On the relationship between the conditional mean and volatility of stock returns: A latent VAR approach ☆

Michael W. Brandt[a],*, Qiang Kang[b]

[a] *The Fuqua School of Business, Duke University and NBER, Durham, NC 27708, USA*
[b] *School of Economics and Finance, University of Hong Kong, Hong Kong*

Abstract

We model the conditional mean and volatility of stock returns as a latent VAR process to study their contemporaneous and intertemporal relationships in a flexible statistical framework and without relying on exogenous predictors. We find a strong and robust negative correlation between the innovations to the conditional moments leading to pronounced countercyclical variation in the Sharpe ratio. We document significant lead-lag correlations between the moments that also appear related to business cycles. Finally, we show that although the conditional correlation between the mean and volatility is negative, the unconditional correlation is positive due to these lead-lag correlations.
© 2003 Elsevier B.V. All rights reserved.

*Corresponding author. Tel.: +1-919-660-1948; fax: +1-919-660-8038.
E-mail address: mbrandt@duke.edu (M.W. Brandt).

1. Introduction

The relationship between the mean and volatility of returns is a central issue in finance. Many asset pricing models imply a proportional relationship between expected excess returns and risk in the cross-section and over time. Unfortunately, the empirical relationship between the mean and volatility of returns, and in particular stock returns, is much less regular. It has proven difficult to explain all cross-sectional differences in expected stock returns through differential risks (e.g., Fama and French, 1996). Furthermore, the evidence on the time-series relationship between the conditional mean and volatility of stock returns is inconclusive and depends on the model and exogenous predictors used to draw inferences (e.g., Harvey, 2001). In this paper, we examine more closely the time-series relationship between the conditional mean and volatility using a latent vector autoregressive (VAR) process for the two moments. This latent VAR approach allows us to study the contemporaneous and intertemporal relationships between expected returns and risk in a flexible statistical framework and, more importantly, without relying on exogenous predictors.

It is clear from both the volatility ratio tests of LeRoy and Porter (1981) and the long-horizon autoregressions of Fama and French (1988) that expected returns are time-varying. Predictive regressions have been used widely to capture the time variation of expected returns with an expanding set of predictors, such as the dividend yield, short rate, term premium, and default premium. However, there is increasing concern over predictive regressions in light of statistical issues (Boudoukh and Richardson, 1993; Stambaugh, 1999; Ferson et al., 2000; Valkanov, 2002) and their lack of robustness and out-of-sample predictive power (Bossaerts and Hillion, 1999; Ang and Bekaert, 2001; Goyal and Welch, 2002).

Time-varying volatility is even more firmly established and is typically modelled in an ARCH or stochastic volatility (SV) framework; see Bollerslev et al. (1992) and Ghysels et al. (1996) for literature surveys. In ARCH models, the conditional variance of returns depends deterministically both on lagged squared returns and lagged variances, while in SV models the conditional variance is a stochastic process. The fact that in an ARCH model all randomness is ex post observable (through returns) implies that the volatility realizations can be recovered from the data. As a result, ARCH models are easy to estimate and quite popular. In SV models, in contrast, the innovations to the volatility process are random and the volatility realizations are therefore unobservable. Estimating an SV model involves integrating out the latent volatilities. The computation of this high-dimensional integral is analytically difficult and computationally intensive, causing SV models to be less popular than ARCH models despite their statistical and economic appeal.[1]

---

[1] Not until recently have effective methods for estimating SV models been developed. These econometric methods include quasi-maximum likelihood (QML) (Harvey et al., 1994; Alizadeh et al., 2002), simulated maximum likelihood (SML) (Danielsson, 1994; Sandmann and Koopman, 1998), Markov chain Monte Carlo (Jacquier et al., 1994; Kim et al., 1998), and the simulated method of moments (SMM) (Duffie and Singleton, 1993; Andersen and Sorensen, 1994).

Given the empirical evidence on both time-varying expected returns and volatility, it is only natural to examine the comovements of the two moments through time. Unfortunately, the findings on this relationship are mixed. Using volatility-in-mean models (essentially ARCH models in which the variance enters contemporaneously in the mean return), French et al. (1987) and Campbell and Hentschel (1992) estimate a positive correlation, while Campbell (1987), Breen et al. (1989), Nelson (1991), and Glosten et al. (1993) all report a negative correlation. Koopman and Uspensky (1999) find evidence of a weak negative relationship with an SV-in-mean model but a weak positive relationship with an ARCH-based volatility-in-mean model. Harrison and Zhang (1999) employ Gallant and Tauchen's (1989) semiparametric method to document a positive relationship between the conditional mean and volatility at long horizons but none at short horizons (such as a month). Finally, Harvey (2001) uses exogenous predictors and finds that the correlation between the moments generally depends on the model and the conditioning information used in the mean and variance estimation.

We examine more closely the time-series relationship between the conditional moments through a latent VAR model. The first equation of the VAR describes the dynamics of the conditional mean. It captures the temporary component in the permanent and temporary components model of Fama and French (1988) and Lamoureux and Zhou (1996), in which stock prices are governed by a random walk and a stationary process, respectively. The second equation of the VAR describes the dynamics of the conditional volatility. It nests the standard SV model studied by Wiggins (1987), Andersen and Sorensen (1994), Jacquier et al. (1994), and Kim et al. (1998), among others.

The latent VAR allows us to study the contemporaneous and intertemporal relationships between the conditional mean and volatility in a flexible statistical framework and without relying on exogenous predictors. With respect to the contemporaneous relationship, the mean equation of the VAR resembles the volatility-in-mean model used by French et al. (1987) and others to measure the contemporaneous effect of volatility on expected returns. In addition to this volatility-in-mean effect, there is a symmetric mean-in-volatility effect in the volatility equation. If a change in the volatility has a contemporaneous effect on the mean, a change in the mean also has a contemporaneous effect on the volatility. With regard to the intertemporal relationship, the VAR allows for lag adjustments of the volatility to changes in the mean, and similarly of the mean to changes in the volatility, which Whitelaw (1994) claims to be salient features of the data. Analogous to the contemporaneous effects, we refer to these lag adjustments as lag-volatility-in-mean and lag-mean-in-volatility effects.

Whitelaw (1994) considers a VAR model similar to ours, except he uses *fitted* moments from a set of first-stage predictive regressions as proxies for the unobservable conditional mean and volatility. The obvious drawback of using fitted moments is that their dynamics are determined by the joint conditional distribution of the first-stage predictors. Thus, with only slight model misspecification, such as omitted variables, the dynamics of the fitted moments do not need to correspond to those of the true moments. Another problem with using fitted moments is that even if

the predictive models for the conditional mean and volatility are well specified, the effect of errors in variables which are induced by the first-stage regressions is not trivial to quantify in a VAR. Our approach is not subject to these critiques because we treat the conditional mean and volatility as latent state variables.

We use simulated maximum likelihood (SML) based on importance sampling to estimate the parameters of the VAR model and to extract estimates of the latent mean, volatility, and implied Sharpe ratio time series. We then examine the contemporaneous and intertemporal relationship between the conditional mean and volatility through the parameter estimates, extracted series, and impulse response functions of the estimated VAR.

Among our empirical results, the following are particularly interesting:[2]

- The contemporaneous correlation between the innovations to the conditional mean and volatility is negative and statistically significant (according to asymptotic standard errors and corroborated by Monte Carlo experiments), implying negative volatility-in-mean and mean-in-volatility effects.
- The negative contemporaneous correlation between the mean and volatility generates substantial variation in the conditional Sharpe ratio that is distinctly countercyclical. *Every* business cycle since 1946 is associated with an almost monotonic rise in the Sharpe ratio from the peak to the trough of the cycle.
- Both lag adjustments, the lag-volatility-in-mean and the lag-mean-in-volatility effects, are positive and statistically significant, according to the asymptotic and finite sample standard errors. Furthermore, the timing of the lag adjustment appears related to business cycles as well. Whenever the economy comes off the peak of a cycle, the conditional volatility rises immediately. The conditional mean, however, increases only gradually as the economy moves from the peak to the trough of the cycle. Therefore, the volatility appears to lead the mean through the recession by about six months. The mean reaches a high at the trough of the cycle, shortly after which the volatility drops again to a normal level. As a result, the increase in the mean is associated with a subsequent drop in the volatility.
- Despite the large and negative contemporaneous correlation between the innovations to the conditional mean and volatility, which measures the *conditional* (on the lagged mean and lagged volatility) correlation between the moments, the *unconditional* contemporaneous correlation between the mean and volatility is large and positive due to the strong lag-volatility-in-mean effect. Unconditionally, times of high expected returns are associated with high volatility. We argue that this difference between the conditional and unconditional correlations may explain the disagreement in the literature about the contemporaneous correlation.

---

[2] We draw inferences with asymptotic standard errors. However, since the asymptotics of our estimator are questionable (see Section 3.4), we corroborate the evidence on the key results with Monte Carlo experiments.

The remainder of this paper is structured as follows. Section 2 sets up the latent VAR for the conditional mean and volatility, derives the implied dynamics of the Sharpe ratio, explains the link to equilibrium asset pricing models, and outlines the hypotheses we want to test within this framework. We describe our econometric approach in Section 3. Section 4 then presents our empirical results in six subsections: Section 4.1 describes the data, Section 4.2 reports the parameter estimates and the results of the hypothesis tests, Section 4.3 analyzes the extracted latent processes, Section 4.4 interprets the impulse response functions, Section 4.5 tries to reconcile the disagreement in the literature about the sign of the contemporaneous correlation between the conditional mean and volatility, and Section 4.6 incorporates exogenous predictors as a robustness check. We conclude in Section 5 with a summary of our main results.

2. Theory

2.1. Return dynamics

Let y_t be continuously compounded excess returns with time-series dynamics

$$y_t = \mu_{t-1} + \sigma_{t-1}\varepsilon_t \quad \text{with } \varepsilon_t \sim \text{N}[0, 1], \tag{1}$$

where μ_t and σ_t denote the conditional mean and volatility of excess returns, respectively. Both the conditional mean and volatility are unobservable quantities, which we assume to evolve jointly as a first-order VAR process in logs:

$$\begin{bmatrix} \ln \mu_t \\ \ln \sigma_t \end{bmatrix} = d + A \begin{bmatrix} \ln \mu_{t-1} \\ \ln \sigma_{t-1} \end{bmatrix} + \eta_t \quad \text{with } \eta_t \sim \text{MVN}[0, \Sigma], \tag{2}$$

where d is a 2×1 coefficient vector that relates to the long-term means of the two latent state variables, A is a 2×2 coefficient matrix, and Σ is a 2×2 covariance matrix (symmetric and nonnegative definite). Specifically, we write

$$d = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}, \quad A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad \text{and} \quad \Sigma = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \quad \text{with}$$

$$b_{12} = b_{21} = \rho\sqrt{b_{11}b_{22}}. \tag{3}$$

As long as the VAR is stationary, closed-form solutions to the unconditional mean and covariance matrix of the conditional moments are:

$$\text{E}\begin{bmatrix} \ln \mu_t \\ \ln \sigma_t \end{bmatrix} = (I - A)^{-1}d \quad \text{and}$$

$$\text{vec}\left(\text{Cov}\begin{bmatrix} \ln \mu_t \\ \ln \sigma_t \end{bmatrix}\right) = (I - (A \otimes A))^{-1}\text{vec}(\Sigma), \tag{4}$$

where the $\otimes$ symbol denotes a Kronecker product.

The key parameters of the latent VAR are the transition matrix A and the contemporaneous correlation coefficient ρ. The diagonal elements of A reflect the persistence of the conditional moments, and the off-diagonal elements capture the intertemporal feedback of the conditional volatility to the conditional mean and vice versa. The coefficient ρ represents the contemporaneous correlation between the *innovations* to the conditional moments.

We entertain three assumptions about the correlation between the return innovations ε_t and the conditional moment innovations η_t. The first assumption is that the return innovations are uncorrelated with the conditional moment innovations, or, $\mathrm{Corr}[\varepsilon_t \eta_s] = [0,0]'$ for all s and t. Second, we assume that the return innovations are contemporaneously correlated with *either* the conditional mean *or* the conditional volatility innovations, but not both. Formally, either $\mathrm{Corr}[\varepsilon_t \eta_t] = [\rho_\mu, 0]'$ or $\mathrm{Corr}[\varepsilon_t \eta_t] = [0, \rho_\sigma]'$. The third assumption is that the return innovations are contemporaneously correlated with both conditional moment innovations, so that $\mathrm{Corr}[\varepsilon_t \eta_t] = [\rho_\mu, \rho_\sigma]'$.

2.2. Interpretation

We model the conditional moments as exponential Gaussian processes. For the mean, this modelling choice guarantees a positive risk premium, which is a sensible assumption for the market portfolio and has been used in similar contexts by Bekaert and Harvey (1995) and De Santis and Gerard (1997,1998). For the volatility, the implied log-normality is consistent with the recent empirical findings of Andersen et al. (2001).

Our latent VAR model fuses and generalizes the permanent and temporary components model of Fama and French (1988) and the standard SV model. The first equation of the VAR describes the dynamics of the conditional mean as

$$\ln \mu_t = d_1 + a_{11} \ln \mu_{t-1} + a_{12} \ln \sigma_{t-1} + \eta_{1t}, \quad \text{with } \eta_{1t} \sim \mathrm{N}[0, b_{11}]. \tag{5}$$

This equation captures the temporary component of Fama and French's permanent and temporary components model, in which stock prices are governed by a random walk and a stationary autoregressive process, respectively. In fact, with $a_{12} = 0$, this model of the temporary component is the same as that of Lamoureux and Zhou (1996) (except in logs). The second equation of the VAR describes the volatility dynamics as

$$\ln \sigma_t = d_2 + a_{21} \ln \mu_{t-1} + a_{22} \ln \sigma_{t-1} + \eta_{2t}, \quad \text{with } \eta_{2t} \sim \mathrm{N}[0, b_{22}] \tag{6}$$

and nests the standard SV model. With $a_{21} = 0$, Eq. (6) is the SV model estimated by Wiggins (1987), Andersen and Sorensen (1994), Jacquier et al. (1994), Kim et al. (1998), and many others.

The mean equation also resembles the volatility-in-mean model used by French et al. (1987) and Glosten et al. (1993) to explore the contemporaneous relationship between the conditional mean and volatility. To illustrate the connection to the

volatility-in-mean model, we project η_{1t} onto η_{2t}, such that

$$\eta_{1t} = \beta_1 \eta_{2t} + \xi_{1t} \quad \text{with } \xi_{1t} \sim N[0, (1-\rho^2)b_{11}], \tag{7}$$

where ξ_{1t} is uncorrelated with η_{2t} by construction and

$$\beta_1 = \frac{\text{Cov}[\eta_{1t}, \eta_{2t}]}{\text{Var}[\eta_{2t}]} = \rho\sqrt{\frac{b_{11}}{b_{22}}}. \tag{8}$$

We then rewrite Eq. (6) as

$$\eta_{2t} = \ln \sigma_t - d_2 - a_{21} \ln \mu_{t-1} - a_{22} \ln \sigma_{t-1} \tag{9}$$

and substitute it and Eq. (7) into Eq. (5) to get

$$\begin{aligned}
\ln \mu_t = {} & (d_1 - \beta_1 d_2) + (a_{11} - \beta_1 a_{21}) \ln \mu_{t-1} \\
& + (a_{12} - \beta_1 a_{22}) \ln \sigma_{t-1} + \beta_1 \ln \sigma_t + \xi_{1t}.
\end{aligned} \tag{10}$$

This reduced-form equation for the conditional mean is a generalized version of the usual volatility-in-mean model, in which the conditional mean depends contemporaneously on the conditional volatility.[3] The volatility-in-mean effect depends on the projection coefficient β_1 and therefore indirectly on the correlation coefficient ρ.

