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

Jan Antell and Mika Vaihekoski†

April 29, 2016

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 (1973, 1980) model and a long-standing risk-return puzzle: the price of market risk has often turned out to be negative and insignificant. Comparing the price of market risk estimates from the new and the traditional testing approaches, results from the new testing approach on US data give strong support positive relationship between conditional variance and equity premium.

Keywords: conditional asset pricing, expected return, variance, risk-return trade-off, risk aversion, volatility-feedback, equity premium

JEL classification: F3, G12, G15.

*We are grateful for the helpful comments from seminar and conference participants at the 2016 Eastern Finance Association 52nd Annual Meeting in Baltimore, USA, 22nd Annual Conference of the Multinational Finance Society in Halkidiki, Greece, 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 Jyväskylä, and University of Turku. An earlier version of the paper was circulated under the title "Expected return and variance: Lambda is alive, significant and positive."

†In alphabetical order. Jan Antell is with the Hanken School of Economics, Department of Finance and Statistics. Address: P.O. Box 479, 00101 Helsingfors, Finland. E-mail: jan.antell@hanken.fi. Mika Vaihekoski is with the Turku School of Economics, University of Turku. Address: Turku School of Economics, Department of Accounting and Finance, 20014 University of Turku, Finland. E-mail: mika.vaihekoski@utu.fi. This research was finished, in part, while Vaihekoski was visiting Scancor at Stanford University.

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 a dividend discount model in the spirit of Campbell and Hentschel (1992) with the selected conditional asset pricing model to study the model and the risk-return trade-off. Our approach differs from traditional testing approaches in a sense that our approach relates realized returns directly 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

[1] The framework in Campbell et al. (1987) is obviously closely related to our study. However, they assume that the variables in the pricing model behaves in autoregressive manner (AR(1) used as an example) which differs from our setup. Also the empirical model derived in Guo and Whitelaw (2006) to study the Merton (1973) ICAPM is closely related to our model, although the motivation and the scope are different. The model derived herein has also some resemblance to the Ang and Liu (2004) valuation framework, and to Callen and Lyle (2014).

[2] The new 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.

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 methodology 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) shows that if investors cannot hold the true market portfolio, rational investors focus not only on systematic risk but also on non-systematic risk. This implies that the total volatility puzzle can be attributed to using a suboptimal proxy for the market portfolio. 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 (see Harvey and Siddique, 2000; Lanne and Saikkonen, 2006; Theodossiou and Savva, 2015).

Two closely related explanations are 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 motive an empirical model where the realized equity premium is related both to conditional as well as unexpected variance. 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 statistically significant simultaneously.

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) because it allows for time-variation in the variance process instead of using a constant measure of the variance. A further development was the GARCH-in-mean specification of Bollerslev et al. (1988), which allows a comprehensive procedure to test the risk-return relationship. Furthermore, researchers have argued that the relationship is asymmetric rather than linear. This variation can easily be addressed by asymmetric GARCH models (see, e.g., Glosten et al., 1993; Bekaert and Wu, 2000). Ultimately, an enormous number of different specifications in the GARCH family have been proposed, including multivariate extensions and non-normal distributions.

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 2013).

Alongside the development of new econometric estimation techniques for the conditional variance, other approaches also have 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 volatility 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 (2014a) provide potential explanations for why the risk-return trade-off 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.

Our new 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 lambda is not significant if one only links realized returns at time t to the contemporaneous conditional variance as is done by most 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 during period $t + 1$ is positively related to the conditional variance at the end of the period. However, this is unlikely to be the case. Our model has also some resemblance to Guo and Whitelaw (2006). They connect a log linearization to Merton's (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 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 volatilities over a short sample from 1983(11) to 2001(5), Guo and Whitelaw find that the price of risk 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. 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. To avoid issues in two-stage estimation, 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. The results show that the lambda estimates from the new model are consistently significant, positive, less sensitive to the sample period, and higher than the lambdas estimated using the traditional approach.

The remainder of the paper is organized as follows. Section 2 presents the theoretical background, a new model for the return-variance relationship, and the empirical research methodology. Econometric issues related to the estimation of the models also are discussed. Section 3 introduces the data used in this paper. Section 4 shows the empirical results together with some additional robustness analysis. Section 5 presents the conclusions and offers suggestions for further research.

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

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

where $E\left[r_{i,t+1}^{e}\left|\Omega_t\right.\right]$ and $E\left[r_{m,t+1}^{e}\left|\Omega_t\right.\right]$ 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 $E\left[r_{m,t+1}^{e}\left|\Omega_t\right.\right] Var\left(r_{m,t+1}|\Omega_t\right)^{-1}$ 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 equals 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 a 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

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

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 approaches

The theoretical model (3) has a number of empirical implications. To test the model, one must provide empirical proxies for expected returns and conditional variance. 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 systematically 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. 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). Its popularity is based on the fact that it 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 common approach to estimate the lambda is the univariate GARCH(1,1)-in-Mean model in which one combines equation (4) with the assumption that $\varepsilon_{m,t+1} \sim iid\left(0, \sigma^2_{m,t+1}\right)$ and the following process for the conditional variance:

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

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, for example, allowing 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 return measurent intervals typically used in asset pricing tests, and that therefore this approach is not the best approach for empirical tests of conditional asset pricing models that allow for time-varying parameters and for expectations. 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 with 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; Munk, 2013). The same conclusion can be also 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.

There are several explanations for the low estimates of lambda. For example, empirical tests of the equation (3) are often based on an implicit assumption that an increase in the variance affects the risk premium, which applies to investment periods across all time horizons – meaning that the term structure of the cost of capital is flat – and, as a result, one does not take into account the mean convergence in the variance. A flat term structure for the cost of capital implies that the required rate of return increases for the next period, and the one after that, and so on. However, in this situation, using Gordon's dividend discount model and reasonable parameter values, it is easy to show that if the volatility increases to 20 percent, it leads to a return shock that should be closer to minus 54.44 percent, *ceteris paribus*. Obviously and intuitively, this is not realistic. Investors do not expect higher volatility to last forever. Thus, they might increase their required rate of return for some limited number of periods. Over time, one can expect the required rate of return to revert to the long-term mean. This implies a term structure for the risk premium (c.f., Feunou et al., 2014).

The solution suggested in the literature is to analyze the unexpected shock to the variance separately from the theoretical relationship. This volatility-feedback effect offers an alternative way to estimate lambda. Following French et al. (1987) adding the unexpected variance into equation (4) gives us the following model:

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

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 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 $r^e_{m,t+1} = \mu + \delta_m \sigma^2_{m,t+1} + \gamma_m \sigma^2_{r,m,t+1} + \varepsilon_{m,t+1}$. As $\delta_m = \lambda_m - \gamma_m$, an estimate for lambda can be calculated as the sum of δ_m and γ_m. We call this equation as the *volatility-feedback approach* to estimating lambda. Although, 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 kind of realized returns would be observed, given that the asset pricing model is correct. To investigate this, we create a model in the spirit of Campbell and Hentschel (1992).[3] 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. We do assume that investors price stocks by discounting future cash flows, here taken to be dividends, or at least that the dividend discount model can be used to match the price observed in the market. 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. Hence, the first dividend is paid at time $t + 1$. Now the price of the security at time t can be stated as

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

$$P_t = (P_{t,t+1} + D_{t,t+1}) exp\left(-r_{t,t+1}\right), \tag{7}$$

where $r_{t,t+1}$ expresses the continuously compounded required rate of return for the period from time t to t+1 whereas $P_{t,t+1}$ and $D_{t,t+1}$ represent the expected price and dividend occurring at the end of the period, all conditional on information available at time t.

