the vector of weights w_t to maximize the objective function $\Omega_t(w_t)$, subject to the restrictions above:

$$\begin{aligned} \max_{w_t} \quad & \Omega_t(w_t) \\ \text{s.t.} \quad & \sum_{i=1}^{N} w_{i,t} = 1 \\ & w_{i,t} \geq 0 \qquad \forall i \in N \end{aligned} \tag{6.5}$$

For the MV case, the objective function is the one-week Sharpe Ratio (SR):

$$\Omega_t(w_t) = \frac{w_t^\top \mathbb{E}_t[r_{t+1}]}{\sqrt{w_t^\top \mathbb{E}_t[\Sigma_{t+1}] w_t}} \tag{6.6}$$

$\mathbb{E}_t[r_{t+1}]$ is the conditional one-step-ahead expected factor return and $\mathbb{E}_t[\Sigma_{t+1}]$ is the conditional one-step-ahead variance-covariance matrix.

For the CDB case, the objective function is the one-week CDB:

$$\Omega_t(w_t) = \mathrm{CDB}_t^q(w_t) \tag{6.7}$$

We perform MV and CDB optimization using in-sample simulated distributions from our dynamic copula model. Each week, we simulate 10,000 returns from the estimated copula model and use those outcomes as the conditional distribution of factor returns. We also perform MV optimization based on sample means and covariances, in which case $\mathbb{E}_t[r_{t+1}]$ and $\mathbb{E}_t[\Sigma_{t+1}]$ are constant and equal to the full sample estimators.

7 Results from mean-variance optimization

In Figure 11, we present the optimized weights over time in the five-factor universe (four-factor when one factor is excluded). The left hand column presents weights of factors when we consider the inclusion and exclusion of HML, while the right hand column presents weights of factors when CMA is included and excluded. The dynamic weights are based on one-week-ahead forecasts from the copula model, while the static lines represent the weights based on sample estimators of the mean-variance inputs.

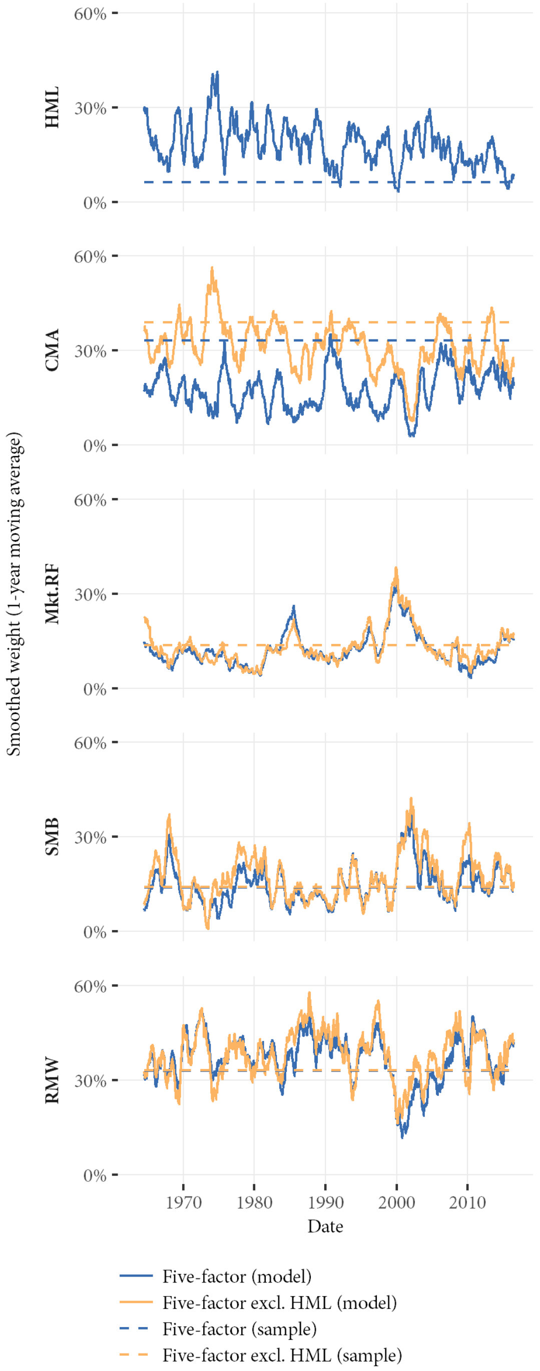
Accompanying the graph, Table 6 presents average MV optimal weights based on dynamic inputs of expected returns and covariances from the copula model. This table also gives average weight differences between models, as well as a number of performance measures that are based on the realized returns. The CDB statistic is based on MV optimal weights and uses the copula model as the distribution of returns. Please note that all performance measures are calculated in-sample and are not indicative of out-of-sample investing based on a copula model; they should be interpreted only in relative terms, in order to determine how much a portfolio is impacted by the exclusion of a factor.[14]

We begin by examining the left column of plots in Figure 11 that includes and excludes HML in the five-factor universe. First, we note that the weight of HML is clearly not zero when introduced in the investible universe. This appears to be the case for both the sample estimate of means and covariances, as well as the dynamic estimates from the copula model. Both inputs suggest that HML does in fact improve the tangency portfolio, as the optimal portfolio has a positive weight on the factor. The inclusion of HML leads to an increase in realized SR, which goes from 1.48 to 1.64. Furthermore, including HML leads to a less risky portfolio, with slightly lower average VaR and average ES and a higher degree of tail diversification, as measured by CDB (see Table 6).

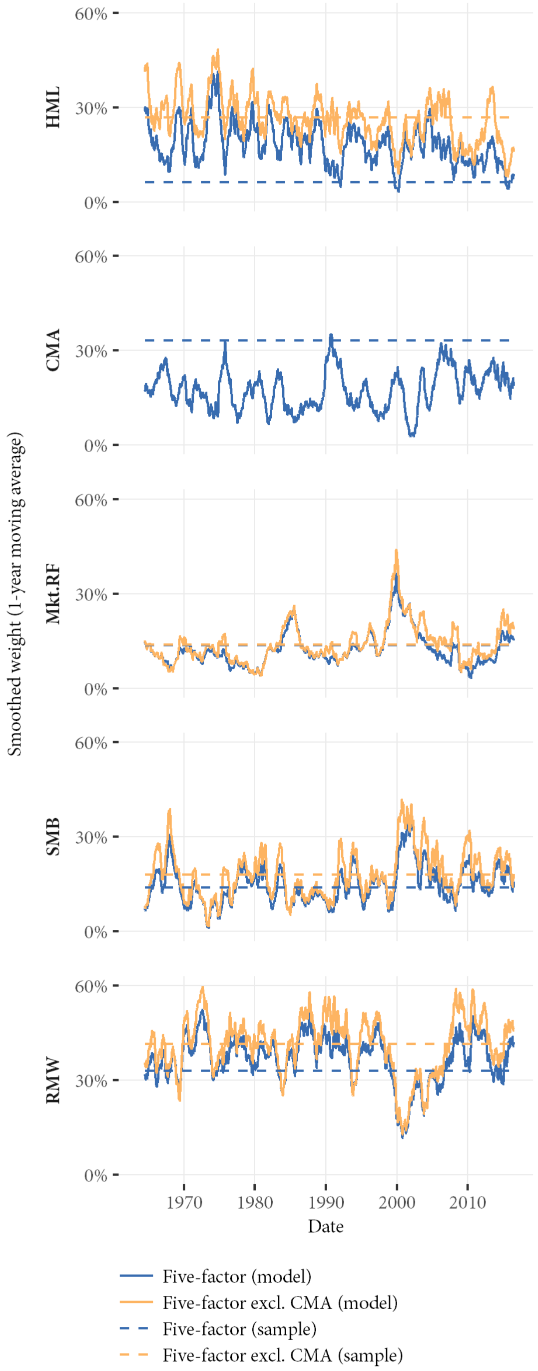
Although we impose the additional restriction of no negative weights, this finding does stand in contrast to the conjecture in Fama and French (2015) that HML should not improve the tangency portfolio. While the unconstrained tangency portfolio may or may not include HML, the simple restriction of no negative weights makes HML an important part of the optimal portfolio. However, we notice that this is less clear if we do not use dynamic inputs. For the full five-factor model, static sample estimators result in a lower allocation to HML (6.2%) than the dynamic inputs from the copula (average 18.2%), and a much smaller difference in realized SR, which goes from 1.25 to 1.27 (see Table 8 (Appendix E) for sample results).

Second, from the left panel of Figure 11, we note that the dynamic weight of HML seems to be highly similar to the decrease in weight in CMA, while all the remaining factors seem to stay very close to their original weights when moving to the full five-factor model. In other words, the weight that is attributed to HML is drawn nearly directly from the weight of CMA, in each period.

[14] A similar table of performance measures and average weight differences using in-sample sample estimators of expected returns and covariances is found in Table 8 (Appendix E).



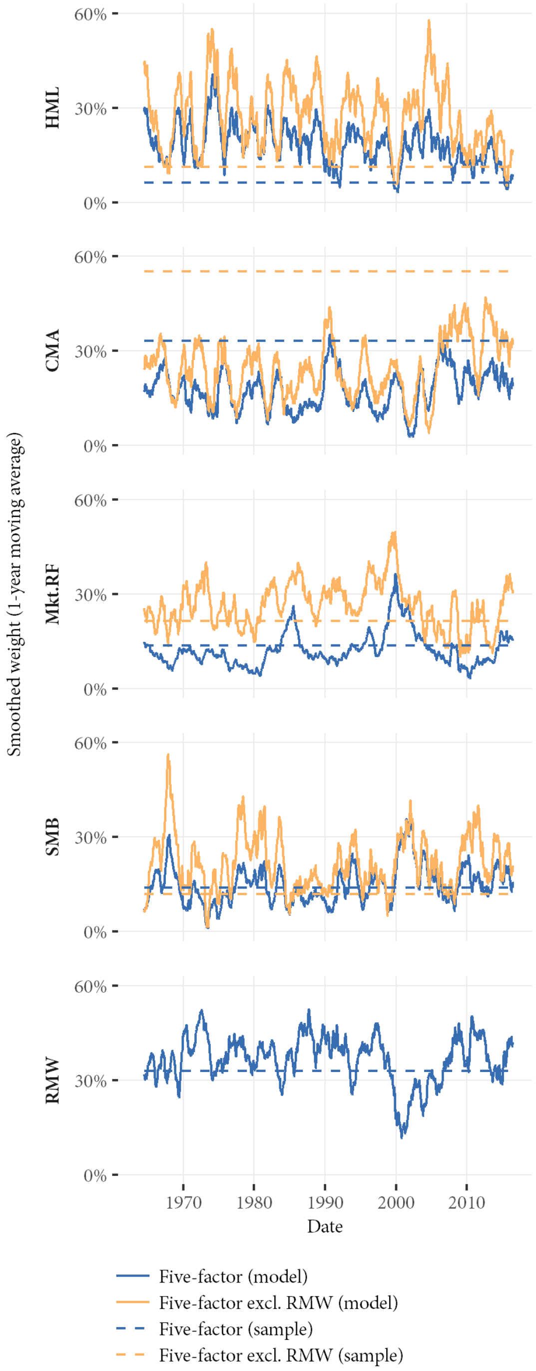
(a) Excluding HML



(b) Excluding CMA

Figure 11: Mean-variance optimal weights with five factors

Smoothed as 1-year moving averages for better legibility. Optimization constrained to fully invested portfolios with non-negative weights. Left hand panel including and excluding HML, right hand including and excluding CMA. Based on one-week-ahead forecasts from the dynamic symmetric *t* copula model 1963–2016.



