Expected and realized returns in conditional asset pricing models: A new testing approach

Antell, Jan; Vaihekoski, Mika

Published in:
Journal of Empirical Finance

DOI:
10.1016/j.jempfin.2019.04.001

Publication date:
2019

Document Version
Peer reviewed version, als known as post-print

Expected and realized returns in conditional asset pricing models: A new testing approach

Abstract

We develop a new approach for testing conditional asset pricing models that avoids the issues in using realized returns as a proxy for expected returns. Testable restrictions are developed by asking what realized returns we would observe, given the pricing model under scrutiny. The new reverse testing approach is used to test the Merton ICAPM and a long-standing risk-return puzzle: the price of market risk has often turned out to be insignificant and at times even negative. The results from the new testing approach on US data give strong support for a positive relationship between conditional variance and the equity premium.

Keywords: conditional asset pricing, equity premium, risk aversion, risk-return trade-off, volatility-feedback.
JEL classification: G12.

1. INTRODUCTION

Tests of asset pricing models evolved from the evaluation of their unconditional cross-sectional implications into tests of their conditional time series implications in the late 1980s (Ferson, 2003). However, as tests of conditional implications focus on period-by-period return properties instead of long-term averages, the tests come with a cost. Using realized returns as a proxy for expected returns is a concern in the conditional tests but not in the unconditional tests as there are a number of reasons to believe that realized returns are not adequate proxies for the conditional expected returns (see, e.g., Brav et al., 2005). For example, Greenwood and Shleifer (2014) document that investors possess high expectations on future returns when rational expectations asset pricing models suggest a low return. A common solution has been to estimate jointly an expectations model – the typical choice being a linear one (Harvey, 2001). However, it suffers from the same problem – the choice of forecasting variables are selected ex post to have predictive power over realized returns. Although a lot of work on finding better proxies for the expected returns have been done, the suggested solutions (e.g., the use of surveys) are not often suitable for tests of conditional asset pricing models.

In this paper, we introduce a new approach to test conditional asset pricing models which avoids the issues in using realized returns as a proxy for expected returns as well as the need to use an empirical expectations model. We turn the tables and ask what realized returns we would observe, given the asset pricing model for the expected returns. Using this insight, we derive a simple but innovative model for the realized returns that combines the dividend discount model of Campbell and Hentschel (1992) with the selected conditional asset pricing model to study the model and the risk-return trade-off. This reverse (flipped) testing approach differs from traditional testing approaches in a sense that it relates realized returns to the *change* in the risk-free rate, in the expected dividends, and in the risk premiums rather than to the *level* of or the *surprise* in the variables.[1]

We use the new approach to test one of the simplest, yet fundamental pricing equations, the Merton (1973, 1980) model.[2] The model suggests that a representative investor must receive a certain amount of positive compensation for her investment, commonly referred to as price of market risk, or *lambda*, for a unit increase in variance. Because the model is applicable to any security and hence also to a market portfolio, Merton's model suggests a positive relationship between the expected return of a market portfolio and the variance of the market, all conditional on available information. However, empirical evidence on this relationship has been mixed, even a long-standing puzzle (sometimes labeled the *total volatility puzzle*). Although some studies have found empirical support for the relationship between return and variance (see, e.g., Ghysels, Santa-Clara and Valkanov, 2005), there is also a great deal of evidence that the relationship is non-significant, with lambda estimates being too small and at times even negative, particularly in shorter samples, and sensitive to methodo-

[1] The framework of Campbell and Hentschel (1992) – generalized e.g. in Campbell et al. (1997) – is obviously closely related to our study. However, our models differ slightly; Campbell et al. show how (real) realized returns can be written as the sum of the expected return for the same period and of the changes in expectatations of future (real) returns and dividends. As an example, they assume that the expected returns behave in an autoregressive manner. We do not separate the expected return terms which allows us to test the theoretical model directly. The empirical model derived in Guo and Whitelaw (2006) to study the Merton (1973) ICAPM is also related to our model, although the motivation and the scope are different.

[2] The reverse testing approach can in principle be used to test any conditional asset pricing model and for any number of assets. The selected model is merely chosen to demonstrate the differences to the traditional testing approach.

logy and sample period (reviews of the studies can be found, e.g., in Bali, 2008, and in Gonzales, Nave and Rubio, 2012).

A number of alternative explanations have emerged. The first line of explanation is based on the idea that the measure of the market portfolio is not adequate (Merton, 1987). This leads to a situation where investors are compensated for holding imperfectly diversified portfolios; hence, the standard model relating market returns to market variance could be missing a source of risk that is driving the puzzle (see Malkiel and Xu, 2006).

A related explanation for the total volatility puzzle is that the simple one-factor asset pricing model is wrong. In line with this reasoning, the inconclusive results are due to some missing risk or investment opportunity hedge factors (e.g., Guo and Whitelaw, 2006; Kim and Nelson, 2014; Feunou et al., 2014). Work along this line has considered, for example, the unpredictable part of the variance (French, Schwert and Stambaugh, 1987) or skewness (Theodossiou and Savva, 2016). The former has been motivated by the leverage effect (Black, 1976) and the volatility-feedback effect (Pindyck, 1984). They both explain why variance and realized return can move in opposite directions. The former states that a negative shock in the market causes the overall leverage to increase, leading to higher volatility. The latter is based on the idea that a positive unexpected shock to volatility leads to a higher risk premium which implies a negative realized return. French et al. (1987) use intuition to motivate an empirical model where the realized equity premium is related both to conditional as well as unexpected variance (later more formally motivated in Campbell, 1991). They find support for the volatility-feedback effect over the leverage effect. However, although their results support the importance of the expected and unexpected variance, they are not simultaneously statistically significant.

The second line of explanation for the puzzle suggests that the variance measures are inadequate and hence should be improved. Suggestions include developments in econometric modeling techniques to model the conditional variance as well as using forward-looking implied volatility measures. The first wave of improvements came with the introduction of the (G)ARCH specification by Engle (1982) and Bollerslev (1986). Ultimately, an enormous number of different specifications in the GARCH family have been proposed, including asymmetric extensions and non-normal distributions (see, e.g., Glosten et al., 1993; Bekaert and Wu, 2000).

A more recent econometric development in estimating variance followed with the mixed data sampling methods (henceforth MIDAS) introduced by Ghysels et al. (2005). MIDAS allows one to combine data of different frequencies. This method is especially suitable for studying the risk-return tradeoff because it allows combining daily data for more accurate variance estimation with lower-frequency data to model the long-term risk-return relationship, thus alleviating problems with noisy short-term returns. Studies utilizing this method to evaluate the relationship between volatility and future returns are rather scarce as of today but include, e.g., Gonzales, Nave, and Rubio (2012) and Ghysels et al. (2005 and 2016).

Alongside the development of new econometric estimation techniques for the conditional variance, other approaches have also been proposed, the most notable being the use of implied volatility calculated from option prices. A number of stock and derivatives exchanges have started to calculate these implied volatility measures (cf., e.g., CBOE's Volatility Index, VIX). Because the implied volatility measure is by construction forward-looking, some researchers have argued for its use over conventional historical measures (for a review, see, e.g., Poon and Grander, 2003). As a result, researchers have also used implied volatilities in studying the relationship between variance and market premia (see, e.g., Guo and Whitelaw, 2006; Santa-Clara and Yan, 2010).

Although there have been clear improvements in variance estimation, generally only certain parts of the puzzle have been explainable, not all of it, and neither under all circumstances nor over short horizons. For example, Hedegaard and Hodrick (2016) provide potential explanations for why the risk-return tradeoff cannot be observed particularly over short horizons. They note that market microstructure frictions, non-synchronous portfolio investment decisions, and individual stock illiquidity can drive the results. Conversely, Hibbert, Daignler and Dupoyet (2008) argue in favor of a behavioral explanation for the negative return-volatility relationship.

The reverse testing approach provides an alternative explanation for the total volatility puzzle that also helps to explain why many earlier models have not been able to fully provide one. We argue that many of the earlier efforts to uncover the conditional return-variance relationship yield susceptible estimates of lambda and have shortcomings that can be circumvented by the empirical model implied by the new approach. First, the new model explains why the lambda estimate is not significant if one only links realized returns to the contemporaneous conditional variance as is done in many previous studies. Second, the model explains why empirical estimates of lambda are by necessity too small unless properly adjusted. Third, the model helps us to understand why the estimation results are affected by the time interval used to measure returns. Finally, comparison of the traditional and the new model reveals why it is possible to find a negative risk-return relationship with the traditional approach in certain sample periods and return horizons and why this is not the case with the new approach.

Applying our new approach to study the Merton (1973, 1980) model yields a model that resembles the volatility-feedback model. However, the models and the results are not the same. In fact, the tests are closely related only if the unexpected realized variance is positively related to the conditional variance at the end of the period. However, this is not necessarily the case. Our model has also some resemblance to Guo and Whitelaw (2006). They connect a log linearization to Mertons (1973) ICAPM, and include both market variance (the risk component), and the covariance with investment opportunities (the hedge component) in their model, alongside with shocks to the risk (i.e. volatility feedback) and hedge components. Further, their model contains shocks to the risk-free rate and dividends, although they are not explicitly estimated in their empirical specification, but left in the error term. Using implied variance over a relatively short sample from 1983(11) to 2001(5), Guo and Whitelaw find that the price of risk estimate is positive, statistically significant, and reasonable in magnitude. They also find that the correlation between the risk component and the hedge component is negative, a result that may explain the weak results using traditional approaches which exclude the hedge component. However, in their condensed models, and in models for checking the robustness, the results are more ambiguous.

Empirically, we compare the new approach to estimate the price of market risk against the approaches used in the literature. We use both traditional measures of volatility such as those based on (asymmetric) GARCH models and new models in the spirit of MIDAS. For robustness (e.g. to avoid errors in the variables issues), we also use a readily available, forward-looking variance measure based on the option implied VIX volatility index. Tests are conducted using US stock market returns for 1928 to 2013.

2. THEORETICAL BACKGROUND

2.1. Merton model for the return-risk relationship

The capital asset pricing model CAPM postulates that the excess return on any security can be determined by

$$E\left[r^{e}_{i,t+1}\left|\Omega_t\right.\right]=\beta_{i,t+1}\left(\Omega_t\right)E\left[r^{e}_{m,t+1}\left|\Omega_t\right.\right], \tag{1}$$

where $E\left[r^{e}_{i,t+1}\left|\Omega_t\right.\right]$ and $E\left[r^{e}_{m,t+1}\left|\Omega_t\right.\right]$ are expected excess returns on security i and the market portfolio, conditional on investors' information set Ω_t available at time t. Because the conditional beta, $\beta_{i,t+1}\left(\Omega_t\right)$, is defined as $Cov\left(r_{i,t+1},r_{m,t+1}\left|\Omega_t\right.\right)Var\left(r_{m,t+1}\left|\Omega_t\right.\right)^{-1}$, where Cov(.) is the conditional covariance between security i and the market and Var(.) is the conditional market variance, we can use equation (1) to define the ratio $E\left[r^{e}_{m,t+1}\left|\Omega_t\right.\right]Var\left(r_{m,t+1}|\Omega_t\right)^{-1}$ as $\lambda_{m,t+1}$, a measure commonly labeled as the conditional price of market risk or reward-to-risk; it measures the compensation the representative investor must receive for a unit increase in the variance of the market return. Under certain assumptions (e.g., power utility), it can be shown that this lambda term is equal to the aggregate relative risk aversion measure.

