Examining the Risk-Return Relationship Using the MF2-GARCH Model

Master's thesis

to obtain the academic degree "Master of Science" (M.Sc.)

submitted

to the Examination Board for the Master's program

Economics

the

Faculty of Economics and Social Sciences of the

Ruprecht-Karls-University Heidelberg

Year

2025

Hussam Elhamy Hamed Shaker **Elnashar**, born in

Kuwait City, Kuwait on 23.05.1998

Summary

This thesis investigates whether the multiplicative factor multi-frequency component GARCH (MF2-GARCH) model of Conrad & Engle (2025) supports earlier findings that the long-term component of market volatility holds more predictive power for market premia than overall volatility.

I create a combined "MF2-GARCH-in-mean" model, placing the components of MF2-GARCH volatility in a risk-return specification similar to those of Maheu & McCurdy (2007) and estimating the parameters for historical data on U.S. market premia. Following Maheu & McCurdy (2007), I test different specifications, varying the choice of components included in the model. To further test the viability of the model, I run several thousand Monte Carlo simulations and record the average biases and standard deviations of the MF2-GARCH-in-mean estimates on this simulated data.

The results show that the long-term component of MF2-GARCH market volatility does better to explain variation in market premia across several specifications and that its coefficients are consistently positive and highly significant. The Monte Carlo simulations show that the parameter estimates of the MF2-GARCH-in-mean model have only moderate bias.

1   Introduction

In his seminal paper, Merton (1973) develops an intertemporal version of the Capital Asset Pricing Model (CAPM). By assuming that the risk-free rate, expected excess returns and volatilities follow stochastic processes, he derives equilibrium asset-pricing relations that link assets' expected excess returns to covariances with the market and investors' hedge demands against changes in the investment opportunity set.

When applying this framework to the market (rather than individual assets), he shows that excess market return can be written as a linear combination of its conditional variance, and its covariance with a vector of state-variables which predict changes in the future investment opportunity set[1]:

$$E_{t-1}r_{M,t} - r_{f,t} = \gamma_M \sigma_{M,t-1}^2 + \gamma_F \sigma_{MF,t-1}$$

where at time t:

- $r_{M,t}$ is the market return,

- $r_{f,t}$ is the risk free rate,

- $\sigma_{M,t-1}^2$ is the conditional market return variance in the previous period,

- and $\sigma_{MF,t-1}$ is the covariance of market return in the previous period with a vector of state-variables that predict future investment opportunities

γ_M describes the amount of expected return per unit of market variance which the investor demands. The second term, $\gamma_F \sigma_{MF,t-1}$, is frequently described as the "hedging component", as it describes the proportion of excess market return which investors demand to hedge against changes in future investment opportunities.

Merton (1980) shows that under one of two simplifying assumptions, the hedging term loses explanatory power/no longer demands a premium and can therefore be safely ignored. These assumptions are that:

1. there is no change in the investment opportunity set over time, so that $\gamma_F = 0$, or

---

[1]Restated with simplified coefficients for readability

2. investors are myopic in their preferences (or have logarithmic utility), such that they only care about the next instant's consumption

Under either assumption, γ_M captures the investor's entire expected return per unit of risk, making it equivalent to the coefficient of relative risk aversion by definition.

It is widely accepted that relative risk aversion is positive because investors dislike risk and therefore demand compensation for bearing it. Consequently, when examining the relationship between excess market return and market volatility, we should expect to find it to be positive.

Demonstrating a statistically significant and positive risk-return tradeoff has important implications for both asset allocation and portfolio optimization. In the long run, allocation decisions are guided by investors' risk tolerances and return objectives. A positive long-term risk-return relationship implies that achieving higher return targets requires accepting greater volatility over long horizons. Short-term allocation involves exploitation of perceived mispricings and market risk conditions. If there is no robust relationship between returns and risk in the short-run, this endeavour becomes less viable.

This thesis aims to find that positive risk-return relationship by modelling risk (in this case, conditional variance) using MF2-GARCH.

2 Literature Review

2.1 Background

Early efforts to quantify the risk-return relationship frequently returned negative and/or statistically insignificant results, however. Various methods have been explored in an effort to find the expected positive result.

Guo & Whitelaw (2006) develop an empirical model relating expected return to risk and the hedging component. They find that omission of the hedge component by previous researchers is partly responsible for the apparent negative relationship between expected return and the risk component, for which they find a positive and

significant coefficient.