Our model extends the volatility-in-mean model in two ways. First, unless $|\rho| = 1$, there are two sources of randomness in the conditional mean process, as opposed to only one, which breaks the overly restrictive one-to-one link between the conditional mean and volatility imposed by the usual volatility-in-mean model. Second, not only the current but also the lagged volatility affects the conditional mean, whereas the usual volatility-in-mean model involves only a contemporaneous effect. Analogous to the volatility-in-mean effect, we label the effect of the lagged volatility the *lag-volatility-in-mean effect*. The lag-volatility-in-mean effect depends on $a_{12} - \beta_1 a_{22}$. If this term is positive (negative), the lagged conditional volatility has a positive (negative) effect on the conditional mean.

Likewise, we can obtain the reduced-form volatility dynamics:

$$\begin{aligned}
\ln \sigma_t = {} & (d_2 - \beta_2 d_1) + (a_{22} - \beta_2 a_{12}) \ln \sigma_{t-1} \\
& + (a_{21} - \beta_2 a_{11}) \ln \mu_{t-1} + \beta_2 \ln \mu_t + \xi_{2t},
\end{aligned} \tag{11}$$

where

$$\beta_2 = \frac{\text{Cov}[\eta_{1t}, \eta_{2t}]}{\text{Var}[\eta_{1t}]} = \rho\sqrt{\frac{b_{22}}{b_{11}}} \tag{12}$$

and

$$\eta_{2t} = \beta_2 \eta_{1t} + \xi_{2t} \quad \text{with } \xi_{2t} \sim N[0, (1-\rho^2)b_{22}]. \tag{13}$$

---

[3] The standard volatility-in-mean model relates the mean return to the volatility (or variance), while our model relates the *log* mean to the *log* volatility. Let $q = d\mu/d\sigma$ be the usual volatility-in-mean coefficient. Then, our coefficient on the log-volatility-in-log-mean $d \ln \mu / d \ln \sigma = (d\mu/d\sigma)(\sigma/\mu) = q\sigma/\mu$ has the same sign as the usual volatility-in-mean coefficient, since both μ and σ are positive.

This reduced-form volatility model captures the persistence of volatility through the term $a_{22} - \beta_2 a_{12}$. In addition, it incorporates the effects of the mean on the volatility, both at the contemporaneous and intertemporal levels, through the coefficients β_2 and $a_{21} - \beta_2 a_{11}$, respectively. Analogous to the volatility-in-mean and lag-volatility-in-mean effects, we label the coefficient on the contemporaneous mean the *mean-in-volatility effect* and the coefficient on the lagged mean the *lag-mean-in-volatility effect*. The correlation ρ between the mean and volatility innovations is again vital in determining the sign and magnitude of the mean-in-volatility and lag-mean-in-volatility effects (through β_2).

The mean-in-volatility effect mirrors the volatility-in-mean effect. Since $\beta_1\beta_2 = \rho^2$, the two effects share the same sign. However, unless $|\rho| = 1$, the mean-in-volatility and volatility-in-mean effects are not perfect reciprocals of each other. The ratio β_1/β_2 measures the variance of the mean innovations relative to that of the volatility innovations.

The motivation for including the non-contemporaneous lag-volatility-in-mean and lag-mean-in-volatility effects, $a_{12} - \beta_1 a_{22}$ and $a_{21} - \beta_2 a_{11}$, respectively, is mostly empirical. Whitelaw (1994) documents significant lead-lag interactions between the conditional mean and volatility. Although it is possible that his results are an artifact of the exogenous predictors used to measure the time variation of the moments, it is equally possible that the lead-lag interactions are a systematic but until-now overlooked feature of the relationship between expected returns and risk. In fact, Lettau and Ludvigson (2001) document similar lead-lag interactions using different exogenous predictors, which is evidence in favor of the second possibility.

Finally, the three assumptions about the correlation between the return innovations and the conditional moment innovations serve to capture and potentially distinguish two popular explanations of asymmetric volatility. Asymmetric volatility refers to the empirical finding that increases in volatility tend to be associated more often with large negative returns than with equally large positive returns. The two popular explanations of asymmetric volatility are the *leverage effect* and the *volatility feedback effect*. The leverage effect states that when the value of a firm drops given a large negative return, the leverage of the firm and the associated probability of bankruptcy increase, causing the equity claims to become riskier. The volatility feedback effect attributes the asymmetric volatility to the equilibrium response of the conditional mean to changes in volatility. Suppose a large negative return is associated with an increase in the conditional volatility (in the spirit of an ARCH model) and the price of risk is positive, so that the increase in the conditional volatility leads to an increase in the conditional mean. Holding future dividends fixed, the expected return can only increase by a drop in the current stock price, resulting in an even larger negative return. In contrast, for a large positive return that also leads to an increase in both conditional moments, the drop in the stock price associated with the increase in the expected return partially offsets the positive return. As a result, the increase in volatility is associated with a large negative return (enhanced by the increase in the conditional mean) or a small positive return (partially offset by the increase in the conditional mean).

The assumption that the return innovations are independent of the conditional moment innovations, $\rho_\sigma = \rho_\mu = 0$, rules out both the leverage and volatility feedback effects. A negative correlation between the return and conditional volatility innovations, $\rho_\sigma < 0$, is consistent with the leverage effect. It can also be consistent with the volatility feedback effect, but only if the contemporaneous correlation between the conditional moments is positive, $\rho > 0$. If the contemporaneous correlation between the conditional moments is negative, the volatility feedback effect implies a positive correlation between the return and conditional volatility innovations, $\rho_\sigma > 0$, because negative returns are offset and positive returns are enhanced by the decrease in the conditional mean associated with the increase in the conditional volatility. Furthermore, since in our model the conditional mean can vary independently of the conditional volatility (unless $|\rho| = 1$), a more direct way to capture the intuition of the volatility feedback effect is through a negative correlation between the return and conditional mean innovations, $\rho_\mu < 0$, which directly ties increases (decreases) in the conditional mean to negative (positive) returns. Finally, the case in which $\rho_\mu < 0$ and $\rho_\sigma < 0$ allows for both effects, regardless of the sign and magnitude of the contemporaneous correlation between the conditional mean and volatility.[4]

2.3. Implied sharpe ratio dynamics

The latent VAR implies that the conditional Sharpe ratio is log-normally distributed and that it follows a two-factor autoregressive (AR) process. We define the log Sharpe ratio as

$$\ln S_t = \ln \mu_t - \ln \sigma_t = T' \begin{bmatrix} \ln \mu_t \\ \ln \sigma_t \end{bmatrix}, \quad \text{where } T = \begin{bmatrix} 1 \\ -1 \end{bmatrix}. \tag{14}$$

Eq. (2) then implies that

$$\begin{aligned} \ln S_t &= T'd + T'A \begin{bmatrix} \ln \mu_{t-1} \\ \ln \sigma_{t-1} \end{bmatrix} + T'\eta_t \\ &= (d_1 - d_2) + (a_{11} - a_{21}) \ln \mu_{t-1} + (a_{12} - a_{22}) \ln \sigma_{t-1} + (\eta_{1t} - \eta_{2t}), \end{aligned} \tag{15}$$

which shows that the variation in the log Sharpe ratio is driven by the mean reversion of the two conditional moments and by the *difference* between the conditional moment innovations, $\eta_{1t} - \eta_{2t}$. Given the joint normality of η_t, the difference between the innovations is normally distributed and so is the log Sharpe ratio.

It follows that the Sharpe ratio is stochastic unless $\eta_{1t} - \eta_{2t}$ is zero. In particular, it is stochastic even if the innovations to the mean and volatility are perfectly correlated, with $|\rho| = 1$. The reason is that even if the innovations are perfectly correlated, their magnitudes are different, unless $b_{11} = b_{22}$, and therefore the

---

[4] We can disentangle ρ_μ and ρ_σ from ρ by factoring the innovations η_t into a common component due to the correlations with the return innovations ε_t and the components that are orthogonal to the return innovations but are correlated with each other by $\tilde{\rho} = (\rho - \rho_\mu \rho_\sigma)/\sqrt{(1 - \rho_\mu^2)(1 - \rho_\sigma^2)}$.

difference between the innovations is nonzero. For example, suppose $\rho = 1$. If $b_{11} > b_{22}$, the Sharpe ratio increases whenever the mean increases, because the volatility increases by less than the mean. Alternatively, if $b_{11} < b_{22}$, the Sharpe ratio decreases whenever the mean increases, because the volatility increases by more than the mean. The Sharpe ratio is only nonstochastic if $|\rho| = 1$ *and* $b_{11} = b_{22}$. Of course, even in this case the Sharpe ratio is still time-varying due to the mean reversion of the two conditional moments (because $\ln \mu_{t-1}$ and $\ln \sigma_{t-1}$ are time-varying).

Substituting $\ln \mu_{t-1} = \ln S_{t-1} + \ln \sigma_{t-1}$ or $\ln \sigma_{t-1} = \ln \mu_{t-1} - \ln S_{t-1}$ into Eq. (15) illustrates that the log Sharpe ratio follows an AR process:

$$\begin{aligned} \ln S_t = {}& (d_1 - d_2) + (a_{11} - a_{21}) \ln S_{t-1} \\ & + (a_{11} + a_{12} - a_{21} - a_{22}) \ln \sigma_{t-1} - (1 - \beta_1)\eta_{2t} + \xi_{1t} \end{aligned} \tag{16}$$

or

$$\begin{aligned} \ln S_t = {}& (d_1 - d_2) + (a_{22} - a_{12}) \ln S_{t-1} \\ & + (a_{11} + a_{12} - a_{21} - a_{22}) \ln \mu_{t-1} + (1 - \beta_2)\eta_{1t} - \xi_{2t}, \end{aligned} \tag{17}$$

respectively. In fact, if $a_{11} + a_{12} - a_{21} - a_{22} = 0$, the log Sharpe ratio follows a univariate AR process, which is fairly standard in the term structure and portfolio choice literatures. A sufficient condition for this case is that $a_{12} = a_{21} = 0$ and $a_{11} = a_{22}$, or in words, that both the conditional mean and volatility are univariate AR processes with the same persistence levels. Otherwise, if $a_{11} + a_{12} - a_{21} - a_{22} \neq 0$, the latent VAR introduces a second factor into the Sharpe ratio dynamics, namely the lagged conditional volatility in Eq. (16) or the lagged conditional mean in Eq. (17). It is important to recognize, however, that although the own-lag coefficients depend on which conditional moment is used as the second factor ($a_{11} - a_{21}$ versus $a_{22} - a_{12}$), the overall time-series dynamics of the log Sharpe ratios (including the persistence level) are the same in Eqs. (16) and (17) because, given the definition $\ln S_t = \ln \mu_t - \ln \sigma_t$, the two equations are mathematically equivalent.

2.4. Link to equilibrium models

To gain some economic intuition about the Sharpe ratio dynamics, consider the first-order condition $E_t[M_{t+1}(R_{t+1} - R_t^f)] = 0$ of a representative investor with marginal utility growth M_{t+1} (the stochastic discount factor or pricing kernel). Using the definition of covariance and the fact that $E[M_{t+1}] = 1/R_t^f$, we can rewrite this first-order condition as

$$E_t[R_{t+1} - R_t^f] = -R_t^f \operatorname{Cov}_t[M_{t+1}, R_{t+1} - R_t^f] \tag{18}$$

and solve for the Sharpe ratio:[5]

$$\frac{E_t[R_{t+1} - R_t^f]}{\operatorname{Std}_t[R_{t+1} - R_t^f]} = -R_t^f \operatorname{Std}_t[M_{t+1}] \operatorname{Corr}_t[M_{t+1}, R_{t+1} - R_t^f]. \tag{19}$$

---

[5] This definition of the Sharpe ratio is not quite comparable to $S_t = \mu_t/\sigma_t$ because it involves gross returns as opposed to continuously compounded returns. However, the difference is negligible for short return horizons.

This last equation illustrates that the log Sharpe ratio inherits the dynamics of the log risk-free rate, the log volatility of marginal utility growth, and the log correlation of marginal utility growth with returns (assuming this correlation is negative). Furthermore, if we parameterize marginal utility growth $M_{t+1} = v'(C_{t+1})/v'(C_t)$ and expand around the current consumption level C_t, we get $\text{Std}_t[M_{t+1}] \simeq RRA_t \, \text{Std}_t[(C_{t+1} - C_t)/C_t]$, which shows that the volatility of marginal utility growth depends to a first-order approximation on relative risk aversion $RRA_t = -C_t v''(C_t)/v'(C_t)$ and the volatility of consumption growth.

There is an extensive literature on the dynamics of the risk-free rate. In addition, there are numerous equilibrium asset pricing models in which the level of relative risk aversion, the volatility of consumption growth, or the correlation between marginal utility growth and returns vary in a slowly mean-reverting fashion. Models with changing risk aversion include Constantinides (1990), Campbell and Cochrane (1999), Bekaert and Grenadier (2001), and Brandt and Wang (2003). Models with varying volatility of consumption growth include Abel (1988), Kandel and Stambaugh (1991), Gennotte and Marsh (1993), and Bansal and Yaron (2003). Finally, Whitelaw (2000) formulates a model with changing correlation between marginal utility growth and returns. Our statistical approach can be interpreted as a reduced-form approximation of these equilibrium asset pricing models that can accommodate *two* independent sources of time variation in Eq. (19).