(c) Excluding RMW

Figure 11: Mean-variance optimal weights with five factors (cont.)

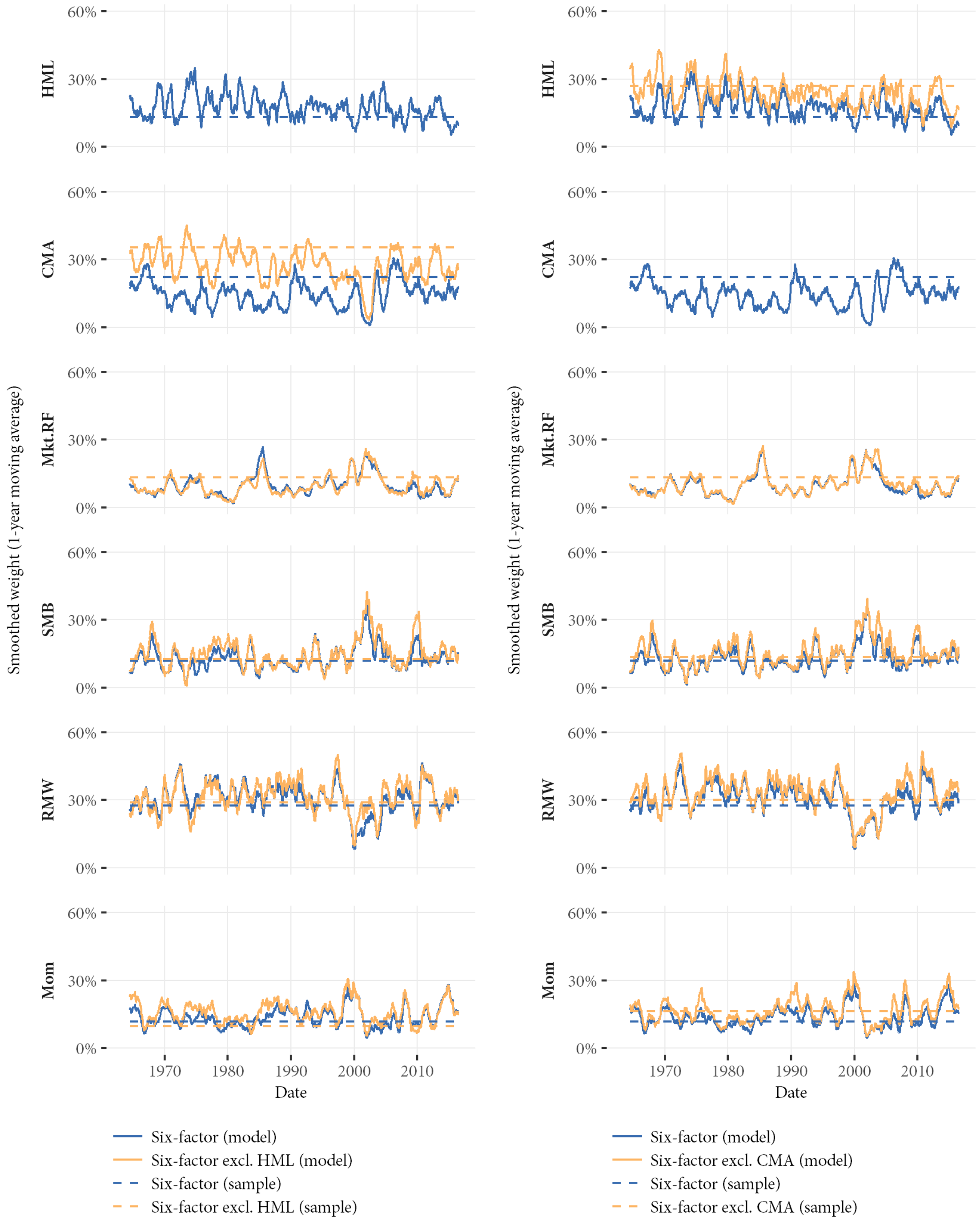
Our interpretation is that in a five-factor model excluding HML (effectively a four-factor model), CMA proxies for HML, which is why CMA absorbs nearly all the weight. HML also seems to proxy for CMA to a lesser extent, as shown by the right hand column of plots in Figure 11, which include and exclude CMA. When CMA is included, the weight of HML is lowered, but so are the weights of SMB and RMW. The proxying behavior of HML and CMA is expected, as the factors are highly correlated and zero-cost portfolio regressions in this thesis (Table 2), as well as in Fama and French (2015) and Asness et al. (2015), indicate that the main explanatory variable for HML is CMA, and vice versa.

Third, from the third panel of Figure 11, presented on a separate page, we find that among the two newly introduced factors CMA and RMW, the latter seems to have a much more substantial impact on the mean-variance portfolio. RMW receives weights of approx. 35% in the five-factor model, which is nearly twice the allocation of any other factor. Furthermore, the inclusion of RMW leads to a large improvement of portfolio performance measures. For dynamic weights in the five-factor model, RMW makes the realized SR go from 1.28 to 1.64. At the same time, RMW leads to large decreases in the average VaR and average ES risk measures, both of which are nearly halved when RMW is included. This highlights the large benefits of including the RMW strategy in a factor portfolio, which are expected due to RMW's low or negative factor loadings in the zero-cost portfolio regressions (see section 4).

We now move on to the six-factor model universe (effectively five-factor models when one factor is excluded), with weights in Figure 12 and figures in the right hand panel of Table 6. As the optimal weights and performance results are in general qualitatively similar to the five-factor model, we will only comment on key changes.

In the expanded universe, Sharpe Ratios are generally very high, regardless of whether a factor is excluded. Surprisingly, the highest SR is realized when CMA is excluded. However, the difference is small. Another notable change is that the allocation to HML based on static sample inputs is now higher (13.2% compared to 6.2%), which could be due to the clearer recognition of momentum differences between HML and CMA firms as Mom is included (cf. the discussion on zero-cost regressions in section 4).

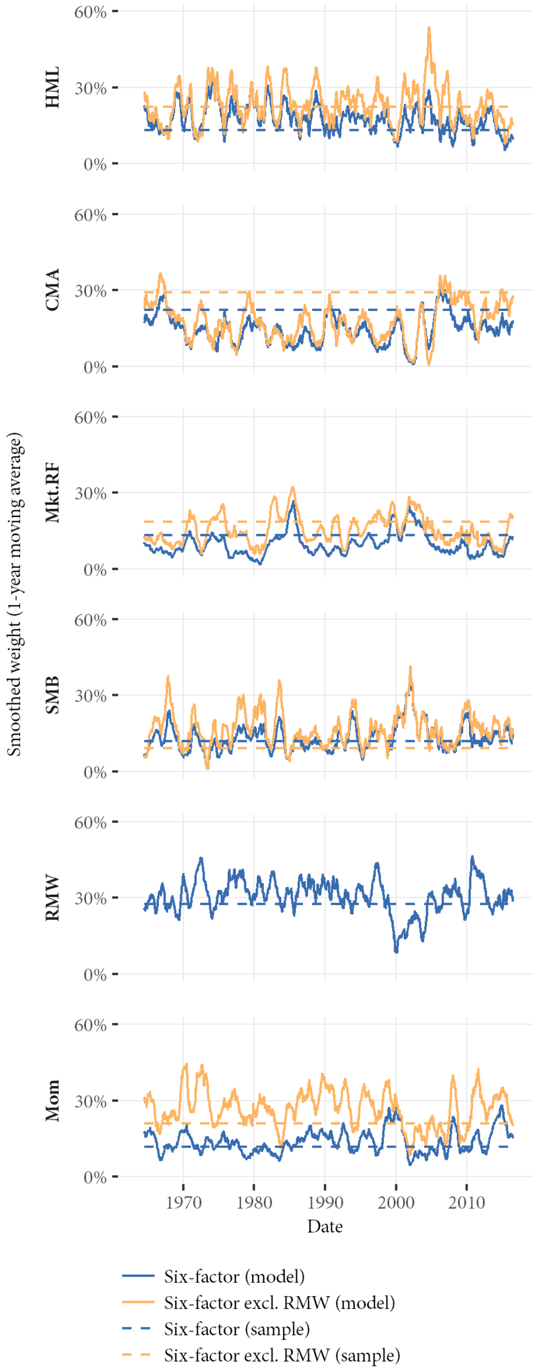
In summary, we examine weights and portfolio performance measures for optimal MV portfolios and find no reason to believe that mean-variance investing in the HML factor is dead or fully subsumed by the remaining factors, as the optimal portfolios include HML and have higher realized Sharpe Ratios, as well as lower realized risk measures. We also find that the inclusion of RMW is substantially more important, and leads to a large weight allocation to the factor as well as a substantial improvement on all performance measures.



(a) Excluding HML **(b) Excluding CMA**

Figure 12: Mean-variance optimal weights with six factors

Smoothed as 1-year moving averages for better legibility. Optimization constrained to fully invested portfolios with non-negative weights. Left hand panel including and excluding HML, right hand including and excluding CMA. Based on one-week-ahead forecasts from the dynamic symmetric *t* copula model 1963–2016.



(c) Excluding RMW

Figure 12: Mean-variance optimal weights with six factors (cont.)

Table 6: Mean-variance optimization with dynamic copula model (1963–2016)

Average weights are averages of dynamic MV optimal weights based on means and covariances from the dynamic symmetric *t* copula model. Differences in average weights are expressed relative to the full five- and six-factor models. Performance measures are based on realized returns. SR is the annualized Sharpe Ratio. VaR, ES and CDB are all based on the one-week-ahead 5% lower tail of the return distribution, which is given by simulations from the copula model. Differences in CDB are to be read as column model minus row model and its associated standard errors (in parentheses) are computed taking the copula model as given.