Following Campbell and Hentschel (1992), equation (7) is linearized by taking the logarithm of both sides and imposing a first-order Taylor series expansion around the average logarithmic dividend-price ratio:

$$ln\left(P_t\right) \approx k_1 + \rho ln\left(P_{t,t+1}\right) + (1-\rho)\, ln\left(D_{t,t+1}\right) - r_{t,t+1}, \tag{8}$$

where $k_1 \equiv -ln\left(\rho\right) - (1-\rho)\, ln\left(1/\rho - 1\right)$ and $\rho \equiv 1/(1+\exp(\overline{d-p}))$.[4] Lower-case letters refer to the logarithm of the variable throughout the paper. Because dividends cannot be negative, ρ is positive and less than one by definition. Furthermore, if the dividend-price ratio remains constant over time, rho is equal to the stock price divided by the sum of the stock price and the dividend. For any reasonable values of rho (> 0.9) it is easy to see that k_1 is also positive. Campbell, Lo, and MacKinlay (1997) suggest that ρ should be 0.997 for monthly data.

Using repeated replacements and imposing the terminal condition as in Campbell and Hentschel (1992), we can write the logarithm of the price as a function of the future dividends and discount rates

$$ln\left(P_t\right) \approx \frac{k_1}{1-\rho} + (1-\rho)\sum_{i=1}^{\infty}\rho^{i-1} d_{t,t+i} - \sum_{i=1}^{\infty}\rho^{i-1} r_{t,t+i}, \tag{9}$$

where $d_{t,t+i} = ln\left(D_{t,t+i}\right)$ is the logarithm of the expected dividend at time $t+i$ $(i > 0)$ and $r_{t,t+i}$ is the continuously compounded required rate of return for period t+i, both conditional on information available at time t. Deriving a similar expression for the log price at t+1, we get

$$ln\left(P_{t+1}\right) \approx \frac{k_1}{1-\rho} + (1-\rho)\sum_{i=1}^{\infty}\rho^{i-1} d_{t+1,t+1+i} - \sum_{i=1}^{\infty}\rho^{i-1} r_{t+1,t+1+i}. \tag{10}$$

Next, we utilize the fact that the continuously compounded *realized return* at time t+1 can be written using a first-order Taylor log-linearization as

$$r_{t+1} \approx k_1 + \rho ln(P_{t+1}) - ln(P_t) + (1-\rho) ln(D_{t+1}), \tag{11}$$

[4] See Internet Appendix ?? for more detailed derivation.

where $D_{t,t+1}$ is the realized dividend at time $t+1$. Inserting log prices (equations (9) and (10)) into equation (11), we get the following expression for realized returns:

$$r_{t+1} \approx k_1 + \rho \left[\frac{k_1}{1-\rho} + (1-\rho) \sum_{i=1}^{\infty} \rho^{i-1} d_{t+1,t+1+i} - \sum_{i=1}^{\infty} \rho^{i-1} r_{t+1,t+1+i} \right] - \left[\frac{k_1}{1-\rho} + (1-\rho) \sum_{i=1}^{\infty} \rho^{i-1} d_{t,t+i} - \sum_{i=1}^{\infty} \rho^{i-1} r_{t,t+i} \right] + (1-\rho)\, ln\left(D_{t+1}\right). \quad (12)$$

Rearranging and collecting the constants together, the equation can be written as

$$r_{t+1} \approx k_2 + (1-\rho) \left[\sum_{i=1}^{\infty} \rho^{i} \left(d_{t+1,t+1+i} - \rho^{-1} d_{t,t+i} \right) \right] + \rho^{-1} \sum_{i=1}^{\infty} \rho^{i} \left(r_{t,t+i} - \rho r_{t+1,t+1+i} \right), \quad (13)$$

where $k_2 = (1-\rho)\, d_{t+1}$. From this point forward our setup differs slightly from Campbell and Hentschel (1992). From now on, we focus on the market portfolio. Using equation (3) as our candidate asset pricing model for the conditional expected (required) returns for the market portfolio and assuming here for convenience that the price of market risk is constant, we can rewrite the last term of equation (13) as follows:

$$\begin{aligned} &\rho^{-1} \sum_{i=1}^{\infty} \rho^{i} \left(r_{t,t+i} - \rho r_{t+1,t+1+i} \right) \\ &= \rho^{-1} \sum_{i=1}^{\infty} \rho^{i} \left(r_{ft,t+i} - \rho r_{ft+1,t+1+i} + \lambda_m \left(\sigma^2_{t,t+i} - \rho \sigma^2_{t+1,t+1+i} \right) \right) \\ &= \rho^{-1} \sum_{i=1}^{\infty} \rho^{i} \left(r_{ft,t+i} - \rho r_{ft+1,t+1+i} \right) + \rho^{-1} \lambda_m \sum_{i=1}^{\infty} \rho^{i} \left(\sigma^2_{t,t+i} - \rho \sigma^2_{t+1,t+1+i} \right). \end{aligned} \quad (14)$$

As a result, our model differs 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) *level* of the conditional variance. We also focus on market risk, whereas Campbell and Hentschel focus on innovations in the dividends.

Using 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 can calculate conditional multistep forecasts i periods ahead using the conditional variance for the next period. Here, 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 follows:

$$\sigma^2_{t,t+i} = \phi^{i-1}\sigma^2_{t,t+1} + \sigma^2\left(1-\phi^{i-1}\right), \tag{15}$$

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 term in the parentheses of equation (14) can be rewritten using the fact that equation (15) implies that

$$\sigma^2_{t,t+i} - \sigma^2_{t+1,t+1+i} = \phi^{i-1}\left(\sigma^2_{t,t+1} - \sigma^2_{t+1,t+2}\right) + \sigma^2(1-\rho)\left(1-\phi^{i-1}\right). \tag{16}$$

Inserting (16) into (14), we can rewrite equation (13) for the realized returns as

$$\begin{aligned} r_{m,t+1} \approx k_2 + (1-\rho)&\left[\sum_{i=1}^{\infty}\rho^i(d_{t+1,t+1+i} - \rho^{-1}d_{t,t+i})\right] \\ &+ \rho^{-1}\sum_{i=1}^{\infty}\rho^i\left(r_{ft,t+i} - \rho r_{ft+1,t+1+i}\right) \\ &+ \lambda_m\rho^{-1}\sum_{i=1}^{\infty}\rho^i\left[\phi^{i-1}\left(\sigma^2_{t,t+1} - \rho\sigma^2_{t+1,t+2}\right) + \sigma^2(1-\rho)\left(1-\phi^{i-1}\right)\right]. \end{aligned} \tag{17}$$

After some modifications, the equation can be written in the form

$$\begin{aligned} r_{m,t+1} \approx k_2 + (1-\rho)&\left[\sum_{i=1}^{\infty}\rho^i(d_{t+1,t+1+i} - \rho^{-1}d_{t,t+i})\right] \\ &+ \rho^{-1}\sum_{i=1}^{\infty}\rho^i\left(r_{ft,t+i} - \rho r_{ft+1,t+1+i}\right) \\ &+ \lambda_m[\left(\sigma^2_{t,t+1} - \sigma^2_{t+1,t+2}\right)\cdot\varphi_{\Delta\sigma} + \sigma^2\cdot\varphi_\sigma], \end{aligned} \tag{18}$$

where

$$\varphi_{\Delta\sigma} = \rho^{-1}\phi^{-1}\sum_{i=1}^{\infty}\rho^i\phi^i = \frac{1}{1-\rho\phi}, \tag{19}$$

and

$$\varphi_{\sigma} = \rho^{-1}(1-\rho)\left[\sum_{i=1}^{\infty}\rho^i - \phi^{-1}\sum_{i=1}^{\infty}\rho^i\phi^i\right] = \left(1-\frac{1-\rho}{1-\rho\phi}\right). \tag{20}$$

Henceforth $\varphi_{\Delta\sigma}$ and φ_{σ} are collectively called sigma multipliers. In practice, 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.