2.5. Hypotheses

Based on the latent VAR model for the conditional mean and volatility, we examine two main hypotheses about the relationship between the two moments. The first hypothesis is about their contemporaneous relationship:

H1: *No contemporaneous relationship exists between the conditional mean and volatility. Formally, in the context of our model, $\beta_1 = \beta_2 = 0$. From Eqs. (10) and (11), both the volatility-in-mean effect β_1 and the mean-in-volatility effect β_2 depend on the correlation ρ. As a result, this hypothesis is equivalent to $\rho = 0$. If the innovations to the moments are uncorrelated, neither the volatility-in-mean effect nor the mean-in-volatility effect exists, and the two moments are contemporaneously unrelated.*

The second hypothesis is about the intertemporal relationship between the conditional mean and volatility. There are two forms of intertemporal relationship in our model: Granger causality and the reduced-form lag adjustments. Granger causality refers to whether one variable helps predict future realizations of another through the off-diagonal elements of the transition matrix A. Despite its name, Granger causality refers to a forecasting relationship and does not identify a structural relationship or a direction of causality. In contrast, the lag-volatility-in-mean and lag-mean-in-volatility effects do shed light on the structural relationship between the two conditional moments, indicating whether the lagged volatility enters the reduced-form mean dynamics (10) and whether the lagged mean enters the

reduced-form volatility dynamics (11), respectively. These lag adjustments depend not only on the off-diagonal elements of A but also on the correlation ρ. The second hypothesis is:

H2: *There exists no lag-volatility-in-mean and/or lag-mean-in-volatility effect, meaning $a_{12} - \beta_1 a_{22} = 0$ and/or $a_{21} - \beta_2 a_{11} = 0$. If we fail to reject this hypothesis, the mean adjusts instantaneously to changes in the volatility and vice versa. Otherwise, there exists an intertemporal relationship between the two moments through the lag adjustments.*

3. Econometric approach

Eqs. (1) and (2) make up a state-space model. In the terminology of state-space models, the first equation is the measurement or observation equation and the second equation is the transition or state equation. To draw inferences about both the parameters of the VAR and the realizations of the latent moments from the observed returns, we need to solve a sequence of filtering and smoothing problems. Filtering generates the one-step-ahead forecasts of the latent variables $\mathrm{E}[\ln \mu_t, \ln \sigma_t | y_1, \dots, y_t]$ and the corresponding forecast variances $\mathrm{Var}[\ln \mu_t, \ln \sigma_t | y_1, \dots, y_t]$, which in a linear Gaussian state-space model are used to construct the likelihood function. Smoothing yields the full-information forecasts of the latent variables $\mathrm{E}[\ln \mu_t, \ln \sigma_t | y_1, \dots, y_T]$ and the corresponding forecast variances $\mathrm{Var}[\ln \mu_t, \ln \sigma_t | y_1, \dots, y_T]$. Unfortunately, the filtering and smoothing problems generated by our latent VAR model are nonstandard because of the nonlinearities in the measurement equation. As a result, the standard Kalman filter (designed for linear Gaussian state-space models) cannot be used directly in the estimation of our model. We instead rely on a simulation-based method for nonlinear and non-Gaussian state-space models.

It is convenient for our econometric discussion to reformulate the model slightly. We define $m_t = \ln \mu_t$ and $v_t = \ln \sigma_t$, so that $\mu_t = \exp\{m_t\}$ and $\sigma_t = \exp\{v_t\}$. We then de-mean the latent processes and, with some abuse of our previous notation, let m_t and v_t denote the *de-meaned* moments. Finally, we define $s_t = [m_t, v_t]'$, $Z_1 = [1, 0]'$, and $Z_2 = [0, 1]'$. With this revised notation, we write Eqs. (1) and (2) as follows:

$$y_t = \bar{\mu}\mathrm{e}^{m_{t-1}} + \bar{\sigma}\mathrm{e}^{v_{t-1}}\varepsilon_t = \bar{\mu}\mathrm{e}^{Z_1' s_{t-1}} + \bar{\sigma}\mathrm{e}^{Z_2' s_{t-1}}\varepsilon_t \quad \text{with } \varepsilon_t \sim \mathrm{N}[0, 1] \tag{20}$$

and

$$s_t = A s_{t-1} + \eta_t \quad \text{with } \eta_t \sim \mathrm{MVN}[0, \Sigma]. \tag{21}$$

We collect the unknown parameters of the transformed state-space model into a single parameter vector

$$\psi = [a_{11}, a_{12}, a_{21}, a_{22}, b_{11}, b_{22}, \rho, \bar{\mu}, \bar{\sigma}]', \tag{22}$$

and denote the histories of the unobserved conditional mean and volatility by

$$\theta = [s_0, s_1, \ldots, s_{T-1}]'. \tag{23}$$

3.1. Estimation

We estimate the latent VAR model by SML. In particular, we numerically construct the likelihood function of the model using an importance sampling approach similar to the one Durbin and Koopman (1997) and Sandmann and Koopman (1998) use for estimating SV models.

3.1.1. Motivation

The likelihood function of the model is

$$
\begin{aligned}
\mathcal{L}(\psi) &= p(y \mid \psi) \\
&= \int p(y, \theta \mid \psi) \, d\theta \\
&= \int p(y \mid \theta, \psi) p(\theta \mid \psi) \, d\theta.
\end{aligned}
\tag{24}
$$

The first line of Eq. (24) defines the likelihood function as the density of the observed returns y given the parameters ψ. The second line expresses this density as the marginal density of the joint distribution of the observed returns y and the unobserved state variables $\theta = [m, v]$. Finally, in the third line we write the joint density of y and θ as the product of the conditional density of y given θ and the marginal density of θ.

The point of this algebra is to illustrate that for a given parameter vector ψ, the likelihood function $\mathcal{L}(\psi)$ can be viewed as an *expectation* of the conditional density $p(y|\theta, \psi)$ with respect to the marginal density $p(\theta|\psi)$. With T data points and two latent state variables, this expectation involves a $2T$-dimensional integral. The high dimensionality of this integral and the nonlinearities in the conditional density $p(y|\theta, \psi)$ make the likelihood function practically impossible to compute analytically.

In principle, the latent state histories θ can be sampled directly from the density $p(\theta|\psi)$ and Monte Carlo integration can then be used to solve the integral in Eq. (24). In practice, however, this brute-force simulation approach is grossly inefficient. We can greatly improve the efficiency of the Monte Carlo integration through the use of importance sampling. To explain how importance sampling works in our context, we rewrite the likelihood function in Eq. (24) as follows:

$$
\begin{aligned}
\mathcal{L}(\psi) &= \int p(y \mid \theta, \psi) \frac{p(\theta \mid \psi)}{g(\theta \mid y, \psi)} g(\theta \mid y, \psi) \, d\theta \\
&= \mathrm{E}_g \left[p(y \mid \theta, \psi) \frac{p(\theta \mid \psi)}{g(\theta \mid y, \psi)} \right],
\end{aligned}
\tag{25}
$$

where $g(\theta|y, \psi)$, the so-called importance density, is a conditional density of θ that is not the same as the conditional density $p(\theta|y, \psi)$ generated by the model,

and $\mathrm{E}_g[\cdot]$ denotes an expectation taken with respect to this importance density.

If $g(\theta|\psi) = p(\theta|\psi)$, the distribution of the latent θ is the same under the two models (though the distribution of y given θ may be very different), and

$$g(\theta \,|\, y, \psi) = \frac{g(y, \theta \,|\, \psi)}{g(y \,|\, \psi)} = \frac{g(y \,|\, \theta, \psi)\, g(\theta \,|\, \psi)}{g(y \,|\, \psi)} = \frac{g(y \,|\, \theta, \psi)p(\theta \,|\, \psi)}{g(y \,|\, \psi)}, \qquad (26)$$

which implies that

$$\frac{p(\theta \,|\, \psi)}{g(\theta \,|\, y, \psi)} = \frac{g(y \,|\, \psi)}{g(y \,|\, \theta, \psi)}, \qquad (27)$$

where $g(y|\psi) \equiv \mathscr{L}_g(\psi)$ is the likelihood function of the model $g(y|\theta, \psi)$. Finally, substituting Eq. (27) into Eq. (25), we have

$$\mathscr{L}(\psi) = \mathrm{E}_g\left[p(y \,|\, \theta, \psi)\frac{\mathscr{L}_g(\psi)}{g(y \,|\, \theta, \psi)}\right] = \mathscr{L}_g(\psi)\mathrm{E}_g\left[\frac{p(y \,|\, \theta, \psi)}{g(y \,|\, \theta, \psi)}\right]. \qquad (28)$$

In summary, the true likelihood function $\mathscr{L}(\psi)$ can be expressed as the product of the likelihood function $\mathscr{L}_g(\psi)$ of the model $g(y|\theta, \psi)$, which we call the approximating model, and a correction factor $\mathrm{E}_g[w(\theta, \psi)]$, where for notational convenience we define

$$w(\theta, \psi) = \frac{p(y \,|\, \theta, \psi)}{g(y \,|\, \theta, \psi)}. \qquad (29)$$

The correction factor characterizes the departure of the true likelihood from the likelihood of the approximating model as the average distance between the conditional densities $p(y \,|\, \theta, \psi)$ and $g(y \,|\, \theta, \psi)$. If the two conditional densities are close to each other for all θ, the function $w(\theta, \psi)$ is close to one and so is its expectation with respect to $g(\theta|y, \psi)$.

The importance density $g(\theta \,|\, y, \psi)$ can in principle be any density of θ, given y and ψ, that is positive everywhere, satisfies the restriction $g(\theta \,|\, \psi) = p(\theta \,|\, \psi)$, and for which $\mathrm{E}_g[w(\theta, \psi)]$ exists. The main idea of our econometric approach is to choose an approximating model for which we can conveniently compute both the likelihood function $\mathscr{L}_g(\psi)$ and the correction factor $\mathrm{E}_g[w(\theta, \psi)]$. We can then evaluate the likelihood function $\mathscr{L}(\psi)$ using Eq. (28).

Before we present the approximating model, note that in choosing an importance density for computational convenience alone we depart from the usual approach in the literature of considering also the statistical properties of the resulting ratio of densities $w(\theta, \psi)$. Geweke (1989) shows that for the variance of this ratio to be finite, the tails of the density in the numerator must be fatter than the tails of the density in the denominator. In contrast to other applications of importance sampling, our choice of approximating model is likely to not satisfy this condition. We return to this issue in Section 3.4.

3.1.2. Approximating model

We choose $g(\theta \,|\, y, \psi)$ to be the conditional density of a *linear Gaussian* state-space model that closely approximates our nonlinear model. For example, in the case

$\text{Corr}[\varepsilon_t, \eta_t] = 0$, we replace the nonlinear measurement Eq. (20) with the linear measurement equation

$$Y_t \equiv \begin{bmatrix} y_t \\ \ln(y_t - \bar{\mu})^2 \end{bmatrix} = H' s_{t-1} + w_t, \quad \text{with } w_t \sim \text{MVN}[c_{t-1}, R_{t-1}], \tag{30}$$

where

$$H = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \quad c_t = \begin{bmatrix} c_{1t} \\ c_{2t} \end{bmatrix}, \quad R_t = \begin{bmatrix} R_{1t} & \rho_w \sqrt{R_{1t} R_{2t}} \\ \rho_w \sqrt{R_{1t} R_{2t}} & R_{2t} \end{bmatrix}, \tag{31}$$

and the parameters c_t and R_t are calibrated such that the densities $g(y \mid \theta, \psi)$ of the linearized model approximate well the true densities $p(y \mid \theta, \psi)$ (the appendices provide more details). The virtue of this linear measurement equation is that, together with the linear transition Eq. (21), it implies a linear Gaussian state-space model for which we can analytically evaluate the likelihood function $\mathscr{L}_g(\psi)$ using the Kalman filter algorithm.

3.1.3. Likelihood evaluation

Given a realization of the state history θ, we can easily evaluate the conditional densities $p(y \mid \theta, \psi)$ and $g(y \mid \theta, \psi)$ and hence also the function $w(\theta, \psi)$. However, θ is unobserved and the expectation $\text{E}_g[w(\theta, \psi)]$ cannot be computed analytically. We therefore use simulations to compute this correction factor. Specifically, we sample N state histories θ^i, for $i = 1, 2, \ldots, N$, from the importance density $g(\theta \mid y, \psi)$ and use these samples to construct the ratios

$$w^i = \frac{p(y \mid \theta^i, \psi)}{g(y \mid \theta^i, \psi)} \tag{32}$$

and to compute an estimate of the correction factor, $\text{E}_g[w(\theta, \psi)]$, as

$$\bar{w} = \frac{1}{N} \sum_{i=1}^{N} w^i. \tag{33}$$

We then use Eq. (28) to construct an estimate of the likelihood function $\mathscr{L}(\psi)$ as

$$\hat{\mathscr{L}}(\psi) = \mathscr{L}_g(\psi) \bar{w}. \tag{34}$$

Since $\bar{w} \to \text{E}_g[w(\theta, \psi)]$ as $N \to \infty$, $\hat{\mathscr{L}}(\psi)$ is a consistent approximation of $\mathscr{L}(\psi)$.

As is usually the case, it is more convenient to work with the *log* likelihood. Taking logs of Eq. (34) gives

$$\ln \hat{\mathscr{L}}(\psi) = \ln \mathscr{L}_g(\psi) + \ln \bar{w}. \tag{35}$$

However, $\ln \hat{\mathscr{L}}(\psi)$ is a slightly biased estimator of $\ln \mathscr{L}(\psi)$ because $\text{E}[\ln \bar{w}] \neq \ln \text{E}_g[w_i]$ due to Jensen's inequality. Durbin and Koopman (1997) and Shephard and Pitt (1997) therefore suggest adding the following term to the log likelihood function to correct the bias from the log transformation (up to order $N^{-3/2}$):

$$\ln \hat{\mathscr{L}}(\psi) = \ln \mathscr{L}_g(\psi) + \ln \bar{w} + \frac{s_w^2}{2N\bar{w}^2}, \tag{36}$$

where

$$s_w^2 = \frac{1}{N-1} \sum_{i=1}^{N} (w^i - \bar{w})^2. \tag{37}$$

Finally, we define our SML estimator, $\hat{\psi}$, as the parameter values that maximize this bias-corrected approximation of the log likelihood function.

One drawback of constructing the log likelihood of our model as the sum of the log likelihood of the approximating model and the log correction factor is that we cannot directly compute the contribution of each observation y_t to the likelihood function. This is because both terms in the sum are evaluated for the whole sample y. As a result, we cannot compute outer-product based standard errors to check the validity of the second-derivative based standard errors we use for drawing asymptotic inferences.

3.2. State variable extraction

Analogous to the simulation-based construction of the correction factor, we compute the smoothed estimates of the unobserved state variables $s_t = [m_t, v_t]'$ as follows:

$$\mathrm{E}[s_t \mid y, \psi] = \frac{1}{N} \sum_{i=1}^{N} w^i s_t^i, \tag{38}$$

where s_t^i is the tth element of θ^i from the ith draw from the importance density.

The smoothed estimates of the state variables s_t are based on the entire sample of returns y instead of on the returns up to date t (filtered estimates). Unfortunately, our estimation approach does not readily produce filtered estimates of the state variables because the draws θ^i are obtained from the full-sample importance density $g(\theta \mid y, \psi)$. The practical implication is that our approach is more suitable for ex post description of the data than for "on-line" signal extraction in a portfolio choice context, for example.

3.3. Identification and numerical implementation

As with most nonlinear models, it is difficult, if not impossible, to prove that the parameters of our model are uniquely identified. In order to minimize the chance of reporting a local or nonunique global maximum, we implement the following procedure. We optimize the likelihood from a variety of sensible (according to the literature and the data) initial parameter guesses for $\bar{\mu}$, $\bar{\sigma}$, A, b_{11}, and b_{22}, and for a grid of correlations ρ, ρ_μ, and ρ_σ, each ranging independently from -0.8 to 0.8. For each set of initial parameter values, we iterate between applying a simplex algorithm and a gradient-based algorithm, where each optimization round is started with the outcome of the previous round, until the criterion function does not improve from one optimization round to the next. We choose as our estimates the parameters that correspond to the greatest likelihood and verify that the parameters that correspond

to slightly inferior local maxima are qualitatively similar to our estimates. The reason for considering a grid of initial values for the correlations is that, intuitively, nonzero correlations are the most likely cause of identification problems. We are particularly concerned about the case $\rho_\mu \neq 0$ and $\rho_\sigma \neq 0$, but have not experienced any clear signs of identification problems (although the likelihood function is not particularly sensitive to the value of ρ_μ).