	Five (four) factor models				Six (five) factor models			
	All	Excl. HML	Excl. CMA	Excl. RMW	All	Excl. HML	Excl. CMA	Excl. RMW
Average weights								
Mkt.RF	12.4	13.2	14.2	26.4	9.9	9.9	10.4	15.5
SMB	15.0	17.4	18.3	21.8	13.4	15.3	15.6	16.8
HML	18.2		26.3	27.2	17.6		24.1	23.1
CMA	17.6	31.4		24.6	14.6	27.5		17.8
RMW	36.8	38.0	41.2		30.6	31.4	33.3	
Mom					13.9	15.9	16.6	26.8
Difference weights (column minus All)								
Mkt.RF		0.8	1.8	14.0		0.1	0.6	5.7
SMB		2.4	3.3	6.8		1.9	2.2	3.4
HML		−18.2	8.1	9.0		−17.6	6.5	5.5
CMA		13.8	−17.6	7.0		12.9	−14.6	3.2
RMW		1.2	4.4	−36.8		0.8	2.7	−30.6
Mom						2.0	2.7	12.8
Performance								
Mean (%)	6.48	6.42	7.28	8.88	7.18	7.34	8.12	9.94
SD (%)	3.95	4.34	4.72	6.93	4.05	4.39	4.49	5.83
SR	1.64	1.48	1.54	1.28	1.77	1.67	1.81	1.70
Avg. VaR (%)	0.59	0.68	0.73	1.14	0.56	0.64	0.66	0.93
Avg. ES (%)	0.83	0.94	1.02	1.58	0.79	0.90	0.93	1.31
Avg. CDB	82.78	79.38	77.98	68.03	84.66	81.44	81.68	74.75
Difference CDB (column model minus row model)								
All		−3.40	−4.81	−14.75		−3.22	−2.98	−9.91
		(0.17)	(0.22)	(0.37)		(0.14)	(0.15)	(0.28)
Excl. HML			−1.41	−11.35			0.24	−6.69
			(0.29)	(0.41)			(0.22)	(0.32)
Excl. CMA				−9.95				−6.93
				(0.43)				(0.31)

8 Results from CDB optimization

We begin by comparing the optimal CDB weights to the optimal MV weights in Figure 13.[15] The difference between weights under CDB and MV optimization are found to be fairly small – especially for Mkt.RF, and to a lesser extent for HML and CMA. It rather seems that the CDB optimal weights are less erratic in their pattern, which is likely to be due to the fact that MV optimization takes into consideration the conditional expected return, while CDB optimization only concerns tail risk. The most notable difference between CDB and MV weights is that CDB seems to allocate less into Mom. Again, this is coherent, as the CDB optimization does not reward the factor for its high expected return.

Next, we study the CDB over time with CDB optimized weights, where we experiment with excluding one of the factor HML, CMA and RMW at a time. The results are given in Figure 14, and are based on a Value-at-Risk cut-off of 5%.[16] We proceed with a number of interesting results that emerge from this picture:

First, we note that regardless of whether momentum is included or not, factor strategies appear to offer high levels of diversification. In absolute terms, all strategies fluctuate in the 80–95 range for the majority of the studied time period. This means that ES is relatively close to VaR, i.e. there is limited tail risk.

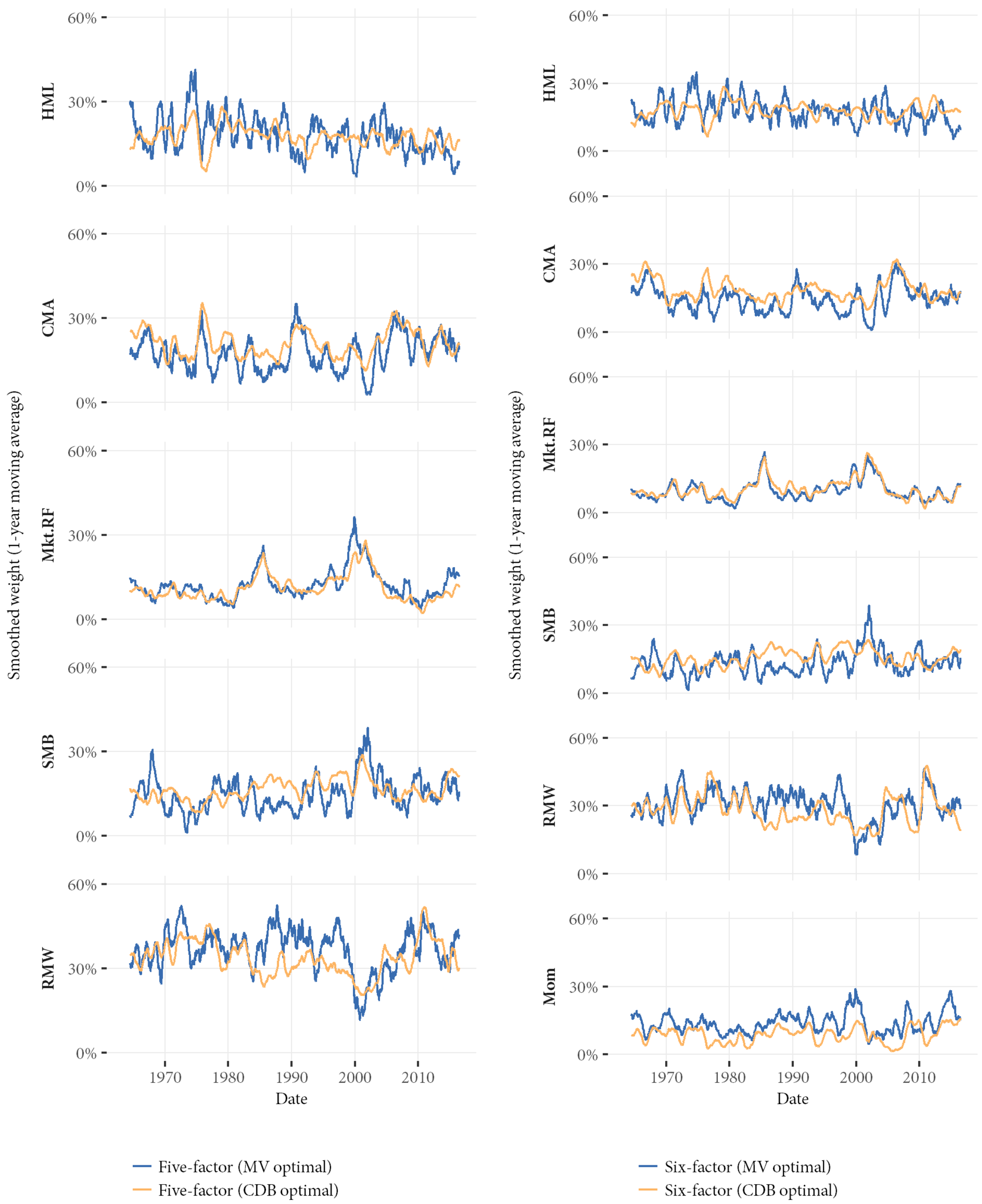
Second, there are notable dips in the diversification benefit measure. The dips represent times when diversification is relatively hard to come by, and roughly coincide in the five- and six-factor models. Interestingly, the periods of low diversification do not seem be stock market crises, as the CDB measure remains relatively high during the bubble of 1999–2000 bubble and the recession of 2007–2009.

Third, the level decreases in diversification benefit of removing HML or CMA seem quite small. Furthermore, this decrease is highly similar; at certain times, portfolios including HML are more diversified and vice versa, but no pattern emerges. However, we note that the exclusion of RMW is dramatically different. Without RMW, the level decrease in CDB is substantial and the dips become much more pronounced and frequent.

For comparison, we also report the CDB over time with MV optimized weights in Figure 15. This allows us to study how well tail diversified an MV portfolio is. In comparison to the optimal CDB in Figure 14, we note that the MV optimization leads to considerably lower tail diversification. Changing to MV optimization leads to a level shift as well as more frequent and deeper dips. Interestingly, the change to MV weights makes diversification benefits worse especially for the model where RMW is excluded. We find it interesting for MV oriented factor investors that including RMW significantly limits the tail risk in MV portfolios. In a number of periods, incl. 1972–1974, 1976–1978, 1980–1982, and 1990–1993, a number of severe declines in tail diversification are avoided almost entirely if RMW is included.

[15] The weight graphs for CDB optimal weights, excluding one factor at a time, are also reported, but have been relegated to Figure 18 and Figure 19 (Appendix E).

[16] In unreported results, the lower cut-off value of 1% is found to give qualitatively similar results.



(a) Five-factor universe **(b) Six-factor universe**

Figure 13: Comparison of CDB optimal weights and MV optimal weights

Smoothed as 1-year moving averages for better legibility. Optimization constrained to fully invested portfolios with non-negative weights. Based on one-week-ahead forecasts from the dynamic symmetric t copula model 1963–2016.

Figure 14: 5% Conditional Diversification Benefit (CDB) for CDB optimal weights

Five- (without Momentum) and six-factor universes. CDB lines have been smoothed as quarterly moving averages for better legibility. See subsection 6.1 for computational details.

Figure 15: 5% Conditional Diversification Benefit (CDB) for MV optimal weights

Five- (without Momentum) and six-factor universes. CDB lines have been smoothed as quarterly moving averages for better legibility. See subsection 6.1 for computational details.

Table 7: CDB optimization with dynamic copula model (1963–2016)

Average weights are averages of dynamic CDB optimal weights based on simulations of the return distribution from the dynamic symmetric *t* copula model. Differences in average weights are expressed relative to the full five- and six-factor models. Performance measures are based on realized returns. SR is the annualized Sharpe Ratio. VaR, ES and CDB are all based on the one-week-ahead 5% lower tail of the return distribution. Differences in CDB are to be read as column model minus row model and its associated standard errors (in parentheses) are computed taking the copula model as given.

	Five (four) factor models				Six (five) factor models			
	All	Excl. HML	Excl. CMA	Excl. RMW	All	Excl. HML	Excl. CMA	Excl. RMW
Average weights								
Mkt.RF	11.1	10.5	11.5	19.2	10.5	10.1	10.6	15.5
SMB	16.6	18.3	19.1	22.6	15.8	17.9	18.1	19.2
HML	17.4		30.1	26.7	18.1		28.7	24.9
CMA	21.2	35.0		31.6	18.7	32.2		24.1
RMW	33.8	36.2	39.3		28.1	31.8	32.6	
Mom					8.8	8.1	10.0	16.2
Difference weights (column minus All)								
Mkt.RF		−0.6	0.4	8.1		−0.4	0.2	5.1
SMB		1.7	2.5	6.0		2.1	2.3	3.4
HML		−17.4	12.7	9.3		−18.1	10.6	6.8
CMA		13.9	−21.2	10.4		13.5	−18.7	5.4
RMW		2.4	5.5	−33.8		3.7	4.4	−28.1
Mom						−0.7	1.3	7.5
Performance								
Mean (%)	2.77	2.94	2.85	3.37	3.37	3.39	3.41	3.90
SD (%)	2.42	2.49	2.69	3.89	2.37	2.52	2.56	3.49
SR	1.14	1.18	1.06	0.87	1.42	1.34	1.33	1.12
Avg. VaR (%)	0.46	0.48	0.52	0.78	0.45	0.47	0.49	0.69
Avg. ES (%)	0.61	0.64	0.69	1.04	0.60	0.64	0.66	0.92
Avg. CDB	90.42	89.29	89.19	83.42	91.33	90.24	90.51	86.62
Difference CDB (column model minus row model)								
All		−1.13	−1.23	−7.01		−1.09	−0.82	−4.71
		(0.02)	(0.03)	(0.11)		(0.02)	(0.03)	(0.10)
Excl. HML			−0.10	−5.88			0.27	−3.62
			(0.04)	(0.11)			(0.04)	(0.10)
Excl. CMA				−5.78				−3.89
				(0.11)				(0.10)

In summary, we find that the high similarity of HML and CMA indicates that tail diversification benefits are not dramatically improved by including both of the factors, which is coherent with the fact that they are closely related and overlap. However, this does not mean that both factors should not be considered jointly, as doing so could still could improve the conventional risk-return trade-off in a mean-variance setting. The RMW factor, on the other hand, is shown to be very important for tail diversification purposes and should be considered by all factor investors concerned with tail risk.

