Intertemporal CAPM and the Cross-Section of Stock Returns

Joseph Chen[†]
University of Southern California

First Version : March 2000
Current Version : June 2003

Working Paper, Comments Welcome

JEL Classifications: C13, C32, G12
Keywords: GARCH, intertemporal CAPM, VAR

[*]A previous version of this paper circulated under the title, "Can the Intertemporal CAPM explain the cross-section of stock returns?" The author would like to thank Geert Bekaert, Harrison Hong, and Ken Singleton, whose consistent guidance and support made this paper possible. I am also indebted to Andrew Ang, Robert Hodrick, Ming Huang, Selo Imrohoroglu, Jun Liu, David Ng, Jun Pan, Lasse Pedersen, and Jeremy Stein for numerous insightful discussions. The author is also grateful of the seminar participants at Cornell University, Harvard University, Massachusetts Institute of Technology, Ohio State University, Stanford University, University of Illinois and the Western Finance Conference for helpful comments. Any errors or omissions are the responsibility of the author.

[†]Marshall School of Business at USC, Hoffman Hall 701, Los Angeles, CA 90089-1427; ph: (213)740-6509; fax: (213)740-6650; e-mail: joe.chen@marshall.usc.edu; WWW: http://www-rcf.usc.edu/∼josephsc.

Intertemporal CAPM and the Cross-Section of Stock Returns

Abstract

This paper develops an asset pricing model based on Merton's (1973) Intertemporal Capital Asset Pricing Model in which an asset earns a risk premium if it performs poorly when prospects of investing for the future turn sour. The model is develop in a framework with time-varying expected market returns and time-varying market volatilities to reflect the changes in the investment opportunity set of the economy. Assuming the pricing kernel depends only on the consumption growth rate and the aggregate market return, Campbell's (1993, 1996) technique of substituting out aggregate consumption using the aggregate budget constraint delivers two key insights. First, the risk premium of a factor ought to be restricted by the amount of information the factor conveys about expected future market returns or volatilities: if a factor provides no information about the future, it should not command any risk premia. Second, risk premia across all factors should be linked to each other through the willingness of investors to bear consumption risk. I model the investment opportunity set using a multivariate VAR-GARCH model with non-Gaussian innovations, and estimate this system to examines if the historically high returns associated with the book-to-market effect can be explained as compensation for risk exposures to changes in the investment opportunity set. The estimates suggest that the forecasting ability of the book-to-market effect is not sufficient to explain its historically high return.

1 Introduction

In this paper, I develop an intertemporal asset pricing model based on Merton's (1973) Intertemporal Capital Asset Pricing Model (ICAPM) in which an asset earns a risk premium if it performs poorly precisely when prospects of investing for the future turn sour–i.e., an adverse change in the investment opportunity set. I estimate this model to examine whether the historically high return associated with the book-to-market effect can be explained as compensation for risk exposures to adverse changes in the investment opportunity set. Unlike most research that aim to explain the book-to-market effect, this paper begins with a pricing kernel that can be derived from the preferences of a representative investor, and contains two components–the consumption growth rate and the market return. I use Campbell's (1993, 1996) method of approximating the aggregate budget constraint to substitute out the consumption growth rate.[1] The model, which uses time-varying expected market returns and time-varying market volatilities to describe the changes in the investment opportunity set, delivers two key insights. First, controlling for the market risk, any factor that reflects the changes in the investment opportunity set ought to forecast future market returns or forecast future market volatilities. Second, risk premia across such factors should be linked to each other through a parameter that can be interpreted as the willingness of investors to bear risk. These observations impose restrictions on factor pricing models that have gone largely unexplored.[2]

Over the last twenty years, a large volume of empirical work has documented a variety of ways in which the cross-section of stock returns can be predicted by publicly available information. One well documented pattern is the book-to-market effect (*bk-mkt*). The effect documented by Rosenberg, Reid, and Lanstein (1985) and Fama and French (1992) has remained rather robust throughout the years. There is a consensus that this predictablility pattern is not explained by the traditional Capital Asset Pricing Model (CAPM).[3] However, there is considerable debate over exactly what underlying economic explanations apply to these patterns. Broadly speaking, there are two competing explanations: the risk-based explanations and the nonrisk-based explanations.

[1] See Restoy and Weil (1998), Hodrick, Ng, and Sengmueller (1999), Campbell and Viceira (2002), and Ng (2002) for other applications of Campbell's (1993, 1996) method.

[2] Studies that do stress these restrictions include Campbell (1996), but he does not allow for time-varying volatilities and does not explore if it explains the book-to-market effect.

[3] In contrast, Ang and Chen (2003) find evidence that the effect is consistent with the CAPM with time-varying risk exposures, except among smaller stocks for which the market may be less efficient.

The proponents of nonrisk-based explanations argue that such predictability patterns reflect market inefficiencies. For instance, Lakonishok, Shleifer, and Vishny (1994) argue that the book-to-market effect reflects investors incorrectly over-extrapolating past earnings growth into the future and over-valuing companies that have performed well in the past. More recently, a number of theory papers argue that in the presence of limits to arbitrage, simple forms of bounded rationality can simultaneously reconcile many such patterns (Barberis, Shleifer, and Vishny, 1998; Daniel, Hirshleifer, and Subramanyam, 1998; Hong and Stein, 1999). Furthermore, Chen, Hong, and Stein (2001, 2002) stress that, in the presence of short-sales constraints, differences of opinion among investors impact asset returns. What these papers have in common is the idea that human limitations influence asset prices when market frictions limit arbitrage forces.

On the other hand, risk-based explanations contend that the book-to-market effect reflects some systematic risk that the CAPM has failed to capture. In particular, many authors have argued that such cross-sectional predictability patterns reflect compensation for risk along the lines of Merton's (1973) ICAPM (for example, Fama and French, 1993, 1995, 1996).[4] Under this reasoning, these effects are interpreted as mimicking portfolios whose returns are correlated with the relevant state variables representing time-variations in the investment opportunity set of the economy. For instance, value stocks may perform poorly during recessionary periods when the overall prospects for the economy in the future is poor.[5]

For the most part, recent empirical research in support of the risk-based explanations has focused on determining how these predictability patterns relate to macroeconomic variables and business cycle fluctuations that are thought to reflect changes in future investment prospects. Perez-Quiros and Timmermann (2000) demonstrate that returns on small firms–which tend to be value firms–are more volatile during economic recessions when investors may be more sensitive to risk. Liew and Vassalou (2000) show that the book-to-market effect help forecast future rates of economic growth. Furthermore, Ferson and Harvey (1999), Lettau and Ludvigson (2001b), and Vassalou (2003) show that accounting for macroeconomic risk reduces the information content of the book-

[4] Lettau and Ludvigson (2001b), Vassalou (2003), and Petkova and Zhang (2003) find that the book-to-market effect is consistent with a conditional version of the CAPM. However, the conditional CAPM is not derived from a dynamically optimizating behavior of a representative agent.

[5] Gomes, Kogan, and Zhang (2002) and Brennan, Wang, and Xia (2003) illustrate such mechanism by using cash-flow valuation models with time-varying investment opportunities.

to-market effect. These studies provide evidence of a link between cross-sectional predictability and forward-looking macroeconomic variables. In addition, Moskowitz (2003) examines the forecasts of future asset covariances using a GARCH model and finds that return on small stocks contain information about future covariances. His work suggests that these effects may also reflect news about future market volatilities.

However, since these studies do not use a theoretical model, they leave unanswered the question of whether their findings are consistent with an underlying economic model–can a rational equilibrium model deliver the magnitude of risk premia associated with these patterns? Fama (1991) and Cochrane (2001) criticize the use of the ICAPM as a "fishing license" to allow a variety of factors into a pricing model without verifying whether such factors are consistent with the optimization behavior of economic agents. Recognizing this problem, two papers that are closest to this paper in spirit, Brennan, Wang, and Xia (2003) and Campbell and Vuolteenaho (2002), build models based on Merton's (1973) ICAPM in which only factors that forecast future investment opportunities are admitted.[6] However, even if a factor model empirically explains the cross-section of stock returns, it does not necessarily tell us much about the underlying economic mechanism that is driving the results.[7] Specifically, even for an economically well-motivated factor, a factor pricing model does not tell us how much factor risk premium we should expect.[8] MacKinlay (1995) suggests that imposing restrictions from a theoretical model helps to protect against admitting wrong models. This paper makes this natural next step.

The goal of this paper is to develop more rigorous empirical tests of the hypothesis that cross-sectional predictability patterns can be explained by an intertemporal CAPM with changes in the investment opportunity set. There is substantial evidence of time-variations in the investment opportunity set of the economy. For instance, the aggregate stock market return has been shown to be forecastable.[9] In particular, the aggregate

[6] Brennan, Wang, and Xia (2003) use factors that forecast real interest rates and the maximal Sharpe ratio, while Campbell and Vuolteenaho (2002) decompose the beta of a stock into one component reflecting news about the market's future cash flows and another component reflecting news about the market's discount rates.

[7] For example, Daniel and Titman (1997) and Ferson, Sarkissian, and Simin (1999) point out that factor models may appear to explain the cross-section of stock returns, even if the true model is based on firm characteristics rather than on economic risk factors within an equilibrium framework.