We are also concerned about the accuracy of the approximation of the likelihood function because a global maximum of an approximated likelihood function is only as good as the approximation. Therefore, we use a large simulation size of $N = 10,000$ (typical applications of SML with importance sampling use no more than 500 simulations) and, following the suggestion of Durbin and Koopman (1997), also use the method of antithetic variates to further reduce the variance of the simulation errors.[6] In addition, we verify for $\rho_\mu = \rho_\sigma = 0$ that increasing the number of simulations to 50,000 does not significantly change the SML estimates. Unfortunately, estimating the other specifications with more simulations is infeasible due to the dimensionality of the grid of starting values.

3.4. Finite simulation and sample size properties

The asymptotic distribution of our estimator for a finite simulation size N is unknown. To make matters worse, our estimator suffers from the critique of Shephard (2000), that the ratio of densities $w(\theta, \psi)$ may not have a finite variance. If the density in the denominator has thinner tails than the density in the numerator (which is likely to be the case because we are approximating an unconditionally fat-tailed SV model with a Gaussian model), extreme draws of θ^i may cause w^i to be unboundedly large. If the ratio of densities does not have a finite variance, the central limit theorem does not apply to the Monte Carlo integral $\bar{w}$ and there is no way to tell how far a given $\bar{w}$ is from its expectation.

To determine the severity of this finite simulation size problem, we follow the suggestion of Shephard (2000) to simulate a long series of w^i ratios and then compute the variances of w^i for the first 500 through 75,000 simulations. If this recursive variance is very sensitive to individual draws and does not converge to some constant level for very large N, the problem is serious. If, however, the recursive variance settles down quickly to a constant level, we can be more confident in our estimator for a finite simulation size.

Fig. 1 presents two recursive variance plots for the model with $\rho_\mu = \rho_\sigma = 0$. To see how the properties of the density ratios differ across the likelihood surface, we evaluate the ratios in Panel A at a representative set of starting values and in Panel B at the estimates presented in the empirical section. In both cases, the variance of the density ratios is very sensitive to individual draws for relatively small N. Furthermore, for the initial parameters, the variance is not only much larger in magnitude than for the SML estimates (by a factor of about 300), but it also keeps

---

[6] Durbin and Koopman (1997) show how to compute the contribution of the simulations to the asymptotic variance of the SML estimates. In our case, this contribution is less than 1%.
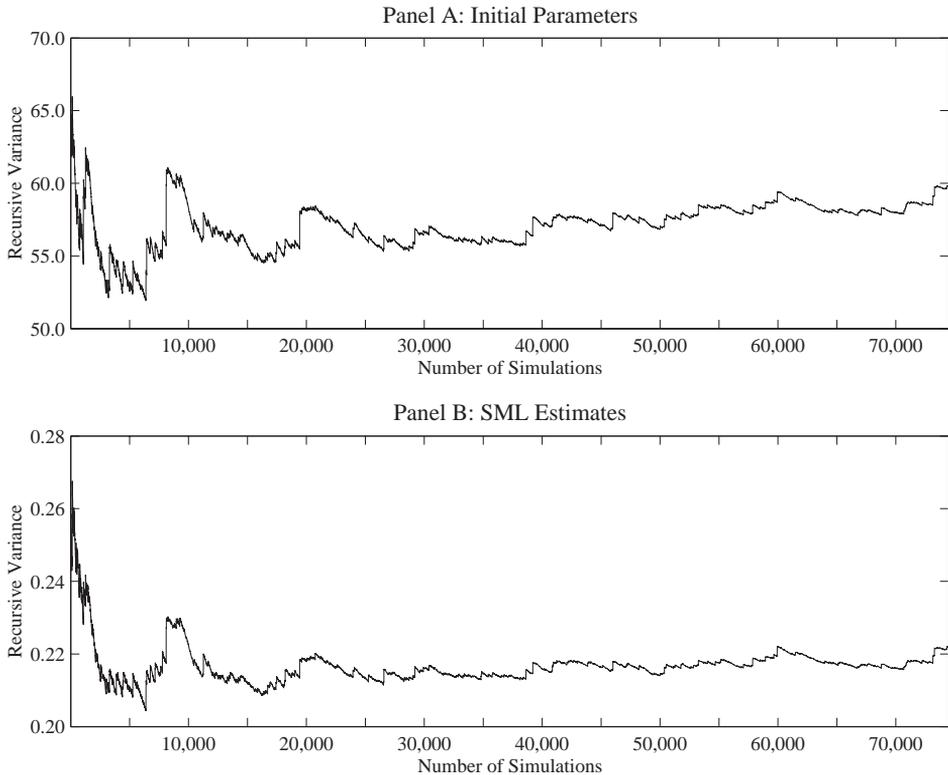
Fig. 1. Recursive variance of the density ratio. This figure plots the variance of the density ratio $w(\theta, \psi) = p(y \mid \theta, \psi)/g(y \mid \theta, \psi)$ with $\rho_{\mu} = \rho_{\sigma} = 0$, using 500 through 75,000 successive simulated values of the latent states θ^i. In Panel A, the parameters ψ are set to representative starting values of the simulated maximum likelihood (SML) estimator. In Panel B, the ratios are evaluated at the SML estimates.

drifting upward as N increases. This confirms Shephard's suspicion that the variance may be unbounded. For the SML estimates, in contrast, the variance seems to be fairly stable after 10,000 simulations (at least in this particular set of simulations), fluctuating from 0.209 to 0.227. This also confirms Shephard's observation in the context of SV models that the ratio of densities is much better behaved at the SML estimates than at other points on the likelihood surface.

Although these recursive variance plots are only indicative of the (lack of) problems induced by evaluating the correction factor with a finite number of simulations, we proceed by assuming that the asymptotic distribution of the estimator is Gaussian as usual. In doing so, we are likely to understate the true standard errors of our estimates for two reasons. First, the sampling variation of the Monte Carlo integral for a finite N (which is $1/N$ times the variance in Panel B of Fig. 1) increases the standard errors. Second, the poor sampling properties of the density ratios may severely deteriorate the asymptotic approximation.

As an alternative way to get a sense for the finite sample properties of the SML estimator, we conduct a Monte Carlo experiment. In particular, we repeatedly estimate the model with $\rho_\mu = \rho_\sigma = 0$ and with $N = 5000$ from 500 independent samples of $T = 636$ monthly returns simulated from the model with parameter values that correspond to the estimates presented in the empirical section. Table 1 presents the results. It shows the true parameter values in the first column and describes the sampling distribution of the corresponding parameter estimates in the remaining columns. The average and median estimates are all close to the true parameter values, suggesting that the estimator is relatively unbiased. Furthermore, except for a_{21} and a_{12}, the standard deviations of the estimates across samples are small. In particular, the correlation between the innovations to the conditional moments ρ, which is central to our first hypothesis, is measured very precisely with a true value of -0.634, an average estimate of -0.636, and a standard deviation of only 0.071 across samples. The estimates of a_{21} and a_{12}, which are important for our second hypothesis, are less precise with true values of -0.053 and 0.108 and estimates ranging from -0.295 to 0.181 with a mean of -0.064 and from -0.109 to 0.320 with a mean of 0.109, respectively. Finally, notice that some of the sampling distributions are quite skewed, especially for the diagonal elements of A and for ρ, raising serious concerns about asymptotic normal approximations.

4. Empirical results

4.1. Data

We study monthly returns on the value-weighted CRSP index in excess of the one-month Treasury bill rate from January 1946 through December 1998 (636 observations). For the purposes of robustness and specification testing, we also consider three information variables known to be correlated with expected returns and volatility: the short rate, term premium, and default premium.[7] The short rate is the yield of a one-month Treasury bill. The term premium is the yield spread of a ten-year Treasury bond and a one-year Treasury bill. Finally, the default premium is the yield spread of corporate bonds with Moody's Baa and Aaa rating. Table 2 presents summary statistics of the data. Fig. 2 plots the series and indicates the dates of all business cycle peaks (dashed lines) and troughs (dotted lines) obtained from the National Bureau of Economic Research (NBER).

4.2. Parameter estimates and hypotheses tests

Table 3 presents SML estimates of the latent VAR with the three different assumptions about the correlation between the return innovations and the

---

[7] In particular, subsets of these predictors are used by Keim and Stambaugh (1986), Campbell (1987), Fama and French (1989), and Ang and Bekaert (2001) to model expected returns and by Campbell (1987), Schwert (1989), Glosten et al. (1993), and Whitelaw (1994) to model volatility.

Table 1
Finite sample properties of simulated maximum likelihood estimator
This table describes the sampling distribution of the simulated maximum likelihood estimates of the model:

$$y_t = \bar{\mu}e^{m_{t-1}} + \bar{\sigma}e^{v_{t-1}}\varepsilon_t = \bar{\mu}e^{Z_1's_{t-1}} + \bar{\sigma}e^{Z_2's_{t-1}}\varepsilon_t \quad \text{with } \varepsilon_t \sim N[0,1]$$

and

$$s_t = As_{t-1} + \eta_t \quad \text{with } \eta_t \sim \text{MVN}[0, \Sigma],$$

where

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} b_{11} & \rho\sqrt{b_{11}b_{22}} \\ \rho\sqrt{b_{11}b_{22}} & b_{22} \end{bmatrix}, \quad \text{and} \quad \text{Corr}[\varepsilon_t, \eta_t] = 0.$$

The results are based on 500 independent samples of $T = 636$ returns simulated from the model with the parameters in the first column.

	True	Average	Standard deviation	Skewness	Excess Kurtosis	Max	Percentiles					Min
							95%	75%	50%	25%	5%	
a_{11}	0.8589	0.8588	0.0626	−0.2422	−0.6821	0.9787	0.9514	0.9047	0.8633	0.8134	0.7465	0.7097
a_{21}	−0.0529	−0.0643	0.0961	0.1959	−0.1858	0.1814	0.1123	0.0005	−0.0743	−0.1661	−0.2318	−0.2945
a_{12}	0.1084	0.1089	0.0910	−0.0908	−0.3689	0.3197	0.2478	0.1581	0.1099	0.0735	−0.0430	−0.1087
a_{22}	0.9226	0.9146	0.0423	−0.3584	0.1328	0.9902	0.9727	0.9352	0.9041	0.8771	0.8344	0.7833
b_{11}	0.0076	0.0081	0.0041	0.0049	−0.5830	0.0208	0.0183	0.0131	0.0089	0.0071	0.0037	0.0021
b_{22}	0.0553	0.0513	0.0061	−0.0707	0.0163	0.0709	0.0603	0.0554	0.0510	0.0476	0.0416	0.0379
ρ	−0.6336	−0.6357	0.0713	0.4844	−0.3324	0.3350	−0.5063	−0.5883	−0.6667	−0.7019	−0.7643	−0.7951

Table 2

Data

This table presents descriptive statistics of monthly returns on the value-weighted CRSP index in excess of the one-month Treasury bill rate as well as of the short rate, term premium, and default premium from January 1946 through December 1998. The short rate is the yield on a one-month Treasury bill, the term premium is the yield spread of a ten-year Treasury bond and a one-year Treasury bill, and the default premium is the yield spread of Moody's Baa- and Aaa-rated bonds.

	Market index	Short rate	Term premium	Default premium
Mean	0.0063	0.0470	0.0115	0.0090
Std. dev.	0.0413	0.0304	0.0126	0.0042
Max	0.1495	0.1805	0.0431	0.0265
Min	−0.2455	0.0031	−0.0367	0.0027
Median	0.0095	0.0448	0.0115	0.0076
Skewness	−0.6257	1.0177	−0.3442	1.5361
Kurtosis	5.5870	4.4510	4.0337	5.3480
Autocorrelation				
1-month	0.0347	0.9572	0.9582	0.9755
6-month	−0.0637	0.8819	0.7040	0.8435
12-month	0.0398	0.8153	0.5410	0.7087
24-month	0.0271	0.6663	0.2359	0.5263
Correlation with				
Market index	1.0000	−0.1209	0.1293	0.0530
Short rate		1.0000	−0.3344	0.6652
Term premium			1.0000	0.0726
Default premium				1.0000

conditional mean and volatility innovations. Model A assumes that these innovations are uncorrelated. In Models B and C, the return innovations are allowed to be correlated either with the conditional mean or the conditional volatility innovations, respectively. Finally, Model D is unrestricted.

The estimates of $\bar{\mu}$ and $\bar{\sigma}$ (together with the estimates of b_{11} and b_{22}) imply an annualized unconditional risk premium of 7.5–7.8% and a return volatility of 13.2–13.6%, which, not surprisingly, match closely the sample moments in Table 2.[8] All but one of the estimates of the own-lag persistence coefficients a_{11} and a_{22} exceed 0.85 (the exception is a_{11} in Model B), suggesting that both the conditional mean and volatility are highly persistent quantities.

The innovations to the conditional mean and volatility are contemporaneously negatively correlated in all models, with a correlation ρ that ranges from −0.635 to −0.458 and that is statistically significant at any conventional significance level. (The t-statistics are computed with Hessian-based standard error.) This finding implies a strong rejection of hypothesis H1. Furthermore, the negative sign implies that the unconditional return distribution is negatively skewed (consistent with Table 2) and that both the volatility-in-mean and mean-in-volatility effects are negative and significant, with $\beta_1 = \rho\sqrt{b_{11}/b_{22}}$ ranging from −0.235 to −0.155 and $\beta_2 =$

[8] We compute the unconditional moments of returns using the properties of bi-variate log-normal random variables and the unconditional mean and covariance matrix of the log moments in Eq. (4).
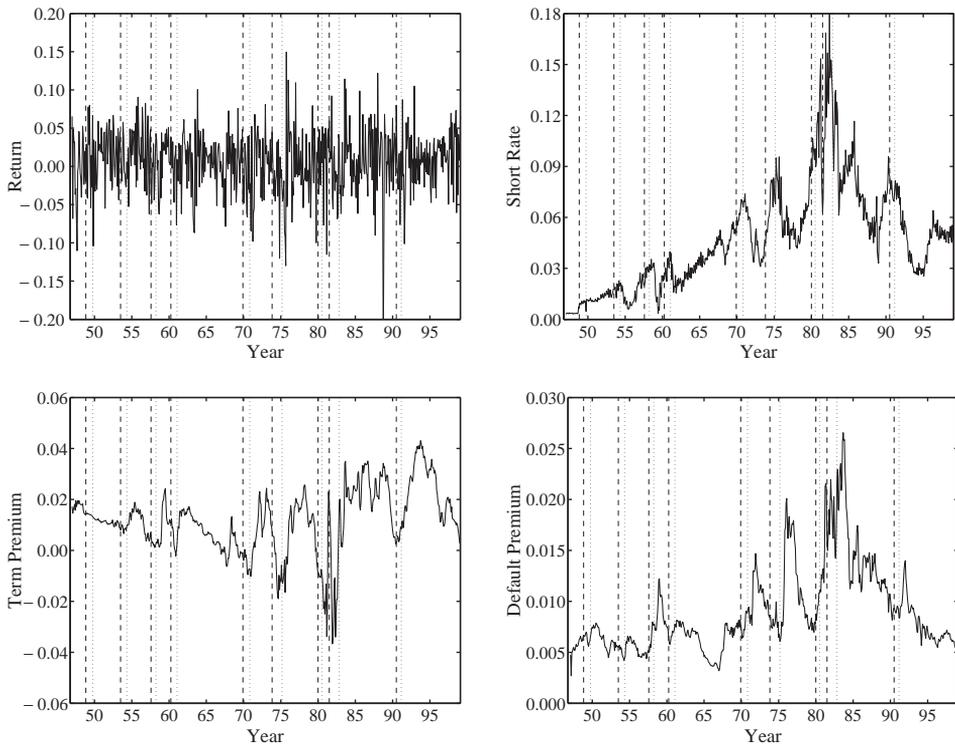
Fig. 2. Returns and predictors. This figure plots the monthly returns on the value-weighted CRSP index in excess of the one-month Treasury bill rate as well as of the short rate, term premium, and default premium from January 1946 through December 1998. The vertical dashed and dotted lines represent the NBER dates of the peak and trough of every business cycle in the sample, respectively.