Merton (1973, 1980) showed that the same conclusion can be achieved using an intertemporal CAPM. Under certain conditions, equilibrium expected returns are related to the (co)variance of market returns and a reward-to-risk term defined as $-U^{''}_{ww}\cdot W\cdot\left(U^{'}_{w}\right)^{-1}$, where U is a utility function for the representative investor, W is wealth, and $U^{'}$ represents partial derivatives of the utility function. In both cases, the equilibrium expected excess returns for any security i can be stated as

$$E\left[r^{e}_{i,t+1}\left|\Omega_t\right.\right]=\lambda_{m,t+1}Cov\left(r_{i,t+1},r_{m,t+1}\left|\Omega_t\right.\right), \tag{2}$$

where the conditional expected excess return $E\left[r^{e}_{i,t+1}\left|\Omega_t\right.\right]$ is linearly related to the time-varying aggregate price of market risk, measured by the parameter $\lambda_{m,t+1}$, and the conditional covariance between the security's return and that of the market, everything conditional on information Ω_t. Because the model is applicable to any security i, and hence also to the market portfolio, the model for the excess return on the market portfolio can be written as

$$E\left[r^{e}_{m,t+1}\left|\Omega_t\right.\right]=\lambda_{m,t+1}Var\left(r_{m,t+1}\left|\Omega_t\right.\right). \tag{3}$$

Equation (3) basically shows that investors must be compensated by a higher expected return if the conditional variance increases. Because subtracting a constant from a random variable does not change the variance, we can rewrite the variance term in the right hand side in excess return form: $Var\left(r^{e}_{m,t+1}\left|\Omega_t\right.\right)$. This equation forms the basis for most of the empirical analysis conducted so far.

2.2. Traditional testing approach

The theoretical model (3) has a number of empirical implications. To test the model, one must provide empirical proxies for expected returns and conditional variances. Typical tests assume that realized returns can be used as a proxy for the expected returns. This is based on a notion of rational expectations, which is commonly used as a basis to test unconditional implications of asset pricing models. Rational expectations imply that although investors' expectations may be wrong in the short run, they are correct on average in the long run, they utilize all information, and they are not consistently biased. Given an estimate for the variance, one typically proceeds to estimate equation (3) under the assumption of constant price of market risk using the following linear model:

$$r^e_{m,t+1} = \mu + \lambda_m \sigma^2_{m,t+1} + \varepsilon_{m,t+1}, \tag{4}$$

where $r^e_{m,t+1}$ is the realized excess market return from time t to t+1, μ is a constant expected to be zero if excess returns are used and the asset pricing model is valid, λ_m is the price of market risk, and $\sigma^2_{m,t+1}$ is the conditional variance for the period from t to t+1, given the information available at time t.[3] We refer to using this equation as the *traditional approach* to estimating lambda.

Empirical research has used a number of alternative approaches to estimate the variance. The simplest is to use the realized squared returns as an estimate for the variance. The most commonly used approach, however, is based on the family of (generalized) autoregressive conditional heteroskedasticity (GARCH) models. Their popularity is based on the fact that they can be used to capture the main stylized features in the volatility of financial assets, namely volatility clustering, time-variation, asymmetry, and non-normality. The most commonly used specification is the univariate GARCH(1,1)-in-Mean model in which one combines equation (4) with the assumption that $\varepsilon_{m,t+1} \sim nid\left(0, \sigma^2_{m,t+1}\right)$ and the following process for the conditional variance:

[3] One could also allow the price of market risk to be time-varying (see, e.g., Harvey, 2001). However, as we focus on the fundamental relationship between risk and return, we assume it to be constant throughout this paper.

$$\sigma^2_{m,t+1} = \omega + \alpha\varepsilon^2_{m,t} + \beta\sigma^2_{m,t}, \tag{5}$$

where the parameters ω, α and β relate to the GARCH(1,1) variance specification. Equation (5) captures time-variation and clustering, and can easily be adjusted to take into account further stylized facts of the variance allowing, for example, for asymmetric responses to return shocks and for alternative distributions.

The traditional approach is, however, problematic because its empirical tests rest on the joint hypothesis of the expectations model and the asset pricing model itself. We believe that realized returns are inadequate proxies for the expected returns, particularly for the relatively short return measurent intervals often used in asset pricing tests, and that this approach therefore is not the best approach for empirical tests of conditional asset pricing models that allow for time-varying expectations. In addition, we argue that the traditional approach implicitly assumes a flat term structure for the risk premium which goes against recent evidence (c.f., Feunou et al., 2014). In fact, we argue that some of the empirical anomalies that have been found with respect to the Merton model are due to the traditional testing approach.

One of the main empirical anomalies with respect to lambda is that the estimates are often too small compared to their *ex ante* expectation, as there are theoretical justifications that lambda should be greater than one but less than five (see e.g. Meyer and Meyer, 2005). The same conclusion can also be drawn by a casual study of equation (3), which indicates that, for a typical long-term average annual volatility (e.g., 15 percent) and market risk premium (e.g., five percent), lambda estimates should be greater than one.

The solution suggested in the literature is to include the surprise to the conditional variance in addition to the theoretical relationship. This volatility-feedback effect offers an alternative way to estimate lambda. Following French et al. (1987), we can add the unexpected variance $\sigma^2_{u,m,t+1}$ into equation (4):

$$r^e_{m,t+1} = \mu + \lambda_m\sigma^2_{m,t+1} + \gamma_m\sigma^2_{u,m,t+1} + \varepsilon_{m,t+1}, \tag{6}$$

where $\sigma^2_{u,m,t+1} = \sigma^2_{r,m,t+1} - \sigma^2_{m,t+1}$. Realized variance, $\sigma^2_{r,m,t+1}$, is often calculated as the sum of daily squared returns within a particular month. In practice, we can estimate lambda easily by augmenting equation (4) with the realized variance and estimating $r^e_{m,t+1} = \mu + \delta_m\sigma^2_{m,t+1} + \gamma_m\sigma^2_{r,m,t+1} + \varepsilon_{m,t+1}$. Since $\delta_m = \lambda_m - \gamma_m$, an estimate for lambda can be calculated as the sum of δ_m and

γ_m. We call this equation the *volatility-feedback approach* to estimating lambda. Although the volatility-feedback approach is a step forward, it still suffers from the use of realized returns as a proxy for expected returns.

2.3. A new model for testing the return-variance relationship

Based on the discussion above, we take a slightly different point of view on estimating the relationship between market variance and the risk premium. Our starting point basically turns the tables and asks the question: What realized returns would one observe, given that the asset pricing model is correct? To investigate this, we create a model in the spirit of Campbell and Hentschel (1992).[4] We analyze realized returns over one period. The length of the period can be chosen freely, but here we assume it to be one month. At first, we do not take a stand on the pricing model or how investors set their discount rates. The model is derived in a continuously compounded world, and thus all rates are continuously compounded returns/dividend growth rates per period.

Our starting point is a dividend-paying security, i.e., an individual stock, a stock portfolio or the overall stock market portfolio. Later, the pricing model under analysis focuses on the market portfolio. The security pays a dividend at the end of each period. From Campbell, Lo, and MacKinlay (1997), we know that the log price of the security at time t can be stated as a function of the future log dividends and continuously coumpounded discount rates using the dynamic Gordon growth model as

$$p_t \approx \frac{k_1}{1-\rho} + \sum_{i=0}^{\infty} \rho^i \left(1-\rho\right) E_t[d_{t+1+i}] - \sum_{i=1}^{\infty} \rho^{i-1} E_t[r_{t+i}], \tag{7}$$

where $k_1 \equiv -ln\left(\rho\right) - \left(1-\rho\right) ln\left(1/\rho - 1\right)$ and $\rho \equiv 1/(1+\exp(\overline{d-p}))$, where $\overline{d-p}$ is the average logarithmic dividend-price ratio. The parameter ρ is positive and less than one by definition. Campbell et al. (1997) suggest that ρ should be 0.997 for monthly data. $E_t[r_{t+i}]$ expresses the continuously compounded required rate of return for the period t+i whereas $E_t[d_{t+i}]$ represent the expected log dividend occurring at the end of the period at time $t+i$, all conditional on information available at time t.

Note that we use lowercase letters for logarithms of variables and that although the formulation in equation (7) differs slightly from Campbell et al.

[4] Originally in Campbell and Shiller (1988). Additionally, see the models in Guo and Whitelaw (2006) and Banerjee, Doran and Peterson (2007).

(1997), it is essentially the same equation. For convenience, we use the following notation from now on: $d_{t,t+i} = E_t[ln(D_{t+i})]$, i.e., the conditional expected value of the log dividend at time $t+i$ ($i > 0$). Similarly, $r_{t,t+i} = E_t[r_{t+i}]$ is the continuously compounded required rate of return for period $t+i$, both conditional on information available at time t.

Using a first-order Taylor log-linearization, the continuously compounded *realized return* at time t+1 can be written as

$$r_{t+1} \approx k_1 + \rho p_{t+1} - p_t + (1-\rho)d_{t+1}, \tag{8}$$

where d_{t+1} is the log of the realized dividend at time t+1. Inserting log prices (equation (7) and a similar expression for the log price at t+1), and rearrenging, we get the following expression for realized returns:

$$r_{t+1} \approx (1-\rho)\sum_{i=0}^{\infty}\rho^i\left(d_{t+1,t+1+i} - d_{t,t+1+i}\right) + \sum_{i=1}^{\infty}\rho^{i-1}\left(r_{t,t+i} - \rho r_{t+1,t+1+i}\right), \tag{9}$$

where we have utilized the fact that $d_{t+1} = d_{t+1,t+1}$.[5] From this point forward our setup differs a bit more from Campbell and Hentschel (1992) as we implement our main insight: different asset pricing models imply different realized returns that can be used to test the model. The testing proceeds by selecting a candidate pricing model, and inserting it into equation (9).

Applying this insight to test the one-factor version of the Merton (1980) model (3) for the conditional expected returns for the market portfolio and assuming that the price of market risk is constant, we can rewrite the last term of equation (9) as follows:

$$\begin{aligned} &\sum_{i=1}^{\infty}\rho^{i-1}\left(r_{t,t+i} - \rho r_{t+1,t+1+i}\right) \\ &= \sum_{i=1}^{\infty}\rho^{i-1}\left(r_{ft,t+i} - \rho r_{ft+1,t+1+i}\right) + \lambda_m \sum_{i=1}^{\infty}\rho^{i-1}\left(\sigma^2_{t,t+i} - \rho\sigma^2_{t+1,t+1+i}\right), \end{aligned} \tag{10}$$

[5] This equation is a slightly rearranged version of equation (3) in Campbell and Hentschel (1992), here for nominal returns. It is easy to show e.g. that the last term of equation (9) can also be written as $\sum_{i=1}^{\infty}\rho^{i-1}\left(r_{t,t+i} - \rho r_{t+1,t+1+i}\right) = E_t[r_{t+1}] - \sum_{i=1}^{\infty}\rho^{i}\left(E_{t+1}[r_{t+1+i}] - E_t[r_{t+1+i}]\right)$.

where $r_{ft,t+i}$ is the risk-free rate for the period $t+i$ given the information at time t.[6] The risk-free rates known at time $t+1$ are defined similarly. In practice, conditional future risk-free rates can be approximated for example with forward rates.

As a result, our model differs slightly from Campbell and Hentschel's specification because our model relates realized returns directly to the *change* in the conditional variance over the period rather than to the (contemporaneous) *expected level* and *unexpected surprise* of the conditional variance.