9 Discussion and conclusion

The first key finding of this thesis is that the classic value factor, HML, is still highly relevant for factor investors. Zero-cost regressions in the five-factor model suggest that HML may be fully explained by the remaining four factors, we find evidence to the contrary when accounting for the momentum factor in a weekly dataset. We believe that the reason why zero-cost regressions indicate that HML does not add value is that the regressions are misspecified. The omission of an important sixth factor strategy, momentum, creates a bias on the factor loadings, and leads to not recognizing the added value of HML. While the role of HML as a unique addition was already defended by Asness et al. (2015), we have pursued the argument and can conclude that optimizations lead to positive allocations to the HML strategy.

We run the mean-variance optimization using dynamic inputs from a copula model, which can generate a conditional distribution of returns that captures the time-variation in dependence. The actual implementation of mean-variance optimizations under the only constraint of non-negative weights gives a significant positive allocation to the HML factor, which improves the realized risk-return trade-off. These results are present, but less pronounced for static sample estimators of means and covariances.

Although variance is the staple measure of risk, we also investigate whether risk beyond the first two moments can provide reasons against the HML factor. The copula model has allowed us to infer the full distribution of returns, and we shift our focus to the tail of the distribution. Here, we find that the diversification benefit of HML is similar to that of CMA. HML can by no means be considered a worse diversifier against tail risk, as measured by expected shortfall of a factor portfolio, than CMA.

Our second key finding is that HML is highly similar, but not quite the same as CMA. On a closer look, differences emerge, and we believe that factor investors should combine both factors with consideration. There is important theoretical and empirical support for an overlap in the stock positions that comprise HML and CMA, which results in a substantially higher correlation for this factor pair than for any other factor pair. The dependence between the two factors is also more stable, and does not exhibit the same pattern of tail dependence as do other factor pairs.

Still, one factor cannot replace the other. HML and CMA exhibit different properties, as HML firms are more profitable and exhibit less momentum, and generate return premia that mean-variance optimization suggests useful. Beyond the risk-reward of the first two moments, analysis of the full return distribution from the copula model shows that the diversification benefit of adding either HML or CMA to a portfolio is not constant over time. Sometimes HML is the better diversifier, sometimes CMA is better, but no pattern emerges as to which factor is better than the other. We see no reason for investors to choose either one or the other, as both provide valuable diversification, and even better, they do so at different times.

We believe that investors should consider the two factors jointly when building portfolios. When one of the two factors is included in a factor portfolio already containing the second, the first factor almost exclusively cannibalizes on the weight from the second. Our findings therefore support the existing theoretical and empirical evidence of an overlap in the firms that comprise the two strategies. All in all, we are wary of factor weighting schemes that suggest pure equal-weights for HML and CMA. While such schemes are valuable for factor investing in general, as they avoid the pitfall of factor (mis)timing, they should be designed in a manner that takes the close link between HML and CMA into account. This pair has a very different dependence from all other pairs – and so the allocation policy to these factors should be different.

A third finding of this thesis is the strength of the profitability factor, RMW. This new factor co-varies negatively with most factors and receives zero or negative factor loadings in zero-cost regressions. The exclusion of RMW in our diversification benefit analysis completely pulls the plug on diversification, making periods of low diversification both more frequent and much more severe. Furthermore, the factor receives high allocations and contributes large improvements to all portfolio performance measures. The fact that there are no strong explanations of the profitability factor as a risk premium makes these findings even more puzzling. Our takeaway is that all funds and investors in the factor space should seriously consider adding this new factor.

During the writing of this thesis, we also considered studying additional emerging factors, in particular the low volatility and betting-against-beta factors. At a first glance, we found highly diversifying characteristics of these factors, which remind us of RMW. It would be especially interesting to study their impact on tail risk, following the CDB analysis. In the end, we did not include them as we hone in on the discussion regarding HML's role in the five- and six-factor models.

A substantial part of this thesis is built upon a rather involved copula model. While we are generally comfortable with the estimation procedure and robustness of the model, we acknowledge that it does lack the power to properly explain asymmetries in tail dependence. A natural route for an extension to a more flexible methodology would be to consider vine copulas in place of the multivariate copula we use. We also believe that the advances of regime switching models could prove fruitful in the factor setting, as such models can more rapidly adjust to shocks.

In unreported results, we have also studied out-of-sample investing with factor timing based on the copula model, but find results to be lackluster without frequent re-estimation. While the copula model can ex post shed light on the roles of different factors, it is not useful for a priori portfolio allocation. This is consistent with the preference of money managers to use static weights, but stands in contrast to the work of Christoffersen and Langlois (2013), upon whose work this thesis is largely based.[17] Out-of-sample factor timing is hard, and please note that we do not purport to create a model for investment uses – our contribution is only possible ex post. While MV and CDB analysis based on dynamic weights may seem counter-intuitive at first, as investors use static weights, we argue that the dynamic analysis is a powerful tool for evaluation purposes.

A final thought regarding factor strategies is that we should be careful in interpreting the factors as long-only strategies. While it is highly likely that some of the factors generate alpha both in the long and short positions, Wang and Yu (2013) shows that the profitability factor is mainly due to alpha on the short positions. Therefore, we emphasize that our findings on factors are applicable only directly to the long-short version of the factors.

[17] See i.a. AQR Capital Management, LLC (2016), BlackRock (2016), MSCI (2015) and Robeco (2014).

References

Aas, K., Czado, C., Frigessi, A., & Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and economics*, *44*(2), 182–198.

Aielli, G. P. (2013). Dynamic conditional correlation: on properties and estimation. *Journal of Business & Economic Statistics*, *31*(3), 282–299.

Anderson, C. W. & Garcia-Feijóo, L. (2006). Empirical evidence on capital investment, growth options, and security returns. *The Journal of Finance*, *61*(1), 171–194.

Ang, A. & Chen, J. (2002). Asymmetric correlations of equity portfolios. *Journal of Financial Economics*, *63*, 443–494.

AQR Capital Management, LLC. (2015). *Building a better equity market neutral strategy*. Retrieved October 11, 2016, from https://www.aqr.com/~/media/files/papers/aqr-building-a-better-equity-market-neutral-strategy.pdf

AQR Capital Management, LLC. (2016). *Resisting the siren song of factor timing*. Retrieved November 25, 2016, from https://www.aqr.com/~/media/files/perspectives/resisting-the-siren-song-of-factor-timing.pdf

Artzner, P., Delbaen, F., Eber, J.-M., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, *9*(3), 203–228.

Asness, C. S., Frazzini, A., Israel, R., & Moskowitz, T. J. (2015). Fact, fiction, and value investing. *Forthcoming, Journal of Portfolio Management, Fall.*

Asness, C. S., Moskowitz, T. J., & Pedersen, L. H. (2013). Value and momentum everywhere. *The Journal of Finance*, *68*(3), 929–985.

Barberis, N. & Huang, M. (2001). Mental accounting, loss aversion, and individual stock returns. *The Journal of Finance*, *56*(4), 1247–1292.

Black, F. (1976). Studies of stock price volatility changes. *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economical Statistics Section*, 177–181.

BlackRock. (2016). *Smart beta guide*. Retrieved November 25, 2016, from https://www.blackrock.com/ca/intermediaries/en/literature/brochure/smart-beta-guide-en-ca.pdf?nc=true&siteEntryPassthrough=true

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327.

Brunnermeier, M. K. (2009). Deciphering the liquidity and credit crunch 2007–2008. *The Journal of Economic Perspectives*, *23*(1), 77–100.

Brunnermeier, M. K. & Pedersen, L. H. (2009). Market liquidity and funding liquidity. *Review of Financial Studies*, *22*(6), 2201–2238.

Chen, X. & Fan, Y. (2006). Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. *Journal of Econometrics*, *135*(1), 125–154.

Chollete, L. & Ning, C. (2012). Asymmetric dependence between aggregate consumption and financial risk. *Unpublished working paper, University of Stavanger and Ryerson University*.

Christoffersen, P., Errunza, V., Jacobs, K., & Langlois, H. (2012). Is the potential for international diversification disappearing? A dynamic copula approach. *Review of Financial Studies*, *25*(12), 3711–3751.

Christoffersen, P. & Langlois, H. (2013). The joint dynamics of equity market factors. *Journal of Financial and Quantitative Analysis*, *48*(05), 1371–1404.

Cooper, M. J., Gulen, H., & Schill, M. J. (2008). Asset growth and the cross-section of stock returns. *The Journal of Finance*, *63*(4), 1609–1651.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, *20*(3), 339–350.

Engle, R. F. & Kroner, K. F. (1995). Multivariate simultaneous generalized arch. *Econometric theory*, *11*(01), 122–150.

Engle, R. F. & Mezrich, J. (1995). Grappling with GARCH. *Risk*, *8*(9), 112–117.

Engle, R. F. & Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *The Journal of Finance*, *48*(5), 1749–1778.

Fama, E. F. & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56.

Fama, E. F. & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, *116*(1), 1–22.

Fisher, T. J. & Gallagher, C. M. (2012). New weighted portmanteau statistics for time series goodness of fit testing. *Journal of the American Statistical Association*, *107*(498), 777–787.

French, K. R. (2016). Kenneth French data library. Retrieved December 6, 2016, from http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, *48*(5), 1779–1801.

Gonçalves, S. & White, H. (2004). Maximum likelihood and the bootstrap for nonlinear dynamic models. *Journal of Econometrics*, *119*(1), 199–219.

Hansen, B. E. (1994). Autoregressive conditional density estimation. *International Economic Review*, 705–730.

Ilmanen, A. (2011). *Expected returns: An investor's guide to harvesting market rewards*. John Wiley & Sons.

Jegadeesh, N. & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of finance*, *48*(1), 65–91.

Jensen, M. C. (1968). The performance of mutual funds in the period 1945–1964. *The Journal of Finance*, *23*(2), 389–416.

Joe, H. (1997). *Multivariate models and multivariate dependence concepts*. CRC Press.

Khandani, A. & Lo, A. (2007). What happened to the quants in august 2007? *Journal of Investment Management*, *5*(4), 29–78.

Khandani, A. E. & Lo, A. W. (2011). What happened to the quants in August 2007? Evidence from factors and transactions data. *Journal of Financial Markets*, *14*, 1–46.

Lakonishok, J., Shleifer, A., & Vishny, R. W. (1994). Contrarian investment, extrapolation, and risk. *The Journal of Finance*, *49*(5), 1541–1578.

Li, W. & Mak, T. (1994). On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity. *Journal of Time Series Analysis*, *15*(6), 627–636.

Liew, J. & Vassalou, M. (2000). Can book-to-market, size and momentum be risk factors that predict economic growth? *Journal of Financial Economics*, *57*(2), 221–245.

Ljung, G. M. & Box, G. E. (1978). On a measure of lack of fit in time series models. *Biometrika*, *65*(2), 297–303.

Longin, F. & Solnik, B. (2001). Extreme correlation of international equity markets. *The Journal of Finance*, *56*(2), 649–676.

McKinsey & Company. (2014). *The trillion-dollar convergence: Capturing the next wave of growth in alternative investments*. Retrieved October 11, 2016, from http://www.mckinsey.com/industries/financial-services/our-insights/the-trillion-dollar-convergence

MSCI. (2015). *The msci diversified multifactor indexes: Maximizing factor exposure while controlling volatility*. Retrieved November 25, 2016, from https://www.msci.com/documents/10199/a49f25c5-982e-40a9-a0da-11ea6118649a

Newey, W. K. & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*(3), 703–708.

Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics*, *108*(1), 1–28.

Patton, A. J. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review*, *47*(2), 527–556.

Patton, A. J. (2012). Copula methods for forecasting multivariate time series. *Handbook of economic forecasting*, *2*, 899–960.

Pedersen, L. H. (2015). *Efficiently inefficient: How smart money invests and market prices are determined*. Princeton University Press.

Petkova, R. & Zhang, L. (2005). Is value riskier than growth? *Journal of Financial Economics*, *78*(1), 187–202.

Politis, D. N. & Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical association*, *89*(428), 1303–1313.

Rémillard, B. (2010). Goodness-of-fit tests for copulas of multivariate time series. *Unpublished working paper*.

Robeco. (2014). *Robeco insight: How factor investing fits into active vs passive*. Retrieved November 25, 2016, from https://www.robeco.com/en/professionals/insights/quantitative-investing/factor-investing/seminar/how-factor-investing-fits-into-active-vs-passive.jsp

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, *6*(2), 461–464.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut Statistique de l'Université de Paris*, *8*, 229–231.

Wang, H. & Yu, J. (2013). Dissecting the profitability premium. In *Afa 2013 san diego meetings paper*.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the Econometric Society*, 1–25.

Zhang, L. (2005). The value premium. *The Journal of Finance*, *60*(1), 67–103.

Appendix A Skewed Student's t distribution and copula density

We use the skewed Student's t distribution in modeling both univariate series as well as for the joint distribution under the copula. We use the definition of Hansen (1994) and the following description is based on Christoffersen and Langlois (2013). A random vector X that follows a multivariate skewed t distribution has the stochastic description

$$X = \sqrt{W}Z + \gamma W \tag{A.1}$$

where γ is a vector of asymmetry parameters, Z follows a standard multivariate normal distribution with correlation matrix Ψ, W follows an inverse gamma distribution $\mathrm{IG}(\frac{\nu}{2}, \frac{\nu}{2})$. Thus, the parameters of the multivariate distribution are degrees of freedom ν, asymmetries γ and an underlying correlation matrix Ψ.

The distribution has expectation and covariance matrix:

$$\mathbb{E}[X] = \frac{\nu}{\nu - 2}\gamma \tag{A.2}$$

$$\mathrm{Cov}(X) = \frac{\nu}{\nu - 2}\Psi + \frac{2\nu^2\gamma\gamma^\top}{(\nu - 2)(\nu - 4)} \tag{A.3}$$

i.e. $\nu \geq 4$ for these to be well-defined. Note that if $\gamma = 0$ (element-wise), X follows a multivariate symmetric t distribution, and additionally if $\nu = \infty$, X follows a multivariate standard normal distribution. Hypotheses $\gamma = 0$ and $1/\nu = 0$ can therefore be used to test for symmetry and normality, respectively.

The copula joint density function c_t always takes the form of the ratio between a joint density function $f_t^c(z_{t+1})$ (i.e. the multivariate normal, symmetric t or skewed t PDF, respectively) of copula shocks z_{t+1}^* and the product of the univariate density functions $f_{i,t}^c(z_{i,t+1})$ (i.e. the univariate normal, symmetric t or skewed t PDF, respectively) of the individual shocks $z_{i,t+1}^*$:

$$c_t(U_{t+1}) = \frac{f_t^c(z_{t+1}^*)}{\prod_{i=1}^N f_{i,t}^c(z_{i,t+1}^*)} \tag{A.4}$$

where the relationship between $z_{i,t+1}^*$ and $u_{i,t+1}$ is governed by the inverse cumulative distribution function, as detailed in the subsequent appendix. Note that if the copula distribution and marginal distributions are the same, the denominator cancels in Equation 5.1 and the copula is directly the joint distribution of the marginal densities.

Appendix B Copula Estimation Procedure

This is a step-by-step description of the procedure used to estimate the copula model, given the set of standardized residuals $\{z_t\}$ from each GARCH model. It is adapted from Christoffersen et al. (2012) and uses the cDCC model of Aielli (2013) (where cDCC stands for *corrected* DCC).

Each week, compute uniform residuals by applying the probability integral transform to standardized residuals from each GARCH model:

$$u_{i,t+1} = \int_{-\infty}^{z_{i,t+1}} f_i(z)dz = \int_{-\infty}^{r_{i,t+1}} f_{i,t}(r)dr \tag{B.1}$$

Note that while the distributions of returns is time-varying due to GARCH dynamics, the distribution of standardized residuals is assumed constant, and in the case of skewed t is parameterized by the shape ν_i and skewness parameter γ_i (estimated as part of the GARCH models).

Transform the uniform residuals into copula residuals $z^*_{i,t+1}$ by applying the inverse cumulative distribution function of the copula to them:

$$z^*_{i,t+1} = F^{-1}_{\nu_c,\gamma_i}(u_{i,t+1}) \tag{B.2}$$

Only under the normal copula will these residuals have expectation zero and unit variance – hence, they are standardized by subtracting the expectation and dividing by the standard deviation of the distribution from Appendix A.

These shocks are now used to fit the corrected DCC process of Aielli (2013). The correction involves the transformation $\bar{z}^*_{i,t+1} = z^*_{i,t+1}\sqrt{q_{ii,t}}$, where $q_{ii,t}$ are the diagonal elements of Q_t and are found by a scalar version of Equation 5.4:

$$q_{ii,t} = (1-\alpha-\beta) + \alpha(\bar{z}^*_{i,t-1})^2 + \beta q_{ii,t-1} \tag{B.3}$$

The corrected shocks are used to estimate the time-invariant component Q:

$$\hat{Q} = \frac{1}{T}\sum_{t=1}^{T} \bar{z}^*_t \bar{z}^{*\top}_t \tag{B.4}$$

Now, the full estimates of $\hat{Q}_t$ are computed using the sample estimate $\hat{Q}$ and the corrected shocks:

$$\hat{Q}_t = (1-\alpha-\beta)\hat{Q} + \beta\hat{Q}_{t-1} + \alpha\bar{z}^*_{t-1}\bar{z}^{*\top}_{t-1} \tag{B.5}$$

The estimated $\hat{Q}_t$ matrices are standardized to estimates $\hat{\Psi}_t$ of the conditional correlation matrices of the copula using Equation 5.5. When fitting the model, we thus choose parameters $\nu_c, \gamma_c, \alpha, \beta$ to generate $\hat{\Psi}_t$ which maximize the log-likelihood of observing copula shocks z^*_t in each period.

Appendix C Univariate diagnostic tests

Autocorrelation test

The autocorrelation test is a weighted Ljung-Box test, following Fisher and Gallagher (2012) and Ljung and Box (1978). Under the null of a correctly specified model with no serial correlation, the weighted Ljung-Box test has been shown to generate results closer to its asymptotic distribution than the standard Ljung-Box test. The test statistic is given by

$$Q_W = T(T+2)\sum_{k=1}^{m} \frac{m-k+1}{m} \frac{\hat{r}_k^2(\hat{\varepsilon}_t/\hat{\sigma}_t)}{T-k} \tag{C.1}$$

where T is the number of observations, $\hat{r}_k^2(\hat{\varepsilon}_t/\hat{\sigma}_t)$ is the squared sample autocorrelation of standardized residuals with lag order k and max lag order m. Under the null, the test statistic is asymptotically distributed $\sum_{k=1}^{m} \chi_k^2 \gamma_k$, where $\{\chi_k^2\}$ are independent chi-squared random variables with one degree of freedom and $\{\gamma_k\}$ are eigenvalues of a weighting matrix. We consider two maximum lag orders, 5 and 10 weeks. The maximum lag length was chosen by visual inspection of the autocorrelation functions for standardized residuals.

Volatility clustering test

For ARCH effects, we use the weighted LM test, following Fisher and Gallagher (2012) and Li and Mak (1994). The test has the null of no autocorrelation in standardized squared residuals from the model, and the test statistic is given by:

$$LM_W = T\sum_{k=b+1}^{m} \frac{m-k+(b+1)}{m} \hat{r}_k^2(\hat{\varepsilon}_t^2/\hat{\sigma}_t) \tag{C.2}$$

where T is the number of observations, b the number of autoregressive lags in the GARCH ($b=1$), $\hat{r}_k^2(\hat{\varepsilon}_t^2/\hat{\sigma}_t)$ is the squared sample autocorrelation of standardized squared residuals with lag order k and max lag order m. Under the null, the test statistic is asymptotically distributed $\sum_{k=1}^{m} \chi_k^2 w_k$, where $\{\chi_k^2\}$ are independent chi-squared random variables with one degree of freedom and $\{w_k\}$ are the weighting parameters ($w = (m-k+(b+1))/m$). The maximum lag length was chosen by visual inspection of the autocorrelation functions for standardized squared residuals.

Leverage effect test

We use the sign bias test of Engle and Ng (1993) to determine whether there are significant leverage effects in the factor returns. Run the regression

$$\hat{z}_t^2 = c_0 + c_1 I_{\hat{\varepsilon}_{t-1}<0} + c_2 I_{\hat{\varepsilon}_{t-1}<0} \cdot \hat{\varepsilon}_{t-1} + c_3 I_{\hat{\varepsilon}_{t-1}\geq 0} \cdot \hat{\varepsilon}_{t-1} + u_t \tag{C.3}$$

where $\hat{z}_t^2$ are the standardized squared residuals of the ARMA-GARCH model, $I_{\cdot}$ are indicator functions that are equal to one when the subscript conditions are true, and $\hat{\varepsilon}_{t-1}$ are the lagged ARMA-GARCH residuals. For the test of negative sign bias (i.e. leverage effect), the null hypothesis is $H_0 : c_2 = 0$, and for the test of positive sign bias (i.e. reverse leverage effect), the null hypothesis is $H_0 : c_3 = 3$. The Wald test statistics are asymptotically distributed χ^2 with one degree of freedom.

Appendix D Stationary bootstrap of copula parameter standard errors

We rely on the multi-step maximum likelihood estimation of the copula model, which takes the standardized residuals of marginal distributions as given in the second step. The first estimation step introduces parameter uncertainty that is not taken into account by the conventional standard errors of the second estimation.[18] We use the stationary block bootstrap method of Politis and Romano (1994) with a block length of 104 weeks (2 years of data) to find reliable standard errors for copula parameters. The procedure is theoretically supported by Gonçalves and White (2004) and implemented as follows (as described in Patton (2012)):

(i) Generate a stationary block bootstrap of the original weekly return data with an expected block length of 104 weeks (note that individual block lengths are random).

(ii) Estimate the copula model of interest and collect the parameter set θ_i.

(iii) Repeat (i)-(ii) S times (we use $S = 100$).

(iv) Use the standard deviation of the distribution of $\{\theta_i\}_{i=1}^{S}$ as the standard error for the parameters.

[18] Here, our model deviates from Christoffersen and Langlois (2013), who use a semi-parametric model that uses the empirical density function, and find standard errors using the analytical approach in Chen and Fan (2006). However, those errors are not valid in a time-varying copula context, as the estimation of means and variances impact the asymptotic distributions of copula parameters (Rémillard, 2010).