[8] Tests of the factor pricing methodology, such as the test of Fama and MacBeth (1973), provides estimates of factor risk premia under the assumption that one has the correct model.

[9] Throughout this paper, I use the term "forecastable" to distinguish time-series predictability from cross-sectional predictability.

market return can be forecasted by its own past returns (Fama and French, 1988a; Poterba and Summers, 1988), the short-term interest rate (Campbell, 1991; Hodrick, 1992), the aggregate dividend yield (Campbell and Shiller, 1988; Fama and French 1988b), the term spread between long-term and short-term interest rates (Fama and French, 1989), and the default spread between low-grade and high-grade corporate bonds (Keim and Stambaugh, 1986). Furthermore, Campbell (1987), Harvey (1989), and Glosten, Jagannathan, and Runkle (1993) have documented that some of these same variables also forecast variations in future market volatilities. Therefore, a natural extension of the static CAPM to an intertemporal setting is to incorporate these forecastable time-variations in expected future market returns and future market volatilities.

This paper develops an intertemporal asset pricing model in a framework in which the conditional means and variances of state variables vary across time to reflect changes in the investment opportunity set. The model begins by positing a pricing kernel in which there are two components: the consumption growth rate and the rate of return on aggregate wealth. A special case covered by this model is the case in which there exists a representative agent with the recursive utility function developed by Epstein and Zin (1989). The aggregate budget constraint imposes restrictions under which a priced asset must either covary with 1) the market return, 2) the changes in the forecasts of future market returns, or 3) the changes in the forecasts of future market volatilities. The basic intuition is that an unexpectedly high level of consumption ought to reflect either an improvement in the forecasts of future market returns or a reduction in the precautionary savings held as a hedge against future uncertainty. The model delivers restrictions under which the risk premia associated with the state variables are linked to each other by the information they convey about the future and the willingness of investors to bear risk. Moreover, the derivation shows that time-variations in market volatility price only those assets that are conditionally co-skewed with the market. This model is estimated to see if the predictability patterns in the cross-section of stock returns reflect these changes in the investment opportunity set.

Many equilibrium models proceed by parameterizing and calibrating the consumption process or the dividend process in the economy. In contrast, this paper infers the consumption process from the data. I estimate the returns process using a vector autoregression (VAR) system to describe the conditional means and a multivariate GARCH model to describe the conditional variances. The consumption process is backed out from the returns process using the aggregate budget constraint. The estimates of this model indicate that the book-to-market effect cannot be explained as as a factor that conveys information about future market returns. Using data covering April 1953 to December 1999, the evidence of forecasting information in the book-to-market effect is small. Moreover, when one imposes the constraint that the risk premia are linked to each other by risk-aversion, the hypothesis that the book-to-market effect is consistent with this intertemporal asset pricing model is rejected.

The results of this paper are most closely related to papers that try to disentangle risk-based explanations from nonrisk-based explanations. MacKinlay (1995) points out that the Sharpe ratios implied by these cross-sectional predictability results appear to be too high. Kirby (1998) notes that the predictability patterns across size portfolios are not inconsistent with circumstances in which time-series forecastability is implied by a pricing model. Lakonishok, Shleifer, and Vishny (1994) point out that value stocks exhibit patterns of investor over-reaction, but find little evidence that value stocks are fundamentally riskier. This paper finds that information content in the cross-sectional predictability of asset returns is not linked to the time-series forecastability of aggregate market returns.

The rest of this paper is organized as follows. Section 2 presents the intertemporal asset pricing model. Section 3 discusses the empirical methods. Section 4 contains a description of the data. Section 5 shows the estimates of the model and analyzes the results. Section 6 offers concluding remarks.

2 An intertemporal asset pricing model

Fama (1970) points out that the single-period CAPM does not apply in a multi-period setting if investor preferences change across time or if the available investment opportunity set changes across time. Merton (1973) develops an intertemporal asset pricing model in which the changes in the investment opportunity set affect future asset returns, which in turn affect consumption. The aspects of the investment opportunity set I study in this paper are the changes in the forecasts of future market returns and the changes in the forecasts of future market volatilities. Given the evidence of time-variation in expected market return and volatility, it seems natural to study their asset pricing implications and see if it explains the book-to-market effect. In order to place some structure on the economy, an underlying economic mechanism I impose is that these variations in the investment opportunity set must all eventually affect consumption at some horizon because the aggregate budget constraint must hold. An increase in expected future return allows investors to consume more today through intertemporal consumption smoothing. In addition, a decrease in future volatilities allows investors to consume more today by reducing their precautionary savings. In an intertemporal setting, a given state variable can earn a risk premium by conveying news regarding changes in the forecasts of future market returns or volatilities. However, the fundamental risk in this setting is consumption risk.

Rather than using changes in the investment opportunity set, one could directly use consumption growth rates to price assets (eg. Hansen and Singleton, 1982). However, consumption may be difficult to measure in practice. To circumvent this difficulty, I follow Campbell's (1993, 1996) method of using the aggregate budget constraint to substitute out the consumption growth rate with current and future market returns. Unfortunately, Campbell's model assumes constant market volatility which is inconsistent with time-variation in expected market return. I extend his work by allowing for time-variations in market volatility and asset covariances. These extensions allow the forecastability of the market to be present in the model and also allow asset risk exposures to be time-varying. After the substitution, the model reveals three sources of risk: the market return, changes in the forecasts of future market returns, and changes in the forecasts of future market volatilities.

To set the stage, consider an economy with rational investors and perfect capital markets in which there are no transaction costs, no short-sales constraints, or any other market imperfections. Assuming further that the law of one price holds, then for every discrete time-period t, there exists a pricing kernel M_{t+1} such that the following Euler equation must hold for any security i with a total rate of return from time t to $t+1$ of $R_{i,t+1}$:

$$E_t[M_{t+1}(1 + R_{i,t+1})] = 1. \tag{1}$$

Assume further that there exists a riskless asset, and let R_i^e denote asset i's return in excess of the risk-free rate. By using a second-order Taylor expansion, the following log-Euler equation must hold:

$$E_t[r_{i,t+1}^e] + \frac{V_{ii,t}}{2} = \text{Cov}_t(r_{i,t+1}^e, -m_{t+1}), \tag{2}$$

where $V_{ii,t} \equiv \text{Var}_t(r^e_{i,t+1})$ and m_{t+1} is the logarithm of M_{t+1}.[10]

Now, suppose that the aggregate market portfolio containing all wealth is tradable.[11] Let W_t denote the level of aggregate wealth at the beginning of time period t, let C_t be the time t consumption level, and let $R_{W,t+1}$ be the total real rate of return on invested wealth (the real market return). Finally, suppose that the pricing kernel can be modeled as

$$M_{t+1} = \beta^{\theta_1} \left(\frac{C_{t+1}}{C_t} \right)^{\frac{-\theta_2}{\sigma}} (R_{W,t+1})^{\theta_3 - 1}, \tag{3}$$

for parameters β, σ, θ_1, θ_2, and θ_3.

The pricing kernel described by Eq. (3) depends only on the consumption growth rate and the aggregate market return. This specification includes Epstein and Zin's (1989) recursive utility function representative agent as a special case. The pricing kernel implied by their recursive utility function takes $\theta_1 \equiv \theta_2 \equiv \theta_3 = \theta$, σ as the elasticity of intertemporal substitution, and β as the subjective discount factor. The power utility function is a particularly special case in which, in addition, $\theta = 1$. In all cases, however, the parameter of interest turns out to be a single parameter, $\gamma = \frac{\theta_2}{\sigma} - \theta_3 + 1$.

The link between consumption and the changes in the investment opportunity set is provided by substituting out the consumption growth rate using the aggregate budget constraint. The intuition behind this substitution is that an increased level of consumption today must be financed by a high current market return, an increased forecast of future market returns (and a higher expected future income), or a lower expectation of future market volatilities (and lower precautionary savings). Let c_t be the logarithm of the real aggregate consumption, let w_t be the logarithm of the real level of aggregate wealth, and

[10] Under an additional assumption that the distribution of M_t and $R_{i,t}$ is conditionally jointly log-normal, this relationship is exact. However, without log-normality, higher-order Taylor expansions would reveal omitted terms that depend on third- or higher-order moments. Harvey and Siddique (2000) and Dittmar (2002) explore the implications for asset prices when these higher-order moments are introduced. However, Ang, Chen, and Xing (2002) argue that these Taylor expansions do not adequately capture deviations from Eq. (2). Since the focus here is on whether changes in the investment opportunity set can explain the cross-section of stock returns, rather than on how much precision is needed in the approximation, this paper begins with the second-order Taylor expansion of Eq. (2).

[11] I create a proxy for the market portfolio using a stock market index. However, Roll (1977) has provided the critique that this proxy may not adequately reflect the true market portfolio since it does not include other forms of wealth, such as human capital. Campbell (1996) and Jagannathan and Wang (1996) attempt to alleviate this concern by incorporating labor income growth rates into the proxy for the market portfolio. An earlier version of this paper also incorporated labor income growth, but it did not produce markedly different results.

let $r_{w,t}$ be the real log-return on the market portfolio. Define $\rho \equiv 1 - \exp(E[c_t - w_t])$, where $E[c_t - w_t]$ is the steady-state log consumption-to-wealth ratio. The parameter ρ is the steady-state ratio of aggregate invested wealth to aggregate total wealth.