Equation (10) is still not directly testable and we need to put more structure into the model. Working first with the second risk premium related term, we use the assumption that the conditional variance is a mean-reverting process (cf., e.g., Engle and Patton, 2001) and that one-step-ahead forecasts can be assessed. We further assume that the conditional variance for any future period $i \geq 1$ can be expressed as a function of the next period's forecast as

$$\sigma^2_{t,t+i} = \phi^{i-1}\sigma^2_{t,t+1} + \sigma^2\left(1-\phi^{i-1}\right), \tag{11}$$

where $|\phi|<1$ is a persistence parameter reflecting the speed of convergence of the conditional variance toward the long-term unconditional variance σ^2. As we expect ϕ to be positive, this model states that if the current variance is below (above) the long-term average, the forecast for the variance also stays below (above) the long-term average, but over time the variance will converge towards the mean. The model also implies that an increase (decrease) in the next period's conditional variance is also reflected in the periods that follow but with decreasing intensity, and in the long term, the variance converges to its long-run mean.

We can write all future conditional variances in terms of next period's conditional variance and the unconditional variance. The last terms in the parentheses of equation (10) can be rewritten using the fact that equation (11) implies that

$$\sigma^2_{t,t+i} - \rho\sigma^2_{t+1,t+1+i} = \phi^{i-1}\left(\sigma^2_{t,t+1} - \rho\sigma^2_{t+1,t+2}\right) + \sigma^2\left(1-\rho\right)\left(1-\phi^{i-1}\right). \tag{12}$$

[6] The same reasoning can be extended to multifactor models. An Internet Appendix shows how the one-factor model in equation (10) can be extended to a general $K+1$ factor model.

Inserting (12) into (10), we can rewrite (9) for the realized *market* returns after some modifications as

$$r_{m,t+1} \approx (1-\rho)\sum_{i=0}^{\infty}\rho^i(d_{t+1,t+1+i} - d_{t,t+1+i}) + \sum_{i=1}^{\infty}\rho^{i-1}(r_{ft,t+i} - \rho r_{ft+1,t+1+i}) \\ + \lambda_m[(\sigma^2_{t,t+1} - \rho\sigma^2_{t+1,t+2})\cdot\varphi_{\Delta\sigma} + \sigma^2{\cdot}\varphi_\sigma], \tag{13}$$

where $\varphi_{\Delta\sigma} = (1-\rho\phi)^{-1}$ and $\varphi_\sigma = \left(1 - \frac{1-\rho}{1-\rho\phi}\right)$. The parameters $\varphi_{\Delta\sigma}$ and φ_σ are collectively called as *sigma multipliers*. In theory, if the variance persistence parameter ϕ equals, say, 0.9 and ρ equals 0.997 for monthly data, $\varphi_{\Delta\sigma}$ equals $1/(1-0.997\cdot 0.9) = 9.74$, and φ_σ equals $(1-\varphi_{\Delta\sigma}\cdot(1-\rho)) = 0.97$. The parameter $\varphi_{\Delta\sigma}$ indicates how much changes in the conditional variance over one period are magnified due to the persistence of variance.

To simplify the model further, we can assume that the interest rate term structure is flat, i.e., the risk-free rate at any given time is the same for all future periods. Using this assumption on the second term on the right of equation (13) gives us $r_{ft,t+1} + (r_{ft,t+1} - r_{ft+1,t+2})\sum_{i=1}^{\infty}\rho^i$. Alternatively, we can assume that all future risk-free rates change from time t to $t+1$, but all changes are equal. Similarly, we want to simplify the series of expected dividends. There are several ways to do this but perhaps the most intuitive approach utilizes the fact that we can rewrite the conditional dividend stream as a function of its growth expectations. This gives us

$$(1-\rho)\sum_{i=0}^{\infty}\rho^i(d_{t+1,t+1+i} - d_{t,t+1+i}) = \sum_{i=0}^{\infty}\rho^i\left(\Delta d_{t+1,t+1+i} - \Delta d_{t,t+1+i}\right) \\ = \sum_{i=0}^{\infty}\rho^i\left(g_{t+1,t+1+i} - g_{t,t+1+i}\right). \tag{14}$$

Now, there are a number of alternative ways to simplify (14), but here we assume that, even though investors can adjust their views on dividend growth from time t to t+1, the difference between the forecast and the new growth rate will converge to zero the further into the future one forecasts the dividends. This assumption is consistent with intuition, because new information at time t+1 is likely to affect investors' expectations regarding the dividend growth in the short run, but this effect is unlikely to last for a long time. As a result, we can simplify equation (14) as

$$\sum_{i=0}^{\infty} \rho^i (g_{t+1,t+1+i} - g_{t,t+1+i}) = (g_{t+1,t+1} - g_{t,t+1}) \left(1 + \sum_{i=1}^{\infty} (\rho\phi_g)^i\right) \tag{15}$$

where $g_{t+1,t+1} - g_{t,t+1}$ represents the change in expectations of future dividend growth rates from time t to t+1, and ϕ_g its converge rate over time (< 1). The sum in the right hand side of equation (15) forms a geometric series which converges and we can calculate its sum.[7] After some modifications, we get the following result:

$$\begin{aligned} r_{m,t+1} \approx & k_2 + (g_{t+1} - g_{t,t+1}) \cdot \varphi_d + (r_{ft} - r_{ft+1}) \cdot \varphi_{rf} \\ & + \lambda_m \left((\sigma^2_{t,t+1} - \rho\sigma^2_{t+1,t+2}) \cdot \varphi_{\Delta\sigma} + \sigma^2 \cdot \varphi_\sigma\right), \end{aligned} \tag{16}$$

where the parameters $\varphi_d = 1 + \frac{\rho\phi_g}{(1-\rho\phi_g)}$ and $\varphi_{rf} = \frac{\rho}{1-\rho}$ can be interpreted as parameters that measure the impact of the change in the dividend growth rates and risk-free rates, respectively. Both of them are by definition positive. The constant k_2 is defined as $k_2 = r_{ft}$. Its value is also expected to be positive when the risk-free rate is positive.

An analysis of equation (16) shows that realized returns should be higher if investors' conditional expectations of the dividend growth rate increase from period t to t+1, ceteris paribus. The same is true if the interest rates decrease. Assuming that the asset pricing model is correct, a decrease in conditional volatility should also lead to higher realized returns. All implications of the model are in line with intuition. It is also quite straightforward to prove that the realized return given by equation (16) equals the expected return given by equation (3) if investors' conditional expectations prove to be right (the proof is provided upon request).

Since equation (16) is an approximation, a relevant question is whether the lambda is sensitive to estimation errors in the variables. Suppose that the expected divided growth rates and risk-free rates are two and three percent per

[7] We could have employed the geometric convergence assumption also with the risk-free rates, i.e., given the change in the risk-free rate for the next period, the changes in the risk-free rates beyond that would converge geometrically to zero. The assumption of geometric convergence is not in fact required either for the risk-free rate series or the dividend growth series as long as their infinitive sum can be written as a linear function of the one-period change. Regardless of the simplifying assumption, equation (16) would ensue, but the theoretical definition of the φ_d and φ_{rf} parameters would be different.

annum, respectively. The lambda is further assumed to be two, rho 0.997, and the variance persistence parameter as well as dividend growth forecast convergence rate are assumed to be 0.9. Conditional volatility is assumed to be 20 per cent per annum. All parameter values are assumed to remain unchanged at times t and t+1. Furthermore, expected and realized log dividend at time t+1 are assumed to be 1/12 dollar. We also assume that the variance exhibits mean-reversion as in equation (11) and that asset-pricing model (3) applies. Using monthly parameter values, we can calculate time series for the expected dividends, conditional variances and required rates of return for each period from t+1 onward, conditional on information available at time t. Then, we can do the same from period t+2 onward, conditional on information available at time t+1. Now, discounting dividends using the required rates, we can derive prices at time t and t+1 for the security as a sum of the discounted dividends. Taking into account the dividend paid at t+1, we calculate the realized return for the security. Using the given values and derived dividends, the realized and expected returns are equal, 0.917 per cent per month.[8]

Now, using equation (16), we can solve for lambda. Obviously, using the parameters above, lambda is two with high accuracy. Using different values for the parameters, lambda is almost unchanged with respect to changes in the values for risk-free rate, level of volatility, variance persistence, and dividend growth convergence rate (the difference is less than 0.001% except for low volatilities; e.g., a volatility of ten per cent leads to a bias +1.016% in lambda). The analysis also reveals that the lambda is most sensitive to the changes in expectations in future divideng growth rates. If the expected dividend growth rate decreases (increases), for example, 20% from time t to t+1, the lambda is biased downwards by 6.452% (upwards by 6.454%). However, it is very unlikely that growth rate changes would trend except in very short samples.

2.4. Empirical model and estimation

The empirical objective of this paper is to estimate the price of market risk, or lambda, to find out whether there is a positive, statistically significant relationship between risk and return, and to assess whether the lambda is within a theoretically justifiable region. In practice, we also want to compare the estimate of lambda from the traditional approach (λ_m^T) with our estimate (λ_m).

[8] Here calculated for the next 2,000 months.

Therefore, we first estimate the GARCH(1,1)-in-Mean model for the market as given by equations (4) and (5) to get the traditional lambda.

To get our estimate for the lambda, we write equation (16) in excess-return form as

$$\begin{aligned} r^e_{m,t+1} = b_1 + b_2[(\sigma^2_{t,t+1} - \rho\sigma^2_{t+1,t+2}) \cdot \varphi_{\Delta\sigma} + \sigma^2 \cdot \varphi_\sigma] \\ + b_3\,(g_{t+1} - g_{t,t+1}) + b_4\,(r_{ft} - r_{ft+1}) + u_{m,t+1}, \end{aligned} \tag{17}$$

where b_1 to b_4 are the coefficients to be estimated. All coefficients (but b_1) are expected to be positive. Since we are estimating the model with excess returns, b_1 is expected to be zero. Note, however, that it cannot be given the same Jensen's alpha interpretation as the constant in standard tests of asset pricing models. The coefficient b_2 corresponds to the lambda. The term $g_{t,t+1}$ is the expected continuously compounded growth rate for the dividends from time t to t+1, conditional on information available at time t. The risk-free rate at time t is given by r_{ft}. The term $\sigma^2_{t,t+1}$ is the variance of the continuously compounded excess market return from time t to t+1, conditional on information available at time t. Variables are defined similarly for time t+1. Other parameters are as defined earlier.

Since the future dividend growth rates are well-known to be difficult to forecast (c.f., Cochrane, 2008) and since our main interest is the estimate for the price of market risk, we also estimate the following simplified version of the model:

$$r^e_{m,t+1} = b_1 + b_2[(\sigma^2_{t,t+1} - \rho\sigma^2_{t+1,t+2}) \cdot \varphi_{\Delta\sigma} + \sigma^2 \cdot \varphi_\sigma] + u_{m,t+1}, \tag{18}$$

where b_1 is expected to account for the mean effect from the components excluded from the model and b_2 is again our estimate for the lambda. This simplied model allows us to study the market risk without taking a stand on how to model changes in investors' views on future dividend growth and interest rates. In the long run, these changes are anyhow expected to be zero on average. In addition, the simplied model can help us to do potentially illustrating comparisons with the traditional approach. Obvisously, the results from the simplied model may suffer from omitted variables bias if the change in risk-free rates and in the dividend growth expectations are correlated with the changes in the conditional variance. However, this is unlikely to be the case as there is nothing indicating anything other than an indirect relationship between changes in dividend growth rate (or interest rate) and changes in market variance.

To estimate the model, we need a proxy for the conditional variance. We use three different proxies, the first two are the conditional variance from GARCH and MIDAS models, and the last one is based on the VIX index. With GARCH we utilize a two-step estimation strategy. In the first step, we estimate the GARCH(1,1) model including only a constant in the mean equation (c.f., e.g., Hedegaard and Hodrick, 2014).[9] In the second step, we estimate equation (17) or (18) using the conditional variance estimates from the first step. Note that we utilize contemporary conditional variance at time $t+1$ for the period ending at time $t+2$ in the mean equation, i.e., $\sigma^2_{t+1,t+2}$. This value corresponds to $\sigma^2_{m,t+2}$ in the GARCH-M specification (equation (5)).