$\rho\sqrt{b_{22}/b_{11}}$ ranging from -1.998 to -1.352. It is clear that there exists a negative contemporaneous trade-off between expected returns and risk, which confirms the results of Campbell (1987), Breen et al. (1989), Nelson (1991), and Glosten et al. (1993), but not those of French et al. (1987) and Campbell and Hentschel (1992).

The off-diagonal elements of the transition matrix a_{12} and a_{21}, which determine the Granger causality embedded in the VAR, are individually not significant at the 5% level (but a_{12} is significant at the 10% level in Models A and C). To determine whether these parameters are also jointly insignificant, we estimate the models under the restriction $a_{12} = a_{21} = 0$ and conduct a likelihood ratio (LR) test of the restriction. The LR test statistics, which are asymptotically χ^2 distributed with two degrees of freedom, all exceed the 1% critical value of 9.21. This rejection is evidence in favor of bidirectional Granger causality between the conditional moments, meaning the lagged volatility helps predict the mean and, to a somewhat lesser degree, the lagged mean helps predict the volatility.

The opposite signs of a_{12} and a_{21} confirm qualitatively Whitelaw's (1994) finding that the cross-autocorrelations between the conditional mean and volatility offset

Table 3

Parameter estimates

This table presents the simulated maximum likelihood estimates of the model:

$$y_t = \bar{\mu}e^{m_{t-1}} + \bar{\sigma}e^{v_{t-1}}\varepsilon_t = \bar{\mu}e^{Z_1's_{t-1}} + \bar{\sigma}e^{Z_2's_{t-1}}\varepsilon_t \quad \text{with } \varepsilon_t \sim N[0,1]$$

and

$$s_t = As_{t-1} + \eta_t \quad \text{with } \eta_t \sim MVN[0, \Sigma],$$

where

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} b_{11} & \rho\sqrt{b_{11}b_{22}} \\ \rho\sqrt{b_{11}b_{22}} & b_{22} \end{bmatrix}, \quad \text{and} \quad \text{Corr}[\varepsilon_t, \eta_t] = \begin{bmatrix} \rho_\mu \\ \rho_\sigma \end{bmatrix}.$$

The estimates are for monthly returns on the value-weighted CRSP index in excess of the one-month Treasury bill rate from January 1946 through December 1998.

Parameters	Model A		Model B		Model C		Model D	
	Estimate	t-Stat	Estimate	t-Stat	Estimate	t-Stat	Estimate	t-Stat
a_{11}	0.8592	17.41	0.8313	9.23	0.8658	11.21	0.8677	10.92
a_{21}	−0.0531	−1.92	−0.0211	−1.43	−0.0885	−1.96	−0.1292	−1.42
a_{12}	0.1081	0.92	0.1168	1.12	0.0861	1.01	0.0947	0.71
a_{22}	0.9237	15.23	0.9110	16.32	0.8973	15.32	0.9086	18.49
b_{11}	7.60×10^{-3}	4.24	6.43×10^{-3}	2.20	5.96×10^{-3}	3.19	4.68×10^{-3}	2.76
b_{22}	0.0554	13.02	0.0561	12.22	0.0614	5.55	0.0591	5.22
ρ	−0.6345	−4.32	−0.4577	−3.21	−0.5584	−5.80	−0.5621	−5.96
ρ_μ	—	—	−0.0866	−0.49	—	—	−0.0517	−0.77
ρ_σ	—	—	—	—	−0.2541	−4.04	−0.2430	−3.67
$\bar{\mu}$	6.50×10^{-3}	3.31	6.48×10^{-3}	5.94	6.24×10^{-3}	5.11	6.24×10^{-3}	4.87
$\bar{\sigma}$	0.0377	10.32	0.0385	5.60	0.0382	8.79	0.0382	9.01
$\beta_1 = \rho\sqrt{b_{11}/b_{22}}$	−0.2350	−6.23	−0.1550	−4.11	−0.1740	−5.98	−0.1582	−3.99
$\beta_2 = \rho\sqrt{b_{22}/b_{11}}$	−1.7131	−8.34	−1.3519	−6.34	−1.7923	−9.32	−1.9975	−10.92
$a_{12} - \beta_1 a_{22}$	0.3252	3.82	0.2580	3.11	0.2422	3.47	0.2384	3.82
$a_{21} - \beta_2 a_{11}$	1.4188	5.21	1.1028	4.35	1.4623	4.11	1.6040	2.42
$a_{11} + a_{12} - a_{21} - a_{22}$	0.0967	3.33	0.0684	2.54	0.1431	2.31	0.1830	2.92
$\ln \mathscr{L}$	1,150.98		1,151.11		1,152.94		1,153.06	

each other through time. An increase in the mean is followed by a decrease in the volatility, while an increase in the volatility leads to a subsequent increase in the mean. Unfortunately, we cannot directly compare our estimates to Whitelaw's results because our VAR is for log moments instead of levels. However, we can approximate the off-diagonal elements of a VAR in levels by $d\mu/d\sigma = (d \ln \mu/d \ln \sigma)(\mu/\sigma) \simeq a_{12}\bar{\mu}/\bar{\sigma}$ and $d\sigma/d\mu = (d \ln \sigma/d \ln \mu)(\sigma/\mu) \simeq a_{21}\bar{\sigma}/\bar{\mu}$. Whitelaw's estimates (for a shorter sample period) are $d\mu/d\sigma = 0.118$ and $d\sigma/d\mu = -0.107$. Our approximations range from 0.014 to 0.021 and from −0.784 to −0.160, respectively.

The lag-volatility-in-mean and lag-mean-in-volatility effects are both positive and significant at the 5% level (in most cases even at the 1% level), with estimates of $a_{12} - \beta_1 a_{22}$ ranging from 0.238 to 0.325 and those of $a_{21} - \beta_2 a_{11}$ ranging from 1.103 to 1.604. (The asymptotic standard errors of $a_{12} - \beta_1 a_{22}$ and $a_{21} - \beta_2 a_{11}$ are

computed using the delta method.) Therefore, both parts of hypothesis H2 are decisively rejected and the reduced-form models in Eqs. (10) and (11) feature lag adjustments. This means that the mean and volatility adjust not only instantaneously but also with a lag to changes in the other moment.

The rejections of the hypotheses H1 and H2 are based on asymptotic standard errors. However, as stressed in Section 3.4, the asymptotics of our estimator are questionable for a number of reasons. To substantiate the evidence on the contemporaneous correlation and lag adjustments, we conduct a Monte Carlo experiment under each null. Specifically, we estimate Model A under each of the restrictions $\rho = 0$, $a_{12} = \beta_1 a_{22}$, and $a_{21} = \beta_2 a_{11}$. We then simulate from each model 500 independent samples of $T = 636$ monthly returns and construct sampling distributions of the restrictions under the null. The results corroborate our asymptotic inferences. Under the null of $\rho = 0$, the estimated correlation is less than -0.5 in only two of the 500 samples. Likewise, under the null $a_{12} = \beta_1 a_{22}$ or $a_{21} = \beta_2 a_{11}$, the estimate of $a_{12} - \beta_1 a_{22}$ or $a_{21} - \beta_2 a_{11}$ exceed their empirical estimates in less than 3% and 1% of the samples under the null, respectively. We conclude that the rejections of H1 and H2 are not attributed to invalid asymptotic approximations.

The estimates of $a_{11} + a_{12} - a_{21} - a_{22}$ range from 0.068 to 0.183 and all have t-statistics greater than two, suggesting (based on the asymptotics) that the Sharpe ratio follows a *two*-factor process. If we use the conditional mean as the second factor, the own-lag persistence of the Sharpe ratio is $a_{22} - a_{12} \simeq 0.80$; if we use the volatility as the second factor, it is $a_{11} - a_{21} \simeq 0.90$. In either case, the Sharpe ratio is highly persistent, which reflects our finding above that both the conditional mean and volatility are highly persistent.

The signs of the estimated correlations ρ_μ between the return and mean innovations (Models B and D), and ρ_σ between the return and volatility innovations (Models C and D), are consistent with the volatility feedback and leverage effects, respectively. Both correlations are negative, which implies that increases in the mean and volatility are associated with negative returns. However, only the estimates of ρ_σ are significant and, judging by the Akaike information criterion, only Model C yields an improvement in fit to the fully restricted Model A. (The Akaike information criterion is defined as $\text{AIC} = \ln \mathscr{L} - \dim[\psi]$.) Since the estimates of ρ_μ are insignificant and the volatility feedback explanation is consistent with a negative ρ_σ only if ρ is positive, we conclude that our results favor the leverage explanation of asymmetric volatility.

Finally, the estimates of the latent VAR and, more importantly, of the reduced-form models in Eqs. (10) and (11), are remarkably similar across the four correlation structures. We therefore focus in the remainder of the paper on Model C, the best-fitting model according to the Akaike information criterion. However, we verified that all qualitative statements we make below are robust to this specification choice.

4.3. Time-series of the conditional moments

Given the parameter estimates in Table 3 and the index returns, we extract estimates of the conditional mean, volatility, and Sharpe ratio. Before we study the

time series of these estimates, we first consider as a point of reference the time series of the returns themselves. It is clear from Table 2 that the returns are noisy, with an annualized volatility of 14% and a first-order autocorrelation of only 0.035. It is therefore not surprising that in Fig. 2 the noise disguises the persistence of the conditional mean. The persistence of the conditional volatility is more apparent. Judging by the frequency and magnitude of large price changes, there are pronounced and prolonged periods of high and low volatility.

The left column of Fig. 3 presents the smoothed estimates of the annualized (*not* logged) mean (first row), volatility (second row), and Sharpe ratio (third row). Each plot also shows as vertical lines the dates of all business cycle peaks (dashed lines) and troughs (dotted lines) in our sample. It is immediately clear that both the conditional mean and volatility vary through time. The annualized mean has a standard deviation of 0.9% and ranges from 4.8% in 1965 to almost 10% in 1988. The annualized volatility has a standard deviation of 4.5% percent and ranges from less than 5% in 1965 to over 30% in 1987. Consistent with the estimates of a_{11} and a_{22}, the conditional mean reverts faster to its long-term average than the conditional volatility. In fact, periods of high or low volatility appear to last as long as a few *decades*.

To gauge the statistical significance of the variation of the conditional mean and volatility, we estimate restricted models with either constant mean or constant volatility and then perform LR tests of these restrictions. Both constrained models can be rejected at the 5% level (asymptotic *p*-values of 2.3% for the constant-mean model and less than 0.5% for the constant-volatility model). To measure economic significance and to relate our results to the literature on time-varying moments, we also compute the statistics $1 - \mathrm{Var}[y_t - \mu_t]/\mathrm{Var}[y_t]$ and $1 - \mathrm{Var}[y_t^2 - \sigma_t^2]/\mathrm{Var}[y_t^2]$, which we interpret as predictive regression R^2's for returns and squared returns, respectively. The values of these statistics, 2.1% for returns and 35.7% for squared returns, are comparable to the results of predicting monthly returns with exogenous predictors and squared returns with ARCH models.

The variation of the conditional mean is distinctly countercyclical. In eight of the nine completed business cycles in the sample, the conditional mean rises almost monotonically from the peak (dotted vertical lines) to the trough (dashed vertical lines) of the cycle (the only exception is the 1948–1949 recession). Although other studies have found similar countercyclical patterns in the conditional mean (Fama and French, 1989), the strength of our finding is astonishing given the fact that we are not using exogenous predictors. The problem with using exogenous predictors to identify cyclical patterns is that the NBER may use the same predictors (or variables that are highly correlated with these predictors) to ex post date the peaks and troughs of the business cycles. Furthermore, it is interesting that although the periods between the peaks and troughs of the business cycles are aligned with the expected return cycles, the sometimes much longer periods between a trough and the following peak often contain more than one large swing in expected returns.

The conditional volatility exhibits a somewhat different cyclical pattern. In particular, the volatility is almost uniformly high and at times decreasing during recessions, when the conditional mean is rising, and low during expansions.

However, this business cycle pattern is distorted by low-frequency swings in volatility that last for decades.

To quantify the patterns in the variation of the conditional moments, we compute the differences between the means and volatilities at the trough and preceding peak of each business cycle. To smooth some of the month-to-month fluctuations in the moments, we work with the average moments over the three months straddling the peak or trough. For the mean, all but one of the differences are positive and the average difference is 1.1%. If the means were independent and identically distributed (iid), which is a very conservative assumption given their persistence, the standard error of this difference is $\sqrt{2} \times 0.9/\sqrt{9} = 0.4\%$, indicating that the average difference is significant at the 5% level. For the volatility, all differences are negative and the average difference is -4.9% with a standard error (under the conservative iid assumption) of $\sqrt{2} \times 4.5/\sqrt{9} = 2.1\%$. In addition, we compute correlations between the moments and the three exogenous conditioning variables. The univariate correlations of the conditional mean with the short rate, term premium, and default premium are 0.192, -0.021, and 0.498, respectively, and a multivariate R^2 is 0.319. The correlations of the conditional volatility with the same variables are 0.252, -0.238, and 0.256, and the R^2 is 0.132. These statistics illustrate at least two facts. First, the mean is highly correlated with the empirically and theoretically countercyclical default spread. Second, the volatility is much less correlated with the predictors than the mean, which is fairly standard in the literature (as long as the predictors exclude lagged volatilities).

The only seemingly counterfactual feature of our estimates is that the conditional mean is weakly positively correlated with the short rate, while predictive return regressions suggest a negative correlation. We conjecture that this result is due to the fact that our estimates are strongly cyclical and do not exhibit a pattern that resembles the almost monotonic rise of the short rate from the 1940s through the early 1980s and then subsequent drop through the late 1990s (see Fig. 2). To verify this conjecture, we relate the conditional mean to the difference between the short rate and its one-year moving average (which essentially detrends the short rate). Consistent with predictive return regressions, the resulting univariate correlation is -0.22. Furthermore, we run a multivariate regression of the conditional mean on all three predictors (which also detrends the short rate if the predictors are cointegrated). The regression coefficient on the short rate is -0.012 with a *t*-statistic of -8.46.

Together, the countercyclical variation of the mean and the increase in volatility during recessions leads to countercyclical variation of the conditional Sharpe ratio. The Sharpe ratio rises from the peak to the trough of *every* completed business cycle in the sample. This countercyclical variation of the Sharpe ratio is consistent with the intuition from habit formation models (Constantinides, 1990; Campbell and Cochrane, 1999). At the peak of a business cycle, consumers are enjoying consumption levels far above their historically built-up "habits" which makes them relatively risk-tolerant. As a result, a low expected reward per unit of risk, or low Sharpe ratio, is required for consumers to invest in stocks at the peak of a cycle. At the trough of a cycle, in contrast, consumption levels are approaching the habits (or

perceived subsistence levels), which makes consumers relatively risk averse. For consumers to invest in stocks at the trough, the Sharpe ratio must be high.