As shown in Appendix A, a log-linear approximation of the aggregate budget constraint gives three factors. The first factor is the market return, $r_{w,t}$. The second is the change in the forecasts of future market returns, $r_{h,t}$, given by

$$r_{h,t+1} \equiv (E_{t+1} - E_t) \sum_{s=1}^{\infty} \rho^s r_{w,t+s+1}. \tag{4}$$

Note that $r_{h,t}$ is the change in the expectation of exponentially weighted sums of future market returns, where the weights are powers of ρ. The third factor, $r_{\nu,t+1}$, is the change in the exponentially weighted forecasts of future market variances:

$$r_{\nu,t+1} \equiv (E_{t+1} - E_t) \sum_{s=1}^{\infty} \rho^s \text{Var}_{t+s}[r_{w,t+s+1} + r_{h,t+s+1}]. \tag{5}$$

Let $\gamma = \frac{\theta_2}{\sigma} - \theta_3 + 1$. Appendix A shows that the conditional asset pricing relation derived from the pricing kernel in Eq. (3) is

$$\begin{aligned} E_t[r^e_{i,t+1}] + \frac{V_{ii,t}}{2} &= \gamma V_{iw,t} + (\gamma - 1) V_{ih,t} - \frac{1}{2}(\gamma - 1)^2 V_{i\nu,t}, \qquad (6)\\ \text{where} \quad V_{iw,t} &\equiv \text{Cov}_t(r^e_{i,t+1}, r_{w,t+1}),\\ V_{ih,t} &\equiv \text{Cov}_t(r^e_{i,t+1}, r_{h,t+1}),\\ \text{and} \quad V_{i\nu,t} &\equiv \text{Cov}_t(r^e_{i,t+1}, r_{\nu,t+1}). \end{aligned}$$

This result is an extension of Campbell (1993). In the special case of the recursive utility representative agent (and the power utility representative agent), the parameter γ corresponds to the coefficient of relative risk-aversion, hence I interpret this parameter as the willingness of economic agents to bear consumption risk. Even though the consumption growth rate and the aggregate market return may each carry different coefficients in the pricing kernel, once the aggregate budget constraint is imposed, a degree of freedom is lost and only one parameter, γ, enters the asset pricing relation. In the recursive utility function case, the other parameters of the model describe the subjective discount factor and the elasticity of intertemporal substitution. These parameters affect the risk-free rate and the consumption choice, but not the cross-section of asset returns.

Eq. (6) illustrates the intuition behind Merton's (1973) ICAPM. The term $\frac{V_{ii,t}}{2}$ is the Jensen's inequality term that arises when log-returns rather than simple returns are used. In this model, the sources of risk premia for an asset are the contemporaneous conditional covariances of its return with 1) the market, $V_{iw,t}$, 2) the changes in the forecasts of future market returns, $V_{ih,t}$, and 3) the changes in the forecasts of future market volatilities, $V_{i\nu,t}$. The term $V_{iw,t}$ is the time-varying beta. From this term alone, an asset's exposure to the current market return carries a price of risk equal to the risk-aversion parameter, γ. This term corresponds to the single-period conditional consumption CAPM component.

By allowing the expected return of the market to vary across time, a new source of risk is introduced: changes in the forecasts of future market returns. The factor $r_{h,t+1}$ reflects these changes in the forecasts, while the term $V_{ih,t}$ is an asset's covariation with this factor. Since future market returns are linked to consumption by the aggregate budget constraint, the coefficient on this term also depends upon the risk-aversion parameter, γ. One special case is that of a log-utility representative agent with $\gamma \equiv 1$, in which time-variations in the expected returns yield no additional risk premia, since investors behave myopically. However, when $\gamma > 1$, the representative agent demands compensation for holding assets that performs poorly precisely when the forecasts for future market returns fall.

By allowing the volatility of the market to vary across time, a source of risk not found in Campbell (1993) is revealed. For all values of $\gamma \neq 1$, the representative agent demands compensation for the risk that an asset will perform poorly precisely when the future becomes less certain. Increased uncertainty induces risk-averse investors to reduce their current consumption in order to increase their precautionary savings. The factor $r_{h,t+1}$ summarizes the changes in the uncertainty of future returns, while the term $V_{i\nu,t}$ reflects an asset's covariation with these changes. This term carries a negative coefficient of $\frac{-1}{2}(\gamma - 1)^2$, which also depends upon the risk-aversion parameter through the budget constraint. Interestingly, as long as the pricing kernel consists of consumption growth rates and aggregate market returns, the aggregate budget constraint reduces the degree of freedom in the asset pricing relation to a single parameter.

3 Empirical methodology

The model derived in the previous section illustrates that, with changes in the investment opportunity set, assets may be priced if they covary with 1) the real rate of market return, 2) the changes in the forecasts of future market returns, or 3) the changes in the forecasts of future market volatilities. The empirical literature on the forecastability of the market and the literature on the time-varying volatility provide evidence that there are state variables that reflect these changes. I estimate and examine the state variable dynamics using a VAR-GARCH model to describe the patterns of the conditional means and the conditional variances found in the data.[12] The risk-aversion parameter, γ, is estimated from the restriction that the model should price the market portfolio. Implicitly, there also is a consumption process that is derived from the asset returns through the aggregate budget constraint.

3.1 Econometric specification

To obtain forecasts of future market returns, I follow Campbell (1996) and Hodrick, Ng, and Sengmueller (1999) and use a first-order VAR process to describe the conditional means of the state variables.[13] By inferring the dynamics of the economy from the data, I abstract away from explaining how the time-variations in the conditional distributions of asset returns are formed in the economy in the first place. The elements of the VAR system are the state variables that reflect news about the current real market return, forecasts of future market returns, or forecasts of future volatilities. I also estimate a multivariate GARCH(1,1) process to describe the dynamics of market volatility. This VAR-GARCH structure specifies the first two conditional moments, but it does not assume that the innovations are conditionally normally distributed. Appendix B details the implications for the asset pricing relation of the model developed in Section 2 when the state variables follow such VAR-GARCH structure.

I proxy for the real rate of return on the market portfolio using the real rate of return on the aggregate financial wealth portfolio. In the econometric system, the real market return is constructed from two observable variables. Let $\mathbf{f_t}$ denote a K-element vector

[12] Certainly, there are alternative ways of modelling conditional means and variances, which I leave for future research. However, modelling the conditional means and variances with a VAR-GARCH model turns out to be particularly tractable.

[13] I conduct specification tests for the VAR order in Section 5.1.

of all variables. Let the first element of $\mathbf{f_t}$ be the logarithm of the rate of return on a stock market index in excess of the risk-free rate. Let the second element of $\mathbf{f_t}$ be the logarithm of the risk-free rate minus the rate of inflation. The sum of the first and the second elements of $\mathbf{f_t}$ is a proxy for the real rate of return on the market, $r_{w,t}$.

Let the remaining elements in $\mathbf{f_t}$ be the assets to be priced or state variables that forecast future market returns or volatilities. The asset to be priced is the portfolio representing the book-to-market effect (*bk-mkt*). Note that this variable not just an asset to be priced, but that it may contain some additional orthogonal information about the future investment opportunity set. Other state variables are macroeconomic variables that prior research has found to have the ability to forecast future market returns: the dividend yield, the term spread, and the default spread.[14] The assets to be priced may also covary with these state variables. Therefore,

$$\mathbf{f_t} = \begin{pmatrix} \{\text{excess stock market return}\} \\ \{\text{risk-free rate minus inflation rate}\} \\ \{\text{book-to-market effect}\} \\ \{\text{dividend yield}\} \\ \{\text{term spread}\} \\ \{\text{default spread}\} \end{pmatrix}. \tag{7}$$

Denote $\mathbf{ei}$ as the K-dimensional vector for which the i-th element is one and the remaining elements are zeros: eg. $\mathbf{e1} \equiv (1, 0, \ldots, 0)'$. The real rate of market return can then be expressed as

$$r_{w,t} = (\mathbf{e1} + \mathbf{e2})'\mathbf{f_t}. \tag{8}$$

Even though the economic agents' forecasts of future returns are not directly observable, a VAR provides an estimate of such forecasts. Consider a first-order K-dimensional VAR system defined by

$$(\mathbf{f_{t+1}} - \mu) = \mathbf{A}(\mathbf{f_t} - \mu) + \epsilon_{\mathbf{t+1}}, \tag{9}$$

where $\mu \equiv E[\mathbf{f_t}]$, $\mathbf{A}$ is a $K \times K$ parameter matrix, and $\epsilon_{\mathbf{t+1}}$ is a K-dimensional vector of innovations in the state variables. To describe the dynamics of the second-order moments, let $\mathbf{V}_t$ denote the variance-covariance matrix of next-period innovations, ϵ_{t+1}, conditional

[14] The relative Treasury bill rate, which is the T-bill rate minus its one-year backward moving average, has been found by Campbell (1991) and Hodrick (1992) to forecast future returns. I do not include the relative Treasury bill rate, since the real rate of return on the T-bill is already included.

on the information set at time t: $\mathbf{V}_t = E_t(\epsilon_{t+1}\epsilon'_{t+1})$. $\mathbf{V}_t$ may be time-varying, and its dynamics are specified below. While $\epsilon_{\mathbf{t+1}}$ is mean zero, it need not be conditionally normal.