To provide an estimate for lambda, we must estimate $\varphi_{\Delta\sigma}$, φ_{σ}, and the unconditional variance, σ^2. To estimate $\varphi_{\Delta\sigma}$ and φ_{σ}, we use their definitions. This requires an estimate for the speed of conditional variance returning to its long-term mean, i.e., the ϕ parameter and the dividend-to-price-related ρ parameter. The latter can be easily calculated from the data, but the former utilizes the results from the model for the conditional variance. Assuming that the conditional variance follows a GARCH(1,1) process, we can write the i-step ahead forecasts for the conditional variances as a combination of the next period's conditional variance and a long-term (unconditional) level, i.e.,

$$\sigma^2_{t,t+1+i} = (\alpha+\beta)^i \sigma^2_{t,t+1} + \omega \frac{1-(\alpha+\beta)^i}{1-\alpha-\beta}, \tag{19}$$

where α, β, and ω are the GARCH parameters. Now, assuming that the GARCH parameters remain constant, our variance convergence speed parameter ϕ is the sum of α and β. The unconditional variance can be estimated as $\omega/(1-\alpha-\beta)$.

Our second proxy for the variance is obtained by mixing data of different frequencies using MIDAS techniques. It is a compromise between the need for lower-frequency data for modeling the risk-return relationship and higher-frequency data for modeling the variance. It is well known that the accuracy of the variance estimates improves with higher data frequency, whereas it is not the case for the mean. As before, we first use the traditional approach to estimate

[9] We choose this specification for simplicity and conformability with the VIX-index which is a measure defined outside the model to be tested.

lambda after which the specification (18) is estimated. Following Ghysels et al. (2005 and 2016), we write equation (4) for the lower frequency (here monthly) excess market return $r_{m,t+1}$ as follows:

$$r^{e}_{m,t+1} = \alpha + \lambda^{T}_{m} h^{MIDAS}_{t+1} + e_{m,t+1},\ e_{m,t+1} \sim Distr\left(0, h^{MIDAS}_{t+1}\right), \tag{20}$$

where we have defined $\sigma^2_{m,t+1} = h^{MIDAS}_{t+1}$ to be the conditional variance for the period from time t to t+1, estimated using higher frequency data (here daily) up to time t with MIDAS. *Distr* refers to some probability distribution, often the normal distribution, but not necessarily. The variance is modeled using the MIDAS on high frequency returns $r_{m,t}$:

$$h^{MIDAS}_{t+1} = C \sum_{d=0}^{D-1} w_{D-d}\left(\theta^{D}\right) r^{2}_{m,daily,t+1-c-d}, \tag{21}$$

where $w_d(\theta^D)$ is a polynomial weighting structure for daily observations. The equation belongs to a group of distributed lag (DL) models. The parameter C is a scaling constant that refers to the average number of trading days in a month; here it converts daily variance into a monthly one. Most of the research have used a value of 22 for C and we follow this approach. D is the number of lagged daily observations used to estimate the monthly variance. It should be chosen such that the specification captures sufficient number of lags, yet being feasible to estimate. In practice, the parameters of the weight function restrict the effective number of lags to less than 200 (Ghysels, 2015). Here we select it to be 30.[10]

The parameter c is the number of lags from which the high frequency regressors start. It is added into equation (21) to highlight the fact that here we need to lag daily observations by one month i.e. from month-end $t+1$ to match the month-end t. The estimation then uses D high frequency (daily) observations from the end of month t backwards to provide an estimate of the conditional variance for the month ending $t+1$. In practice, the value for c has to be set for the average number of trading days in a month. We set c to be 23 for the full sample and 22 for the post-1990 sample.

A number of polynomial weighting structures can be used (for more infor-

[10] The estimation is done using version 1.1 of the Matlab routines provided by Professor Eric Ghysels on his website.

mation, see Ghysels et al., 2007). Here we use the normalized beta probability density function with a zero last lag. The weights w given on past daily observations are calculated as follows:

$$w_i\left(D, \theta_1, \theta_2\right) = \frac{x_i^{\theta_1 - 1}\left(1 - x_i\right)^{\theta_2 - 1}}{\sum_{j=1}^{D} x_j^{\theta_1 - 1}\left(1 - x_j\right)^{\theta_2 - 1}}, \tag{22}$$

where $x_i = (i-1)/(D-1)$. For a reasonably large D, the sum of the weights is very close to one. Specification (22) ensures that all weights are positive, guaranteeing a positive variance estimate. The shape parameters θ_1 and θ_2 are estimated jointly with the rest of the parameters and allow for a rich spectrum of weighting schemes. The variance estimator of French et al. (1987) has some similarities to specification (22). However, it gives equal weights to the observations.

To estimate equation (17) or (18) in the MIDAS framework, we follow the two-step procedure as before with the GARCH approach, i.e., we first derive our estimate for the conditional variance and then plug it into equation (17) or (18) for the second step. To calculate the $\varphi_{\Delta\sigma}$ and φ_{σ} parameters, we assume here that the variance follows a mean reverting AR(1) process given by

$$h_{m,t+1} = \phi_0 + \phi_1 h_{m,t} + \varepsilon_{m,t+1}. \tag{23}$$

To calculate forecasts for the conditional variance, the speed of convergence to the unconditional variance is simply ϕ_1, and the unconditional variance is $\phi_0/(1-\phi_1)$.

Our third proxy for the conditional variance is based on implied volatilities calculated from options prices observed in the market. As the variance is readily available, this approach can be estimated in one step to avoid any potential EIV and endogeneity issues. We start by estimating the traditional lambda based on realized returns as proxies for the expected returns. Using the notation in equation (4), we define $\sigma^2_{m,t+1} = IV^2_{m,t+1}$ as the squared implied volatility observed at time t for the period t+1. Now, we can estimate the traditional lambda using the following model:

$$r^e_{m,t+1} = \alpha_m + \lambda^T_m IV^2_{m,t+1} + \varepsilon_{m,t+1}. \tag{24}$$

To estimate our model using implied volatilities, we need to calculate the required sigma multipliers, $\varphi_{\Delta\sigma}$ and φ_{σ}. To do this, we first run an AR(1) model for the implied variance to estimate the variance persistence parameter and unconditional variance as with the MIDAS estimation and then proceed similarly.

Utilizing a two-step estimation strategy to estimate our simplified model (18) or the full model (17) raises the question of whether there might be biases in our second-step estimator for the lambda because the independent variable is subject to an errors-in-the-variables (EIV) problem. Following earlier studies, we argue that the potential measurement error in the variance decreases due to the long sample period (cf. Shanken, 1992) and, as a result, the lambda estimates are not systematically distorted. For example, Hedegaard and Hodrick (2014) use a four-step procedure in a multivariate setup.[11] They conduct a simulation study to conclude that the parameters of interest are well-behaved, and that their standard errors are correctly estimated. Moreover, as the VIX index is a variance measure defined outside the model to be tested, we have analogously chosen to model the variance as a GARCH or MIDAS process containing only a constant in the mean equation.

3. DATA

3.1. Variables

We estimate our models using two sets of data. For the GARCH estimation, we utilize monthly returns for the US stock market and a risk-free rate of return from January 1928 to December 2013, i.e., 1,032 months of data. For the MIDAS as well as for the volatility-feedback estimation, we complement the monthly data with daily return observations for the same period. Consequently, the beginning of the sample period matches closely to that of Ghysels et al. (2005), but the sample period extends several years beyond, including the financial crisis that peaked in autumn 2008 and winter 2009.

We use the month-end CRSP value-weighted total return as a proxy for the market return. For the MIDAS estimation, we complement the dataset with daily returns of the CRSP index. When estimating the volatility-feedback

[11] First, they estimate univariate GARCH(1,1) models for all assets, with only constant only in the mean equation. Second, the standardized residuals from step one are used to get correlations from a DCC model. Third, a conditional covariance matrix is constructed based on the variances from step one and the correlation matrix from step two. Finally, the risk-return relationship is estimated in step four.

model, we use the sum of daily squared returns as a proxy for the realized variance. The return includes dividends and is adjusted for splits and issues. The risk-free rate for month $t+1$ is based on the one-month holding period return on US Treasury bills closest to one month at the end of month t. These data are also from the CRSP database. The excess return is obtained as the difference between the market return and the risk-free rate of return. Continuously compounded returns in decimal format are used throughout this study unless otherwise stated.

For the full model, we also need a measure for the change in the risk-free interest rate level. Here, we proxy the risk-free interest rate level with the long-term US government bond yield taken from the Ibbotson SBBI (2014). In addition, we need a measure for the change in the expected dividend growth rates. To create a proxy for this change, we first calculate the dividends paid in monetary terms in each twelve-month period during the sample. The dividend for a given twelve-month period is obtained by multiplying the CRSP price index a year ago with the difference between the total return and price index returns in the twelve-month period, that is, e.g., $D_t = (R_t - R_t^{ex}) \times P_{t-12}$, where returns are percentage annual returns (c.f., Cochrane, 2008). In the second step, we calculate the realized logarithmic annual change in the log dividends, i.e. $g_t = (d_t - d_{t-12})$, and use it as a proxy for the future growth rate of dividends ($g_{t,t+1}$). The same approach is used to calculate our proxy for time $t+1$ dividend growth rate (g_{t+1}). This proxy, however, produces a potential endegeneity issue as the contemporary realized dividend is also used to calculate the realized return at time $t+1$. To avoid this, we adjust D_{t+1} by subtracting the contemporary monthly dividend and adding back the dividend paid in the previous month.

When using implied volatilities, we utilize a readily available volatility index from an options exchange – the Volatility Index calculated by the Chicago Board Options Exchange for the US market. The updated index labeled "VIX" is available from the beginning of 1990 onwards. The VIX is based on the 30-day implied volatility per annum calculated from the options traded for the stocks included in the S&P500. VIX values are based on averaging observations from put and call options over a wide range of strike prices, and the index measures the volatility per annum (CBOE, 2009). Our sample period starts in January 1990 and ends in December 2013, providing us with 288 monthly observations. The VIX is accessed on the CBOE's website.

3.2. Descriptive analysis

Table 1 provides descriptive statistics for the monthly and daily variables. Panel A uses data for the entire sample period (January 1928–December 2013). Panel B provides similar descriptive statistics for the period overlapping with the VIX data (January 1990–December 2013). In addition to the series in Panel A, Panel B includes VIX data and their values squared.

The mean monthly risk premium over the entire sample period is 0.477 per cent per month (or 5.72% per annum), with a volatility of 5.44 per cent per month (18.84% p.a.). The descriptive statistics for the subsample in Panel B show that both the average return and particularly the volatility have been lower in recent decades. The volatility is 4.46 per month (15.45% p.a.). It is clearly lower than the average market expectation of 20.20% p.a. as given by the VIX index. The average dividend growth rate is 4.24% for the full sample and 5.44% for the subsample. Government bond yields have been, on average, 5.09% and 5.43% for the full and subsample, respectively.

Almost all of the series are non-normally distributed according to the Jarque-Bera (1987) test for normality. The monthly risk premia are negatively skewed and show much less kurtosis in the post-1990 subsample than over the entire sample period. As expected, the monthly risk premium shows a fairly low, albeit significant, positive first-order autocorrelation. The dividend growth rate shows high autocorrelation (0.697) – as expected due to overlapping dividend observations used to calculate the growth rate – as do the government bond yield series (0.996).