Kim et al. (2004) implement Markov-switching market volatility, finding a negative and significant volatility feedback effect (implying a positive relationship between market volatility and the equity premium). Their analysis provides evidence for volatility regime-switching and therefore informs future analysis of the risk-return relationship, suggesting that regime-switching should be accounted for.

Lundblad (2007) finds a positive tradeoff over a long time period, but comes to the conclusion that capturing the positive risk-return relationship requires much larger samples than are normally used. He also recognizes that volatility regime-switching occurs in crises.

What has motivated this thesis is that more recently, several papers have found success when measuring or modeling risk over a longer horizon. One core issue in exploring the risk-return relationship is that variance (or risk/volatility) is not directly observable. The question of how to model market volatility is salient as a result.

Guo & Neely (2007) find that the risk-return relationship is positive and significant when risk is measured by the persistent/long-term component of the Component-GARCH (CGARCH) model (Engle & Lee, 1999). The CGARCH model separates total conditional variance into two additive components. The CGARCH(1,1) specification, for example, would look as follows:

$$r_t = \mu + \epsilon_t$$

$$\sigma_t^2 = q_t + \alpha(\epsilon_{t-1}^2 - \sigma_{t-1}^2) + \beta(\sigma_{t-1}^2 - q_{t-1})$$

$$q_t = \omega + pq_{t-1} + \varphi(\epsilon_{t-1}^2 - \sigma_{t-1}^2)$$

where:

- r_t and μ are the return at time t and its mean, respectively

- $\epsilon_t|I_{t-1} \sim N(0, \sigma_t^2)$ is the return innovation

- σ_t^2 is the total conditional variance

- q_t is the persistent/long-term variance

3

- $(\epsilon_{t-1}^2 - \sigma_{t-1}^2)$ is the usual GARCH innovation (news)

The MF2-GARCH model, which is at the center of this thesis, similarly separates variance into a short- and long-term component, but does so multiplicatively.

Similarly, Ghysels et al. (2005) use a mixed data sampling (MIDAS) approach to investigate this relationship, finding that short-term windows of measurement yield insignificant or even negative volatility coefficients, and that increasing the MIDAS window to the medium-term (3-4 months) flips the volatility coefficient to positive and statistically significant, with declining coefficients and model fit for windows larger than 6 months.

Maheu & McCurdy (2007) propose a parsimonious volatility model that allows different components (short- and long-term) to decay at different rates, using it to estimate the relationship between excess market return and market return volatility. They use a realized volatility (RV) approach, and their volatility components are weighted sums of past RV values. All of their specifications show evidence of a positive risk-return relationship, and they find higher equity premium fits for models which price the smooth component of volatility in the conditional mean. I use a risk-return specification in this thesis that is similar to their univariate specifications.

2.2 Contribution

Given that the long-term component of volatility appears to hold more predictive power for return, in this thesis I introduce MF2-GARCH volatility to the univariate risk-return framework of Maheu & McCurdy (2007).

In the resulting combined MF2-GARCH-in-mean specification, the conditional mean of excess market return is modeled using the short- and long-term components of MF2-GARCH volatility. I estimate four risk-return specifications, two where the conditional mean is regressed on only one of the components, one where it is regressed on both, and one where it is regressed on overall conditional variance.

The novelty in this process is in the way that MF2-GARCH models the long-term component of volatility. The MF2-GARCH model takes advantage of the empirical fact that rolling window moving averages of the standardized forecast errors of one-

component GARCH models behave counter-cyclically and have predictive power for future standardized forecast errors (Conrad & Engle, 2025). Simple GARCH models do not accurately capture these counter-cyclical movements. MF2-GARCH does so explicitly.

The question of including a constant/intercept in the regression of return on risk is also material. Christoffersen et al.'s (2006) findings show that estimating a proportional model with no intercept strengthens evidence of a positive risk-return relationship. For this reason, I estimate a proportional (no-intercept) and non-proportional version of each specification and, following Guo & Neely (2007), I implement a likelihood ratio test to see whether including an intercept adds to the explanatory power of the model.

As touched upon above, there is evidence of volatility "regime-switching" in equity markets, especially during crises. Ghysels et al. (2016) use a "flight to safety" indicator variable to exclude crisis periods from their analysis of the risk-return relationship using the MIDAS approach. They find a significant and positive relationship between return and their MIDAS estimator during the "normal" regime, but a reversal of this relationship in the "crisis" regime.