Fig. 3 offers a more visual description of how the lead-lag interactions between the conditional mean and volatility arise. As the economy comes off the peak of a business cycle, the volatility rises almost immediately. The mean, however, increases only gradually as the economy moves from the peak to the trough of the cycle. Therefore, the volatility appears to lead the mean through the recession. The mean then reaches a high at the trough of the cycle, shortly after which the volatility drops back to its lower expansion level. As a result, the increase in the mean is associated with a subsequent drop in the volatility.

4.4. Impulse response functions

To better understand the dynamic behavior of the conditional moments, we now turn to the impulse response functions of the VAR. The impulse response functions show how a one standard deviation shock to one of the innovations of the VAR affects contemporaneously the conditional mean and volatility and then propagates to future realizations of the moments through the dynamic structure of the VAR. Since our model describes the dynamics of the log mean and log volatility, we transform the impulse response functions of the VAR into the implied impulse response functions for the mean and volatility (without logs).

As long as the innovations of the VAR are uncorrelated, interpreting the impulse response functions is straightforward. The innovation to the mean is η_{1t} and the innovation to the volatility is η_{2t}. However, the innovations of the VAR are usually correlated, which means that η_{1t} and η_{2t} have a common component that acts on both the mean and volatility, and two orthogonal components that act only on the mean or the volatility, respectively. Since we cannot identify all three components separately, we follow the convention to attribute the common component of the innovations to η_{1t} (simply because it belongs to the first equation of the VAR) and to remove this common component from η_{2t}, so that this second innovation is by construction orthogonal to the first. Mechanically, we use a Cholesky decomposition of the error covariance matrix to orthogonalize the innovations of the VAR. The impulse response functions then trace the effects of a one standard deviation shock to the common component (which contemporaneously acts on the mean and, to the extent that the original innovations are correlated, also on the volatility) and to the orthogonal component (which by construction acts contemporaneously only on the volatility).

The six plots in Fig. 4 illustrate the impulse responses for the unrestricted VAR (solid lines) and the restricted VAR with $a_{12} = a_{21} = 0$ (dashed lines). The first column of plots shows the effects of a one standard deviation mean innovation at date t on the mean (first row), the volatility (second row), and the Sharpe ratio (third row) at dates $t, t + 1, \ldots, t + 60$. Similarly, the second column shows the effects of a one standard deviation orthogonalized volatility innovation on the current and future mean, volatility, and Sharpe ratio.
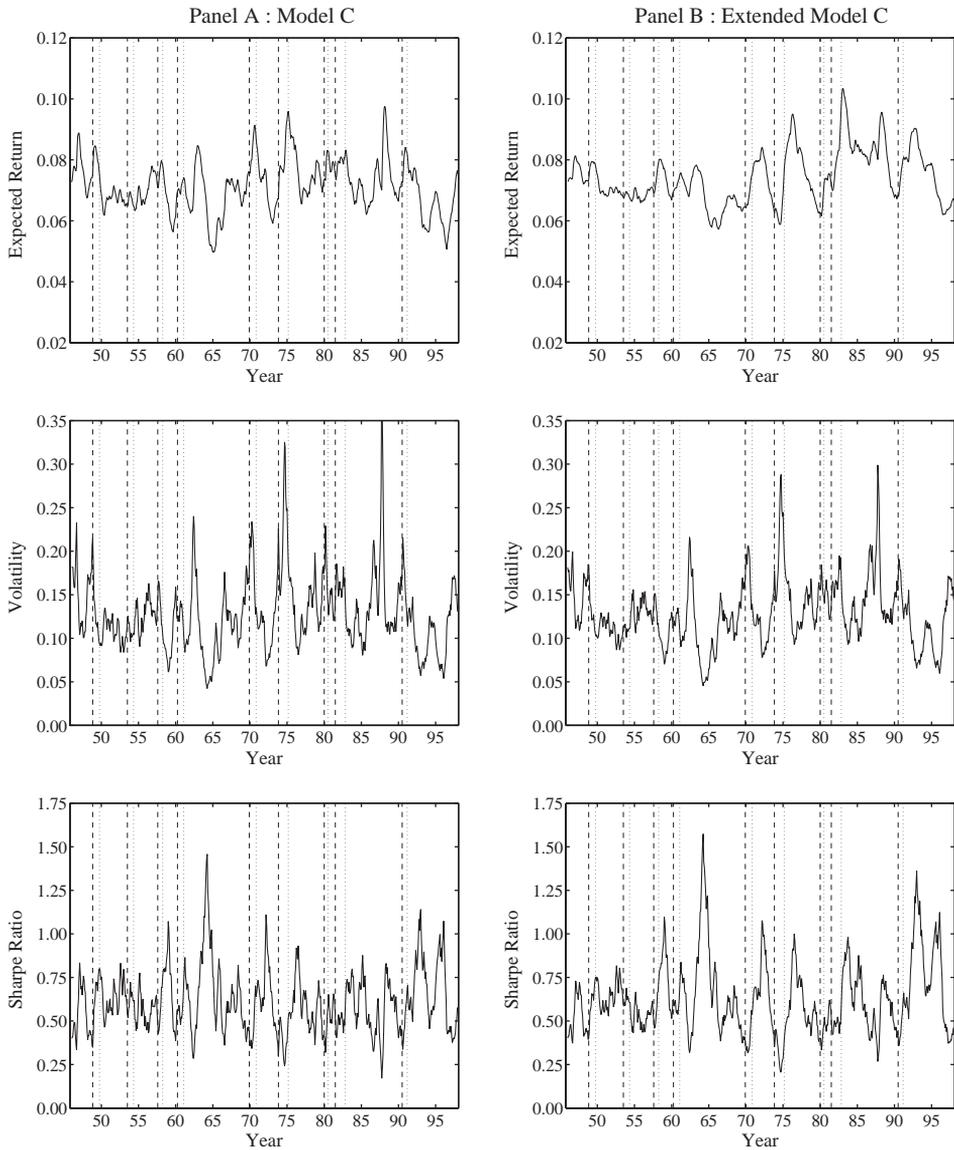
Fig. 3. Estimates of the conditional mean, volatility, and Sharpe ratio. This figure plots the annualized full-information (or smoothed) estimates of the conditional mean $E[\mu_t \,|\, y_1, \ldots, y_T]$, volatility $E[\sigma_t \,|\, y_1, \ldots, y_T]$, and Sharpe ratio $E[\mu_t \,|\, y_1, \ldots, y_T]/E[\sigma_t \,|\, y_1, \ldots, y_T]$. Panels A and B correspond to Model C without conditioning variables and the extended Model C with the short rate, term premium, and default premium as the conditioning variables, respectively. The vertical dashed and dotted lines represent the NBER dates of the peak and trough of every business cycle in the sample, respectively.

Panel A : Response to Mean Innovation

Panel B : Response to Orthogonal Volatility Innovation
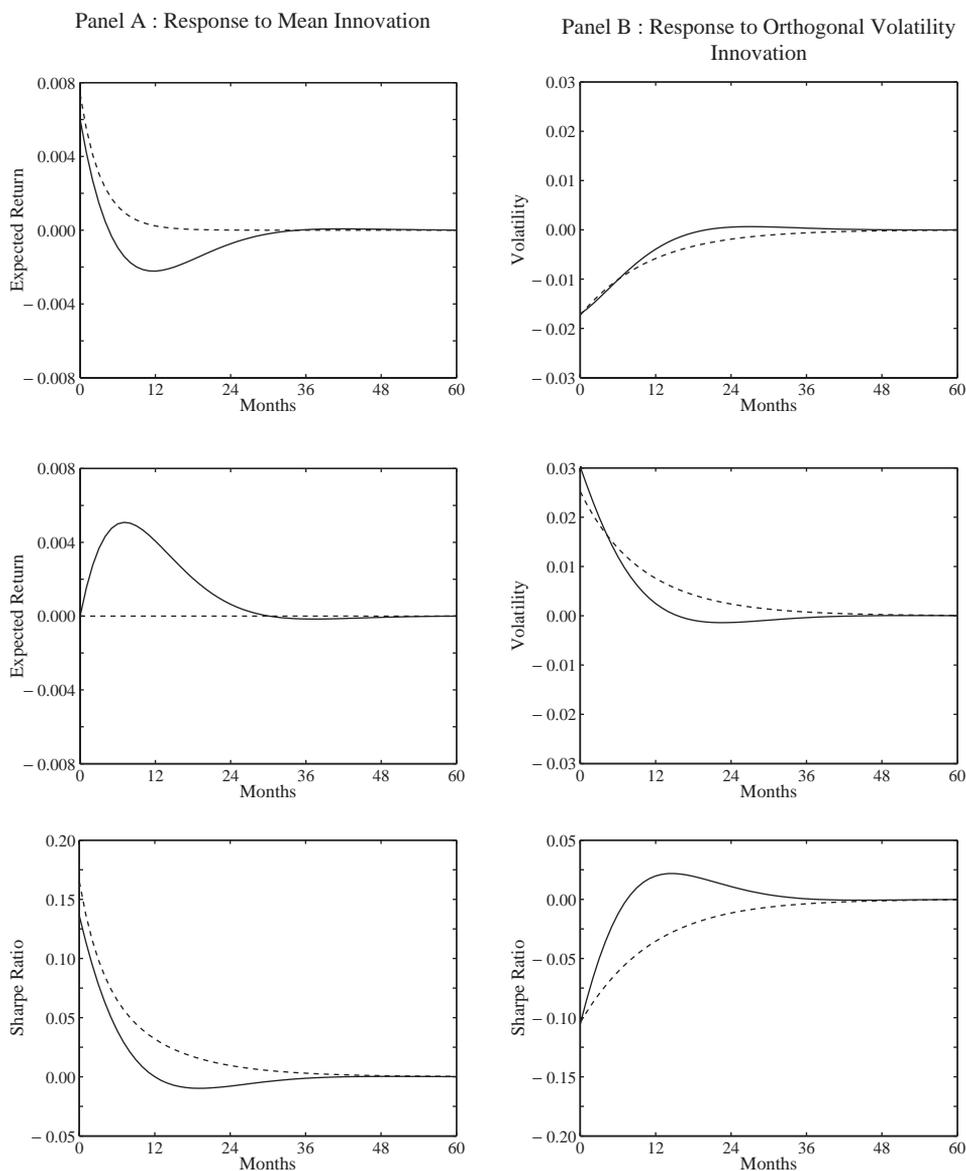

Fig. 4. Impulse response functions. This figure plots the contemporaneous and lagged responses of the conditional mean, volatility, and Sharpe ratio to a one standard deviation mean innovation (in Panel A) and a one standard deviation orthogonalized volatility innovation (in Panel B) for Model C. The solid and dashed lines correspond to unrestricted and restricted ($a_{21} = a_{12} = 0$) estimates.

Consider first the impulse responses for the unrestricted VAR. The first column of plots illustrates clearly the negative contemporaneous correlation between the conditional mean and volatility. A one standard deviation mean innovation

results in a contemporaneous 0.66% increase in the annualized mean, a 1.76% decrease in the annualized volatility, and a 0.14 increase in the annualized Sharpe ratio. Thereafter, the mean fully mean-reverts within four months and then overshoots its long-term average by dropping as much as an additional 0.29% over the next 24 months. The explanation for this overshooting of the mean after a positive mean innovation is the lag adjustment of the mean to the change in volatility. The volatility keeps dropping, although at a decreasing rate, for about 24 months after the mean innovation. This persistent drop in volatility causes the mean to fall due to the positive and relatively large Granger causality term a_{12} as well as the positive lag-volatility-in-mean effect $a_{12} - \beta_1 a_{22}$. The three plots also illustrate clearly the persistence of the conditional moments. One year after the shock, the conditional mean and volatility are still 0.22% and 0.68% away from their respective long-term averages, respectively.

The second column illustrates that much of the variation of the conditional volatility is contemporaneously independent of the variation of the conditional mean. A one standard deviation orthogonalized volatility innovation immediately leads to a 3% increase in the annualized volatility. Since the volatility is highly persistent, it returns to its long-term average only about 18 months after the innovation. The lag adjustment of the conditional mean to changes in the volatility is also more pronounced for the orthogonalized volatility innovation because by construction the mean does not react contemporaneously to this innovation. The response of the mean to the orthogonalized volatility innovation is hump-shaped. The mean does not respond contemporaneously but rises immediately after the innovation and peaks six months later at a level of 0.5% above its long-term average. The mean then mean-reverts to its long-term average over the following 30 months. Notice that even after the volatility fully mean-reverts, the mean is still relatively high with a normal level of risk. This results in an above-average Sharpe ratio from 12 through 36 months after the orthogonalized volatility innovation.

An intuitive way to illustrate the role of the lag adjustments in the dynamics of the conditional mean and volatility is to compare the impulse responses of the restricted VAR with $a_{21} = a_{21} = 0$ (dashed lines) to those of the unrestricted VAR (solid lines). Consider first the responses of the restricted VAR to a one standard deviation mean innovation. Although the initial effect of the innovation is virtually the same as in the unrestricted VAR, the mean fully mean-reverts within one year and, more importantly, does not overshoot its long-term average. The reason for this difference in mean responses is that in the restricted VAR, the decrease in volatility (from the mean innovation) does not feed back to the mean in future periods. This lack of feedback is even more transparent for a one standard deviation orthogonalized volatility innovation, where in the restricted VAR the mean *never* responds to the innovation. It is therefore clear from Fig. 4 that the lag adjustments of the mean to changes in volatility and, to a lesser extent, of the volatility to changes in the mean, play an important role in the dynamics of the conditional moments.

4.5. A closer look at the contemporaneous correlation

There is a surprising amount of disagreement about even the sign of the contemporaneous correlation between the conditional mean and volatility of stock returns. Our results contribute to this empirical debate by documenting a strong negative correlation between the innovations to the conditional mean and volatility, which in turn implies negative and substantial volatility-in-mean and mean-in-volatility effects. However, on the face of it, our analysis is subject to the caveat that we model the *log* moments, while most of the literature relates the mean and volatility directly (without logs).

To establish a tighter link to the literature, notice that the log moments $\ln \mu_t$ and $\ln \sigma_t$ are bivariate normally distributed. The moments μ_t and σ_t are therefore bivariate log-normally distributed. It follows from the bivariate log-normal distribution that

$$\text{Corr}[\mu_t, \sigma_t | \mu_{t-1}, \sigma_{t-1}] = \frac{\exp\{\rho\sqrt{b_{11}b_{22}}\} - 1}{\sqrt{\exp\{b_{11}\} - 1}\sqrt{\exp\{b_{22}\} - 1}} \simeq \rho \qquad (39)$$

for sufficiently small b_{11} and b_{22}. In words, we can legitimately approximate the correlation between the conditional mean and volatility by the correlation between the corresponding log moments. Given the parameter estimates in Table 3, the approximation error is only 0.012 (-0.557 with logs versus -0.545 without logs). We conclude that despite the log transformation, our empirical results can be interpreted readily in the context of the literature on the contemporaneous correlation between the conditional mean and volatility.