4. EMPIRICAL RESULTS

4.1. Results with GARCH variance

We begin our analysis by studying the price of risk (lambda) using the traditional approach. We compare the results from the traditional approach with those obtained using the new approach developed in this paper. The former approach is based on the underlying assumption that realized returns are good proxies for expected returns, whereas the latter does not require this. We begin the analysis with the generalized autoregressive conditional heteroskedasticity-in-mean approach (GARCH-M). The quasi-maximum likelihood approach (QML) is used in the estimation.[12] Estimations are conducted

[12] More details of the estimations are provided upon request.

using monthly data from January 1928 to December 2013. Table 2 presents the results.

First, we estimate the pricing model (3) as it has been typically estimated in the literature, i.e., equations (4) and (5). We begin with the standard GARCH(1,1)-M specification. Panel A provides the results. The price of market risk is estimated to be 0.697, which is positive as expected by the theory; however, it is not significantly different from zero, with a t-value of 0.851. It is also lower than one would expect, but in line with earlier studies (cf., e.g., a value of 1.060 with a t-value of 1.292 in Ghysels et al., 2005). In addition, the explanatory power of the traditional model is low, with an adjusted R-squared of -0.4%.

There could be several reasons for the empirical estimation not confirming a significant relationship between returns and variance. One potential explanation could be that the conditional return is not normally distributed. Therefore, we run the model assuming a t distribution instead of the normal distribution. However, this does not materially change the results. The explanatory power of the model drops slightly, and the estimate for the price of risk is even lower than before: 0.628 with a t-value of 0.728. Residual diagnostics (not reported) show that both models are able to capture the heteroskedasticity dynamics properly. However, the normality assumption is rejected. Overall, there are no major differences between the diagnostics of the models.

Another potential explanation for the insignificant lambda estimate could be asymmetry in the variance process (cf., e.g., Bekaert and Wu, 2000; Cappiello et al., 2006), indicating that the variance response of negative shocks differs from that of positive shocks. To test this, we utilize the GJR-GARCH model by Glosten, Jaganathan, and Runkle (1993) and replace equation (5) with the following GJR-GARCH(1,1)-M model:

$$\sigma^2_{m,t+1} = \omega + \alpha\varepsilon^2_{m,t} + \gamma\varepsilon^2_{m,t}I_{m,t} + \beta\sigma^2_{m,t}, \tag{25}$$

where $I_{m,t} = 1$ if $\varepsilon_{m,t} < 0$, and zero otherwise. In practice, the gamma parameter, γ, captures the effect of negative shocks. We estimate the model using conditional normality and the t-distribution. The results are reported in Panel A.

The degrees of freedom for the t-distribution is estimated to be 7.031 with a t-value of 116.299 (not reported), meaning that the tails of the distribution are fatter than is commensurate with the normal distribution. The conditional volatility is asymmetric, with a positive response to negative shocks (the gamma parameter estimate is statistically significant at the 5 per cent level). However, the explanatory power of the model does not materially increase, and the price of risk estimate remains non-significant. In fact, the lambdas are even lower than before. As a result, it is fair to conclude that the traditional approach, when estimated with the commonly used GARCH-in-mean approach, does not seem able to find a statistically significant (positive) relationship between variance and return.

Next, we use the volatility-feedback approach to estimate lambda. The results are reported in Panel B of Table 2. The estimations are done similar to Panel A, but realized variance is added as an explanatory variable into the mean equation (c.f., equation (6)). The reported lambda is the sum of the estimates for the delta and gamma parameters. Its t-value is based on the Wald-test on the null hypothesis that their sum is zero. The results are similar to those found in French et al. (1987). The estimated gammas – measuring the impact of (unexpected) realized variance on realized returns – are negative and highly significant. The explanatory power of the model is also clearly higher than before. However, the evidence goes straight against the Merton model. None of the lambda estimates are significant which clearly suggests that, despite the significant volatility-feedback effect, taking the effect into account cannot help us find support for a positive relationship between the conditional risk premium and variance.

Finally, we turn to the new model introduced in this paper. We estimate first our full model (17) and then our simplified model (18) utilizing the same GARCH processes as before and again with conditional normality and t-distribution assumed. The estimation is conducted in two stages. Note that when we are utilizing the GJR-GARCH-specification, the variance persistence parameter is the sum of α, β, and half the asymmetry parameter, γ. The unconditional variance can be stated as $\omega/(1-\alpha-\beta-\gamma/2)$. The results for the full model are reported in Panel C of Table 2.[13]

In line with the results in Panels A and B, the results in Panel C show that the estimated variance process parameters are significant in almost all cases.

[13] Because the return series shows signs of autocorrelation, we also test for autocorrelation in the residuals of our model. As there are indications of first-order autocorrelation, we use the Newey-West (1987) adjustment for autocorrelation and heteroskedasticity when calculating the standard errors for the parameters using the OLS. Thus, all reported t-values for the mean equation parameter estimates are calculated with the adjustment.

However, in contrast to the results shown in Panels A and B, the lambda (b_2) estimates are all significant and within a reasonable range. With the standard GARCH process and normality assumed, the estimate is 0.240 (t-value 1.658). Interestingly, for the t distribution, the estimate increases to 0.569 (t-value 1.895). Utilizing the GJR-GARCH approach, the lambda estimate increases even further, first to 0.806 (under normality) and then to 1.630 (under the t distribution). The last two results are highly significant. The results from the simplified version of the model (not reported) support these results. The price of market risk estimate is highly significant except for the case of standard GARCH under the assumption of normality. The constant b_1, on the other hand, is not significant as expected, giving further support for the model.

Our second explanatory variable, the change in the dividend growth rate (with parameter b_3), is also statistically clearly significant as the model implies. The result is in line with Lettau and Ludvigson (2005) who find evidence that the expected dividend growth covaries with the expected returns. Our third explanatory variable, the change in the risk-free rate, has also a statistically significant effect (b_4).[14] All coefficient estimates are positive as suggested by our model. The explanatory power of the model is also considerably higher than it is for the traditional or for the volatility-feedback models – especially allowing for asymmetry in the variance process seems to improve the overall explanatory power of the model.

The results give strong support for a positive relationship between conditional equity premium and variance. One obvious question could be to test whether the variance shock offers explanatory power over the change in the conditional variance as in equation (18). To test this question, we first calculate the variance shock as the difference between realized variance and the conditional variance. Then we add this variable into our model (equation (18)) and re-estimate the model with GJR-variance and t-distribution. The results (not reported) show that the variance shock is not significant. The volatility-feedback parameter estimate is -0.446 (t-value -0.608) and lambda is almost unchanged (1.765 with t-value of 8.361). As a result, we feel confident to conclude that the volatility-feedback effect is not a sufficient explanation for the total volatility puzzle.

[14] To remove suspicions of endogeneity, we also estimated the model using realized market returns instead of the excess returns. The results are similar to those reported here.

4.2. Empirical results with MIDAS

Next, we turn to MIDAS estimation. First, we test the asset pricing model using the traditional approach. The estimation is based on equations (20), (21), and (22). As we have multiplied the squared daily returns with 22 to arrive at a per-month form, we can interpret the coefficient for the high-frequency terms in equation (20) as the price of market risk. The results are reported in Panels A and B of Table 3, respectively.

The results show that the traditional lambda estimate -0.358 (t-value -1.072) is negative and statistically not significant. Parameter θ_1 is significant and close to one. This implies that the weighting structure is mostly determined by θ_2 – a value higher than one implies less weight to older observations than newer ones. However, the parameter estimate is not significant. This implies that our model for high-frequency data (daily returns) may not be the best one. Hence, we also estimate the model using the normalized exponential Almon lag polynomial. Again, the lambda estimate is negative (-0.341 with t-value -1.312), but neither of the MIDAS parameters are significant (not reported). Our lambda estimate is close to Ghysels et al. (2016; contains corrected results for the 2005 article). Using simple returns (in contrast to our continuously compounded returns), a sample from 1928 to 2011, and normalized exponential Almon lag polynomial weighting with high frequency lags equal to 260 days, they estimate the lambda to be -0.0212 (t-value -0.0224).

It is evident from the results that the traditional approach is not able to provide a significant and theoretically plausible positive lambda estimate. Thus, we turn to the volatility-feedback version of the model, i.e. equation (6). The results are reported in Panel A of Table 3. As with the GARCH-variance, the volatility-feedback effect is statistically highly significant. The lambda estimates are still negative, this time highly significant. The same applies to the MIDAS parameters. Overall, the fit of the model is clearly higher, but the results are unsatisfactory as far as the Merton model is concerned.

Finally, we continue with the new approach and test our model. However, to test the model, we must obtain conditional variances. To this end, we proceed again with the two-step approach. First, we estimate the MIDAS model as before, but with one main difference: we use squared monthly returns as the dependent, lower-frequency variable. In practice, we estimate the following model:

$$r_{m,t+1}^2 = \alpha + \beta h_{t+1}^{MIDAS} + \varepsilon_{m,t+1},\ \varepsilon_{m,t+1} \sim Distr\left(0, h_{t+1}^{MIDAS}\right), \qquad (26)$$

where h_{t+1}^{MIDAS} is the contemporary variance estimate based on the daily squared returns. As a result, the beta parameter can be interpreted as a high-frequency response, linking daily and monthly squared returns. It should be close to one if daily variances can be used as a proxy for monthly variance. In the second stage, we use the variance estimates from the first stage (fitted monthly variances from the model) to test equation (17) using the OLS. For the regression, we estimate first the variance persistence parameter using equation (23) and use it to calculate the sigma multipliers.

The results are reported in Panel B of Table 3. The adjusted R-squared is for the tested model (17). Interestingly, the conditional variance process does not show the same level of persistence (the AR(1) parameter is estimated to be 0.503) when compared with the estimates for the GARCH model (cf. Table 2). This could be due to the MIDAS approach, which arguably may be able to track changes in the variance more quickly due to the use of higher-frequency data.

The results show the lambda estimate (1.507 with t-value 4.870) to be statistically significant at the one per cent level. This provides further support for the model introduced in this paper. The lambda estimate is in line with earlier estimates from the GJR-GARCH model. The results from the simplified model confirm the significance and positivity of the lambda estimate (1.696 with t-value 4.819). Again, the change in the dividend growth rate (b_3) and the change in the risk-free rate (b_4) are found statistically significant. The explanatory power of the model is somewhat lower than that of the GJR-GARCH (21.0 vs 32.8 and 40.9 percent). The explanatory power seems to be more in line with that of non-asymmetric GARCH modeled variance which might reflect the fact that MIDAS may not be able to handle asymmetric responses as well as GJR-GARCH.

4.3. Empirical results with the VIX

Our third alternative proxy for the conditional variance is the implied variance based on option prices observed in the market. In practice, we utilize squared values for the VIX data, but convert them into a monthly measure by dividing the values by twelve. The estimation is conducted using monthly data from January 1990 to December 2013. For comparison, we have also estimated the results using GJR-GARCH and MIDAS for the same period. We first estimate the model using the traditional approach. The estimation is conducted using equation (24). Results are reported in Panel A of Table 4. The results show that the traditional price of market risk estimate is -0.127 and that the estimate is not significant (t-value -0.092). In addition, the explanatory power of the model is low. The results from the GJR-GARCH and MIDAS are similar.

Next we estimate the volatility-feedback model. As the variance is readily available, we can directly estimate equation (6). The results in Table 4 show that, although the volatility-feedback effect is highly significant and the explanatory power of the model is significantly higher than before (the adjusted R^2 is 24.3 per cent), the price of market risk is still not significant and its estimate is negative (-0.638 with t-value -1.104). The results from the GJR-GARCH and MIDAS are basically similar although the former gives an unsignificant positive estimate for the price of risk and the latter significant, yet negative estimate for the price of risk. The results merely go on to show that there is some type of relationship between ex post returns and ex ante variance which the model captures.