The intuition is that crises cause investors to move capital to safe haven assets of lower long-term volatility. Following their example, I use a binary dummy variable to control for crisis periods in this thesis. The exact dates of the crisis periods are informed by the recession periods defined by the National Bureau of Economic Research (NBER). I also show parameter estimates without the crisis dummy variable to illustrate its effect.

The parameter estimates that I obtain show that the short-term component of MF2-GARCH volatility only has a significant (and sometimes negative) coefficient in limited scenarios. On the other hand, the long-term component shows significance across all specifications. It is also positively related to return in all cases. In some cases, the crisis period dummy variable also points to components of volatility losing importance during crisis periods, which is consistent with flight to safe haven assets by investors.

Finally, I use Monte Carlo simulations to generate daily market premia and esti-

mate the bias of the model over many iterations to examine the bias of the model's parameter estimates. For each specification, I generate a sample of 30,240 trading days, (or 120 trading years) of data. This is motivated by Lundblad (2007)'s finding that small samples can significantly distort the estimated relationship between risk and return.

3   Econometric Model

3.1   The MF2-GARCH Model

Volatility here has two multiplicative components: short- (h_t) and long-term (τ_t)

$$\sigma_t^2 = h_t \tau_t$$

Daily stock returns are written as:

$$r_t = \mu_t + \sigma_t Z_t$$

$$= \mu_t + \sqrt{h_t \tau_t} Z_t$$

where μ is the unconditional mean of returns and Z_t are return innovations (assumptions about these innovations are detailed below).

The short-term component is modeled as a GJR-GARCH(1,1) process:

$$h_t = (1 - \phi) + (\alpha + \gamma 1_{\{r_{t-1}<0\}})\frac{(r_{t-1} - \mu)^2}{\tau_{t-1}} + \beta h_{t-1} \tag{1}$$

where $\phi = \alpha + \frac{\gamma}{2} + \beta$.

To ensure that this process is covariance stationary, MF2-GARCH relies on the assumption that the parameters satisfy the following inequalities:

- $\alpha > 0$

- $\alpha + \gamma > 0$

- $\beta > 0$

- $\phi = \alpha + \frac{\gamma}{2} + \beta < 1$

6

and that Z_t:

- is i.i.d.

- has a symmetric density with $E(Z_t) = 0$ and $E(Z_t^2) = 1$

- is such that Z_t^2 has a nondegenerate distribution with $E(Z_t^4) < \infty$

While it is not necessary to assume that Z_t follows a standard normal distribution, I do so to derive the log-likelihood function (Section 4.1) and to generate shocks in the Monte Carlo simulations (Section 4.3), as this satisfies the above assumptions.

Following Engle (2009a), Conrad & Engle (2025) define $V_t = \frac{(r_t - \mu)^2}{h_t}$ as the squared "deGARCHed returns". These represent the standardized volatility forecast errors from Equation (1). The long term component, τ_t, is specified as a multiplicative error model (MEM) equation for the conditional expectation of $V_{t-1}^{(m)}$, the moving average of the standardized errors V_t:

$$\tau_t = \lambda_0 + \lambda_1 V_{t-1}^{(m)} + \lambda_2 \tau_{t-1} \tag{2}$$

$$V_{t-1}^{(m)} = \frac{1}{m} \sum_{j=1}^{m} V_{t-j} = \frac{1}{m} \sum_{j=1}^{m} \frac{(r_{t-j} - \mu)^2}{h_{t-j}} \tag{3}$$

This is because of the empirical fact that a rolling window moving average of the past daily standardized forecast errors of one-component GARCH models has predictive power for future volatility (more specifically, squared returns). The errors are predictable and counter-cyclical, with one-component GARCH models underpredicting volatility in economic recessions and overpredicting it in expansions.

This is the MF2-GARCH-rw-m variant, where rw-m stands for "rolling window of length m". Conrad & Engle (2025) introduce another variant which allows for variable weighting schemes to be applied to different values of Vt. Nonetheless, they find that across various subsamples, the flat weighting scheme of the MF2-GARCH-rw-m is consistently preferred when considering the resulting Bayesian Information Criterion (BIC) values. Allowing for a flexible weighting scheme only increases the standard errors of the parameter estimates. For this reason, I employ MF2-GARCH-rw-m alone.