Under the VAR, estimates for the left-hand side of the asset pricing relation given in Eq. (6) can be readily obtained. Suppose that the asset to be priced is the i-th element in $\mathbf{f_t}$ and its excess return is denoted $r^e_{i,t+1}$. Then, from Eq. (9),

$$E_t[r^e_{i,t+1}] = (\mu + \mathbf{A}(\mathbf{f_t} - \mu))'\mathbf{ei}. \tag{10}$$

The Jensen's inequality term is given by

$$\frac{V_{ii,t}}{2} = \frac{1}{2}\mathbf{ei}'\mathbf{V}_t\mathbf{ei}. \tag{11}$$

The VAR also provides an estimate of the first term on the right-hand side of Eq. (6) as

$$V_{iw,t} = (\mathbf{e1} + \mathbf{e2})'\mathbf{V}_t\mathbf{ei}. \tag{12}$$

Turning to the second term in Eq. (6), under the VAR, the forecasts at time t of $\mathbf{f_{t+s}}$ are represented by

$$E_t[\mathbf{f_{t+s}} - \mu] = \mathbf{A}^s(\mathbf{f_t} - \mu). \tag{13}$$

Therefore, the change in the forecasts at time $(t+1)$ of $\mathbf{f_{t+s+1}}$ due to the shock $\epsilon_{\mathbf{t+1}}$ is

$$(E_{t+1} - E_t)\mathbf{f_{t+s+1}} = \mathbf{A^s}\epsilon_{\mathbf{t+1}}, \tag{14}$$

which is the *impulse-response* of $\mathbf{f_{t+s+1}}$ to the innovation $\epsilon_{\mathbf{t+1}}$. Therefore, the change in the s-period ahead forecast of the real market return can be written as

$$(E_{t+1} - E_t)r_{w,t+s+1} = (E_{t+1} - E_t)(\mathbf{e1} + \mathbf{e2})'\mathbf{f_{t+s+1}} = (\mathbf{e1} + \mathbf{e2})'\mathbf{A^s}\epsilon_{\mathbf{t+1}}. \tag{15}$$

Note that

$$\begin{aligned} r_{h,t+1} &= (E_{t+1} - E_t)\sum_{s=1}^{\infty}\rho^s r_{w,t+s+1}, \\ &= (\mathbf{e1} + \mathbf{e2})'\sum_{s=1}^{\infty}\rho^s\mathbf{A}^s\epsilon_{\mathbf{t+1}}, \\ &= (\mathbf{e1} + \mathbf{e2})'\rho\mathbf{A}(\mathbf{I} - \rho\mathbf{A})^{-1}\epsilon_{\mathbf{t+1}} = \lambda_{\mathbf{h}}{}'\epsilon_{\mathbf{t+1}}, \end{aligned} \tag{16}$$

where $\lambda_{\mathbf{h}}{}'$ is defined as $(\mathbf{e1} + \mathbf{e2})'\rho\mathbf{A}(\mathbf{I} - \rho\mathbf{A})^{-1}$. Finally, note that

$$V_{ih,t} = \lambda_{\mathbf{h}}{}'\mathbf{V}_t\mathbf{ei}, \tag{17}$$

which may vary across time as $\mathbf{V}_t$ varies.

Eqs. (11), (12), and (17) depend upon the conditional variance-covariance matrix $\mathbf{V}_t$. I assume that $\mathbf{V}_t$ follows a GARCH(1,1) process with the BEKK representation of Engle and Kroner (1995):

$$\mathbf{V}_t = \mathbf{P}\mathbf{P}' + \boldsymbol{\Delta}\mathbf{V}_{\mathbf{t-1}}\boldsymbol{\Delta}' + \boldsymbol{\Psi}\epsilon_{\mathbf{t-1}}\epsilon'_{\mathbf{t-1}}\boldsymbol{\Psi}', \tag{18}$$

where $\mathbf{P}$ is a lower-triangular K-dimensional matrix and $\boldsymbol{\Delta}$ and $\boldsymbol{\Psi}$ are $K \times K$ dimensional matrices.

In a multivariate setting, restrictions are usually imposed on the parameters of the GARCH process in order to prevent over-parameterization. The parameters of the multivariate GARCH are $\mathbf{P}$, $\boldsymbol{\Delta}$, and $\boldsymbol{\Theta}$. The only restriction imposed on $\mathbf{P}$ is that it must be a lower-triangular matrix. For the matrices $\boldsymbol{\Delta}$ and $\boldsymbol{\Theta}$, I impose the following elements to be zeros:

$$\boldsymbol{\Delta} = \begin{pmatrix} \delta_{11} & 0 & 0 & 0 & 0 & 0 \\ \delta_{21} & \delta_{22} & 0 & 0 & 0 & 0 \\ \delta_{31} & 0 & \delta_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \boldsymbol{\Theta} = \begin{pmatrix} \theta_{11} & 0 & 0 & 0 & 0 & 0 \\ \theta_{21} & \theta_{22} & 0 & 0 & 0 & 0 \\ \theta_{31} & 0 & \theta_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{19}$$

This specification is similar to the structure in Bekaert and Wu (2000) in which the conditional variance of the excess market return depends only upon the past variance and innovations of itself. However, the conditional variance (and covariances) of the real risk-free rate, and *bk-mkt* may depend upon their own past variances and innovations, as well as upon the past variances, the past covariances, and the past innovations of the market. Since there is no clear evidence in the literature of persistence in shocks to volatilities of the dividend yield, the term spread, or the default spread, I assume that there is no GARCH structure associated with the last three state variables. On the other hand, there are no restrictions imposed on the parameters of the VAR structure, μ and $\mathbf{A}$. Hence, there are 6 parameters in μ, 36 parameters in $\mathbf{A}$, 21 parameters in $\mathbf{P}$, 5 parameters in $\boldsymbol{\Delta}$, and 5 parameters in $\boldsymbol{\Theta}$, for a total of 73 parameters in this six-dimensional system[15] .

[15] During the estimation, the system has a total of 561 observations with six variables to identify 73 parameters in the VAR-GARCH models.

With $\mathbf{V}_t$ specified as in Eq. (18), the final term on the right-hand side of Eq. (6) can be calculated. This term is $V_{i\nu,t}$, where $r_{\nu,t} = (E_{t+1} - E_t) \sum_{s=1}^{\infty} \rho^s \text{Var}_{t+s}[r_{w,t+s+1} + r_{h,t+s+1}]$. Appendix B shows that under a GARCH model,

$$V_{i\nu,t} \equiv V_{i\nu} = (\mathbf{e1} + \mathbf{e2} + \lambda_{\mathbf{h}})' S_{i\mathbf{V},t} (\mathbf{e1} + \mathbf{e2} + \lambda_{\mathbf{h}}), \tag{20}$$

where

$$S_{i\mathbf{V},t} = \text{vec}^{-1} \left[(\mathbf{I} - \rho(\mathbf{\Delta} + \mathbf{\Psi}) \otimes (\mathbf{\Delta} + \mathbf{\Psi}))^{-1} \text{vec}(\mathbf{\Psi} E_t[\epsilon^e_{i,t+1} \epsilon_{\mathbf{t+1}} \epsilon_{\mathbf{t+1}}{}'] \mathbf{\Psi}') \right].$$

The operator $\otimes$ denotes the Kronecker product, vec^{-1} denotes the inverse of the *vec* operator[16], and $\epsilon^e_{i,t+1} = r^e_{i,t+1} - E_t[r^e_{i,t+1}]$.

The term $S_{i\mathbf{V},t}$ reflects the conditional covariance between the return on asset i and the changes in the forecasts of future $\mathbf{V}_{t+s}$. The term $V_{i\nu,t}$ captures the part of $S_{i\mathbf{V},t}$ that relates to future market volatilities. Note that these terms depend upon $E_t[\epsilon^e_{i,t+1} \epsilon_{\mathbf{t+1}} \epsilon_{\mathbf{t+1}}{}']$, which is a matrix of conditional co-skewness between $r^e_{i,t+1}$ and the elements of $\mathbf{f_t}$. The elements of this matrix are all zeros if the innovations are conditionally normally distributed, an assumption that is not made here.[17]

Combining these results, under the distributional assumption of a VAR-GARCH, the asset pricing relation from Eq. (6) can be rewritten as

$$E_t[r^e_{i,t+1}] + \frac{V_{ii,t}}{2} = \gamma V_{iw,t} + (\gamma - 1) V_{ih,t} - \frac{1}{2}(\gamma - 1)^2 V_{i\nu,t}, \tag{21}$$

where

$$E_t[r^e_{i,t+1}] = (\mu + \mathbf{A}(\mathbf{f_t} - \mu))' \mathbf{ei},$$

$$V_{ii,t} = \mathbf{ei}' \mathbf{V}_t \mathbf{ei},$$

$$V_{iw,t} = (\mathbf{e1} + \mathbf{e2})' \mathbf{V}_t \mathbf{ei},$$

$$V_{ih,t} = \lambda_{\mathbf{h}}{}' \mathbf{V}_t \mathbf{ei},$$

and

$$V_{i\nu,t} = (\mathbf{e1} + \mathbf{e2} + \lambda_{\mathbf{h}})' S_{i\mathbf{V},t} (\mathbf{e1} + \mathbf{e2} + \lambda_{\mathbf{h}}).$$

The term $\lambda_{\mathbf{h}}{}' = (\mathbf{e1} + \mathbf{e2})' \rho \mathbf{A} (\mathbf{I} - \rho \mathbf{A})^{-1}$, and $S_{i\mathbf{V},t}$ is defined by Eq. (21).