Now, for simplicity and insight, 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 maturities. For example, at time t, the risk-free rate $r_{ft,t+i}$ is the same for all values of i.[5] Similarly, we also assume that dividends grow at a constant growth rate (for a certain period)—or that if they do not, one can derive an implicit growth rate that yields the prevailing price for the stock. Thus, future dividends can be written as a function of time, the last observed dividend and their growth rate. Assuming further that the growth rate is equal over time, we can write the dividend term in equation (18) as the sum of a constant and the following equation:

$$(1-\rho)\sum_{i=1}^{\infty}\rho^i\left(d_{t+1,t+1+i}-d_{t,t+i+1}\right) = (1-\rho)\sum_{i=1}^{\infty}\rho^i\left(d_{t+1}+(g_{t+1}\cdot i)-d_{t,t+1}-(g_t\cdot i)\right), \tag{21}$$

where g_t and g_{t+1} are the expected future dividend growth rates, conditional on information known at times t and $t+1$.[6] Assuming that the conditional dividend growth rates start to converge in the long run (or alternatively that the dividends converge in the long run), we can simplify equation (21). This assumption is consistent with the intuition, because new information at time t+1 may affect investors' expectations regarding the dividends in the short run, but this effect is unlikely over the long term. Assuming further that dividend d_{t+1} paid at time t+1 is equal to its expected value $d_{t,t+1}$, and inserting the results into equation (18), we get the following result:

[5] A light variation would be to assume that the interest rate changes take place by parallel shifts in the term structure.

[6] Here we utilize $\sum_{i=1}^{\infty}\rho^i(d_{t+1,t+1+i}-\rho^{-1}d_{t,t+i}) = -d_{t,t+1}+\sum_{i=1}^{\infty}\rho^i\left(d_{t+1,t+1+i}-d_{t,t+i+1}\right)$.

$$r_{m,t+1} \approx k_3 + (g_{t+1} - g_t) \cdot \varphi_d + (r_{ft} - r_{ft+1}) \cdot \varphi_{rf} + \lambda_m \left((\sigma^2_{t,t+1} - \sigma^2_{t+1,t+2}) \cdot \varphi_{\Delta\sigma} + \sigma^2 \cdot \varphi_\sigma \right), \tag{22}$$

where φ_d and φ_{rf} can be interpreted as parameters that measure the effect of the change in the dividend growth rate and risk-free rate, respectively. Both of them are by definition positive. The constant k_3 is defined as

$$k_3 = (1 - \rho)\,(d_{t+1} - d_{t,t+1}) + r_{ft}. \tag{23}$$

The values of parameter k_3 can be either positive or negative because the first term (dividend forecast error) is either positive or negative and the last term is positive.

Analyzing equation (22) shows that realized returns are higher if investors' conditional expectations of the long-term 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 (22) equals the expected return given by equation (3) if investors' conditional expectations prove to be right (proof can be found from Internat Appendix **??**).

We can study whether our derivation of the lambda can be used to explain some of the anomalies observed in the market. There are two important sub-questions. The first question is related to whether the lambdas calculated using the new approach differ from the lambdas based on the traditional approach and, if they do, how. The second question is related to whether time aggregation has an effect on the result. To answer the first question, we solve for the price of market risk at time $t{+}1$ from equation (22):

$$\lambda_m \approx \frac{r_{m,t+1} - k_3 - (g_{t+1} - g_t) \cdot \varphi_d - (r_{ft} - r_{ft+1}) \cdot \varphi_{rf}}{\left(\sigma^2_{t,t+1} - \rho\sigma^2_{t+1,t+2}\right) \cdot \varphi_{\Delta\sigma} + \sigma^2 \cdot \varphi_\sigma}. \tag{24}$$

Most of the earlier studies using, for example, the GARCH-M framework, estimate a rational expectations lambda (labeled henceforth as λ^T_m or traditional lambda) as a ratio of mean realized return to realized variance for the same period, i.e., $\lambda^T_m = \overline{r_m}/\sigma^2$. It is easy to see why the estimate for λ^T_m can be negative if the mean realized market return has been negative during the sample period. Conversely, taking expectations of equation (24), we can see that a negative average realized return need not lead to a negative lambda, because the denominator can also be negative during the same period. Obviously, if the sample period is sufficiently long, the lambda estimates in both approaches converge as the denominator converges to the unconditional variance.

To answer the second question, we can see that λ_m^T is in principle unaffected by the time aggregation if the variance time-aggregate linearly as returns. However, this is not the case if the returns show, e.g., 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. This does not necessarily happen for lambdas estimated with the approach presented here if the variance persistence parameter varies for different return aggregation periods.

Finally, we can also study the sensitivity of our lambda to measurement errors, for example, due to a short sample. To this end, we assume that the expected divided growth rates and risk-free rates are two and three percent per annum, respectively. The lambda is further assumed to be two, rho 0.997, and the variance persistence parameter 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 dividend at time $t{+}1$ are assumed to be 1/12 dollar. We also assume that the variance exhibits mean-reversion as in equation (15) 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 $t{+}1$. 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 2,000 months in the analysis, the realized and expected returns are equal, 0.917 per cent.

Now, using equation (24), we can solve for lambda. Obviously, using the parameters above, lambda is two with high accuracy. Allowing for errors in the underlying parameters provides us an opportunity to analyze the sensitivity of the lambda to the underlying parameters. Using reasonable values for the parameters, lambda is almost unchanged with respect to changes in the values for dividend growth rate, risk-free rate, level of volatility, and variance persistence (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 our lambda is most sensitive to the relationship between conditional volatilities. If the conditional volatility decreases (increases), for example, 2.5% from time t to $t{+}1$ (corresponding to a sample mean realized monthly return of –3.47% and +5.39%, respectively), our lambda is biased downwards by 1.268% (upwards by 3.815%). However, it is very unlikely that variance 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 size of 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 of λ_m. 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 (22) in excess-return form as

$$r_{m,t+1}^e = b_1 + b_2[(\sigma_{t,t+1}^2 - \sigma_{t+1,t+2}^2) \cdot \varphi_{\Delta\sigma} + \sigma^2 \cdot \varphi_\sigma] + b_3 (g_{t+1} - g_t) + b_4 (r_{ft} - r_{ft+1}) + u_{m,t+1}, \quad (25)$$

where b_1 to b_4 are the coefficients to be estimated. All coefficients are expected to be positive. b_2 is our estimate for the lambda. The term g_t is the expected continuously compounded growth rate for the dividends over one period starting from time t onwards, conditional on information available at time t. The risk-free rate at time t is given by r_{ft}. The term $\sigma_{t,t+1}^2$ is the variance of the continuously compounded market return from time t to t+1, conditioned on information available at time t. Parameters are defined similarly for time t+1. Other parameters are as defined earlier.

Since our main interest is the estimate for the price of market risk, our baseline model is the following shortened version of the model under the assumption that changes in both the interest rate level and the dividend growth rates are of lesser importance:

$$r_{m,t+1}^e = b_1 + b_2[(\sigma_{t,t+1}^2 - \sigma_{t+1,t+2}^2) \cdot \varphi_{\Delta\sigma} + \sigma^2 \cdot \varphi_\sigma] + u_{m,t+1}, \quad (26)$$

where b_1 is also expected to account for the mean effect from the components excluded from the model and b_2 is again our estimate for the lambda. Note that in our specifications (25) and (26), the constant b_1 cannot be given the same Jensen's alpha interpretation as the constant in standard tests of asset pricing models.