The MF2-GARCH specification markedly outperforms the nested GJR-Garch, Spline-GARCH, GARCH-MIDAS-RV and log-HAR in out-of-sample forecasts of volatility, especially in the long-term (Conrad & Engle, 2025). This makes it desirable for applications that require long-horizon forward-looking forecasts of volatility.

3.2 Maheu-McCurdy Univariate Risk-Return Specification

Maheu & McCurdy (2007) introduce a basic risk-return model in which the conditional mean of the excess market return is related to both the conditional variance of market return as well as one, some, or all of the components of variance:

$$\mu_t = \delta_0 + \delta_1 \sigma^2_{t,(q)} + \sigma_{t,(k)} z_t, \quad z_t \sim N(0,1) \tag{4}$$

where $\sigma_{t,(k)}$ is the (square root of) the conditional variance at time t given by a k-component volatility model and $\sigma^2_{t,(q)}$ is a weighted sum of q of its components.

While they use a realized variance (RV) approach to model risk in their original work, I use MF2-GARCH volatility and its components instead.

3.3 MF2-GARCH-in-Mean

While the original MF2-GARCH specification treats the mean component (μ) unconditionally as a parameter to be estimated, I instead use the conditional mean, μ_t. Here the conditional mean is given by the Maheu-McCurdy specification in Equation (4). The squared demeaned return in Equation (1), $(r_{t-1} - \mu)^2$, becomes:

$$(r_{t-1} - \mu_t)^2 = r_{t-1} - \delta_0 - \delta_1 \sigma^2_{t-1,(q)}$$

and as this thesis uses excess market return, r_{t-1} is the daily market premium for period $t-1$.

In Equation (4), the conditional variance and its components are estimated with MF2-GARCH-rw-m. As in the original paper by Conrad & Engle (2025), I estimate all models for values of m from 20 up to 160 and determine the optimal value for m as the one that minimizes the BIC value.

I investigate if the components of the MF2-GARCH model support the finding that long-term volatility is a better determinant of excess market return.

For $\sigma^2_{t,(q)}$ I first include just one of the short and long-term components (q=1):

$$\mu_t = \delta_0 + \delta_{1,s}h_t$$

$$\mu_t = \delta_0 + \delta_{1,l}\tau_t$$

Then I include both (q=2):

$$\mu_t = \delta_0 + \delta_{1,s}h_t + \delta_{1,l}\tau_t$$

I also estimate a specification in which the overall conditional variance (as estimated by MF2-GARCH) is the sole regressor:

$$\mu_t = \delta_0 + \delta_1\sigma^2_t$$

For all four of these specifications, I estimate a no-intercept (proportional) variant, where δ_0 is forced to be zero, and a non-proportional variant, where it is not.

3.4 Controlling for Crisis Periods/Final Specification

Previous attempts to determine the risk-return relationship which account for volatility regime-switching during crisis periods have found that the significance and direction of the relationship varies based on the regime, as mentioned in Section 2.1. Including crisis periods in the sample can thus lead to a breakdown of the linearity of the risk-return relationship.

In addition, Danielsson et al. (2018) use multi-year deviations of volatility from the trend to predict banking crises. Their dummy variables show strong significance, which supports the notion that volatility clustering is more pronounced during crises. This particularly affects the viability of a constant window size (m) for the moving average/long-term volatility component of the MF2-GARCH model.

For these reasons, I include a binary dummy variable in the risk-return specification as an indicator for periods of crisis. The final model thus becomes:

$$\mu_t = \delta_0 + \theta_0 D_t + (\delta_1 + \theta_1 D_t)\sigma^2_{t,(q)} \tag{5}$$

$$h_t = (1 - \phi) + (\alpha + \gamma 1_{\{r_{t-1}-\mu_{t-1}<0\}})\frac{(r_{t-1} - \mu_{t-1})^2}{\tau_{t-1}} + \beta h_{t-1} \tag{6}$$

9

where D_t is the dummy variable, θ_0 and θ_1 are its coefficients, and the long term component τ_t is as previously defined by Equation (2) and (3).

The exact crisis period dates (where $D_t = 1$) are described in Section 4.2 and presented in detail in Table 3. I also repeat the estimation of all specifications without controlling for crisis periods to demonstrate the effect of doing so.