The implicit consumption process can also be derived under the VAR-GARCH. From

[16] If $\mathbf{X}$ is an $m \times n$ matrix, vec($\mathbf{X}$) is an $mn \times 1$ vector obtained by stacking the columns of $\mathbf{X}$.

[17] Restoy and Weil (1998) point out that in their AR-GARCH settings, returns carry information about the conditional means, while only the squares of returns carry information about the conditional variances.

Eq. (A.17) of Appendix A, the aggregate budget constraint yields

$$\begin{aligned}\Delta c_{t+1} - E_t[\Delta c_{t+1}] &= (\mathbf{e1}+\mathbf{e2})'\epsilon_{\mathbf{t+1}} + (1-\sigma\frac{\theta_3}{\theta_2})\lambda_{\mathbf{h}}{}'\epsilon_{\mathbf{t+1}} - \frac{1}{2}(\gamma-1)(1-\sigma\frac{\theta_3}{\theta_2})r_{\nu,t+1}, \\ \text{where} & \\ r_{\nu,t+1} &= \sum_{s=1}^{\infty}\rho^s(\Delta+\Psi)^{s-1}\Psi(\epsilon_{t+1}\epsilon'_{t+1}-\mathbf{V}_t)\Psi'(\Delta'+\Psi')^{s-1}.\end{aligned} \tag{22}$$

Eq. (22) decomposes the unexpected shocks to the growth rate of real consumption, Δc_{t+1}. The first term is the impact on consumption due to unexpected real market returns. The second term relates to the news about future income and represents changes in consumption due to consumption smoothing. The final term relates to the news about future uncertainty and represents the changes due to changes in precautionary savings. The term $r_{\nu,t+1}$ is the weighted sum of the changes in the forecasts of future volatility under the GARCH structure. In the log-utility representative agent case where $\gamma = 1$ (or equivalently $\sigma\frac{\theta_3}{\theta_2} = 1$), Eq. (22) shows that economic agents immediately consume all unexpected real increase in aggregate wealth.

3.2 Estimating and testing the model

I test the model by looking at an unconditional implication of Eq. (21):

$$E[r_i^e] + \frac{V_{ii}}{2} = \gamma V_{iw} + (\gamma-1)V_{ih} - \frac{1}{2}(\gamma-1)^2 V_{i\nu}, \tag{23}$$

$$\begin{aligned}\text{where} \quad E[r_i^e] &= \mu'\mathbf{ei}, \\ V_{ii} &= \mathbf{ei}'\mathbf{V}\mathbf{ei}, \\ V_{iw} &= (\mathbf{e1}+\mathbf{e2})'\mathbf{V}\mathbf{ei}, \\ V_{ih} &= \lambda_{\mathbf{h}}{}'\mathbf{V}\mathbf{ei}, \\ V_{i\nu} &= (\mathbf{e1}+\mathbf{e2}+\lambda_{\mathbf{h}})'S_{i\mathbf{V}}(\mathbf{e1}+\mathbf{e2}+\lambda_{\mathbf{h}}), \\ \text{and} \quad S_{i\mathbf{V}} &= \text{vec}^{-1}\left[(\mathbf{I}-\rho(\mathbf{\Delta}+\mathbf{\Psi})\otimes(\mathbf{\Delta}+\mathbf{\Psi}))^{-1}\text{vec}(\mathbf{\Psi}E[\epsilon_i^e\epsilon\epsilon']\mathbf{\Psi}')\right],\end{aligned}$$

where $\mathbf{V} \equiv E[\mathbf{V_{ff,t}}]$ denotes the time-series average of the conditional covariance matrix of $\mathbf{f_t}$ and $E[\epsilon_i^e\epsilon\epsilon']$ denotes the time-series average of the conditional co-skewness of $r_{i,t}^e$

with the elements of $\mathbf{f_t}$.[18] The term $\mathbf{V}$ is estimated using[19]

$$\mathbf{V} \equiv E[\mathbf{V_{ff,t}}] = \text{vec}^{-1}\left[(\mathbf{I} - \rho(\mathbf{\Delta} + \mathbf{\Psi}) \otimes (\mathbf{\Delta} + \mathbf{\Psi}))^{-1}\text{vec}(\mathbf{P}\mathbf{P}')\right]. \tag{24}$$

The term $E[\epsilon_i^e \epsilon\epsilon']$ is estimated as follows. Letting $\hat{\Xi}$ denote the vector of all VAR parameter estimates, the residuals are estimated by inverting Eq. (9) as,

$$\hat{\epsilon}_{\mathbf{t+1}}(\hat{\Xi}) = (\mathbf{f_{t+1}} - \hat{\mu}) - \hat{\mathbf{A}}(\mathbf{f_t} - \hat{\mu}). \tag{25}$$

The i, j, k-th co-skewness, $s_{ijk} = E[\epsilon_{i,t}\epsilon_{j,t}\epsilon_{k,t}]$, is estimated as

$$\hat{s}_{ijk}(\hat{\Xi}) = \frac{1}{T}\sum_{1}^{T} \epsilon_{i,t}(\hat{\Xi})\epsilon_{j,t}(\hat{\Xi})\epsilon_{k,t}(\hat{\Xi}). \tag{26}$$

Defined this way, these skewness statistics are functions of the VAR parameter estimates.

The asset pricing relation described by Eq. (23) depends upon the parameters of the VAR-GARCH process, the unconditional co-skewness $E[\epsilon_i^e \epsilon\epsilon']$, the approximation constant ρ, and the risk-aversion parameter γ. This model is estimated as follows. First, the parameters of VAR-GARCH are estimated by Quasi-Maximum Likelihood Estimation (QMLE). Given QMLE estimates of the VAR-GARCH process, the estimates of co-skewness are obtained from the residuals. As in other studies that use log-linear approximations of the representative agent's budget constraint (Campbell 1996; Hodrick, Ng, and Sengmueller 1999), the approximation constant is exogenously specified. Given these parameter estimates, the risk-aversion parameter γ is estimated from the moment condition that the model should price the market portfolio. Finally, given these estimates, I test whether the model prices *bk-mkt*.

Even if the innovations, ϵ_{t+1}, are not conditionally normally distributed, Bollerslev and Wooldridge (1992) have shown that the parameters of the VAR-GARCH can be consistently estimated by a QMLE. A common practice in the estimation of multivariate GARCH models is to numerically calculate the score, $\mathbf{s_t}$, of the maximum likelihood function. For numerical stability, I analytically calculate the scores (see Appendix C). However, since the innovations are not necessarily normally distributed, the usual

[18] Note that the average conditional moments are typically not the same as the unconditional moments.

[19] Both $\mathbf{V}$ and $E[r_i^e]$ have been estimated using an alternative method of taking their sample counterparts, $E[r_i^e] = \frac{1}{T}\sum_{t=1}^{T} E_t[r_{i,t+1}^e]$ and $\mathbf{V} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{V_{ff,t}}$, where $E_t[r_{i,t+1}^e]$ and $\mathbf{V_{ff,t}}$ are the estimated time series of the VAR. This method produces negligible differences in the results.

maximum likelihood standard errors cannot be used. Hence, the standard errors are calculated using Hansen's (1982) Generalized Method of Moments (GMM), including the Newey and West (1987) adjustment for serial-correlations. The optimal order of Newey-West lags is estimated in accordance with Andrews (1991).

The other parameters of the model are ρ and γ. Since the approximation constant, ρ, is very difficult to estimate in practice, I exogenously specify it as $\rho = 0.9949$, but this choice not arbitrary. Since ρ is the steady-state consumption-to-wealth ratio, this value corresponds to an economy that consumes 6% of its total wealth each year, which is reasonable. This is also the value that was used by Campbell (1996) and Hodrick, Ng, and Sengmueller (1999).

Note that once the estimates of the VAR-GARCH model, $E[\epsilon_i^e \epsilon \epsilon']$, and ρ are available, the only undetermined variable of Eq. (23) is γ. Since the asset pricing relation has to be satisfied for all investable portfolios, and in particular for the market portfolio, an estimate for γ can be drawn from the moment condition implied by Eq. (23), where asset i is the market. I use the moment condition

$$E[r_i^e] + \frac{V_{ii}}{2} = \gamma V_{iw} + (\gamma - 1)V_{ih} - \frac{1}{2}(\gamma - 1)^2 V_{i\nu}, \quad \text{where} \quad i = \textit{emkt}, \tag{27}$$

to identify γ. Combined with the QMLE where the score functions are the orthogonality conditions, this is an exactly identified GMM estimation.