To estimate the model, we need a proxy for the conditional variance. We use three different proxies, the first one being the conditional variance from a GARCH model. We utilize a two-step estimation strategy. First, we estimate the GARCH model including only a constant in the mean equation. In the second step, we estimate equation (25) or (26) 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., $\sigma_{t+1,t+2}^2$. This value corresponds to $\sigma_{m,t+2}^2$ 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 definitions (19) and (20). 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.,

$$\sigma^2_{t,t+1+i} = (\alpha+\beta)^i \sigma^2_{t,t+1} + \omega \frac{1-(\alpha+\beta)^i}{1-\alpha-\beta}, \tag{27}$$

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 lambda after which the specification (26) is estimated. Following Ghysels et al. (2005 and 2013), we write equation (4) for lower frequency (here monthly) excess market return $r_{m,t+1}$ as follows:

$$r^e_{m,t+1} = \alpha + \lambda^T_m h^{MIDAS}_{t+1} + e_{m,t+1},\ e_{m,t+1} \sim Distr\left(0, h^{MIDAS}_{t+1}\right), \tag{28}$$

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 $r_{m,t}$:

$$h^{MIDAS}_{t+1} = 22 \sum_{d=0}^{D-1} w_d\left(\theta^D\right) r^2_{m,daily,t-d}, \tag{29}$$

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 number 22 is a scaling constant that refers to the average number of trading days in a month; it converts daily variance into a monthly one. D is chosen such that the specification captures sufficient lags but is feasible to estimate. In practice, the parameters of the weight function restrict the effective number of lags to less than 200 (Ghysels, 2014). Here we select it to be 30 to attain convergence in all estimations using the same weighting approach.

A number of polynomial weighting structures can be used (for more information, 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:

$$w_i\left(D,\theta_1,\theta_2\right)=\frac{x_i^{\theta_1}\left(1-x_i\right)^{\theta_2-1}}{\sum_{i=1}^{D}x_i^{\theta_1}\left(1-x_i\right)^{\theta_2-1}}, \tag{30}$$

where $x_i = (i-1)/(D-1)$. For a reasonably large D, the sum of the weights is very close to one. Specification (30) 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 (30). However, it gives equal weights to the observations. In specification (30), the estimated values of θ_1 and θ_2 implicitly allocate less weight to older observations than to newer ones.

To estimate equation (25) or (26) 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 (25) or (26) for the second step. To calculate the $\varphi_{\Delta\sigma}$ and φ_σ parameters, we assume that the variance follows an AR(1) process given by

$$h_{m,t+1}=\phi_0+\phi_1 h_{m,t}+\varepsilon_{m,t+1}. \tag{31}$$

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 earlier, we start by estimating the traditional lambda based on realized returns as proxies for the expected returns. Using the notation in equation (5), we define $\sigma^2_{m,t+1}=IV^2_{m,t}$ as the squared implied volatility observed at time t for the period from t to t+1. Now, we can estimate the traditional lambda using the following model:

$$r^e_{m,t+1}=\alpha_m+\lambda^T_m IV^2_{m,t}+\varepsilon_{m,t+1}. \tag{32}$$

A slight variation of the model (32) is used in Dennis, Mayhew, and Stivers (2006), who use the innovation in the implied volatility measure to study its effect on the returns. They find a substantially negative relationship between the innovation in the systematic market volatility and stock returns. They also find that the commonly observed asymmetry in the return-volatility relationship is mostly driven by market-wide effects rather than by security-specific ones. Another variation is in Banerjee, Doran, and Peterson (2007), who develop a model in which the realized returns are related to the level and the innovations in the variance. They find a statistically significant and positive lambda using S&P500 index returns and the VIX from 1987 to 2005. The results are more significant if two-month returns are used, which is in line with the earlier discussion on the role of time aggregation.

To estimate our model (26) using implied volatilities, we define $\sigma^2_{t+1,t+2} = IV^2_{m,t+1}$ as the squared implied volatility observed at time t+1 for the period from t+1 to t+2. Before we can estimate the equation, 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.

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 estimation as well as for the volatility-feedback approach 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 model, we use the sum of daily returns squared 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 the past twelve months. 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. In the second step, we calculate the logarithmic annual change in the dividends and use it as a proxy for the future growth rate of dividends. Thus, we use the past annual dividend growth rate to forecast future dividend behavior.

As stated earlier, we also test the model using implied volatilities. In practice, we utilize a readily available volatility index from an options exchange. Our primary measure is the Volatility Index calculated by the Chicago Board Options Exchange for the US market. The original CBOE Volatility Index (”VXO”) is available on a daily basis from 1986 to present day. However, CBOE updated the original index and created a theoretically more suitable index labeled ”VIX”[7] that 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). In our empirical analysis, we utilize the VIX (squared and divided by twelve). 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 in this study. 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 on average 4.46 per month (15.45% p.a.). It is clearly lower than the market expectation given by the VIX because it gives an average volatility forecast of 20.20% p.a. The average dividend growth rate is 4.49% for the full sample and 4.92% for the subsample. Government bond

[7] The VXO was based on S&P100 stocks and had some shortcomings in its calculation methodology (cf. Carr and Wu, 2006).

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 (0.895) as expected due to overlapping dividend observations used to calculate the growth rate, as do the government bond yield series (0.980).

4. EMPIRICAL RESULTS

4.1. Baseline model with GARCH estimates

We begin our analysis by studying the price of risk (lambda) using the traditional approach, i.e., the generalized autoregressive conditional heteroskedasticity-in-mean approach (GARCH-M). 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 utilize the quasi-maximum likelihood approach (QML) in the estimation.[8] Table 2 presents the results. Estimations are conducted using monthly data from January 1928 to December 2013.

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 0.686, which is positive as expected by the theory; however, it is not significantly different from zero, with a t-value of 0.839. 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 -0.3%.

There could be several reasons for the empirical estimation not confirming a significant relationship between expected returns and variance. One potential explanation could be that the conditional return is not normally distributed. Therefore, we run the model assuming a conditional t distribution instead of the normal distribution. The results are reported in Panel A. However, this does not improve the model. The explanatory power of the model drops slightly, and the estimate for the price of risk is even lower than before: 0.617 with a t-value of 0.715. Residual diagnostics (not reported) show that both models are able to

[8] More detailed information about the estimations in this paper is available upon request.

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:

$$\sigma_{m,t+1}^2 = \omega + \alpha\varepsilon_{m,t}^2 + \gamma\varepsilon_{m,t}^2 I_{m,t} + \beta\sigma_{m,t}^2, \tag{33}$$

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 results do not provide strong support for either of the distributions, although the degrees of freedom for the t-distribution is estimated to be 7.012 with a t-value of 6.314, 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 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 parameter 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. Reported lambda is the sum of the delta and gamma parameters. Their t-values are calculated using Wald-test on the hypothesis that their sum is zero. The results are similar to those found in French et al. (1987). Gammas – measuring the impact of realized variance on realized returns – are found negative and highly significant. The explanatory power of the model is also clearly higher than before than before. However, the evidence goes straight against the Merton model. None of the lambdas are significant which clearly suggest that, despite the significant volatility-feedback effect, taking the effect into account cannot help us to find support for the positive relationship between conditional risk premium and variance.

Finally, we turn to the new estimation model introduced in this paper. We estimate our baseline model utilizing the same GARCH processes as before and again with conditional normality and t-distribution assumed. The estimation is conducted in two stages. First, we estimate the parameters for the GARCH (or GJR-GARCH) process with only a constant in the mean equation. We use the results to calculate the variance persistence parameter ϕ, as indicated by equation (31), and the sigma-multipliers as indicated by equations (19) and (20). 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 $\omega/(1-\alpha-\beta-\gamma/2)$. In the second step, we run a linear regression model according to equation (26).[9] The results are reported in Panel C of Table 2.

Utilizing a two-step estimation strategy to estimate our baseline model (26) or the full model (25) 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 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 that, as a result, the lambda estimates are not systematically distorted. For example, Hedegaard and Hodrick (2014b) use a four-step procedure in a multivariate setup. First, they estimate univariate GARCH(1,1) models for all assets. 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. The authors conduct a simulation study to conclude that the parameters of interest are well-behaved, and that their standard errors are correctly estimated.

In line with the results in Panels A and B, our results show that the variance parameters are significant in almost all cases. However, in contrast to the results shown in Panels A and B, the lambda estimates are significant except for the case of standard GARCH under the assumption of normality. In this case, the estimate is 0.236 (t-value 1.395). Interestingly, for the t distribution, the estimate increases to 0.536 (t-value 1.842). Utilizing the GJR-GARCH approach, the lambda estimate increases even further, first to 0.900 (under normality) and then to 1.828 (under the t distribution). The results are highly significant in both cases. The explanatory power of the model is also considerably higher than it is for the traditional or for the volatility-feedback models.