Appendix E Additional Figures and Tables

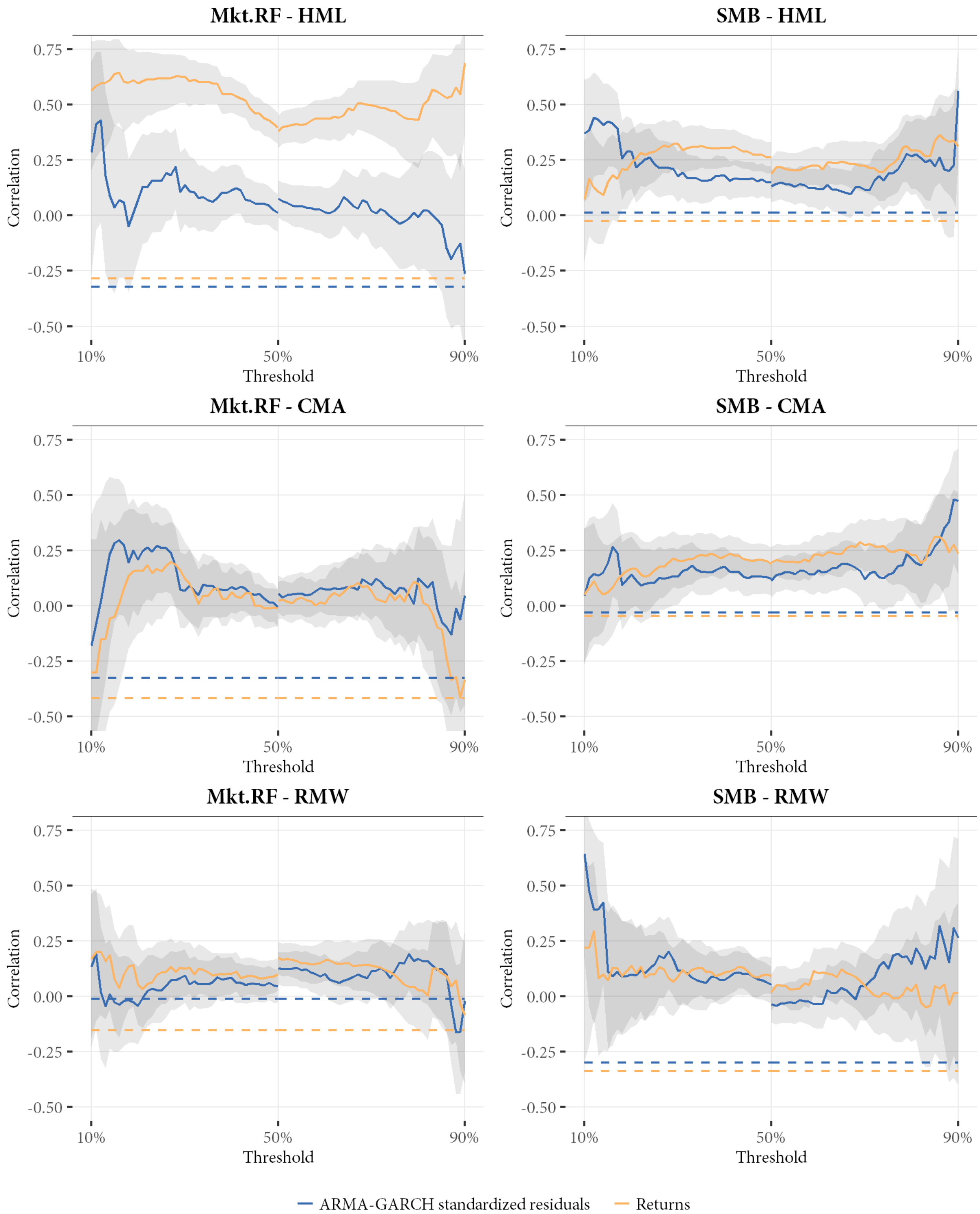


Figure 16: Threshold correlations of returns and ARMA-GARCH standardized residuals

The formula for threshold correlations for a threshold p is given in Equation 5.10. 95% shaded confidence bounds, taking the model as given. The unconditional correlation is given by the dashed line. Based on weekly data 1963–2016.

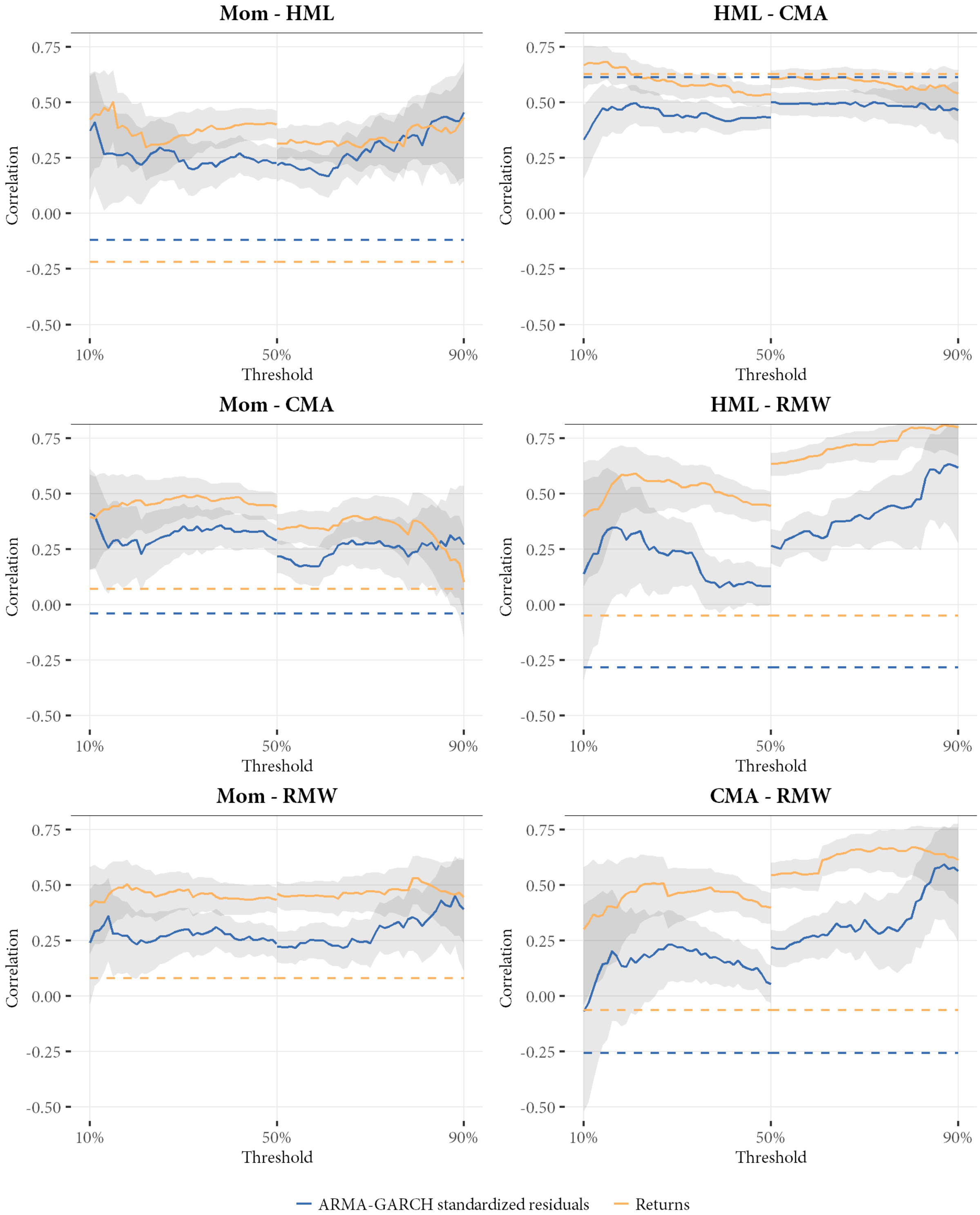


Figure 16: Threshold correlations of returns and ARMA-GARCH standardized residuals (cont.)

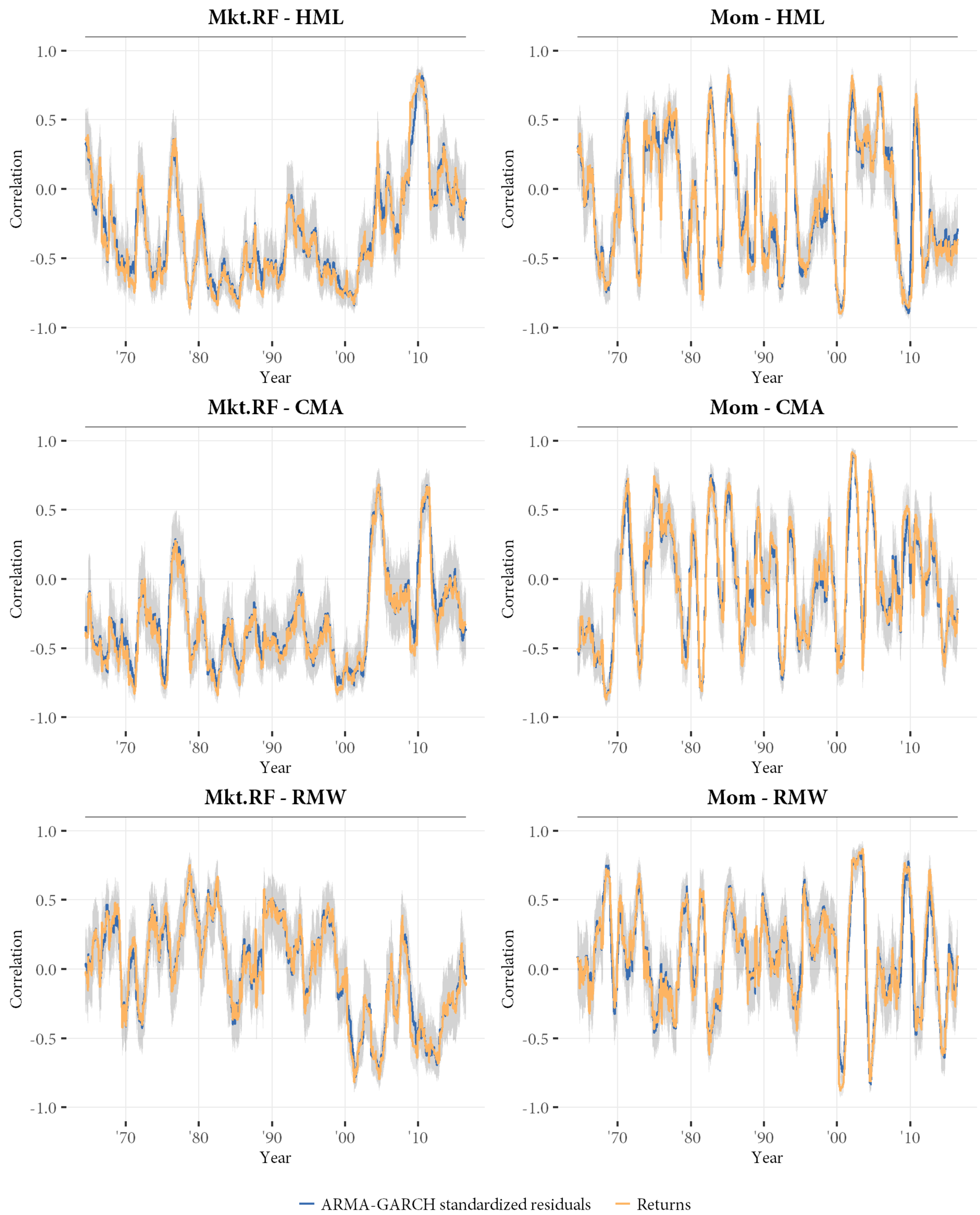


Figure 17: Rolling correlations of returns and ARMA-GARCH standardized residuals

95% shaded confidence bounds, taking the model as given. The unconditional correlation is given by the dashed line. Based on weekly data 1963–2016.