3.3 Comparison with other factor pricing models

In many factor pricing models, the logarithm of the pricing kernel, m_{t+1}, is modeled as a linear combination of K factors, $\mathbf{f_{t+1}}$, where

$$m_{t+1} = -\kappa_{t+1} - \mathbf{\Lambda}'(\mathbf{f_{t+1}} - \mathbf{E_t}[\mathbf{f_{t+1}}]), \tag{28}$$

and the vector $\mathbf{\Lambda}$ is the K-dimensional vector of the *prices of risk* associated with the factors in $\mathbf{f_t}$. The term κ_{t+1} is a normalizing constant to ensure that $R_{f,t+1} = \frac{1}{E_t[M_{t+1}]}$. The log-Euler equation from Eq. (2) can be rewritten as

$$E_t[r_{i,t+1}^e] + \frac{V_{ii,t}}{2} = \mathbf{\Lambda}'\mathbf{V_{if,t}}, \tag{29}$$

where $\mathbf{V_{if,t}} \equiv \text{Cov}_t(r_{i,t+1}^e, \mathbf{f_{t+1}})$.

The asset pricing relation of this paper has two distinguishing features. Note that Eq. (21) can be rewritten as

$$
\begin{aligned}
E_t[r^e_{i,t+1}] + \frac{V_{ii,t}}{2} &= \mathbf{\Lambda}'\mathbf{V_{if,t}} - \tilde{\mathbf{\Lambda}}'\mathbf{S_{iV,t}}\tilde{\mathbf{\Lambda}}, \qquad (30)\\
\text{where} \qquad \mathbf{\Lambda} &= \gamma(\mathbf{e1}+\mathbf{e2}) + (\gamma-1)\lambda_{\mathbf{h}},\\
\text{and} \qquad \tilde{\mathbf{\Lambda}} &= \frac{1}{\sqrt{2}}(\gamma-1)(\mathbf{e1}+\mathbf{e2}+\lambda_{\mathbf{h}}).
\end{aligned}
$$

One feature of Eq. (30) is that it allows for co-skewness to be priced when conditional normality is relaxed. This feature is reflected by the term $-\tilde{\mathbf{\Lambda}}'\mathbf{S_{iV,t}}\tilde{\mathbf{\Lambda}}$: an asset earns a risk premium if it performs poorly when future market volatility rises. The term $\tilde{\mathbf{\Lambda}}$ can be interpreted as the prices of risk associated with this channel.

More importantly, the model places restrictions on the prices of risk: the prices of risk are related to the information content of the state variables and the risk-aversion parameter. These restrictions are less stringent than those implied by the Consumption CAPM of Breeden (1979), since it allows for an additional channel through which assets may be priced: the changes in the investment opportunity set. Nevertheless, news about the future prospects of the market has to tie back into consumption through Eq. (22), since the aggregate budget constraint must be satisfied. The term $\lambda_{\mathbf{h}}$ summarizes how much information the innovations in the state variables contain about the changes in the future prospects of the market.

In general, factor pricing models treat $\mathbf{\Lambda}$ as free parameters to be estimated from the data. For example, Chen, Roll, and Ross (1986) and Ferson and Harvey (1991) find that macroeconomic variables draw economically significant prices of risk, while Fama and French (1993) find that *bk-mkt* command significant prices of risk. More recently, Ferson and Harvey (1999) and Vassalou (2003) argue that *bk-mkt* draw their risk premia from exposures to macroeconomic risk. However, since these models are estimated without constraint, there is no direct link between the prices of risk and a theoretical pricing model (see Cochrane, 2001). One could estimate an unconstrained model that prices the cross-section of stock returns, but this still would not necessarily tell us much about the underlying economic mechanism that is driving the results, since the link is lost.

A theoretical model can provide a link between the prices of risk and the underlying economic mechanism. For example, Hansen and Singleton (1983) studies the asset pricing relation of the Consumption CAPM, where the pricing kernel consists of the consumption growth rate. In that situation, the price of consumption risk depends on the risk-aversion coefficient of a representative agent. In the asset pricing model of this paper, the relevant factors are the current market return, the changes in the forecasts of future market returns, and the changes in the forecasts of future market volatility. These factors are linked to the aggregate consumption growth rate through the budget constraint, and hence their prices of risk are linked to their information content and the risk-aversion parameter. Overall, Eq. (30) indicates that it is not enough to say that there are state variables that proxy for the changes in the investment opportunity set; the prices of risk associated with such state variables ought to take a particular form. The remainder of this paper tests whether this is true.

4 Data

The econometric system I study includes time-series data on a proxy for the real market return; the return on a zero-cost portfolio that reflects the book-to-market effect; and macroeconomic variables that have been found to forecast future stock market returns. The data sources for this study are the Center for Research in Security Prices (CRSP), Standard & Poor's Data Resources, Inc. (DRI, formerly Citibase), the Federal Reserve Economic Data (FRED) at the Federal Reserve Bank of St. Louis, and Kenneth French's website.[20] Data are collected at the monthly frequency from April 1953 through December 1999.

4.1 Description of the data

The first two variables form a proxy for the real rate of return on the market portfolio. The first variable, *emkt* (excess market return), is the rate of return in excess of the one-month Treasury bill of the value-weighted stock index from CRSP. The second variable, *rrf* (real riskfree rate), is the one-month T-bill rate minus the rate of inflation. The rate of inflation is calculated as the change in the seasonally adjusted Consumer Price Index (CPI).[21] The sum of these two variables is used as a proxy for the real rate of market return.

[20] The author thanks Kenneth French for making his data on book-to-market portfolios available on his website at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

[21] The CPI series used is the seasonally adjusted CPI for all urban consumers, including all item such that the 36-month average in 1982-84 equals to 100.

Furthermore, *rrf* is included in the system since the short rate has been found by Ang and Bekaert (2001) to be a robust forecaster of future stock market returns. The next variable is the return on *bk-mkt*, which is a portfolio that goes long the highest quintile book-to-market (value) stocks and shorts the lowest quintile book-to-market (growth) stocks.[22] These returns are all expressed as log-returns.

The final three variables in the VAR systems are state variables that other studies have found to have forecasting ability over future returns on the stock market. The first variable is the logarithm of the dividend yield of the S&P 500 index, *dyld*. The second variable captures the term spread, *term*. It is the logarithm of the yield spread between the 10-year U.S. Treasury bonds and the 1-year U.S. Treasury bills. The final variable, *dflt*, measures the default spread using the logarithm of the yield spread between Baa-rated corporate bonds and Aaa-rated corporate bonds. All of these variables are expressed in continuously compounded annualized terms.

4.2 Summary statistics

Table 1 presents the summary statistics and the contemporaneous correlations of the variables in the VAR system, while Fig. 1 plots the time series of the variables. The average excess return on the stock market portfolio for this period is 7.93% per annum, including the Jensen's inequality term, with an annualized standard deviation of 14.71%. The one-month Treasury bill has averaged 1.32% per annum in real terms, while the average annualized rate of growth of the CPI has been 3.96%. Panel A of Fig. 1 shows the cumulative returns on *emkt* and *rrf* of a $1 investment at the beginning of the sample period.

The entire sample from April 1953 through December 1999 (N=561) shows that the book-to-market effect exists throughout the sample. Panel B of Fig. 1 illustrates the robustness of the effect throughout the sample. The book-to-market strategy has consistently produced annualized returns of 9.44% with annualized standard deviations of 10.55%. Table 1 indicates the presence of slight serial-correlations in *bk-mkt*, as well as

[22] Stocks sorted by market capitalizations and book-to-market ratios are formed according to Fama and French (1992) and obtained from Kenneth French's website. Common stocks listed on NYSE, AMEX, and NASDAQ of companies incorporated in the U.S. are sorted into five quintiles according to their market capitalizations and their book-to-market ratios. The breakpoints for the quintiles are formed based upon stocks listed on the NYSE. Thus, the lowest quintile portfolio contains all stocks that have lower market capitalizations or book-to-market ratios than the 20th percentile NYSE stock.

slight annual seasonality. Furthermore, *bk-mkt* is negatively correlated with the market, which indicates that *bk-mkt* has a negative market beta.

Panels C and D of Fig. 1 show the plots of *dyld*, *term*, and *dflt*. The dividend yield of the S&P 500 has averaged 3.57% annually, while the term spread and the default spread have averaged 0.68% and 0.87% annually, respectively. These three variables are all highly persistent. The dividend yield on the S&P 500 is negatively correlated with the term spread and strongly positively correlated with the default spread.

5 Empirical results

In this section, I begin by presenting the estimates of the VAR-GARCH model described in Section 3. The VAR-GARCH model provides estimates of the forecasting abilities of the state variables. The state variables are *emkt*, *rrf*, *bk-mkt*, *dyld*, *term*, and *dflt*. I first examine the information content of the variables and then consider the asset pricing relation of the model developed in Section 2.

5.1 VAR-GARCH parameter estimates

Table 2 conducts specification tests of the first-order VAR model. Panel A of Table 2 tests for the appropriate order of VAR lags using the Schwarz (1978) selection criterion, and the result supports the first-order specifications. Panel B of Table 2 conducts diagnotic tests for residual serial-correlations at lags one through eleven, and at lags one through twelve, using the l-test of Cumby and Huizinga (1992). Looking at the serial-correlations up to eleven lags, there is some evidence of misspecification for the dynamics of the real risk-free rate. Hodrick (1992) advocates detrending the risk-free rate for this reason, but my model requires the actual real risk-free rate to be modeled. The serial-correlations at the twelfth lag also suggest that there may be some seasonal effect that the VAR model is unable to capture. Hence, the results of this analysis should be taken with this caveat in mind. Panel C of Table 2 performs joint forecastability tests and confirms that the state variables in the VARs can be forecasted by the lagged variables. The eigenvalues of the VAR estimates are within the unit circle, which indicate that the VAR systems are stationary. Finally, the optimal order of Newey-West lag is determined by the procedure of Andrews (1991) to be three.