The results give strong support for the positive relationship between conditional equity premium and variance. One obvious question could be to test whether the variance shock

[9] 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 equations (25) and (26) and for the AR(1) process are calculated with the adjustment.

offers explanative power over the change in the conditional variance as in equation (26). To test this question, we first noted calculated the variance shock as the difference between realized variance and the conditional variance. Then we added this variable into equation (26) and re-estimated the model with GJR-variance process and t-distribution. The results show that the variance shock is not significant. Volatility-feedback parameter value is -0.446 (t-value -0.608) and lambda is almost unchanged (1.765 with t-value of 8.361). We feel confident to conclude that the volatility-feedback effect does not offer a sufficient explanation for the observed returns, and at the same time its application does not provide support for the Merton model.

4.2. Baseline model with MIDAS estimates

Next, we turn to MIDAS estimation. First, we test the asset pricing model using the traditional approach. The estimation is based on equations (28), (29), and (30). 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 (28) as the price of market risk. We test the model using two samples, first from January 1928 and then from January 1990 to December 2013. The MIDAS estimations are also conducted using the maximum likelihood approach.[10] The results are reported in Panels A and B of Table 3, respectively.

The results show that the traditional lambda estimate is negative using both the full and shorter sub-sample. However, the value for the full sample, -0.174, is not statistically significant. Somewhat surprisingly, the post-1990 sample yields a value of -1.189, which differs significantly from zero at the five percent level. However, in contrast to the full sample, in the case of the post-1990 sample, the MIDAS parameters that are related to the high-frequency distribution are not statistically 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 parameter is negative (-1.054 with t-value -3.012) and statistically significant. This time, however, the high-frequency lag structure fits the model better, and both MIDAS parameters are statistically highly significant.

We also want to see whether we are able to replicate the results in Ghysels et al. (2013; contains corrected results for the 2005 article). Using arithmetic returns (in contrast to our continuously compounded returns), a sample from 1928 to 2000, and a similar estimation setup (normalized exponential Almon lag polynomial, lag equal to 260 days), we estimate the lambda to be 0.3829 (t-value 0.8496), which is close to their value of 0.1472 (t-value

[10] In practice, the estimation is based on version 1.1 of the Matlab routines provided by Professor Eric Ghysels on his website.

0.1249). Similar to their result, the value is not significant. The results differ to some degree because our data are not exactly the same. For example, they used the three-month risk-free rate, their daily market return prior to July 1962 is from professor Schwert as opposed to our CRSP-originated market returns, and our estimation setup is somewhat different.

It is evident from the results that the traditional approach is not able to provide a significant lambda estimate. Thus, we continue with the new approach and test our baseline model. However, to test the model, we must obtain conditional variances. To this end, we proceed with a 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:

$$r_{m,t+1}^2 = \alpha + \beta h_{t+1}^{MIDAS} + \varepsilon_{m,t+1},\ \varepsilon_{m,t+1} \sim Distr\left(0, h_{t+1}^{MIDAS}\right), \tag{34}$$

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 parameter 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 (26) using the OLS. For the regression, we estimate first the variance persistence parameter using equation (31) and use it to calculate the sigma multipliers.

The results are reported in Table 3. The Newey-West (1987) adjustment is applied for the t-values. The adjusted R-squared is for the baseline equation (26). Interestingly, the conditional variance does not show the same level of persistence (the AR(1) parameter is estimated to be 0.503 and 0.641 for the full and subsample, respectively) when compared with the estimates for the autoregressive β parameter in the GARCH model (cf. Table 2). This could be due to the MIDAS approach, which arguably could track changes in the variance more quickly due to the use of higher-frequency data.

The results show that for both samples, we again find the lambda estimate to be statistically significant at the one per cent level. This provides further support for the model introduced in this paper.[11] The lambda for the full sample (1.695 with t-value 4.819) is in line with earlier estimates from the GJR-GARCH model, although the explanatory power of the model is lower (10.3 vs 36.7 percent). Conversely, the explanatory power is higher than the model based on GARCH variance estimates alone. Interestingly, the price of risk seems to be higher for the post-1990 sub-period (2.826 with t-value 6.995) than for the whole sample.

[11] We also estimated lambda using the volatility-feedback approach. The results are basically similar to Table 2.

4.3. Baseline model 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. We first estimate the model using the traditional approach. The estimation is conducted using equation (32). Results are reported in Table 4. The results show that the traditional price of market risk estimate is -0.127. In addition, the explanatory power of the model is really low.

Next we test the volatility-feedback model. As the variance is readily available, we can directly estimate (6). The results in Table 4 show that, although the volatility-feedback is highly significant and the explanatory power of the model is significantly higher (adjusted R^2 is 24.3 per cent), the price of market risk is still not found significant and its estimate is negative (-0.638 with t-value -1.105). The results merely shows that there is some type of relationship between ex post returns and ex ante variance.

Finally, we turn to our model. Before we can estimate the model, we estimate the variance persistence parameter using an AR(1) specification. After that, we can estimate the sigma multipliers and proceed with the tests of our baseline model in equation (26). Table 4 shows that the implied variance exhibits a somewhat higher persistence than does the variance implied by the MIDAS but on the other hand it is lower than that implied by the GARCH. The results also provide strong support for our model. The price of market risk (2.636) is statistically highly significant (t-value 7.199). This value is surprisingly close to the value given by the MIDAS estimation for the same period.

The VIX provides us 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 (31) for the conditional variance with equation (26) for the risk premium and using the cross-equation restriction via equations (19) and (20), 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 is almost the same, 2.553, but the t-value is lower (5); however, it remains highly significant.[12]

[12] We also estimated the system with the Generalized Method of Moments (GMM). Orthogonalizing on the constant and lagged values of VIX (equations (26) and (31)) and contemporary values of VIX (equation (31)), and using an iterative process for the weights, we obtained essentially the same results.

4.4. Full model

Finally, we test our full model as given by equation (25). We estimate the model using OLS with conditional variance estimates from the first-pass estimation. For the full sample, we can test the model using the conditional variances based on the GJR-GARCH under the t-distribution and based on MIDAS estimation. For the short 1990-2013 sample, we also utilize the variances based on the VIX index. The results are reported in Table 5.

The results for the full sample (Panel A) show that the lambda is statistically significant and greater than one using both approaches. The explanatory power of the models can also be considered high, in any case higher than for the baseline model. Our second explanatory variable, the change in the dividend growth rate, is statistically clearly significant as the model implies. This result per se is not surprising, because a number of papers have found a relationship between returns and dividends (see, e.g., Campbell and Schiller, 1988), although the model here approaches the question from a slightly different perspective. 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, is also statistically significant (t-values are 2.281 and 2.320 for the two models, respectively), and its coefficient is positive as suggested by our model.

The results for the post-1990 sample show patterns fairly similar to the full sample. The explanatory power of the model is higher than that of the baseline model. Lambdas are significant in all cases, although their values are somewhat lower than those from the baseline model alone (cf. 1.968 vs. 2.826, when MIDAS estimates are used). Interestingly we can again observe that lambda is estimated to be slightly higher for the post-1990 subsample. On the other hand, the change in the risk-free rate is not statistically significant and its sign is unexpectedly negative. This could be driven by the fact that the assumption of parallel shifts in the term structure could be too strong for the sub-period in question. We tested this by having changes in short and long-term risk-free rates separated into two variables. However, there was no material change in the results. As such, a more complex approach may be needed to model changes in the risk-free rates.