4   Method

4.1   Log-Likelihood Function

Given past information set I_{t-1} and assuming that $r_t|I_{t-1} \sim N(\mu_t, \sigma_t^2)$ gives the tth market premium observation the following likelihood function:

$$f(r_t|I_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_t^2}} exp\left[-\frac{(r_t - \mu_t)^2}{2\sigma_t^2}\right]$$

Under the model, $\sigma_t^2 = h_t\tau_t$. The total likelihood, denoted as l, is therefore given by:

$$l = \frac{1}{\sqrt{2\pi h_t\tau_t}} \sum_{t=1}^{T} exp\left[-\frac{1}{2}\frac{(r_t - \mu_t)^2}{h_t\tau_t}\right]$$

where T is the sample size, and the total log-likelihood is given by:

$$L = \frac{1}{2}\sum_{t=1}^{T}[ln(2\pi) + ln(h_t\tau_t) + \frac{(r_t - \mu_t)^2}{h_t\tau_t}]$$

This gives the combined risk-return specification given by Equation (5) and (6) the following total log-likelihood function:

$$L = \frac{1}{2}\sum_{t=1}^{T}[ln(2\pi) + ln(h_t\tau_t) + \frac{(r_t - (\delta_0 + \theta_0 D_t) - (\delta_1 + \theta_1 D_t)\sigma_{t,(q)}^2)^2}{h_t\tau_t}] \qquad (7)$$

In my implementation, I minimize the negative log-likelihood function, $-L$.

4.2   Estimation

The full Python code and data files are provided as attachments to the digital version of this thesis, and links to an online repository can be found in Section A.1 of the appendix. The code is based on the original MF2-GARCH toolbox for Matlab (Conrad & Schoelkopf, 2025).

I estimate all parameters simultaneously using Quasi-Maximum Likelihood Estimation (QMLE) implemented in Python. Minimization of the negative log-likelihood function is done with the optimize.minimize function of the SciPy package, using the sequential least squares programming (SLSQP) option.

Likelihood ratio tests of the proportional variants against non-proportional counterparts are implemented manually with the SciPy package.

4.3   Monte Carlo Simulation

To test the viability of the MF2-GARCH-in-mean model, I generate simulated data using known parameters, assuming that the model holds, and I estimate the parameters over the simulated data.

I do this for every specification. For simplicity, no crisis periods are generated and the crisis indicator dummy variable is excluded. The choice of moving average window size (m) is set at $m = 63$, or three trading months.

For the "true" parameter values, I use averages of the parameter estimates I obtained from fitting the model on real data. The parameter values used in the Monte Carlo simulations are shown in Table 1. The values of τ_t, h_t, and r_t at time $t = 0$ are their average values obtained from fitting the model on real data. The moving average of the daily standardized forecast errors $(V_t^{(m)})$ requires at least m previous data points, and τ_t is a function of $V_{t-1}^{(m)}$. As a result, for all $t < m$, values of τ_t are hard-coded to be equal to the initial τ_0 value, and $V_t^{(m)}$ is simply an average of all past values of V_t.

To avoid dependence on initial value choices in the simulated data, the first 252 data points (one trading year) are discarded as a form of "burn-in". In each simulation, I simulate $T = 30,240$ days of data, the equivalent of 120 trading years. I perform $R = 1,000$ simulations and report the mean and standard deviation of the parameter estimates, mean bias, and average standard deviations.

This is implemented in Python and Monte Carlo simulation features are included in the code referred to above in Section 4.2. Parameter estimation on simulated data is performed as described in Section 4.2.

5 Empirical Results

5.1 Data

I apply the combined specification given in Equation (5) and (6) to U.S. daily market premium data.

The data are downloaded from CRSP. For excess market return, I use the U.S. market premium from the Fama-French 3 Factor library. The data runs from July of 1926 to April of 2025, inclusive. I limit it to a modern subsample starting in January of 1964 as detailed in Section 5.1.1.

The excess market return is calculated by subtracting the one-month Treasury bill rate from the value-weighted CRSP market return for NYSE, AMEX, and NAS-DAQ firms. The resulting values are taken as Rt in the combined MF2-GARCH-in-mean model. The binary crisis indicator variable is added manually (see section 5.1.2 for more details). The data was converted from .csv to .xlsx format for convenience.

Table 2 shows summary statistics for the market premium data. As expected, the premia have a moderately positive mean, a negative skew, a high kurtosis and a weak autocorrelation coefficient.