Table 3 examines in detail the QMLE estimates of the VAR-GARCH parameters for the system. Panel A shows the estimates of the VAR forecasting equations. Fig. 2 plots the time series of the expected excess market returns and real risk-free rates according to the estimates as well as their realized values. According to the estimates, a 1% wider-than-average term spread forecasts about 0.5% higher than average excess market return over the next month, while a high default spread forecasts a high real risk-free rate. A high dividend yield also forecasts a high real risk-free rate, but it fails to forecast statistically significant changes in excess market return. The VARs also indicate that *dyld*, *term*, and *dflt* are persistent variables, which suggests that they may carry longer-horizon forecasting abilities. An ICAPM would suggest that assets with returns that are correlated with these state variables command risk premia, because they convey information about changes in the investment opportunity set.

The estimates of the VARs also reveal that *emkt* contains information about changes in the investment opportunity set. A high excess market return forecasts a lower future dividend yield, as well as a reduction in the default spread over the next period. In turn, a lower dividend yield forecasts a lower future market return, which suggests that the market may have a slowly mean-reverting component. Such longer-horizon forecasts are examined in greater detail below. In contrast, the real risk-free rate is somewhat persistent but contains little forecasting information.

If the returns on the book-to-market effect can be explained as compensation for exposures to changes in expected market returns, they should forecast future market returns. The question now becomes, controlling for other state variables, do these portfolios perform poorly when the prospects for the future turn sour? The VAR parameter estimates suggest otherwise. For the book-to-market effect, a high rate of return on *bk-mkt* forecasts a statistically significantly lower next-period return on *emkt*. Hence it performs well, rather than poorly, when the prospects for the future diminish. In contrast, the results of Liew and Vassalou (2000) suggest that that *bk-mkt* return should be high when the prospects for the future are improving.[23]

Panels B and C of Table 3 show the estimates of the GARCH parameters. These estimates suggest that there are variations across time in volatility. The volatilities of *emkt*, *rrf*, and *bk-mkt* are also all found to be persistent. Furthermore, the off-diagonals

[23] Our results differ, since my estimates are not based upon international data and I control for other state variables that forecast future returns.

of the GARCH parameters suggest that a high shock in *emkt* forecasts a reduction in the covariance of *emkt* with both *bk-mkt* and *mom*. These estimates generate some time-variations in the betas of the assets with the market.

According to the model developed in this paper, an asset may also command risk premia because it forecasts future volatilities. which would be reflected through their conditional skewness and co-skewness. Table 4 shows the estimates of skewness and co-skewness implied by the residuals of the VAR-GARCH models. The entries in Table 4 are expressed as coefficients of (co-)skewness by scaling them by the standard deviations of $\epsilon_{i,t}$, $\epsilon_{j,t}$, and $\epsilon_{k,t}$.

The estimates of skewness of the residuals indicate that excess market returns are negatively skewed. The estimates of co-skewness of *emkt* with *bk-mkt* is statistically significant and suggests that *bk-mkt* becomes more volatile when the market is rising. However, *bk-mkt* fares well when the aggregate market becomes riskier, which is not consistent with the argument that *bk-mkt* earns risk premia because it forecasts changes in future uncertainty.

In summary, the initial examination of the VAR-GARCH parameter estimates suggests that it may be difficult for the changes in the investment opportunity set to explain the book-to-market effect. The estimates of the VAR suggest that, controlling for other state variables, *bk-mkt* performs well (rather than poorly) when the expected next period return on the market is falling. Moreover, co-skewness of the residuals suggests that *bk-mkt* performs well (rather than poorly) when the market becomes riskier over the next period. However, the model derived in Section 2 shows that it is not sufficient to look at the forecasting ability of a factor over the next period, but one must consider the cumulative impacts of the shocks over all future horizons and covariation of an asset with other state-variables with forecasting abilities.

5.2 Cumulative forecasting abilities

Table 5 shows the cumulative weighted sums across all future horizons of the impulse-responses of the real market return, which were derived in Eq. (16) as $\lambda_{\mathbf{h}}$. Changes in the forecasts of future real market returns are exponentially weighted at the rate $\rho = 0.9949$.[24] The rows of this panel show the estimates of $\lambda_{\mathbf{h}}$ from the VAR-GARCH system. These

[24] Other values for ρ in the range of $[0.99, 1.00]$ produce negligible changes in the results.

estimates indicate that only *emkt* and *dyld* carry statistically significant information about changes in cumulative future returns. A 1% increase in *emkt* today forecasts a mean-reversion of about 0.2% in cumulative future returns of *emkt*. A 0.1% increase in *dyld* forecasts a future increase in *emkt* of about 1.5% in cumulative terms. The point estimate of the cumulative effect of a shock in *bk-mkt* is negative. Hence, *bk-mkt* tends to perform well when the cumulative expected return of the market falls.

However, an asset may also command risk premia because it covaries with other factors that are priced, rather than because it carries independent information about current or future returns to wealth. Table 6 shows the average contributions to expected returns from the three sources of risk revealed by the model, controlling for these covariations. Panel A shows the time-series average of the covariations of *emkt*, *rrf*, and *bk-mkt* with the real market return, V_{iw}. These entries in Panel A are expressed as annualized percentage returns for a unit increase in γ. As expected, the excess return on the market commands positive risk premia. However, since the GARCH estimates imply that *bk-mkt* covaries negatively with the real market return, the point estimate suggests that *bk-mkt* should have negative expected returns due to the negative market beta.

Panels B and C show the contributions to the expected returns from the hedging benefits against the changes in the investment opportunity set. Estimates in Panels B and C are the terms $\mathbf{V_{ih}}$ and $\mathbf{V_{i\nu}}$ from Eq. (23), respectively. The entries in Panel B can be interpreted as the average contributions to expected returns for covariations with changes in the forecasts of future market returns, per a unit increase in$(\gamma - 1)$. These estimates show that only the information content in *emkt* carries statistically significant contributions to expected returns through this channel–a high current market return forecasts lower cumulative future market returns. The point estimate for *bk-mkt* is negative. The entries in Panel C are contributions to expected returns for a unit decrease in $\frac{1}{2}(\gamma - 1)^2$ that are attributable to an asset's covariation with the changes in the forecasts of future market volatility. However, these estimates are not statistically or economically significant.

In summary, not only is the estimate of market beta of *bk-mkt* is negative, the evidence of forecasting ability in *bk-mkt* is rather small. The point estimate for *bk-mkt* indicates that it forecasts negative changes in future expected returns and fails to significantly forecast future volatility. Therefore, the book-to-market effect has a negative market beta, and it performs well when the prospects for the future turns sour (a good hedge against adverse changes in the investment opportunity set). The next section conducts a formal hypothesis test of the asset pricing relation.

5.3 Hypothesis tests

The second implication of the model is that the risk premia across factors are linked to each other through the parameter γ, which can be interpreted as a risk-aversion parameter. Identifying this parameter allows formal hypothesis tests of the model to be conducted. As described in Section 3.2, this parameter is identified by imposing the model to price the market portfolio. In other words, I test whether the model jointly prices the equity premium and the book-to-market effect. The estimate of γ from the VAR system is 9.54, with a standard error of 3.12.

This asset pricing relation is tested by examining the time-series averages of the mispricing across time. Table 7 shows the portion of returns in the data that is not explained by various models. The first column shows the time-series averages of the conditional expected returns from the VAR-GARCH estimates, with an adjustment for the Jensen's inequality term. The Wald statistics test whether these effects are statistically significant:

$$\text{H0:} E[r_i^e] + \frac{V_{ii}}{2} = 0. \tag{31}$$

The second column includes a term that controls for the conditional market beta and tests whether a conditional CAPM explains these returns:

$$\text{H0:} \left(E[r_i^e] + \frac{V_{ii}}{2} \right) - \gamma V_{iw,t} = 0. \tag{32}$$

The last column includes terms for the risk premia that come from exposures to the changes in the investment opportunity set, and it tests whether the changes in the investment opportunity set can explain these returns:

$$\text{H0:} \left(E[r_i^e] + \frac{V_{ii}}{2} \right) - \left(\gamma V_{iw} + (\gamma - 1) V_{ih} - \frac{1}{2} (\gamma - 1)^2 V_{i\nu} \right) = 0. \tag{33}$$

It should be noted that these statistics are functions of the estimates of the VAR-GARCH parameters and γ, which means that the standard errors can be calculated using the Delta-method.[25]

[25] However, Bekaert and Hodrick (2001) have shown that when testing the Expectations Hypothesis the

The Wald tests reject the hypothesis that the return on *bk-mkt* can be explained as risk premia for exposures to the market risk and to changes in the investment opportunity set. The return on *bk-mkt* is estimated to be 8.88% per year, but controlling for its negative beta increases the returns to 12.68% per year. Adjusting for exposures to the changes in the investment opportunity set explains very little of the returns; 12.71% per year is not explained. Furthermore, the lower bound of the 95% confidence interval of the returns remaining after controlling for these risks is 9.81%, and the Wald statistic rejects the hypothesis that the book-to-market effect is consistent with the model. In summary, when the constraints implied by the aggregate budget constraint are imposed, so that assets are priced if they can forecast changes in the investment opportunity set and the risk premia across assets are linked to risk-aversion, hypothesis tests reject the assertion that the information content of *bk-mkt* can explain the asset's historically high returns.