4.5. Additional considerations and robustness checks

It is interesting to compare the behavior of different variance estimates because doing so can reveal how they react in different market conditions. Figure 1 shows annualized conditional volatilities (i.e., the square root of the fitted annualized variance) from the GJR-GARCH(1,1)-M and MIDAS estimations using the full sample from 1928 to 2013. The correlation between the series is 0.589. The series show evidence that the MIDAS volatility is able to react much more quickly to market uncertainty, as one would expect due to its use of mixed-frequency data. The MIDAS volatility has a higher standard deviation (8.77% vs. 7.29%), even though the average conditional volatility is lower for the MIDAS series (16.76% vs. 17.23%). When both series are compared to volatility estimates based on the absolute realized returns, both volatility averages are higher than the mean of realized absolute monthly returns (3.941% corresponding to 13.65% per annum) but their standard deviation is lower than that of absolute returns (13.20%). Consequently, the new methods indicate higher volatility, on average, but with lower variation.

[**Place Figure 1 about here**]

In Figure 2, we compare the volatility series estimated using data for the subsample from 1990 to 2013. For comparison, we have also included the values for the VIX. Again, we can see that the series are highly correlated. The highest correlation is between the MIDAS and the VIX (0.849), and the lowest is between MIDAS and the GJR-GARCH series (0.582). Again, the MIDAS series shows higher variability than the GJR-GARCH series (6.20% vs 5.86%), but less than the VIX implied volatility (7.68%). The average volatility implied by the VIX series (20.20%) is also clearly higher than that implied by the other series (14.91% and 14.23% for GJR-GARCH and MIDAS, respectively).

[**Place Figure 2 about here**]

We continue by testing the robustness of the results of the traditional approach, most notably the results in Panel A in Table 2. First, we re-run the tests using simple returns instead of continuously compounded returns. This does not change the results in general. For example, the GARCH(1,1)-M yields a lambda of 0.013 with a t-value of 1.623. The GJR-GARCH(1,1)-M model yields the same results. Second, we modify the distribution of the error. Using the GED-distribution instead of the t-distribution, the results do not change markedly. For example, with the GJR-GARCH-M model, the lambda estimate is 0.080 with a t-value of 0.098. We also tested the model using the skewed t-distribution, and, although the skewness parameter was significant, the lambda estimate was negative (-0.179) and remained insignificant.

Next, we turn to the results for the new approach. Our main focus is on testing the robustness of the results for our baseline model in Panel B of Table 2. As was the case with the traditional approach, using simple returns does not change the results that much. For example, the lambda estimated using the conditional variance from the GJR-GARCH(1,1) model with t-distributed errors becomes 1.676 (t-value 20.994). Similarly, the results in Table 5 are materially unchanged. Using a shorter subsample from 1990 to 2013, the lambda estimate becomes 2.826 (t-value 6.938), which is clearly greater than that estimated using the traditional approach, -0.385 (t-value -0.162).

An obvious question is also whether the results are driven by our sample period. To study this, we use a rolling estimation approach 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. All possible samples are considered leading to more samples for shorter sample periods. At the same time, we analyze whether our results are driven by our specification for the GARCH model. Ultimately, lambdas are estimated using both the traditional as well as the new, reverse testing approach, first with the basic GARCH model with normally distributed errors, and then with the GJR-GARCH model with normally as well as t-distributed errors. For all cases, lambdas and their t-values, among others, are recorded.

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.[13] 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. Obviously, 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.

Second, even with the simplest estimation setup, GARCH-M with normally distributed errors, the traditional approach does not show a single significant lambda estimate using the 70- or 80-year samples; the new approach has nine significant lambdas out of seventy-two. When we allow for the asymmetry in variance, the traditional approach is not doing any better, whereas the new approach finds all of the lambdas to be statistically significant. With 60-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 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 approach is used.

When using the GJR-GARCH model, the traditional approach produces negative lamb-

[13] These include, for example, issues with near singularity or convergence to a local maximum, which typically manifested itself by an estimate that differs radically from the previous as well as the following estimates.

das 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.

We also compare our results with Ghysels et al. (2013). Using MIDAS estimation, they show that the traditional lambda varies considerably depending on the sample period. For example, using subsamples from 1928 to 1963 and from 1964 to 2000, their lambda estimates were -1.0615 and 3.4646, respectively. Neither of the lambdas was significant. Estimating our baseline model with the same subsamples using the MIDAS approach as given in Table 3 gives us lambda estimates of 1.432 (t-value 6.497) and 1.103 (t-value 3.063) for the first subsample and the second subsample. The estimates are quite similar to and in line with the lambda estimate over the whole sample period (1.695), although our rolling estimation does confirm the finding in Ghysels et al. (2013) that lambda estimates vary over time.

A related question is whether our lambda estimate is sensitive to the sampling frequency and the length of the return measurement horizon. To test this, we estimate the traditional and the new models with the GJR-GARCH as shown in Table 2, assuming the t-distribution with one, two, five, and ten day non-overlapping returns over the same sample period. The results from the traditional model show negative estimates for lambda for all but the ten-day returns (e.g., using one-day returns, the lambda estimate is -0.783 with a t-value of -1.118). In all cases, the values are insignificant. The results from the new model, on the other hand, show that lambdas are significant in all cases and that the estimates are aligned. One-day returns produce the lowest value for lambda (0.920 with a t-value of 4.889). Two-day returns yield the highest lambda (2.236 with a t-value of 3.402), followed by five-day returns (1.847 with a t-value of 16.439) and ten-day returns (1.779 with a t-value of 6.679). Overall, our results show that our model is able to yield believable estimates of the lambda even when estimated using short-horizon returns.

Our final robustness check studies the effect of the speed of variance convergence on the results. To test this, we take the results from last row in Table 4 and test different values for the variance persistence ϕ parameter. The results show that if the speed of convergence decreases to 0.7, the lambda estimate increases to 4.081, whereas, conversely, a truly high speed of convergence of 0.95 gives us a lambda estimate of 0.715, ceteris paribus. Yet, the traditional model yields a negative lambda using the same data. This highlights the fact that testing the return-variance relationship should take into account the long-term behavior of the variance.

5. SUMMARY AND CONCLUSIONS

In this paper, we develop a new approach for testing conditional asset pricing models. The 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 (1973, 1980) asset pricing model and combined with the assumption of the mean-reverting conditional variance, it suggests an empirical model that links the realized equity premium to the price of market risk and to changes in the conditional variance and its long-term persistence. In addition, the model implies that the realized returns are also related to the changes in expected dividend growth rate and 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 the positive relationship between conditional variance and equity premium. Price of market risk is found to be close to zero, and at times even negative. On the other hand, the lambda 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 (1973, 1980) model. The results from the new approach 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 give support for the new reverse testing approach.

There are a number of interesting issues that could be studied with the approach introduced in this paper. It allows one to revisit some of the empirical tests on conditional asset pricing models. Namely, the conditional APMs have yet to fulfill their full potential in explaining some of the issues unexplained by the unconditional CAPM. It is also fairly easy to extend the model tested in a number of ways. For example, one could allow for time-varying price of risk or add other priced risk factors into the model. In addition, one could test alternative ways to model the dividend growth rate or risk-free rate and its term structure. These questions are left for future study.

REFERENCES

Ang, A.R and Liu, J. (2004). How to discount cashflows with time-varying expected returns. The Journal of Finance, 59(6), 2745-2783.

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.

Bollerslev, T. and Wooldridge, J. M. (1992) Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. Econometric Review, 11(2), 143-172.

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

Callen, J. and Lyle, M. R. (2014) The term structure of implied cost of equity capital. Rotman School of Management Working Paper No. 1738401. Available at: http://ssrn.

com/abstract=1738401

Campbell, J. Y. and Hentschell, 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

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

Dennis, P., Mayhew, S., and Stivers, C. (2006) Stock returns, implied volatility innovations, and the asymmetric volatility phenomenon. Journal of Financial and Quantitative Analysis, 41(2), 381-406.