Finally, we turn to the new model. We first estimate the variance persistence parameter using an AR(1) specification as in equation (23). Panel B of Table 4 shows that the implied variance exhibits a somewhat higher persistence (0.807) than does the variance implied by the MIDAS but on the other hand it is lower than that implied by the GARCH. After this, we calculate the sigma multipliers and proceed with the tests of our full model (equation (17)). The results with VIX variance again provide strong support for our model. The price of market risk estimate (0.616) is positive, within reasonable range, and statistically highly significant (t-value 6.437). The explanatory power of the model is again clearly higher than that of the traditional and volatility-feedback models.

Corresponding subsample results using GARCH and MIDAS variance confirm the positivity and significance of the estimated lambda parameter. Interestingly, we can observe that lambda is estimated to be higher for the post-1990 subsample than for the full sample using GARCH and MIDAS variance. For example, in case of MIDAS, the lambda estimate is 2.107 for the sample of 1990-2013 where the full sample estimate is 1.507. On the other hand, the parameter estimate for the change in the risk-free rate (b_4) is not statistically significant (except with MIDAS variance) for the subsample and its sign is unexpectedly negative. This could be driven by the fact that the assumption for the interest rate structure is too strong for the sub-period in question. We test this by separating changes in short and long-term risk-free rates into two variables and retesting equation (17) with five variables. However, there is no material change in the results (not reported) – the changes in the short-term interest rates are not found significant all the while the changes in the long-term risk-free rates remain significant. As such, a more complex approach may be needed to model changes in the risk-free rates.

The VIX provides us also with an opportunity to test the model as a joint system of two equations and thus avoiding the issues with two-step estimation. Combining equation (23) for the conditional variance with equation (17) for the risk premium and using the definitions for the sigma multipliers, we can estimate the system, for example, with the seemingly unrelated regression (SUR) method. SUR takes into account heteroskedasticity and contemporaneous correlation in the errors across equations. The results are in line with the results reported in Table 4. The lambda estimate is higher, 1.836, but it is still highly significant (t-value is 5.104). As before, the parameter estimate b_3 is highly significant whereas b_4 is not.[15]

4.4. Additional considerations and robustness checks

An obvious question is whether the results are driven by the sample period. To study this, we use a rolling window estimation to estimate lambdas over all possible sample periods with a fixed length. We begin the estimation with eighty year samples (giving us 72 possible samples, the last one beginning in December 1933) and then shorten the sample period by ten years in each step until we have samples covering only 20 years.

We observe a number of well-known empirical regularities. The results from the GARCH-type estimations are sensitive to the number of observations (length of the sample period) and the sample period itself. A number of obvious convergence issues can be detected for samples shorter than 60 years even for the simplest model. As expected, the issue is aggravated with the use of more complex models. The traditional approach is slightly more sensitive to these issues as it requires that the variance is also included in the mean equation. Obvi-

[15] We also estimated the system with the Generalized Method of Moments (GMM). Orthogonalizing on the constant and lagged values of VIX (equations (17) and (23)), contemporary values of VIX (equation (17)), as well as lagged and contemporary values for the expected dividend growth rate and risk-free rate, and using an iterative process for the weights, we obtained essentially the same results.

ously, a number of these issues could be avoided by fine-tuning the estimation. Nonetheless, the results clearly indicate that using too short sample periods can produce susceptible estimates for lambda, and, at the minimum, one should always conduct robustness checks to guarantee that the results are not driven by the sample in question.

As far as the results go, even with the simplest estimation setup, GARCH-M with normally distributed errors, the traditional model does not show a single significant lambda estimate using the 70- or 80-year samples; with 80-year sample, the new approach (using the simplified equation (18)) has nine significant lambda estimates out of seventy-two. When we allow for asymmetry in the variance, the traditional model is not doing any better, whereas the new approach finds all but one of the lambdas to be statistically significant with 80-year samples. With 70-year samples, the situation is the same. In fact, the sample had to be shortened to 50 years to find even a single significant lambda estimate with the traditional estimation approach. This goes to show that having a longer sample period does not necessarily solve the total volatility puzzle when the traditional model is used.

If one uses the GJR-GARCH variance, the traditional approach produces negative lambda estimates whereas the new approach gives positive ones with only few negative lambdas. For example, all traditional lambdas estimated using 70-year samples starting in July 1939 or later are negative. The new approach does not show even a single negative lambda. Hence, we can conclude that the main empirical result in this paper is not driven by the selection of the sample period nor the choice of the GARCH specification. Moreover, the results are similar for the full model (equation (17)) all the while the parameter estimates for the changes in the dividend growth rate and risk-free rate are also significant.

We also analyze the effect of replacing the constant in the mean equation with equation (4) in the first step when we estimate our model (equation (18)). Otherwise the estimation proceeds as before. The results are again basically the same. This is as expected given the poor performance of the traditional model. For example, using the 80-year samples, the average value for lambda was 1.124, but with the revised mean equation, it is 1.186. There are no major differences in estimates' statistical significance either.

Another question is whether our proxy variables for the sum of the changes in investors' views on future risk-free rates as well as dividend growth rates are justifiable. To study this, we first test whether changes in our proxy, the US government long-term bond yield, are positively related to changes in the short-term interest rate and the intermediate-term bond yield. Using the CRSP three-month interest rates and the Ibbotson intermediate US government bond yields, we find the contemporary correlations to be 0.485 and 0.825. These results show that changes in the long-term bond yield seems to reflect also changes in the short- and intermediate-term interest rates and thus movements in the yield curve. Admittedly, these three interest rates do not represent the whole yield curve and the relationship between the yield curve and the future interest rates is not exactly one-to-one. However, these results give reasonable justification for the assumption that higher long-term bond yields are associated with higher interest rates and vice versa.

Correspondingly, we study whether our dividend growth proxy is able to predict realized dividend growth in the future. Using our proxy variable to forecast realized dividend growth in the future, we find a positive and statistically significant relationship for one, two, and twelve periods. Hence we feel confident that our proxy is able to predict dividend growth and that it can be used as a proxy for changes in investors' views.

A related, yet slightly different question, is whether the results are robust to our choice of using past values of dividend growth as a forecast for the future growth. To test this, we estimate our model in Panel C of Table 2 with a number of alternative proxies for the dividend growth.[16] Overall, the main result continues to hold. In all cases, the parameter estimate for the price of variance risk is positive and significant – the effect on the lambda estimate is typically ± 0.2. The significance of the dividend growth component (i.e., parameter b_3 in equation (17)) seems to be slightly responsive to the proxy used, but this merely reflects the overall difficulty in forecasting dividend growth.[17]

[16] For example, we use the dividend growth over six-month periods instead of twelve months as well as the annual dividends calculated from monthly observations as in Golez (2014). Then we estimate AR(1) and AR(12) models for the original growth rate series and use the predicted dividend growth from the model. Finally, in the spirit of Jagannathan and Liu (2019), we estimate a VAR(1) model for the dividend growth and log price to dividend with and without the log of Shiller's CAPE, and use the model prediction as the dividend growth to re-estimate our model.

[17] The discussion in this paragraph is partly applicable also to our forecast for the risk-free rates. However, contrary to the dividend growth rate, the government long-term bond yield is a commonly agreed measure, and the future risk-free interest rates are interlinked via the term structure. We estimate the last model (GJR-GARCH with t-distributed residuals) in Panel C of Table 2 using the yield on intermediate government bonds as well as having the term related to the risk-free rates broken down into two components: the short-term (ST, 3-month T-bill) rate and the long-term (LT) government bond yield (in effect, in equation (17) we have $b_{4,ST}$ and $b_{4,LT}$). Again the main result is robust, the effect of the interest rate proxy on the price of market risk being small.

We also perform a number of additional robustness checks.[18] For example, we use simple returns instead of continuously compounded returns and different distributional assumptions for the GARCH process. The results are robust in all cases. We also test whether the lambda estimate is sensitive to the sampling frequency and the length of the return measurement horizon. Using one, two, five, and ten day non-overlapping returns over the same sample period, the traditional lambda estimates are all insignificant whereas the results from the new approach, on the other hand, show that lambdas are significant in all cases and that the estimates are aligned. Overall, we are confident that our results are robust.

Finally, we analyze the questions raised in the introduction. The first two questions are related to explaining why the lambdas calculated using the traditional approach differ from the lambdas based on the new approach and why they are typically smaller. Econometrically, the answer is obvious and based on an analogy for omitted variables bias. If the true model is $y_t = b_0 + b_1 x_t + e_t$, where $x_t = z_t - z_{t-1}$ (here: change in the conditional variance corresponding to the simplified model), estimating wrongly the model $y_t = b_0 + b_1 z_{t-1} + e_t$ means that one is estimating $y_t = b_0 + b_1(-x_t + z_t) + e_t$, which typically leads to a lower estimate for the parameter b_1 even if one takes into account the positive correlation between z_t and x_t. Another explanation is based on the fact that the traditional estimation approach is based on an implicit assumption that an increase in the variance affects the risk premium, which applies to investment periods across all time horizons – implicitly assuming that the term structure of the cost of capital is flat. Thus, if the investors take the convergence of the variance back to its long-term level into account, the estimated relationship between realized return to variance is smaller than assumed.

The third and the fourth questions are related to whether the sample period and the choice of the time aggregation can explain some of the different results in the literature as well as whether the same effect carries over to the new testing approach. To answer the third question, we note that the theoretical relationship in equation (3) tells us that at any point in time, lambda can be observed as the ratio of expected excess return divided by the conditional variance, i.e., $\lambda_m = E_t[r^e_{m,t+1}]/\sigma^2_t(r_{m,t+1})$. If one replaces expected return with the realized return – as done in the traditional approach – a negative observation would

[18] A more in-depth analysis of the tests and the results is available upon request.

imply a negative λ_m^T. Conversely, using equation (16) to solve for the price of market risk gives us

$$\lambda_m \approx \frac{r_{m,t+1} - k_2 - (g_{t+1} - g_{t,t+1}) \cdot \varphi_d - (r_{ft} - r_{ft+1}) \cdot \varphi_{rf}}{(\sigma^2_{t,t+1} - \rho\sigma^2_{t+1,t+2}) \cdot \varphi_{\Delta\sigma} + \sigma^2 \cdot \varphi_\sigma}. \tag{27}$$

Now, we can see that observing a negative realized return need not lead to a negative λ_m, because the denominator can also be negative during the same period, ceteris paribus.

If we consider our estimates for lambda, equation (4) gives us the following estimate for the traditional lambda: $\hat{\lambda}_m^T = Cov(\sigma^2_{t,t+1}, r^e_{m,t+1})/Var(\sigma^2_{t,t+1})$. It can become negative if the nominator is negative, i.e., $E[\sigma^2_{t,t+1} r^e_{m,t+1}] < (\sigma^2_m E[r^e_m])$. One can say that the traditional approach works best when the volatility surprises are fairly small as high volatility surprises cause negative outliers into the realized returns which bias the traditional estimate downwards (c.f., Ghysels et al., 2016).

The new approach gives the following estimate for lambda (for tractability, using the simplified equation (18)) $\hat{\lambda} = \varphi_{\Delta\sigma} Cov(\sigma^2_{t,t+1} - \rho\sigma^2_{t+1,t+2}, r^e_{m,t+1})/\varphi^2_{\Delta\sigma} Var(\sigma^2_{t,t+1} - \rho\sigma^2_{t+1,t+2})$. Now, the nominator can be negative if $E[(\sigma^2_{t,t+1} - \rho\sigma^2_{t+1,t+2}) r^e_{m,t+1}] < (1-\rho)(\sigma^2_m E[r^e_m])$, but given a situation where the traditional lambda can be negative (highly negative realised return driven by volatility surprises), we can see that the nominator does not need to be negative if high negative realized returns are associated with an increase in conditional variance.