However, not only can our results be interpreted in the context of this literature, but they actually shed light on *why* there is such disagreement about the sign of the contemporaneous correlation. The key issue is the distinction between the conditional (on $\ln \mu_{t-1}$ and $\ln \sigma_{t-1}$) and unconditional correlation between the moments. The conditional correlation is the same as the correlation between the innovations to the moments and is equal to ρ. The unconditional correlation, in contrast, is computed from the unconditional covariance matrix in Eq. (4). It is clear from this equation that, depending on A, the conditional and unconditional correlations can be quite different.

Fig. 5 illustrates this point by plotting the unconditional cross-autocorrelations (as opposed to the conditional cross-autocorrelations implicit in the impulse response functions) of the mean and volatility implied by the parameter estimates in Table 3, where the ''offset'' represents the number of months the volatility is lagged (negative offset) or led (positive offset) relative to the mean. Each of the unconditional cross-autocorrelations is computed from the corresponding unconditional auto-covariance matrices of the VAR. For example, the unconditional contemporaneous correlation, at an offset of zero, is computed from the unconditional covariance matrix in Eq. (4).

The unconditional cross-correlations between the mean and leads of the volatility of more than 12 months are all negative, with a minimum of -0.09 at an offset of 21 months. In contrast, the correlations are positive for all lagged values of the volatility
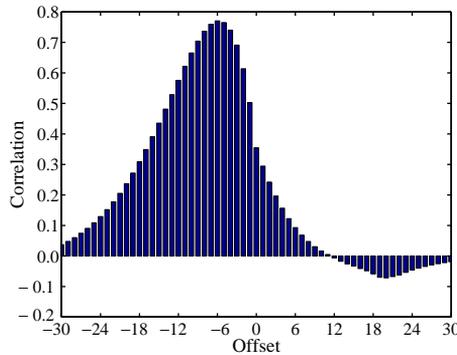
Fig. 5. Cross-autocorrelations. This figure plots the cross-autocorrelations of the conditional mean and volatility, where the offset denotes the number of months the volatility is lagged (negative offset) or led (positive offset) relative to the mean.

as well as for led values up to an offset of 12 months. The correlations decay with both increasingly negative and positive offsets, but the decay is asymmetric due to the differences between the off-diagonal elements of A. The cross-correlations peak at an offset of minus six months with 0.75, which means that unconditionally the volatility leads the mean by six months. It is interesting to note that the average duration of a recession in our sample is approximately nine months. The six-month lead of the volatility therefore seems to relate to the business cycle pattern in the conditional moments that we documented in Section 4.3. The increase in volatility at the beginning of the recession is associated with an (almost certain) increase in the mean approximately six months later, toward the end of the recession.

The most important result for understanding the apparent disagreement about the sign of the contemporaneous correlation is that, although the conditional contemporaneous correlation is large and negative (-0.557), the unconditional contemporaneous correlation (at an offset of zero) is large and *positive* (0.312). This positive unconditional correlation arises both because of the large and positive Granger causality term a_{12}, which implies a large and positive lag-volatility-in-mean effect, and because of the high persistence of the volatility. For example, suppose the volatility increases from a normal 15% to 25% (as the economy enters a recession). Over the following few months the mean increases gradually, reaching its highest level about six months after the increase in volatility. Since the volatility is highly persistent, it is still relatively high (23%, for example) even six months after the initial increase. Therefore, it appears unconditionally that whenever the mean is high the volatility is also high, suggesting a positive correlation between the moments.

This finding implies that the sign and magnitude of the contemporaneous correlation between the conditional mean and volatility depends crucially on whether we measure the conditional or unconditional correlation. It is not surprising then that the results about the contemporaneous correlation in the literature depend on the statistical model and predictors (Harvey, 2001; Koopman and Uspensky, 1999), the return horizon (Harrison and Zhang, 1999), and, most importantly,

whether we measure the correlation in levels or first-differences (French et al., 1987; Whitelaw, 1994).

4.6. Exogenous predictors

On the one hand, the advantage of our latent VAR approach is that it allows us to study the dynamics of the conditional mean and volatility without relying on exogenous predictors. On the other hand, by ignoring the evidence of a correlation between the conditional moments and exogenous predictors, we are potentially discarding some useful information. Discarding information is always undesirable, but it is particularly so in this application because, judging by the vast disagreement in the literature, precisely measuring the relationship between the conditional mean and volatility is difficult.

As a compromise and robustness check, we therefore estimate an extended version of the model in which each moment can also depend on three variables known to be correlated with expected returns and volatility: the short rate, term premium, and default premium (see footnote 7 for references).[9] In particular, we assume that these three variables are truly exogenous and then consider the following model specification:

$$y_t = \bar{\mu}e^{m_{t-1}} + \bar{\sigma}e^{v_{t-1}}\varepsilon_t = \bar{\mu}e^{Z_1's_{t-1}} + \bar{\sigma}e^{Z_2's_{t-1}}\varepsilon_t \quad \text{with } \varepsilon_t \sim \mathrm{N}[0,1] \tag{40}$$

and

$$s_t = Cx_{t-1} + As_{t-1} + \eta_t \quad \text{with } \eta_t \sim \mathrm{MVN}[0,\Sigma], \tag{41}$$

where x_t denotes the *de-meaned* predictors observed at date t.

The extended transition Eq. (41) can be interpreted directly in the context of the predictability (of both the mean and volatility) literature. With $A = 0$, the first equation of the extended VAR resembles a predictive regression (except for the log) and the second equation is an EGARCH model with only exogenous predictors. Without this restriction, both the exogenous predictors x_t and the lagged moments s_{t-1} forecast future realizations of the moments s_t. Furthermore, since the predictors are de-meaned, the coefficient matrix C shows how deviations of the predictive variables from their long-term averages predict deviations of the conditional moments from their long-term averages.

Table 4 reports the estimates of the extended model and also replicates for comparison the results for Model C in Table 3. The estimates of A and Σ are quite similar across the two models. When we add exogenous predictors, the own-lag coefficients a_{11} and a_{22} both increase. The coefficients a_{12} on the volatility in the mean equation and a_{21} on the mean in the volatility equation both decrease in magnitude. The variance of the mean and volatility decrease, which means that the

---

[9] Two common volatility predictors, squared returns (French et al., 1987; Schwert, 1989; Schwert and Seguin, 1991; Andersen et al., 2001) and option-implied volatilities (Christensen and Prabhala, 1998), are not included in this list. Squared returns are already implicitly incorporated in our estimation through the approximating model in Eq. (30) and the availability of option-implied volatilities is too limited to study business cycle patterns in the conditional moments (covering only one business cycle).

Table 4

Parameter estimates with exogenous predictors

This table presents the simulated maximum likelihood estimates of the model:

$$y_t = \bar{\mu}e^{m_{t-1}} + \bar{\sigma}e^{v_{t-1}}\varepsilon_t = \bar{\mu}e^{Z_1' s_{t-1}} + \bar{\sigma}e^{Z_2' s_{t-1}}\varepsilon_t \quad \text{with } \varepsilon_t \sim \text{N}[0, 1]$$

and

$$s_t = Cx_t + As_{t-1} + \eta_t \quad \text{with } \eta_t \sim \text{MVN}[0, \Sigma],$$

where

$$C = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \end{bmatrix}, \quad A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} b_{11} & \rho\sqrt{b_{11}b_{22}} \\ \rho\sqrt{b_{11}b_{22}} & b_{22} \end{bmatrix}, \quad \text{and}$$

$$\text{Corr}[\varepsilon_t, \eta_t] = \begin{bmatrix} 0 \\ \rho_\sigma \end{bmatrix}.$$

The vector of conditioning variables x_t contains the de-meaned short rate, term premium, and default premium. The estimates are for monthly returns on the value-weighted CRSP index in excess of the one-month Treasury bill rate from January 1946 through December 1998. The results for the base model without conditioning variables are reproduced from Table 3.

Parameters	Model C		Extended Model C	
	Estimate	*t*-Stat	Estimate	*t*-Stat
a_{11}	0.8658	11.21	0.8843	11.69
a_{21}	−0.0885	−1.96	−0.0746	−2.91
a_{12}	0.0861	1.01	0.0312	1.40
a_{22}	0.8973	15.32	0.8955	32.47
b_{11}	5.96×10^{-3}	3.19	5.62×10^{-3}	2.23
b_{22}	0.0614	5.55	0.0441	4.47
ρ	−0.5584	−5.80	−0.4512	−2.64
ρ_σ	−0.2541	−4.04	−0.2817	−4.22
$\bar{\mu}$	6.24×10^{-3}	5.11	6.29×10^{-3}	4.48
$\bar{\sigma}$	0.0382	8.79	0.0380	9.87
c_{11}	—	—	−2.6187	−5.22
c_{21}	—	—	0.6550	3.18
c_{12}	—	—	0.3613	2.36
c_{22}	—	—	1.4233	3.04
c_{13}	—	—	2.9969	5.43
c_{23}	—	—	1.6929	5.16
$\ln \mathscr{L}$	1,152.94		1,157.62	

exogenous predictors help explain some of the variation in the moments that was previously left unexplained. Finally, the contemporaneous correlation between the innovations to the mean and volatility decreases in magnitude from −0.557 to −0.451. Together, these differences translate into slightly smaller lag-volatility-in-

mean and lag-mean-in-volatility effects (0.188 versus 0.252 and 1.044 versus 1.471, respectively). However, both effects remain significant at the 1% level.

In the mean equation of the extended VAR, the coefficients on the term premium and default premium (c_{12} and c_{13}) are positive and the coefficient on the short rate (c_{11}) is negative. In the volatility equation, the coefficients on all three predictors (c_{21}, c_{22}, and c_{23}) are positive. The signs of the coefficients are consistent with the results of predictive return and squared return regressions. All estimates are statistically significant at the 5% level and in most cases even at the 1% level. However, it is important to recognize that these results are not comparable to predictive return or squared return regressions. Instead, they correspond to regressions with the *conditional moments* as dependent variables. As we document above, the conditional moments are highly correlated with the predictors (multiple R^2 of 0.32 for the mean and 0.13 for the volatility).

Despite the fact that all of the elements of C are statistically significant, the overall fit of the model does not improve substantially when we add the exogenous predictors. An LR test of the hypothesis $C = 0$ fails to reject with a *p*-value of 0.152. Likewise, a Hausman (1978) test of the hypothesis that the matrix A is the same across the two models fails to reject at conventional significance levels. This failure to reject suggests that the exogenous predictors do not significantly alter the dynamics of the conditional moments.

The plots on the right-hand side of Fig. 3 present estimates of the annualized mean (first row), volatility (second row), and Sharpe ratio (third row), extracted from the extended VAR with exogenous predictors. The time variation of the three moments, especially of the volatility and Sharpe ratio, are remarkably similar across the models with and without exogenous predictors (correlations of 0.73 for the mean, 0.97 for the volatility, and 0.91 for the Sharpe ratio). Specifically, the business cycle pattern in the variation of the conditional moments is qualitatively unaltered by the exogenous predictors. If anything, the extended model further emphasizes the countercyclical variation of the Sharpe ratio. This confirms the statistical evidence above and attests to the robustness of our empirical results.

5. Conclusion

We model the conditional mean and volatility of stock returns as a latent VAR to explore both the contemporaneous and intertemporal relationship between the two moments, as well as the time variation of the implied Sharpe ratio, in a flexible statistical framework and without relying on exogenous predictors. The first equation of the VAR describes the dynamics of the conditional mean and captures the temporary component in the permanent and temporary components model. The second equation describes the dynamics of the conditional volatility and nests the standard SV model. Finally, the implied Sharpe ratio is time-varying in a way that is consistent with a variety of equilibrium asset pricing models.

On the contemporaneous relationship between the conditional mean and volatility, we estimate a strong negative correlation between the innovations to the

two moments, which implies negative and substantial volatility-in-mean and mean-in-volatility effects. On the intertemporal relationship, the reduced-form lag-volatility-in-mean and lag-mean-in-volatility effects are both positive and significant. Together, the contemporaneous and intertemporal relationships generate pronounced countercyclical variation of the Sharpe ratio. *Every* recorded business cycle in the sample is associated with an almost monotonic rise in the Sharpe ratio from the peak to the trough of the cycle. We argue that this countercyclical variation of the Sharpe ratio is consistent with the intuition of habit formation.

We document a systematic pattern in the way the mean and volatility change over the business cycle. Whenever the economy comes off the peak of a business cycle, the conditional volatility rises immediately. The conditional mean, however, increases only gradually as the economy moves from the peak to the trough of the cycle. Thus, the volatility appears to lead the mean (by about six months) through the recession. The mean reaches a high at the trough of the cycle, shortly after which the volatility drops again to a normal level. As a result, the increase in the mean is associated with a subsequent drop in the volatility.

Despite the large and negative contemporaneous correlation between the innovations to the conditional mean and volatility, which measures the conditional (on the lagged mean and volatility) correlation between the moments, the unconditional contemporaneous correlation between the mean and volatility is large and positive due to the strong lag-volatility-in-mean effect. Unconditionally, times of high expected returns are therefore associated with high volatility. We argue that this difference between the conditional and unconditional correlations may explain why there is such disagreement about the sign of the contemporaneous correlation between the two moments.

Appendix A. Choice of approximating model

The measurement Eq. (20) is nonlinear in both m_{t-1} and v_{t-1}. We consider *separate* first-order expansions of this equation around $\mathrm{E}[m_{t-1}] = \mathrm{E}[v_{t-1}] = 0$:

$$y_t = \bar{\mu} + \bar{\sigma} e^{v_{t-1}} \varepsilon_t \tag{A.1}$$

and

$$y_t = \bar{\mu} e^{m_{t-1}} + \bar{\sigma} \varepsilon_t, \tag{A.2}$$

respectively. Both of these equations are still nonlinear in the state variables, but each is only nonlinear in one of the variables. Following convention in the stochastic volatility literature, we log-linearize Eq. (A.1):

$$\tilde{y}_t \equiv \ln(y_t - \bar{\mu})^2 = 2 \ln \bar{\sigma} + 2 v_{t-1} + \ln \varepsilon_t^2, \tag{A.3}$$

where $\ln \varepsilon_t^2$ is log-χ^2 distributed. We further approximate Eq. (A.2) and the log-linear Eq. (A.3) with the following linear equations:

$$y_t = m_{t-1} + w_{1t} \quad \text{with } w_{1t} \sim \mathrm{N}[c_{1t-1}, R_{1t-1}], \tag{A.4}$$

$$\tilde{y}_t = 2v_{t-1} + w_{2t} \quad \text{with } w_{2t} \sim N[c_{2t-1}, R_{2t-1}], \tag{A.5}$$

where w_{1t} and w_{2t} are correlated. Finally, we write these equations in vector form:

$$Y_t \equiv \begin{bmatrix} y_t - c_{1t-1} \\ \tilde{y}_t - c_{2t-1} \end{bmatrix} = H's_{t-1} + w_t \quad \text{with } w_t \sim \text{MVN}[0, R_{t-1}], \tag{A.6}$$

where

$$H = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \quad \text{and} \quad R_t = \begin{bmatrix} R_{1t} & \rho_w \sqrt{R_{1t}R_{2t}} \\ \rho_w \sqrt{R_{1t}R_{2t}} & R_{2t} \end{bmatrix}. \tag{A.7}$$

In the special case $\text{Corr}[\varepsilon_t, \eta_t] = 0$, we can use Eq. (A.6) as the measurement equation and Eq. (21) as the transition equation of the linearized model. In general, the log-linearization of Eq. (A.1) looses all information about the correlation between ε_t and the second element on η_t. However, Harvey and Shephard (1996) show in the context of a more standard stochastic volatility model that this information about the correlation can be recovered from the signs of the return innovations. The following is a straightforward application of Harvey and Shephard's approach to our latent VAR model.