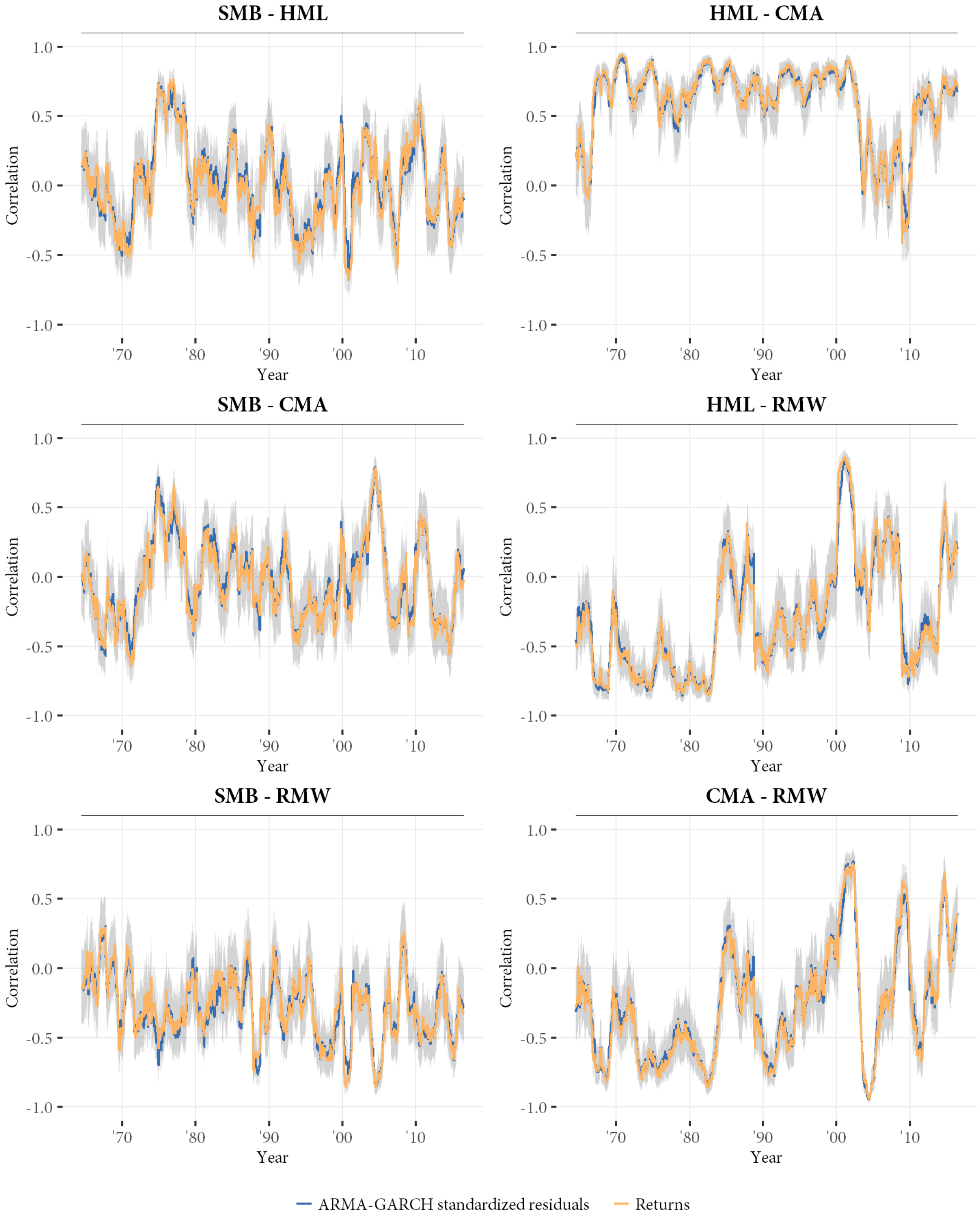
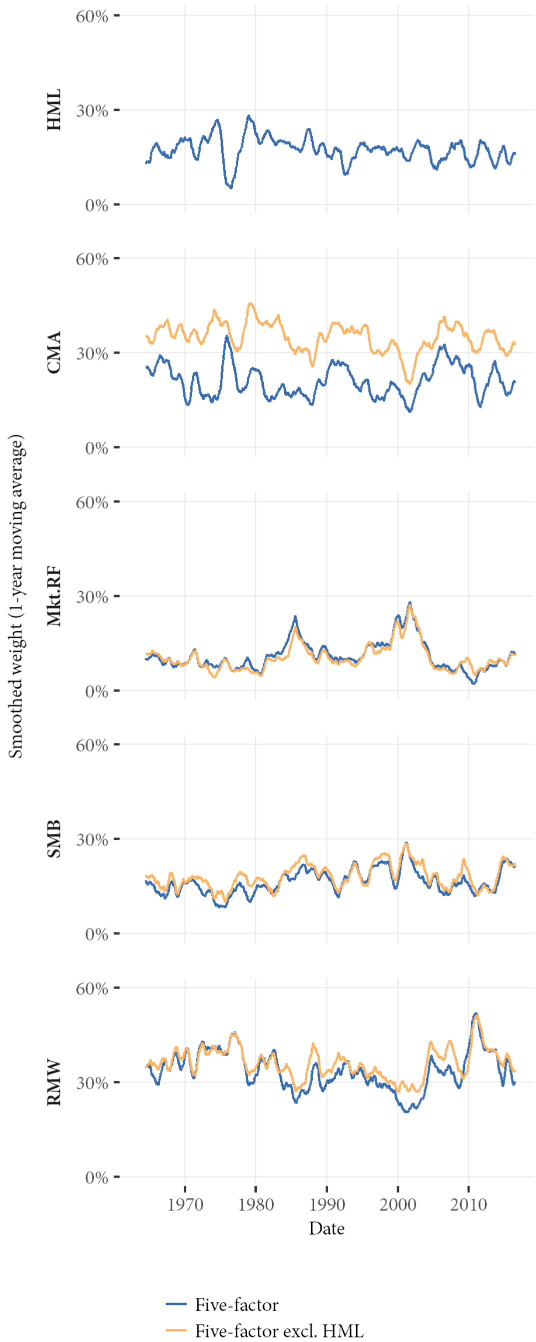


Figure 17: Rolling correlations of returns and ARMA-GARCH standardized residuals (cont.)

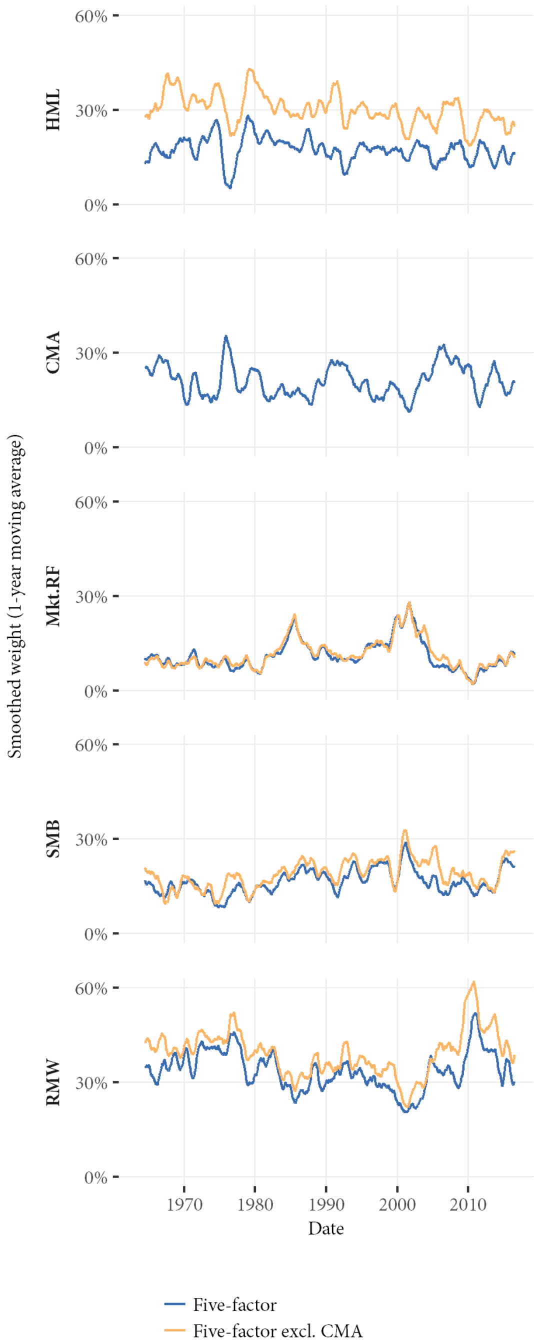
Table 8: Mean-Variance optimization with static sample inputs (1963–2016)

Static weights are the MV optimal weights based on in-sample sample estimators of means and covariances. Differences in average weights are expressed relative to the full five- and six-factor models. Performance measures are based on realized returns. SR is the annualized Sharpe Ratio. VaR, ES and CDB are all based on the one-week-ahead 5% lower tail of the return distribution, which is given by simulations from the copula model. Differences in CDB are to be read as column model minus row model and its associated standard errors (in parentheses) are computed taking the copula model as given.