5.1.1 Sample Dates

While exact dates vary, the mid 1960s mark a clear turning point toward lower volatility in output and inflation in the postwar U.S. economy. They also mark a shift in monetary policy focus. As Hetzel (2013) notes, in this period, the U.S. Federal Reserve began actively adjusting short-term interest rates to smooth business-cycle fluctuations (affecting market volatility), whereas their pre-World War Two focus was to back the dollar with gold and prevent speculative credit booms.

This is public information which informs investor behavior and therefore affects the equity premium which they demand in return for bearing risk, affecting the market risk-return relationship. I therefore limit the data to a "modern" subsample starting from January of 1964, following the example of Ghysels et al. (2005).

5.1.2 Crisis Periods

A binary dummy variable (D_t) is used to exclude the idiosyncratic effects of crisis periods.

To define the specific dates which mark the beginnings and ends of crisis periods in U.S. markets, I use economic recessions as recognized by the U.S. National Bureau of Economic Research (NBER). This is to ensure that the dummy variable truly captures periods of macroeconomic stress rather than benign clusters of high volatility.

A list of these periods and their corresponding start and end dates is provided in Table 3. The value of the dummy variable is 1 in these periods and 0 otherwise.

5.1.3 Choice of moving average window size

I estimate the model for values of m from 20 to 160 and choose the value which minimizes the BIC. For all specifications, this value is $m = 63$, or around 3 trading months.

5.2 Estimates and Interpretations

5.2.1 Parameter estimates

Table 4 reports QMLE parameter estimates from the MF2-GARCH-in-mean model. Each panel presents the proportional (no-intercept) and non-proportional (with-intercept) variants of each specification as well as the corresponding log-likelihood and BIC values. Standard errors in parentheses are Bollerslev-Wooldridge robust standard errors. Each panel also shows the value of the likelihood ratio test (LRT) statistic for the proportional variant against the non-proportional variant. A significant LRT statistic means a rejection of the null hypothesis that the proportional variant has sufficient explanatory power.

For all specifications, the moving average window size (m) which minimizes the BIC value is $m = 63$. The plots of BIC values against m values are shown in Figure 1a to Figure 1h.

In the short-term component only specification (Panel A), the estimate of the

13

coefficient of the short-term component $(\widehat{\delta_{1,s}})$ is positive and significant at the 5% level in the proportional variant and negative and significant at the 1% level in the non-proportional variant. The LRT statistic is highly significant, rejecting the proportional variant in favor of the non-proportional one. The estimate of the intercept in the non-proportional variant $(\widehat{\delta_0})$ is positive and significant at the 1% level. The coefficients of the crisis indicator dummy variable are statistically indistinguishable from zero.

In the long-term component only specification (Panel B), the estimate of the coefficient of the long-term component $(\widehat{\delta_{1,l}})$ is positive and significant at the 1% level in both the proportional and non-proportional variants. The LRT fails to reject the null hypothesis, favoring the proportional variant. The estimate of the intercept in the non-proportional variant $(\widehat{\delta_0})$ is positive but statistically insignificant. When interacted with the long-term component in the proportional specification, the crisis indicator dummy variable has a negative coefficient estimate $(\widehat{\theta_{1,l}})$ which is significant at the 1% level. The other coefficients of the crisis indicator dummy variable are statistically indistinguishable from zero.

In the two-component specification (Panel C), where the short- and long-term components are both additively included, $\widehat{\delta_{1,s}}$ is negative and insignificant in both the proportional and non-proportional variants. $\widehat{\delta_{1,l}}$, on the other hand, is positive in both, and significant at the 5% level in the proportional variant and at the 1% level in the non-proportional one. The LRT again fails to reject the null hypothesis, favoring the proportional variant. The estimate of the intercept is positive but insignificant in the non-proportional variant. When interacted with the short-term component in the proportional specification, the crisis indicator dummy variable has a negative coefficient estimate $(\widehat{\theta_{1,s}})$ which is significant at the 5% level. All of the dummy variable's other coefficient estimates are statistically indistinguishable from zero.

Notably, when the risk-return specification uses overall conditional variance (Panel D), the estimate of the coefficient of volatility $(\widehat{\delta_1})$ is significant at the 1% level in the proportional specification. The LRT is significant at the 10% level, however, pointing to the non-proportional variant being preferred. In the non-proportional

variant, the coefficient becomes statistically indistinguishable from zero. Nonetheless, the highly significant coefficient estimate in the proportional specification suggests that, by including the long-term component, MF2-GARCH volatility captures some additional information that simpler models do not.