To answer the fourth question, we can see that $\hat{\lambda}_m^T$ is in principle unaffected by the time aggregation if the variance time-aggregates linearly as returns do. However, this is not the case if the returns show autocorrelation. If the variance decreases slower than linearly for less-aggregated returns, traditional lambdas, estimated using returns measured over short periods, can be biased downwards in samples with positive average excess returns. This does not necessarily happen for lambdas estimated with the approach presented here if the variance persistence parameter varies for different return aggregation periods.

5. SUMMARY AND CONCLUSIONS

In this paper, we develop a new intuitive approach for testing conditional asset pricing models. This new reverse testing approach avoids the issues that arise when realized returns or their time series forecasts are used as a proxy for the expected returns in asset pricing tests.

When the new approach is applied to the Merton (1980) asset pricing model and combined with the assumption of mean-reverting conditional variance, it suggests an empirical model that links the realized equity premium to the price of market risk and to change in the conditional variance and to its long-term persistence as well as to the changes in the expected dividend growth rate and the risk-free rate.

Empirically, we study the relationship between the conditional equity market risk premium and variance using data for the US stock market from 1928 to 2013. For the empirical estimation of the model, we compare the traditional and the volatility-feedback testing approaches against the new approach introduced in this paper. We utilize and compare three different approaches to model the conditional variance. Our first specification utilizes the commonly used GARCH-M framework. In addition, we utilize the MIDAS approach of Ghysels et al. (2005, 2013) and the implied variance (VIX-index) observed on the options market.

The results show that neither the traditional nor the volatility-feedback approach give support for a positive relationship between conditional variance and equity premium. The price of market risk estimate is close to zero, and at times even negative. On the other hand, the price of market risk estimates from the new model are consistently statistically and economically significant, positive, and higher than those estimated using the traditional approach giving strong support for the Merton (1980) model. The results from the new approach are not dependent on the method used to estimate the variance, and they are also less sensitive to the timing of the sample and its length. In addition, the approach works even on return measurement horizons shorter than one month. We also find support for the importance of the changes in investors' view on dividend growth, but less so on the risk-free rate.

Overall, the results also give support for the new testing approach. Although the reverse (flipped) testing approach is simple and intuitive to use, it comes with a cost, as one has to apply auxiliary assumptions to reach convergence in its components. However, as such, the new approach helps to explain why earlier results may have been unable to uncover the risk-return relationship. It also suggests that we need to revisit some of the earlier empirical results on conditional asset pricing models. A natural extension is to study whether the hedging components are also priced and whether the price of risk is time-varying. However, these questions are left for future study.

Acknowledgements
We are grateful for the helpful comments from an anonymous referee as well as seminar and conference participants at the 2016 Eastern Finance Association 52nd Annual Meeting in Baltimore, USA; 2016 FMA European Conference in Helsinki, Finland; 22nd Annual Conference of the Multinational Finance Society in Halkidiki, Greece; 2016 Annual Conference of the European Financial Management Association in Basel, Switzerland; 22nd International Conference on Forecasting Financial Markets in Rennes, France; Arne Ryde Workshop in Financial Economics in Lund, Sweden; Scancor Seminar at Stanford University, USA; and seminars at Hanken School of Economics, University of Jyvaskyla, and University of Turku.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

Bali, T.G. (2008) The intertemporal relation between expected return and risk. Journal of Financial Economics, 87, 101-131.

Banerjee, P. S., Doran, J. S., and Peterson, D. R. (2007) Implied volatility and future portfolio returns. Journal of Banking and Finance, 31(10), 3183-3199.

Bekaert, G. and Wu, G. (2000) Asymmetric volatility and risk in equity markets. The Review of Financial Studies, 13(1), 1-42.

Black, F. (1976) Studies of stock price volatility changes. Proceedings of the Business Economic Statistical Section, Americal Statistical Association, 177-181.

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

Bollerslev, T., Engle, R. F. and Wooldridge, J. M. (1988) A capital asset pricing model with time-varying covariances. Journal of Political Economy, 96, 116-131.

Brav, A., Lehavy, R., and Michaely, R. (2005) Using expectations to test asset pricing models. Financial Management, 34(3), 31-64.

Campbell, J. Y. (1991) A variance decomposition for stock returns. The Economic Journal, 101, 157-179.

Campbell, J. Y. and Hentschel, L. (1992) No news is good news. An asymmetric model of changing volatility in stock returns. Journal of Financial Economics, 31, 281-318.

Campbell, J. Y., Lo, A. W., and MacKinlay, A. C. (1997) The econometrics of financial markets. USA: Princeton University Press.

Campbell, J. Y. and Shiller, R. J. (1988) The dividend-price ratio and expectations of future dividends and discount factors. The Review of Financial Studies, 1(3), 195-228.

Cappiello, L., Engle, R. F., and Sheppard, K. (2006) Asymmetric Dynamics in the Correlations of Global Equity and Bond Returns. Journal of Financial Econometrics, 4(4), 537-572.

Carr, P. and Wu, L. (2006) A tale of two indices. Journal of Derivatives, 13, 13-29.

CBOE (2009) The CBOE volatility index VIX. Available at: http://www.cboe.com/micro/vix/vixwhite.pdf

Cochrane, J. H. (2008) The dog that did not bark: A defense of return predictability. The Review of Financial Studies, 21(4), 1533-1575.

Corrado, C. and Miller, T. W. Jr. (2006) Estimating expected excess returns using historical and option-implied volatility. The Journal of Financial Research, 29(1), 95-112.

Elton, E. (1999) Expected return, realized return and asset pricing tests. The Journal of Finance, 54(4), 1199-1220.

Engle, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of the United Kingdom inflation. Econometrica, 50(4), 987-1007.

Engle, R.F. and Patton, A.J. (2001) What good is a volatility model? Quantitative Finance, 1(2), 237-245.

Ferson, W. E. (2003) Test of multifactor models, volatility bounds and portfolio performance. In Constantinides, G.M., Harris, M., and Stulz, R. (eds.) Handbook of Economics and Finance. Elsevier.

Feunou, B., Fontaine, J.-S., Taamouti, A., and Tdongap, R. (2014) Risk premium, variance premium, and the maturity structure of uncertainty. Review of Finance, 18(1), 219-269.

French, K. R., Schwert, G. W., and Stambaugh, R. F. (1987) Expected Stock Returns and Volatility. Journal of Financial Economy, 19, 3-29.

Ghysels, E. (2015) Matlab Toolbox for Mixed Sampling Frequency Data Analysis using MIDAS Regressions Models. Unpublished manuscript, available at: http://www.unc.edu/~eghysels/papers/MIDAS_Usersguide_V1.0.pdf

Ghysels, E., Plazzi, A., and Valkanov, R. (2016) The Risk-Return Relationship and Financial Crisis. Unpublished manuscript, available at: http://ssrn.com/abstract=2776702

Ghysels, E., Santa-Clara, P., and Valkanov, R. (2005) There is a Risk-Return Trade-off After All. Journal of Financial Economics, 76, 509-548.

Ghysels, E., Sinko, A., and Valkanov, R. (2007) MIDAS regressions: Further results and new directions. Econometric Review, 26(1), 53-90.

Glosten, L. R., Jagannathan, R., and 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.

Golez, B. (2014) Expected returns and dividend growth rates implied by derivative markets. The Review of Financial Studies, 27(3), 790-822.

Gonzales, M., Nave, J., and Rubio, G. (2012) The cross section of expected returns with MIDAS betas. Journal of Financial and Quantitative Analysis, 47(1), 115-135.

Greenwood, R. and Shleifer, A. (2014) Expectations of Returns and Expected Returns. Review of Financial Studies, 27(3), 714-746.

Guo, H. and Whitelaw, R. F. (2006) Uncovering the Risk-Return Relation in the Stock Market. The Journal of Finance, 61, 1433-1463.

Harvey, C. R. (2001) The specification of conditional expectations. Journal of Empirical Finance, 8(5), 573-637.

Hedegaard, E. and Hodrick, R. J. (2014) Measuring the Risk-Return Tradeoff with Time-Varying Conditional Covariances. NBER Working Paper Series 20245.

Hedegaard, E. and Hodrick, R. J. (2016) Estimating the Risk-Return Trade-off with Overlapping Data Inference. Journal of Banking and Finance, 67, 135-145.

Hibbert, A. M., Daigler, R. T., and Dupoyet, B. (2008) A behavioral explanation for the negative asymmetric return-volatility relation. Journal of Banking and

Finance, 32(10), 2254-2266.

Ibbotson SBBI (2014) 2014 Classic Yearbook. Market Results for Stocks, Bonds, Bills, and Inflation 1926-2013. USA: Morningstar.

Jagannathan, R. and Liu, B. (2019) Dividend Dynamics, Learning, and Expected Stock Index Returns. Journal of Finance, 74(1), 401-448.

Jarque, C. M. and Bera, A. K. (1987) A Test for Normality of Observations and Regression Residuals. International Statistical Review, 55, 163-172.

Kim, Y. and Nelson, C. R. (2014) Pricing stock market volatility: Does it matter whether the volatility is related to the business cycle. Journal of Financial Economics, 12(2), 307-328.

Lettau, M. and Ludvigson, S. C. (2005) Expected returns and expected dividend growth. Journal of Financial Economics, 76, 583-626.

Ljung, G. M. and Box, G. E. P. (1978) On a Measure of Lack of Fit in Time Series Models. Biometrika, 65, 297-303.

Malkiel, B. G. and Xu, Y. (2006) Idiosyncratic risk and security returns. Unpublished working paper available at http://utd.edu/~yexiaoxu/IVOT_H.PDF

Merton, R. C. (1973) An Intertemporal Capital Asset Pricing Model. Econometrica, 41(5), 867-887.

Merton, R. (1980) On estimating the expected return on the market. Journal of Financial Economics, 8, 323-361.

Merton, R.C. (1987) A Simple Model of Capital Market Equilibrium with Incomplete Information. The Journal of Finance, 42(3), 483-510.

Meyer, D. J. and Meyer, J. (2005) Relative risk aversion: What do we know? Journal of Risk and Uncertainty, 31(3), 243-262.

Pindyck, R. S. (1982) Risk, inflation, and the stock market. The Americal Economic Review, 74, 335-351.

Poon, S.H. and Grander, C.W.J. (2003) Forecasting Volatility in Financial Markets: A Review. Journal of Economic Literature, 41, 478-539.

Santa-Clara, P. and Yan, S. (2010) Crashes, Volatility, and the Equity Premium: Lessons from the S&P500 Options. The Review of Economics and Statistics, 92(2), 435-445.

Shanken, J. (1992) On the Estimation of Beta Pricing Models. The Review of Financial Studies, 5(1), 1-33.

Theodossiou, P. and Savva, C. S. (2016) Skewness and the Relation between Risk and Return. Management Science, 62(6), 1598-1609.

Table 1: Descriptive statistics for monthly and daily market returns as well as for the monthly VIX index. The sample periods for monthly data are from January 1928 to December 2013 (1,032 observations) and from January 1990 to December 2013 (288 observations). For daily data, the sample includes 22,682 and 6,049 observations for the full and sub samples, respectively. All returns are continuously compounded, in percentage form, and in excess of the one month risk-free rate if indicated. The dividend growth rate per annum is the continuously compounded growth of the dividends paid during the past twelve months compared to dividends paid a year ago. The long-term government bond yield per annum is taken from the SBBI book. VIX is the option implied volatility per annum. Monthly observations are day-matched using month-end dates for the CRSP index. Normality refers to the p-value for the Jarque-Bera (1987) test for normality. Q(3) is the Ljung-Box (1978) test for autocorrelation. Autocorrelation coefficients and Ljung-Box test statistics significantly (5%) different from zero are marked with an asterisk.