	Five (four) factor models				Six (five) factor models			
	All	Excl. HML	Excl. CMA	Excl. RMW	All	Excl. HML	Excl. CMA	Excl. RMW
Static weights								
Mkt.RF	13.6	13.7	13.8	21.5	13.2	13.4	13.2	18.5
SMB	14.0	14.1	18.0	12.0	12.0	12.7	13.4	9.1
HML	6.2		26.7	11.2	13.2		27.0	22.3
CMA	33.2	39.0		55.3	22.2	35.2		29.2
RMW	33.0	33.2	41.5		27.7	29.0	30.2	
Mom					11.7	9.7	16.2	21.0
Difference weights (column minus All)								
Mkt.RF		0.1	0.2	7.9		0.2	0.0	5.3
SMB		0.2	4.0	−2.0		0.7	1.4	−2.9
HML		−6.2	20.5	5.0		−13.2	13.7	9.1
CMA		5.8	−33.2	22.1		13.0	−22.2	7.0
RMW		0.3	8.5	−33.0		1.4	2.6	−27.7
Mom						−2.0	4.5	9.3
Performance								
Mean (%)	3.71	3.68	3.72	4.14	4.32	4.14	4.56	5.10
SD (%)	2.93	2.94	3.46	4.32	2.99	2.99	3.35	4.19
SR	1.27	1.25	1.07	0.96	1.44	1.38	1.36	1.22
Avg. VaR (%)	0.56	0.57	0.63	0.89	0.57	0.59	0.64	0.84
Avg. ES (%)	0.76	0.78	0.86	1.20	0.79	0.81	0.88	1.16
Avg. CDB	87.34	86.68	86.45	79.15	87.87	87.02	86.77	82.63
Difference CDB (column model minus row model)								
All		−0.66	−0.89	−8.19		−0.85	−1.10	−5.24
		(0.01)	(0.05)	(0.13)		(0.03)	(0.04)	(0.10)
Excl. HML			−0.22	−7.53			−0.22	−4.40
			(0.06)	(0.13)			(0.06)	(0.11)
Excl. CMA				−7.31				−4.14
				(0.14)				(0.10)



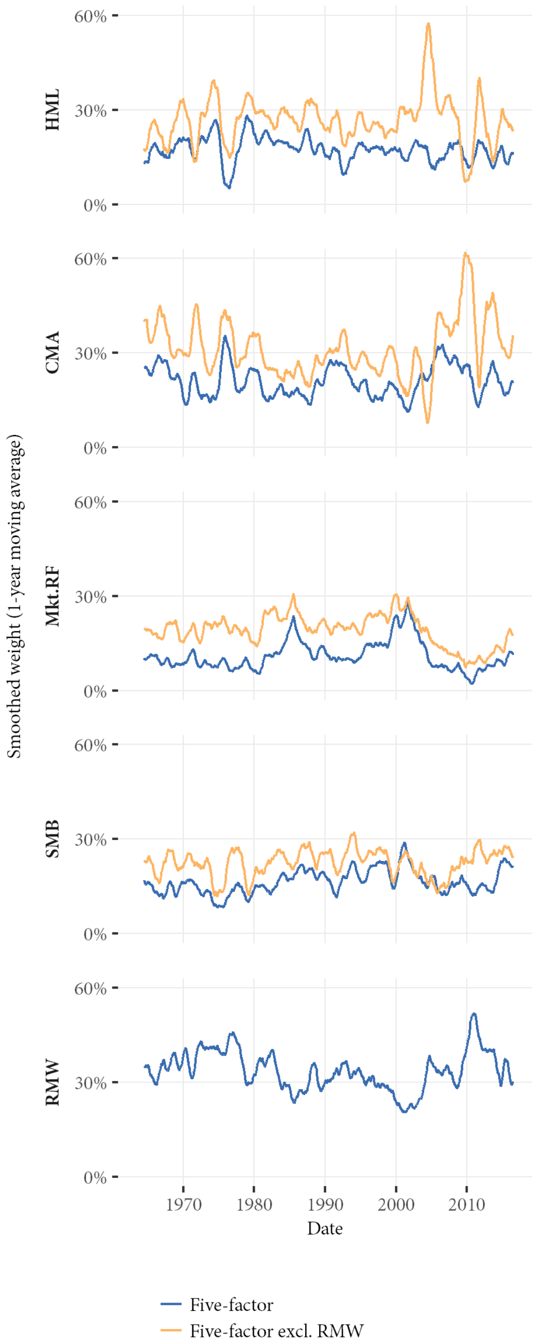
(a) Excluding HML



(b) Excluding CMA

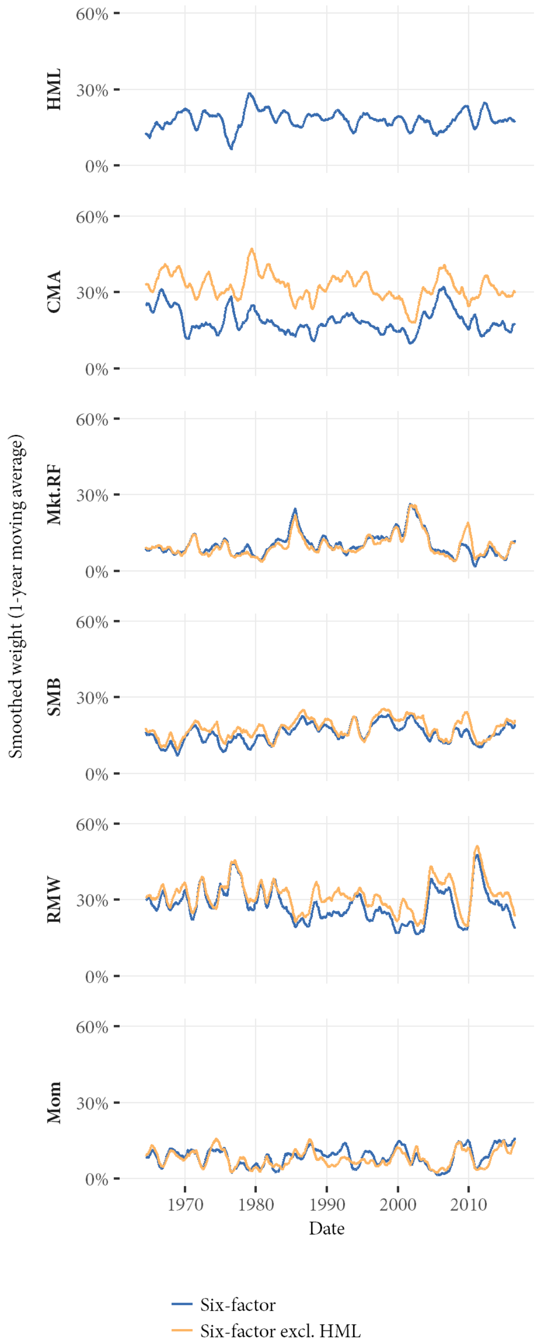
Figure 18: CDB optimal weights with five factors

Smoothed as 1-year moving averages for better legibility. Optimization constrained to fully invested portfolios with non-negative weights. Left hand panel including and excluding HML, right hand including and excluding CMA. Based on one-week-ahead forecasts from the dynamic symmetric t copula model 1963–2016.

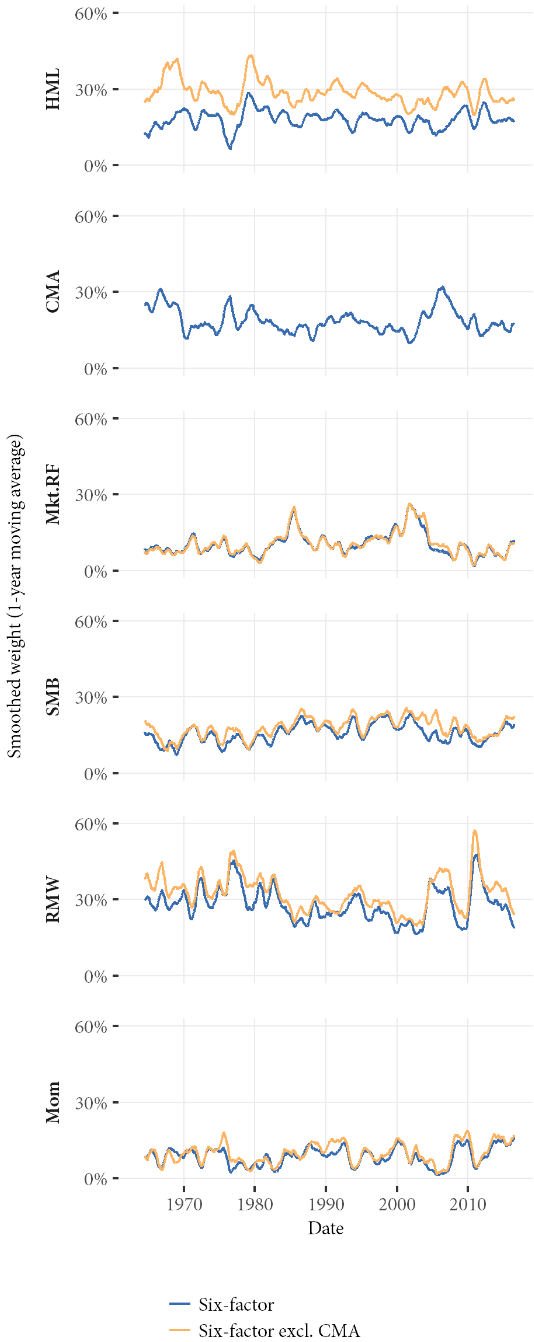


(c) Excluding RMW

Figure 18: CDB optimal weights with five factors (cont.)



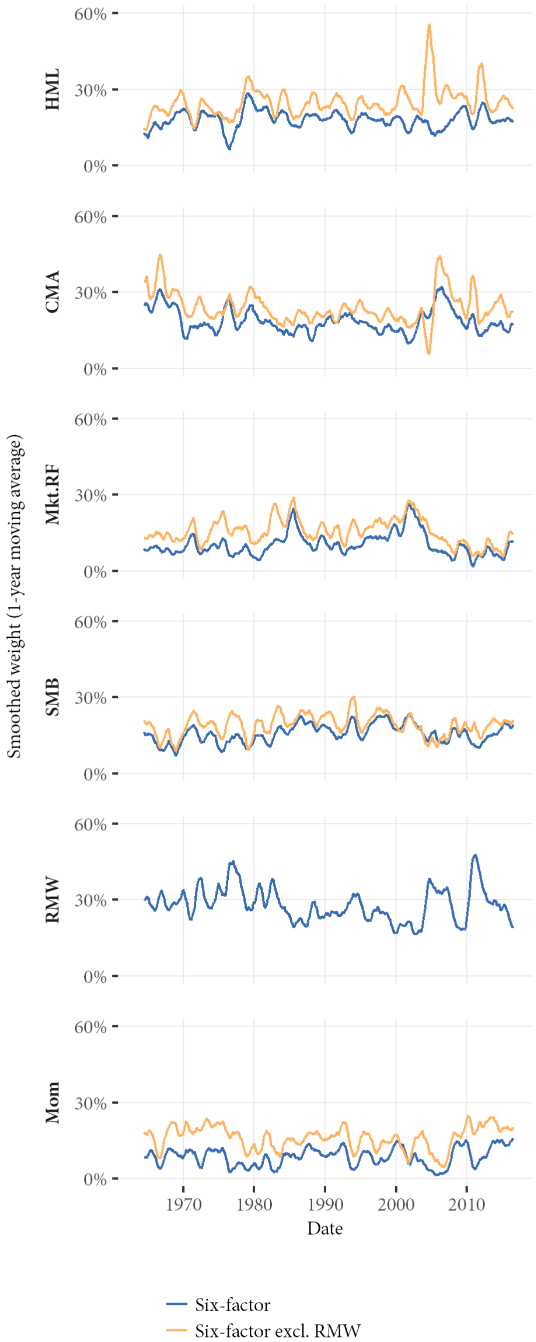
(a) Excluding HML



(b) Excluding CMA

Figure 19: CDB optimal weights with six factors

Smoothed as 1-year moving averages for better legibility. Optimization constrained to fully invested portfolios with non-negative weights. Left hand panel including and excluding HML, right hand including and excluding CMA. Based on one-week-ahead forecasts from the dynamic symmetric *t* copula model 1963–2016.



(c) Excluding RMW

Figure 19: CDB optimal weights with six factors (cont.)