Define $sgn_t = \text{sign}[y_t - \bar{\mu}]$, such that sgn_t is one (minus one) if $y_t - \bar{\mu}$ is positive (negative). Following the steps in Harvey and Shephard (1996), the transition equation conditional on the sign of the returns is:

$$s_t = K\,sgn_t + As_{t-1} + \eta_t^* \tag{A.8}$$

with

$$\begin{bmatrix} w_t \\ \eta_t^* \end{bmatrix} \Bigg| sgn_t \sim \text{MVN} \left[0, \begin{bmatrix} R_{t-1} & G_{t-1} \\ G'_{t-1} & \tilde{\Sigma} \end{bmatrix} \right], \tag{A.9}$$

where R_t is the same as in Eq. (A.7) and

$$K = \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = 0.7979 \begin{bmatrix} \rho_\mu \sqrt{b_{11}} \\ \rho_\sigma \sqrt{b_{22}} \end{bmatrix},$$

$$G_t = \begin{bmatrix} \rho_\mu \sqrt{b_{11}} & \rho_\sigma \sqrt{b_{22}} \\ 1.1061 \rho_\mu \sqrt{b_{11}}\,sgn_t & 1.1061 \rho_\sigma \sqrt{b_{22}}\,sgn_t \end{bmatrix},$$

$$\tilde{\Sigma} = \begin{bmatrix} b_{11} - k_1 & \rho \sqrt{(b_{11} - k_1)(b_{22} - k_2)} \\ \rho \sqrt{(b_{11} - k_1)(b_{22} - k_2)} & b_{22} - k_2 \end{bmatrix}. \tag{A.10}$$

Note that if $\rho_\mu = \rho_\sigma = 0$, we get $K = 0$, $\tilde{\Sigma} = \Sigma$, and $G_t = 0$ for all t.

Appendix B. Calibrating the approximating model

Let $\bar{\theta}_t = E_g[\theta_t]$ be the smoothed value of θ_t for the approximating model obtained through the Kalman filter. Under the assumptions of De Jong and Shephard's (1995)

simulation smoother, simulated values of θ_t are normally distributed around their mean $\bar{\theta}_t$. This suggests that the parameters c_t and R_t of the approximating model should be chosen such that the densities $p(y|\theta, \psi)$ and $g(y|\theta, \psi)$ are as close as possible in the neighborhood of $\bar{\theta}_t$.

To operationalize this idea, we define

$$l(\theta, \psi) \equiv \ln p(y \,|\, \theta, \psi) - \ln g(y \,|\, \theta, \psi)$$

$$= \sum_{t=1}^{T} \ln p(y_t \,|\, s_{t-1}, \psi) - \ln g(y_t \,|\, s_{t-1}, \psi) \tag{B.1}$$

and choose c_t and R_t such that the first- and second-order derivatives of $l(\theta, \psi)$ with respect to s_t are set to zero at $\theta_t = \bar{\theta}_t$. More specifically, we set $\rho_w = 0$ and solve for c_{1t} and R_{1t} from the derivatives of Eqs. (A.2) and (A.4), and for c_{2t} and R_{2t} from the derivatives of Eqs. (A.3) and (A.5).[10]

From Eqs. (A.2) and (A.4) we have

$$p_t^{(m)} \equiv \ln p(y_t \,|\, s_{t-1}, \psi) = -\frac{1}{2}\{\ln 2\pi + 2\ln \bar{\sigma} + \bar{\sigma}^{-2}(y_t - \bar{\mu}e^{m_{t-1}})^2\},$$

$$g_t^{(m)} \equiv \ln g(y_t \,|\, s_{t-1}, \psi) = -\frac{1}{2}\{\ln 2\pi + \ln R_{1t-1} + R_{1t-1}^{-1}(y_t - m_{t-1} - c_{t-1})^2\}.$$

$$\tag{B.2}$$

The first- and second-order derivatives of p_t and g_t with respect to m_{t-1} are

$$\dot{p}_t^{(m)} = \bar{\sigma}^{-2}(y_t - \bar{\mu}e^{m_{t-1}})\bar{\mu}e^{m_{t-1}}, \tag{B.3}$$

$$\dot{g}_t^{(m)} = R_{1t-1}^{-1}(y_t - m_{t-1} - c_{1t-1}), \tag{B.4}$$

$$\ddot{p}_t^{(m)} = \bar{\sigma}^{-2}(y_t - \bar{\mu}e^{m_{t-1}})\bar{\mu}e^{m_{t-1}} - \bar{\sigma}^{-2}(\bar{\mu}e^{m_{t-1}})^2, \tag{B.5}$$

$$\ddot{g}_t^{(m)} = -R_{1t-1}^{-1}. \tag{B.6}$$

Equating the second-order derivatives, $\ddot{p}_t^{(m)} = \ddot{g}_t^{(m)}$, yields

$$R_{1t-1} = -\frac{\bar{\sigma}^2}{(y_t - 2\bar{\mu}e^{m_{t-1}})\bar{\mu}e^{m_{t-1}}}, \tag{B.7}$$

where we take the absolute value of R_{1t} if it is negative. Likewise, equating the first-order derivatives, $\dot{p}_t^{(m)} = \dot{g}_t^{(m)}$, and using Eq. (B.6), we get

$$c_{1t-1} = y_t - m_{t-1} + \frac{(y_t - \bar{\mu}e^{m_{t-1}})}{(y_t - 2\bar{\mu}e^{m_{t-1}})}. \tag{B.8}$$

From Eqs. (A.3) and (A.5) and using the log-χ^2 density, we have

$$\dot{p}_t^{(v)} = \exp\{\tilde{y} - 2\ln \bar{\sigma} - 2v_{t-1}\} - 1, \tag{B.9}$$

$$\dot{g}_t^{(v)} = 2R_{2t-1}^{-1}(\tilde{y}_t - 2v_{t-1} - c_{2t-1}), \tag{B.10}$$

---

[10] The coefficient ρ_w captures the correlation between a normally-distributed and a log-χ^2-distributed random variable. Our Monte Carlo experiments show that this correlation is very small (less than 0.05), and we therefore set $\rho_w = 0$.

$$\ddot{p}_t^{(v)} = -2\exp\{\tilde{y} - 2\ln\bar{\sigma} - 2v_{t-1}\},\tag{B.11}$$

$$\ddot{g}_t^{(v)} = -4R_{2t-1}^{-1},\tag{B.12}$$

and equating the first- and second-order derivatives yields

$$R_{2t-1} = \frac{2}{\exp\{\tilde{y} - 2\ln\bar{\sigma} - 2v_{t-1}\}},\tag{B.13}$$

$$c_{2t-1} = \tilde{y}_t - 2v_{t-1} - \frac{\exp\{\tilde{y} - 2\ln\bar{\sigma} - 2v_{t-1}\} - 1}{\exp\{\tilde{y} - 2\ln\bar{\sigma} - 2v_{t-1}\}}.\tag{B.14}$$

References

Abel, A., 1988. Stock prices under time-varying dividend risk: an exact solution in an infinite-horizon general equilibrium model. Journal of Monetary Economics 22, 375–393.

Alizadeh, S., Brandt, M., Diebold, F., 2002. Range-based estimation of stochastic volatility models. Journal of Finance 57, 1047–1091.

Andersen, T., Bollerslev, T., Diebold, F., Ebens, H., 2001. The distribution of stock return volatility. Journal of Financial Economics 61, 43–76.

Andersen, T., Sorensen, B., 1994. Estimation of a stochastic volatility model: a Monte Carlo study. Journal of Business and Economic Statistics 14, 328–352.

Ang, A., Bekaert, G., 2001. Stock return predictability: is it there? Unpublished working paper, Columbia University.

Bansal, R., Yaron, A., 2003. Risks for the long run: a potential resolution of asset pricing puzzles. Journal of Finance, forthcoming.

Bekaert, G., Grenadier, S., 2001. Stock and bond pricing in an affine economy. Unpublished working paper, Columbia University.

Bekaert, G., Harvey, C., 1995. Time-varying world market integration. Journal of Finance 50, 403–444.

Bollerslev, T., Chou, R., Kroner, K., 1992. ARCH modelling in finance: a selective review of the theory and empirical evidence. Journal of Econometrics 52, 5–59.

Bossaerts, P., Hillion, P., 1999. Implementing statistical criteria to select return forecasting models: what do we learn? Review of Financial Studies 12, 405–428.

Boudoukh, J., Richardson, M., 1993. The statistics of long-horizon regressions. Mathematical Finance 4, 103–120.

Brandt, M., Wang, K., 2003. Time-varying risk aversion and unexpected inflation. Journal of Monetary Economics 50, 1457–1498.

Breen, W., Glosten, L., Jagannathan, R., 1989. Economic significance of predictable variations in stock index returns. Journal of Finance 44, 1177–1189.

Campbell, J., 1987. Stock returns and the term structure. Journal of Financial Economics 18, 373–399.

Campbell, J., Cochrane, J., 1999. By force of habit: a consumption-based explanation of aggregate stock market behavior. Journal of Political Economy 107, 205–251.

Campbell, J., Hentschel, L., 1992. No news is good news: an asymmetric model of changing volatility in stock returns. Journal of Financial Economics 31, 281–318.

Christensen, B., Prabhala, N., 1998. The relation between implied and realized volatility. Journal of Financial Economics 50, 125–150.

Constantinides, G., 1990. Habit formation: a resolution of the equity premium puzzle. Journal of Political Economy 98, 519–543.

Danielsson, J., 1994. Stochastic volatility in asset prices: estimation with simulated maximum likelihood. Journal of Econometrics 64, 375–400.

De Jong, P., Shephard, N., 1995. The simulation smoother for time series models. Biometrika 82, 339–350.

De Santis, G., Gerard, B., 1997. International asset pricing and portfolio diversification with time-varying risk. Journal of Finance 52, 1881–1912.

De Santis, G., Gerard, B., 1998. How big is the premium for currency risk? Journal of Financial Economics 49, 375–412.

Duffie, D., Singleton, K., 1993. Simulated moments estimation of Markov models of asset prices. Econometrica 61, 929–952.

Durbin, J., Koopman, S., 1997. Monte Carlo maximum likelihood estimation for non-Gaussian state space models. Biometrika 84, 669–684.

Fama, E., French, K., 1988. Permanent and temporary components of stock prices. Journal of Political Economy 96, 246–273.

Fama, E., French, K., 1989. Business conditions and expected returns on stocks and bonds. Journal of Financial Economics 25, 23–49.

Fama, E., French, K., 1996. The CAPM is wanted, dead or alive. Journal of Finance 51, 1947–1958.

Ferson, W., Sarkissian, S., Simin, T., 2000. Spurious regressions in financial economics. Unpublished working paper, University of Washington.

French, K., Schwert, W., Stambaugh, R., 1987. Expected stock returns and volatility. Journal of Financial Economics 19, 3–30.

Gallant, R., Tauchen, G., 1989. Seminonparametric estimation of conditionally constrained heterogeneous processes: asset pricing applications. Econometrica 57, 1091–1120.

Gennotte, G., Marsh, T., 1993. Variations in economic uncertainty and risk premiums on capital assets. European Economic Review 37, 1021–1041.

Geweke, J., 1989. Bayesian inference in econometric models using Monte Carlo integration. Econometrica 57, 1317–1339.

Ghysels, E., Harvey, A., Renault, E., 1996. Stochastic volatility. In: Maddala, G., Rao, C. (Eds.), Statistical Methods in Finance, Handbook of Statistics, Vol. 14. North-Holland, Amsterdam, The Netherlands.

Glosten, L., Jagannathan, R., Runkle, D., 1993. On the relation between the expected value and the volatility of the nominal excess returns on stocks. Journal of Finance 48, 1779–1802.

Goyal, A., Welch, I., 2002. Predicting the equity premium with dividend ratios. Unpublished working paper, Emory University.

Harrison, P., Zhang, H., 1999. An investigation of the risk and return relation at long horizon. Review of Economics and Statistics 81, 399–408.

Harvey, A., Ruiz, E., Shephard, N., 1994. Multivariate stochastic variance models. Review of Economic Studies 61, 247–264.

Harvey, A., Shephard, N., 1996. Estimation of an asymmetric stochastic volatility model for asset returns. Journal of Business and Economic Statistics 14, 429–434.

Harvey, C., 2001. The specification of conditional expectations. Journal of Empirical Finance 8, 573–638.

Hausman, J., 1978. Specification tests in econometrics. Econometrica 46, 1251–1272.

Jacquier, E., Polson, N., Rossi, P., 1994. Bayesian analysis of stochastic volatility models. Journal of Business and Economic Statistics 12, 371–417.

Kandel, S., Stambaugh, R., 1991. Asset returns and intertemporal preferences. Journal of Monetary Economics 27, 37–91.

Keim, D., Stambaugh, R., 1986. Predicting returns in the stock and bond markets. Journal of Financial Economics 17, 357–390.

Kim, S., Shephard, N., Chib, S., 1998. Stochastic volatility: likelihood inference and comparison with ARCH models. Review of Economic Studies 65, 361–393.

Koopman, S., Uspensky, E., 1999. The stochastic volatility in mean model: empirical evidence from international stock markets. Unpublished working paper, Free University of Amsterdam.

Lamoureux, C., Zhou, G., 1996. Temporary components of stock returns: what do the data tell us? Review of Financial Studies 9, 1033–1059.

LeRoy, S., Porter, R., 1981. The present-value relation: tests based on implied variance bounds. Econometrica 49, 555–574.

Lettau, M., Ludvigson, S., 2001. Measuring and modelling variation in the risk-return trade-off. Unpublished working paper, NYU.

Nelson, D., 1991. Conditional hetereoskedasticity in asset returns: a new approach. Econometrica 59, 347–370.

Sandmann, G., Koopman, S., 1998. Estimation of stochastic volatility models via Monte Carlo maximum likelihood. Journal of Econometrics 87, 271–301.

Schwert, W., 1989. Why does stock market volatility change over time? Journal of Finance 44, 1115–1153.

Schwert, W., Seguin, P., 1991. Heteroskedasticity in stock returns. Journal of Finance 45, 1129–1155.

Shephard, N., 2000. Comment on Time-series analysis of non-Gaussian observations based on state-space models from both classical and Bayesian perspectives. Journal of the Royal Statistical Society 62, 29–30.

Shephard, N., Pitt, M., 1997. Likelihood analysis of non-Gaussian measurement time series. Biometrika 84, 653–667.

Stambaugh, R., 1999. Predictive regressions. Journal of Financial Economics 54, 375–421.

Valkanov, R., 2002. Long-horizon regressions: theoretical results and applications. Unpublished working paper, UCLA.

Whitelaw, R., 1994. Time variation and covariations in the expectation and volatility of stock market returns. Journal of Finance 49, 515–541.

Whitelaw, R., 2000. Stock market risk and return: an equilibrium approach. Review of Financial Studies 13, 521–547.

Wiggins, J., 1987. Option values under stochastic volatility: theory and empirical estimates. Journal of Financial Economics 19, 351–372.