	Mean	**Std.**	**Skew-**	**Excess**	**J-B**	**Autocorrelation**			
Variable		**dev.**	**ness**	**Kurtosis**	(p-val)	ρ_1	ρ_2	ρ_3	**Q(3)**
Panel A: Jan 1928 - Dec 2013									
monthly CRSP R_m–R_f	0.477	5.44	−0.53	6.51	< 0.001	0.110*	−0.011	−0.081*	19.38*
monthly CRSP (R_m)ˆ2	30.055	85.08	8.28	84.62	< 0.001	0.230*	0.169*	0.261*	155.59*
dividend growth rate p.a.	4.236	17.63	−0.50	0.91	< 0.001	0.697*	0.586*	0.677*	> 999.99*
government bond yield p.a.	5.092	2.63	0.96	0.30	< 0.001	0.996*	0.991*	0.987*	> 999.99*
daily CRSP R_m	0.035	1.08	−0.43	17.40	< 0.001	0.071*	−0.040*	0.004	151.34*
daily CRSP (R_m)ˆ2	1.169	5.13	27.58	1417.12	< 0.001	0.244*	0.273*	0.197*	> 999.99*
Panel B: Jan 1990 - Dec 2013									
monthly CRSP R_m–R_f	0.511	4.46	−0.85	1.60	< 0.001	0.101*	−0.010	0.049	3.73
monthly CRSP (R_m)ˆ2	20.377	35.20	5.13	38.56	< 0.001	0.192*	0.076	0.146*	18.94*
dividend growth rate p.a.	5.438	15.25	−0.17	0.15	0.464	0.845*	0.728*	0.676*	465.71*
government bond yield p.a.	5.433	1.54	0.01	−0.53	0.182	0.980*	0.958*	0.939*	> 999.99*
daily CRSP R_m	0.037	1.15	−0.29	8.10	< 0.001	−0.029*	−0.040*	0.007	15.37*
daily CRSP (R_m)ˆ2	1.323	4.19	12.08	220.56	< 0.001	0.212*	0.381*	0.207*	> 999.99*
VIX	20.195	7.69	1.63	4.10	< 0.001	0.854*	0.711*	0.621*	472.66*
VIXˆ2	466.800	421.52	3.38	16.50	< 0.001	0.809*	0.601*	0.509*	371.84*

Table 2: Empirical results with GARCH variance. Quasi-maximum likelihood (QML) estimates for the constant price of market risk are reported for the traditional (Panel A) and volatility-feedback approaches (Panel B) under different GARCH and GJR-GARCH-in-mean models. The volatility-feedback approach adds the surprise in the variance, measured as the difference between realized variance and conditional variance (coefficient γ_m). Bollerslev-Wooldridge (1992) robust standard errors are used in the case of the normal distribution. In Panel C, parameter estimates are from a two-step procudure. In the first step, the conditional variance is estimated using QML. In the second step, the excess returns are regressed on the difference in the conditional variance as well as on the changes in the risk free rates and conditional dividend growth rates (equation (17)). The Newey-West (1987) adjustment has been used to calculate standard errors for the mean equation. Excess US continuously compounded returns (CRSP total return index) from January 1928 to December 2013 (1,032 observations) are used in the estimation. The adjusted R^2 is for the mean equation. t-values are provided in parentheses. Coefficients significantly (10%, 5% or 1%) different from zero are marked with one, two, or three asterisks, respectively.

	Mean equation			**Variance equation**				
	Constant	λ_m	γ_m	ω	α_1	β_1	$\gamma(\varepsilon_t < 0)$	**Adj.** R^2
Panel A: Traditional approach								
GARCH(1,1)-M	0.005***	0.697		0.000**	0.129***	0.854***		−0.004
+ N-distributed errors	(2.660)	(0.851)		(2.121)	(4.359)	(25.137)		
GARCH(1,1)-M	0.008***	0.628		0.000**	0.128***	0.836***		−0.008
+ t-distributed errors	(3.673)	(0.728)		(2.683)	(4.296)	(25.201)		
GJR-GARCH(1,1)-M	0.006***	0.224		0.000**	0.067*	0.853***	0.093**	−0.002
+ N-distributed errors	(2.765)	(0.261)		(2.387)	(1.655)	(23.961)	(2.014)	
GJR-GARCH(1,1)-M	0.008***	0.110		0.000**	0.033	0.830***	0.142***	−0.005
+ t-distributed errors	(4.009)	(0.132)		(3.243)	(0.870)	(23.128)	(3.100)	
Panel B: Volatility-feedback approach								
GARCH(1,1)-M	0.006***	0.480	−5.688***	0.000***	0.168***	0.795***		0.115
+ N-distributed errors	(3.544)	(0.468)	(−10.518)	(3.954)	(6.251)	(32.673)		
GARCH(1,1)-M	0.006***	0.517	−5.415***	0.000***	0.162***	0.795***		0.120
+ t-distributed errors	(3.442)	(0.539)	(−24.457)	(3.025)	(5.360)	(24.770)		
GJR-GARCH(1,1)-M	0.006***	−0.159	−5.442***	0.000***	0.096***	0.799***	0.126***	0.121
+ N-distributed errors	(3.951)	(−0.189)	(−10.711)	(4.056)	(3.362)	(30.052)	(3.771)	
GJR-GARCH(1,1)-M	0.006***	−0.002	−5.175***	0.000**	0.085***	0.799***	0.125***	0.123
+ t-distributed errors	(3.972)	(−0.002)	(−21.767)	(3.380)	(2.612)	(25.302)	(2.937)	

Table 2 continued.

	Mean equation				Variance equation				
	b_1	b_2	b_3	b_4	ω	α_1	β_1	$\gamma(\varepsilon_t < 0)$	**Adj.** R^2
Panel C: New approach, full model									
GARCH(1,1)-M + OLS	0.004 **	0.240*	0.132***	2.208***	0.000 **	0.129***	0.851***		0.177
+ N-distributed errors	(2.034)	(1.658)	(6.459)	(3.257)	(2.967)	(6.184)	(44.346)		
GARCH(1,1)-M + OLS	0.003	0.569*	0.130***	2.172***	0.000 **	0.136***	0.815***		0.186
+ t-distributed errors	(1.527)	(1.895)	(6.196)	(3.229)	(2.700)	(4.481)	(21.404)		
GJR-GARCH(1,1)-M + OLS	0.003	0.806***	0.105***	1.984***	0.000***	0.065 **	0.854***	0.096 **	0.328
+ N-distributed errors	(1.693)	(8.170)	(5.449)	(3.288)	(3.032)	(2.233)	(41.912)	(2.480)	
GJR-GARCH(1,1)-M + OLS	0.001	1.630***	0.090***	1.804***	0.000 **	0.033	0.830***	0.143***	0.409
+ t-distributed errors	(0.722)	(8.794)	(5.559)	(3.101)	(2.834)	(1.210)	(20.577)	(2.653)	

Table 3: Empirical results with MIDAS variance. Maximum likelihood estimates for the price of market risk (λ_m) as well as for the weighting structure parameters are reported in Panel A for the traditional and volatility-feedback appraches. The volatility-feedback approach adds the surprise in the contemporary variance into the model (coefficient γ_m). The estimation is conducted using monthly returns and daily returns squared where monthly returns are in excess of the risk-free rate. Continuously compounded returns have been used in all estimations. The sample is from from January 1928 to December 2013 (1,032 observations). Reported MIDAS estimates are for the normalized beta density function with a zero last lag. In Panel B, parameter estimates for the new approach are from a two-step procudure. In the first step, a MIDAS estimation has been conducted to get estimates for the conditional variance. Squared monthly returns have been used as the dependent variable. In the second step, fitted variance estimates from the first step have been used to estimate the price of market risk using the OLS as given by equation (17) (full model). The Newey-West (1987) adjustment has been used to calculate standard errors. The required sigma-multiplier have been estimated using an AR(1) model for the conditional variance series (reported alongside MIDAS parameter estimates). The adjusted R-squared is for the mean equation. t-values are provided in parentheses. Coefficients significantly (10%, 5% or 1%) different from zero are marked with one, two, or three asterisks, respectively.

	Mean equation			**MIDAS-parameters**			**AR(1) parameters**		
	Constant	λ_m	γ_m	β	θ_1	θ_2	ϕ_0	ϕ_1	**adj.** R^2
Panel A: Traditional approach									
Traditional approach	0.006***	−0.358			0.957***	8.599			0.001
	(3.011)	(−1.072)			(7.482)	(0.319)			
Volatility feedback approach	0.010***	−1.707***	−4.433***		0.951***	1.133***			0.117
	(5.112)	(−4.514)	(−11.588)		(55.130)	(4.053)			

	Mean equation				**MIDAS-parameters**			**AR(1) parameters**		
	b_1	b_2	b_3	b_4	β	θ_1	θ_2	ϕ_0	ϕ_1	**adj.** R^2
Panel B: New approach										
Full model	0.000	1.507***	0.125***	2.240***	0.964***	0.950***	2.108 **	0.001***	0.503***	0.210
	(0.156)	(4.870)	(7.070)	(3.272)	(21.784)	(170.112)	(9.210)	(7.920)	(8.583)	

Table 4: Empirical results with VIX variance. OLS estimates for the price of market risk are reported for different models. The price of market risk (λ_m) is first estimated using the traditional approach by regressing excess market returns on the conditional variance for the same period (available at the beginning of the period). The volatility-feedback approach adds the surprise in the variance, measured as the difference between realized variance and conditional variance (coefficient γ_m). In Panel B, the price of market risk has been estimated using OLS on the equation (17) (full model). The parameter b_2 corresponds to the price of risk (λ_m) estimate. Parameters required for the sigma-multiplier are estimated by an AR(1) model for the conditional variance. Results from similar estimations using GJR-GARCH and MIDAS variance are also provided. The estimation is conducted using monthly data from January 1990 to December 2013 (288 observations). Market returns are measured by the CRSP total return index. Realized variance is the sum of daily return squared within a month. The implied variance is the VIX-index squared divided by twelve. Returns are continuously compounded and in excess of the risk-free rate. The Newey-West (1987) adjustment has been used to calculate standard errors. t-values are provided in parentheses. Coefficients significantly (10%, 5% or 1%) different from zero are marked with one, two, or three asterisks, respectively.

	Traditional model			**Volatility feedback model**			
	Constant	λ_m	Adj. R^2	Constant	λ_m	γ_m	Adj. R^2
Panel A: Traditional approach							
VIX	0.006	−0.127	−0.003	0.001	−0.638	−6.252***	0.243
	(1.312)	(−0.092)		(0.280)	(−1.104)	(−5.627)	
GJR-GARCH	0.008**	−0.384	−0.006	0.007*	1.889	−4.521***	0.163
+ t-distributed errors	(2.045)	(−0.161)		(1.971)	(0.702)	(−9.817)	
MIDAS	0.009***	−1.296**	0.022	0.014***	−3.070***	−5.978***	0.190
	(2.949)	(−2.524)		(4.893)	(−5.973)	(−7.649)	
					AR(1) parameters		
	b_1	b_2	b_3	b_4	ϕ_0	ϕ_1	**Adj.** R^2
Panel B: New approach							
Full model (VIX)	−0.004	0.616***	0.138***	−0.522	0.001***	0.807***	0.506
	(−1.692)	(6.437)	(5.307)	(−0.378)	(3.385)	(10.905)	
Full model (GJR-GARCH)	0.002	2.478***	0.113***	−0.894			0.554
+ t-distributed errors	(0.794)	(7.122)	(4.797)	(−0.735)			
Full model (MIDAS)	0.002	2.107***	0.221***	−2.831***	0.001***	0.641***	0.324
	(0.789)	(5.205)	(6.970)	(−2.525)	(3.107)	(4.960)	