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 Tédongap, 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. (2014) 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. (2013) The Risk-Return Relationship and Financial Crisis. Unpublished manuscript, available at: http://rady.ucsd.edu/docs/faculty/R_R_Revisted_Apr_12_2013_EG.PDF

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.

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.

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

Hedegaard, E. and Hodrick, R. J. (2014a) Estimating the Risk-Return Trade-off with Overlapping Data Inference. NBER Working Paper Series 19969.

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

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.

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.

Lanne, M. and Saikkonen, P. (2006) Why is it so difficult to uncover the risk-return tradeoff in stock returns? Economics Letters, 92, 118-125.

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) Ideosyncratic 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.

Munk, C. (2013) Financial Asset Pricing Theory. UK: Oxford University Press.

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

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. (2015) Skewness and the Relation between Risk and Return. Forthcoming in the Management Science.

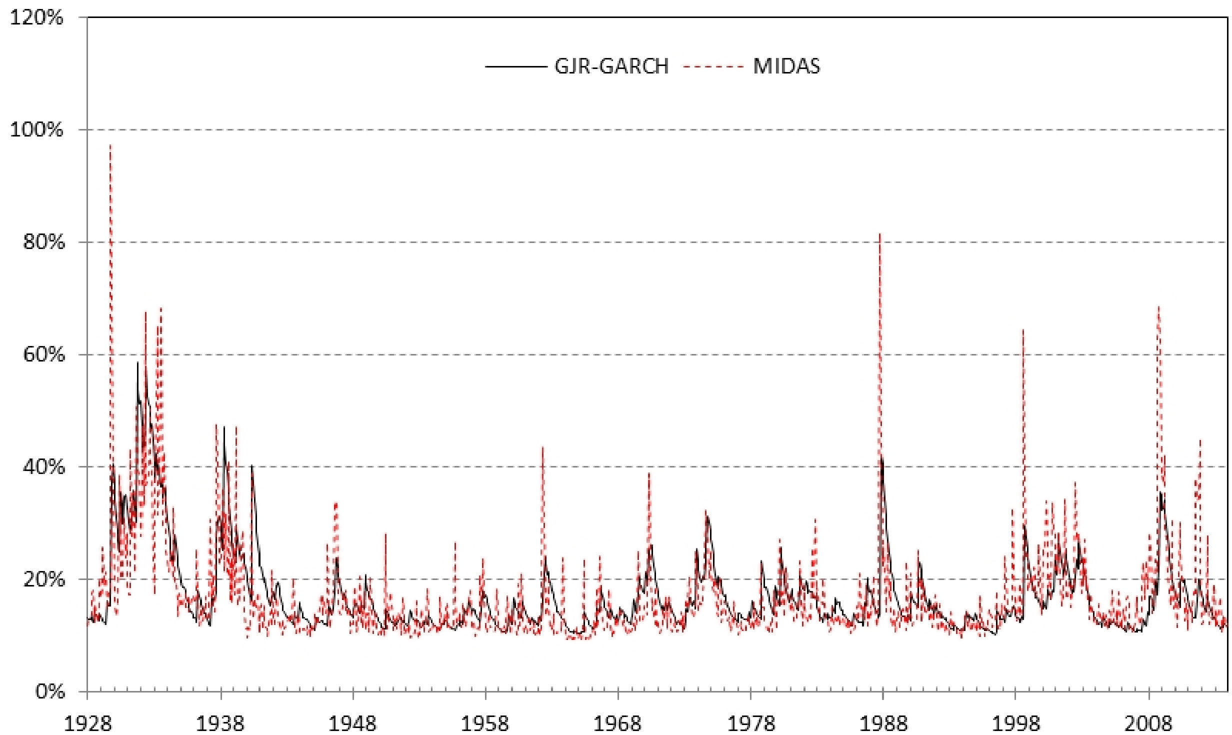

Fig. 1. Conditional volatilities from GJR-GARCH and MIDAS estimations from January 1928 to December 2013. All values are per annum.

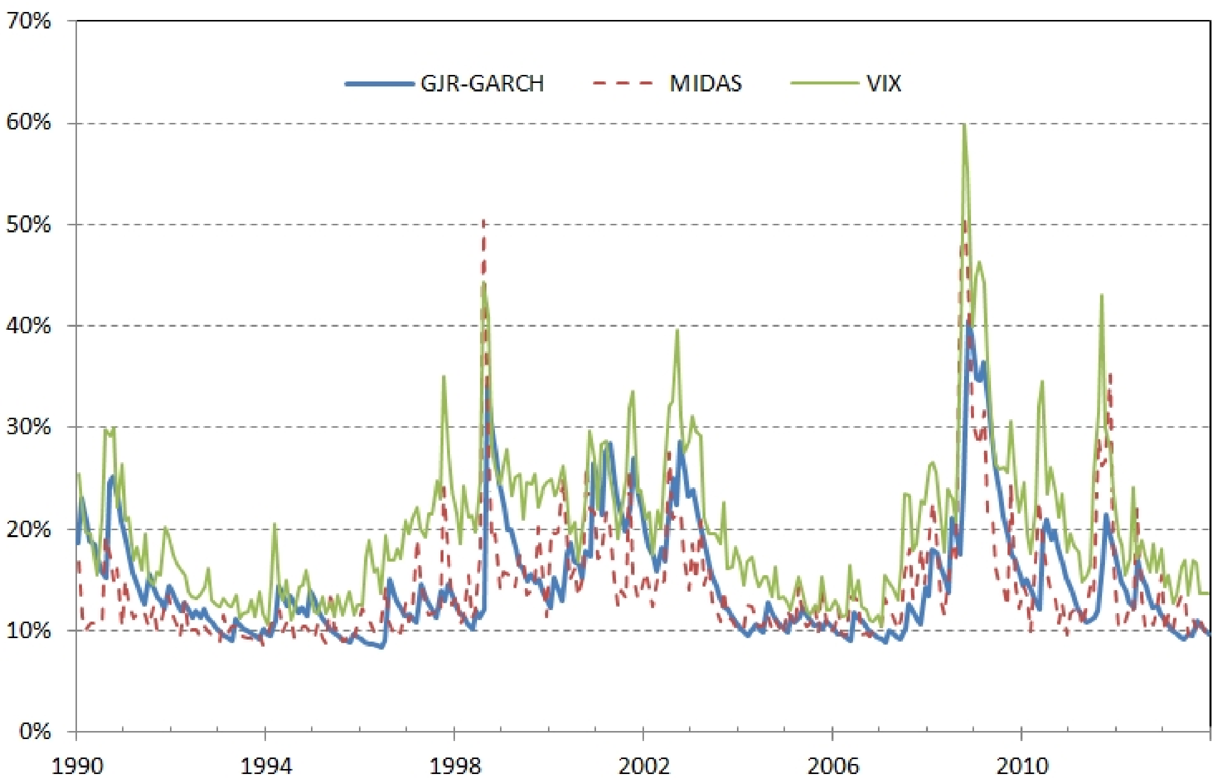

Fig. 2. Conditional volatilities from GJR-GARCH and MIDAS estimations using the 1990-2013 data together with the month-end VIX-index values from January 1990 to December 2013. All values are per annum.

Table 1: Descriptive statistics. Descriptive statistics for monthly and daily market returns as well as for the monthly VIX index. 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. 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. 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.