The model with the best fit (according to BIC value) is the proportional long-term component only specification. Summary statistics for volatility and its components under this specification are provided in Table 5. It should be noted, however, that the MF2-GARCH parameter estimates (α, γ, β, λ_0, λ_1, λ_2) are quite similar across all the specifications, and consequently, so are the summary statistics of conditional variance.

5.2.2   Interpretation

When only the short-term volatility component enters the mean equation (Panel A), its "price of risk" behaves differently depending on whether the intercept is included. In the proportional variant, the positivity and high significance of the short-term coefficient suggests that investors demand higher expected returns for bearing transitory fluctuations in return. However, once an intercept is introduced (non-proportional variant), the coefficient becomes negative and highly significant, and the LRT strongly favors the non-proportional specification. This flip in sign indicates that allowing a baseline premium absorbs much of the average compensation, and what remains implies a discount for short-term swings, possibly reflecting volatility-timing effects. In this case, the negative coefficient may be because of investors who target certain risk levels in their portfolios selling the "market asset" during short-term spikes in volatility in order to rebalance portfolio risk.

By contrast, isolating the long-term volatility component (Panel B) yields a highly significant positive relationship with the premium in both the proportional and non-proportional variants and the LRT does not reject the simpler proportional form, implying that the slow-moving volatility factor commands a stable premium regardless of whether or not we include an intercept. This finding is consistent with the intuition that persistent uncertainty constitutes genuine background risk for investors, who therefore demand a premium that cannot be absorbed by a constant

term. While transitory movements captured by the short-term component can be attributed to noise, longer-term patterns might appear to be more characteristic and indicative of the state of the market, influencing the behavior of investors. The highly significant and negative coefficient estimate for the crisis indicator variable $(\widehat{\theta_{1,l}})$ is consistent with the "flight-to-safety" hypothesis, which would see investors' risk tolerance decrease during crises, as they seek safe assets which have less correlation with market movements.

With both components included (Panel C), the long-run factor again turns out to be the source of compensation. The long-term coefficient retaining positivity and significance and the short-term coefficient losing it entirely suggests that transitory uncertainty contributes no real premium once persistent uncertainty is accounted for. The LRT favoring the proportional variant indicates that the intercept adds little explanatory power when both volatility horizons are jointly modeled, which suggests that the average market premium can be understood as arising entirely from the long-term risk factor.

The final specification based on total conditional variance (Panel D)shows a significant positive coefficient in the proportional model, telling us that aggregate variability matters in determining the premium. However, the LRT again points to the non-proportional variant as superior, and in the non-proportional variant the volatility coefficient becomes insignificant. This and the lower (more desirable) BIC value of the proportional long-term specification both support the notion that MF2-GARCH's long-term component does better at capturing information which relates directly to the market premium.

According to the BIC values, the proportional long-term only model is the most parsimonious and best-fitting specification. That the long-term component alone captures the bulk of the risk-return trade-off, and that this result holds even when compared against richer two-component or total-variance specification, lends strong support to the idea that long-horizon uncertainty is the key determinant of market premia.

In this proportional long-term component specification, we can interpret $\widehat{\delta_{1,l}}$, as the "daily" coefficient of relative risk aversion (CRRA), as explained in Section

1. If we annualize this CRRA by multiplying it by 252 (the number of trading days in a year), we get a value of 12.348. This is slightly high compared to the "traditional" benchmark range of 2-10 ************ which is usually estimated by previous research. This is perhaps explained by the focus of this specification on long-horizon risk, which may be viewed by investors as more enduring rather than transitory, causing them to be more averse to it.

What is interesting to note about the LRT statistics and the option of including an intercept is that the regressand in this model is the market premium, meaning that the risk-free rate (which we can consider the constant component of overall market return) is already subtracted and controlled for. A positive and significant intercept, like the one in the non-proportional short-term component only specification (Panel A), can just mean that the intercept is capturing the average premium, and that the short-term component's inclusion explains deviations from the average premium. If we used the raw market return as a regressand instead, one would expect the intercept to be equal to the risk-free rate.

Finally, the similarity of the underlying MF2-GARCH parameter estimates across all specifications suggests consistent volatility dynamics, meaning that roughly the same information is retained across specification changes, so that they can be reasonably compared.