Variable	Mean	Std. dev.	Skew-ness	Excess Kurtosis	J-B (p-val)	Autocorrelation 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.	4.488	16.33	-0.51	1.32	<0.001	0.895*	0.768*	0.643*	>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.	4.922	14.94	-0.14	0.06	0.614	0.908*	0.787*	0.665*	>999.99*
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: Baseline model 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. Volatility-feedback approach includes realized variance variable in the mean equation. Bollerslev-Wooldridge (1992) robust standard errors are used in the case of the normal distribution. In panel C, price of risk estimates (λ_m) 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 given by the equation (26). Newey-West (1987) standard errors that are robust to heteroskedasticity and autocorrelation up to one lag are used 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. Adjusted R^2 is for the mean equation. t-values are provided below parameter values 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***	0.686		0.000**	0.129***	0.854***		−0.003
+ N-distributed errors	(2.668)	(0.839)		(2.124)	(4.368)	(25.196)		
GARCH(1,1)-M	0.008***	0.617		0.000**	0.127***	0.836***		−0.008
+ t-distributed errors	(3.700)	(0.715)		(2.680)	(4.291)	(25.184)		
GJR-GARCH(1,1)-M	0.006***	0.213		0.000**	0.067*	0.853***	0.093**	−0.002
+ N-distributed errors	(2.772)	(0.249)		(2.392)	(1.654)	(23.025)	(2.000)	
GJR-GARCH(1,1)-M	0.008***	0.099		0.000**	0.035	0.829***	0.143***	−0.005
+ t-distributed errors	(4.020)	(0.119)		(3.241)	(0.866)	(23.096)	(3.101)	
Panel B: Volatility-feedback approach								
GARCH(1,1)-M	0.006***	0.448	−5.609***	0.000***	0.168***	0.795***		0.115
+ N-distributed errors	(3.545)	(0.545)	(−29.352)	(3.231)	(6.248)	(27.467)		
GARCH(1,1)-M	0.006***	0.483	−5.434***	0.000***	0.127***	0.836***		0.120
+ t-distributed errors	(3.438)	(0.502)	(−24.393)	(2.680)	(4.291)	(25.184)		
GJR-GARCH(1,1)-M	0.006***	−0.177	−5.463***	0.000**	0.096***	0.799***	0.093***	0.120
+ N-distributed errors	(4.080)	(−0.256)	(−27.192)	(3.620)	(3.368)	(28.758)	(2.978)	
GJR-GARCH(1,1)-M	0.006***	−0.024	−5.193***	0.000**	0.085***	0.799***	0.125***	0.123
+ t-distributed errors	(3.958)	(−0.029)	(−21.704)	(3.381)	(2.617)	(25.313)	(2.936)	

Table 2 continued.

	Mean equation			Variance equation				
	Constant	λ_m	γ_m	ω	α_1	β_1	$\gamma(\varepsilon_t < 0)$	**Adj.** R^2
Panel C: New approach								
GARCH(1,1)-M + OLS	0.004*	0.236		0.000**	0.129***	0.854***		0.058
+ N-distributed errors	(1.809)	(1.395)		(2.911)	(6.105)	(44.557)		
GARCH(1,1)-M + OLS	0.003	0.526*		0.000**	0.127***	0.837***		0.071
+ t-distributed errors	(1.424)	(1.842)		(2.886)	(4.613)	(27.984)		
GJR-GARCH(1,1)-M + OLS	0.002	0.900***		0.000**	0.066**	0.852***	0.095**	0.250
+ N-distributed errors	(1.405)	(8.912)		(3.058)	(2.383)	(40.446)	(2.445)	
GJR-GARCH(1,1)-M + OLS	0.001	1.828***		0.000**	0.033	0.830***	0.143***	0.367
+ t-distributed errors	(0.606)	(8.816)		(3.001)	(2.086)	(22.297)	(2.899)	

Table 3: Baseline model with MIDAS variance. Maximum likelihood estimates for the price of market risk (λ_m^T) under the traditional apprach as well as weighting structure are reported for the MIDAS estimation. The estimation is conducted using monthly returns and daily returns squared where monthly returns are in excess of the risk-free rate. Daily returns have been multiplied by 22 to state them in per month form. Continuously compounded returns have been used in all estimations. Samples are from January 1928 and from January 1990 to December 2013. Reported MIDAS parameters are for the normalized beta density with a zero last lag. The new approach have been estimated in two steps. In the first step, a MIDAS estimation has been conducted, but the squared monthly returns have been used as the dependent variables. In the second step, fitted variance estimates have been used to estimate the price of market risk (λ^m) using the OLS as given by equation (26). The Newey-West (1987) adjustment has been used to calculate standard errors. Parameters required for the sigma-multiplier are estimated running an AR(1)-model for the conditional variance (reported alongside MIDAS parameters). Adjusted R-square 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.

		$\sigma_\mathbf{t}^\mathbf{2}$	$\Delta\sigma_\mathbf{t}^\mathbf{2}$	**MIDAS-parameters**			**AR(1) parameters**		
	Constant	λ_m^T	λ_m	β	θ_1	θ_2	ϕ_0	ϕ_1	**adj.** R^2
Panel A: 1928-2013									
Traditional approach	0.005***	-0.174			43.404***	300.000***			0.002
	(2.992)	(-0.990)			(4.680)	(4.782)			
New approach	0.000		1.695***	0.964***	0.950***	2.108**	0.001***	0.503***	0.103
	(-0.152)		(4.819)	(21.784)	(170.112)	(9.210)	(7.920)	(8.583)	
Panel B: 1990-2013									
Traditional approach	0.008***	-1.189**			4.621	10.306			0.024
	(2.915)	(-2.421)			(1.438)	(1.060)			
New approach	0.000		2.826***	0.544***	0.943***	1.862***	0.001***	0.641***	0.154
	(0.891)		(6.995)	(16.625)	(138.435)	(5.761)	(3.107)	(4.960)	

Table 4: Baseline model 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 of regressing excess market returns on the conditional variance for the same period (available at the beginning of the period). Then the same regression is augmented by the shock in the variance variable, measured as the difference between realized variance and conditional variance. Finally, the price of market risk is estimated using the new approach where excess returns are regressed on the first difference in the conditional variance as given by equation (26). Parameters required for the sigma-multiplier are estimated running an AR(1) model for the conditional variance. 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. 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.

				AR(1) parameters		
	Constant	λ_m	γ_m	ϕ_0	ϕ_1	**Adj.** R^2
Traditional approach	0.006	−0.127				−0.003
	(1.312)	(−0.092)				
Volatility feedback approach	0.001	−0.638	−6.252***			0.243
	(0.280)	(−1.104)	(−5.627)			
New approach	−0.005**	2.636***		0.001***	0.807***	0.448
	(−2.265)	(7.199)		(3.385)	(10.905)	

Table 5: Results for the full model under the new approach. OLS estimates for the equation (25) are given using different specifications for the conditional variance. Parameter b_2 corresponds to the price of risk parameter (λ_m). In panel A, variance estimates from GJR-GARCH(1,1) and MIDAS are employed using a sample from January 1928 to December 2013 (1,032 observations). In panel B, variance estimates from GJR-GARCH(1,1) and MIDAS as well as the VIX index are employed using a sample from January 1990 to December 2013 (288 observations). Δg is the first difference in the dividend growth rate per annum as given by the continuously compounded growth of the dividends paid during the past twelve months compared to dividends paid a year ago. Δrf is the first difference in the long-term US government bond yield. The Newey-West (1987) adjustment with one lag has been used to calculate the 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.

	Constant b_1	$\boldsymbol{\Delta\sigma_t^2}$ b_2	$\boldsymbol{\Delta g}$ b_3	$\boldsymbol{\Delta rf}$ b_4	**Adj.** R^2
Panel A: 1928-2013					
GJR-GARCH(1,1) + OLS	0.002	1.247***	0.347***	1.187**	0.556
+ t-distributed errors	(1.534)	(8.037)	(9.824)	(2.281)	
MIDAS	0.001	1.150***	0.435***	1.349**	0.455
	(0.738)	(4.396)	(12.573)	(2.320)	
Panel B: 1990-2013					
GJR-GARCH(1,1) + OLS	0.002	2.037***	0.246***	−1.158	0.591
+ t-distributed errors	(1.181)	(6.977)	(6.459)	(−0.976)	
MIDAS	0.001	1.968***	0.387***	−1.138	0.441
	(0.670)	(5.259)	(8.790)	(−1.528)	
VIX	−0.002	0.524***	0.268***	−0.856	0.560
	(−1.011)	(5.420)	(6.232)	(−0.647)	