5.4 Significance of omitted factors

The VAR-GARCH analysis of the book-to-market effect indicates that the forecasting ability of *bk-mkt* is not strong enough to interpret it as a factor that forecasts changes in the investment opportunity set. This leaves open the question as to what other underlying economic mechanism is needed to explain the book-to-market effect. One possibility is that there exists a factor that has been left out of the VAR-GARCH specification or that there is a missing component in the pricing kernel in addition to the consumption growth rate and the aggregate market return. To understand the significance of a missing mechanism, I construct a lower bound for the Sharpe ratio of the mimicking portfolio associated with such an omitted factor. One lower bound is the Sharpe ratio for the portion of the returns on *bk-mkt* and *mom* that is not explained by the model.

Suppose that the true pricing kernel contains another component, X_t:

$$M_{t+1} = \beta^{\theta_1} \left(\frac{C_{t+1}}{C_t} \right)^{\frac{-\theta_2}{\sigma}} \left(R_{W,t+1} \right)^{\theta_3 - 1} \left(X_{t+1} \right)^{\gamma_x}, \tag{34}$$

where γ_x is an unknown parameter. One possible missing component is the "surplus consumption ratio" of Campbell and Cochrane (1999). Bekaert and Grenadier (2001)

Wald statistic tends to over-reject the null hypothesis. To guard against making potentially misleading inferences, an earlier version of this paper used the LM and the DM statistics to supplement the Wald statistic and found lower statistical significance, but qualitatively similar results.

and Barberis, Huang, and Santos (2001) offer the possibility of other state variables that may enter the pricing kernel. What these papers have in common is that they introduce time-variations in investor preferences to generate a new component in the pricing kernel.

There are alternative interpretations for a factor omitted from the VAR analysis. One interpretation is that the omitted factor reflects the higher-order moment terms that was left out of the approximation of the Euler equation. Another interpretation is that it comes from a variable missing in the VAR-GARCH specification, rather than in the pricing kernel. It is quite possible that the econometrician can never observe the true changes in the investment opportunity set. Furthermore, Hansen and Richard (1987) criticize that by conducting unconditional tests the econometrician misses the agents' true conditional information set. Hence, another interpretation for the missing factor is that it reflects the true unobserved changes in the investment opportunity set.

For simplicity, suppose that the omitted factor provides information orthogonal to that in $\mathbf{f_t}$, except possibly with *bk-mkt*. In this case, Appendix D shows that the mimicking portfolio for such a factor, x, would have the property that:

$$\left(\frac{E[r_i^e] + \frac{V_{ii}}{2} - (\mathbf{\Lambda}'\mathbf{V_{if,t}} - \tilde{\mathbf{\Lambda}}'\mathbf{S_{iV}}\tilde{\mathbf{\Lambda}})}{\sqrt{V_{ii}}} \right) \leq \left(\frac{E[x] + \frac{V_{xx}}{2}}{\sqrt{V_{xx}}} \right). \qquad (35)$$

The left-hand side of Eq. (35) is the return on *bk-mkt* adjusted for market risk and changes in the investment opportunity set, divided by its standard deviation. It is the Sharpe ratio of the portion of returns on *bk-mkt* that cannot be explained by the model. The right-hand side of Eq. (35) is the Sharpe ratio of the omitted factor, controlling for the Jensen's inequality term. The interpretation of Eq. (35) is that the Sharpe ratio of the omitted factor must be at least as large as the Sharpe ratio of the unexplained portion of asset returns.

The lower rows of Table 7 show the lower bounds on the Sharpe ratios for such an omitted factor, expressed in annualized terms. The lower bound derived by the *bk-mkt* system is 1.27, with a standard error of 0.14. As a comparison, the *emkt* has an annualized Sharpe ratio of 0.54. Therefore, to justify the return on *bk-mkt*, there must be a missing factor that is as valuable as the aggregate stock market in terms of Sharpe ratios. Ongoing research investigates whether this is a plausible possibility.

In a setting without time-varying expected returns, MacKinlay (1995) argues that risk-based models have difficulty matching the observed data, because these models have bounded Sharpe ratios. MacKinlay shows that maximal Sharpe ratios in the data are too high to be consistent with any reasonable underlying asset pricing model, without accounting for the implications of intertemporal asset pricing. He reasons that a reasonable maximal Sharpe ratio is around 0.6. This paper shows that accounting for changes in the investment opportunity set does little to mitigate this problem.

6 Concluding comments

One possible explanation for the historically high returns found in the cross-section of stock returns is that these returns reflect compensation for risk exposures to adverse changes in the investment opportunity set–an asset may earn a risk premium if it performs poorly when investment prospects for the future turn sour. I develop an intertemporal asset pricing model in which the pricing kernel depends upon the consumption growth rate and the aggregate market return and estimate a system in which expected returns and volatilities are time-varying. Restrictions implied by the model show that an asset that earns risk premium for exposures to adverse changes in the investment opportunity set should forecast future market conditions. Moreover, the model shows that factor risk premia should be linked to a single parameter that can be interpreted as the willingness of investors to bear risk. I investigate the extent to which the historical returns of portfolios generated by the book-to-market effect can be explained as exposures to time-variations in the forecasts of future market returns and future market volatilities. The results of this study indicate that the book-to-market effect cannot be explained using these changes in the investment opportunity set.

Empirically, the main difficulty in using changes in the investment opportunity set to explain asset returns is that such changes are difficult to measure. While several papers document time-variations in expected future market returns and volatilities, the empirical validity of those results are not without debate (see, for example, Ang and Bekaert, 2001). Moreover, book-to-market portfolios exhibit little covariation with such changes. The intertemporal asset pricing model requires assets to either forecast future market returns or covary with other state variables that do so. Moreover, it is not enough for an asset (or a state variable) to forecast future market returns, but risk premia associated with such factors have to be consistent with the willingness of investors to bear risk. These restrictions implied by the model suggest that this class of model will have a difficult time explaining the book-to-market effect.

Of course, as with any empirical study, the results are only as valid as the assumptions of the model. It is possible that there exist other components that are missing from my model, such as state variables reflecting changes in investor preferences, higher-order moments, or variables that are observable to the economic agents but not to the econometrician. Works like Ang, Chen, and Xing (2002) and Ang and Chen (2003) investigates some of these possibilities. Furthermore, this model assumes that capital markets are perfect, that the pricing kernel depends on aggregate consumption growth and aggregate market returns, and that the aggregate budget constraint is binding. Models that incorporate market frictions, such as models in which these aggregation properties do not hold, might also explain the variations in the cross-section of stock returns. While ongoing research investigates these possibilities, this paper suggests that accounting for the changes in the investment opportunity set does little in explaining the cross-section of stock returns.

Appendices

A An implementation of Intertemporal CAPM

This appendix details the derivation of the intertemporal asset pricing model used in this paper. For most parts, I follow Campbell's (1993, 1996) implementation of Merton's (1973) Intertemporal CAPM to substitute out the consumption growth rate using the aggregate budget constraint.[26] The derivation in this paper extends Campbell (1993) to allow for time-varying covariances. This implementation also allows for conditionally non-normal distributions, but the higher-order conditional moments are assumed to be constants across time. The derivation begins with a pricing kernel that has two components: the consumption growth rate and the aggregate market return. This pricing kernel generalizes the pricing kernel implied by the Epstein and Zin (1989) recursive utility function of the representative agent studied by Campbell (1993).

Log-linear approximation of the budget constraint

Following Campbell (1993), the use of the consumption data in the model can be circumvented by taking a log-linear approximation of the aggregate budget constraint. Let W_t denote the aggregate wealth level at the beginning of time period t, let C_t be the time t consumption level, and let $R_{W,t+1}$ be the total rate of return on wealth from time t to $t+1$ (the market return). Furthermore, let the lower-case variables denote the natural logarithms of the upper-case variables. The aggregate budget constraint can be written as

$$W_{t+1} = R_{W,t+1}(W_t - C_t), \tag{A.1}$$

$$\frac{W_{t+1}}{W_t} = R_{W,t+1}(1 - \frac{C_t}{W_t}), \tag{A.2}$$

$$\text{or} \quad \Delta w_t = r_{W,t+1} + \log(1 - \exp(c_t - w_t)). \tag{A.3}$$

Using the function, $f(x) = \log(1 - \exp(x))$, and defining $\rho \equiv 1 - \exp(E[c_t - w_t])$, the first-order Taylor approximation of Eq. (A.3) about $E[c_t - w_t]$ yields

$$\Delta w_{t+1} = k + r_{w,t+1} + [\frac{\rho - 1}{\rho}](c_t - w_t) + o^2(\cdot), \tag{A.4}$$

where $k = -[\frac{\rho-1}{\rho}]E[c_t - w_t]$. The parameter ρ is the steady-state ratio of aggregate invested wealth to aggregate total wealth. The higher-order terms of $o^2(\cdot)$ are the higher powers of the log consumption-to-wealth ratio, $(c_t - w_t)$. Hence, the accuracy of the approximation depends only on the variability of the log consumption-to-wealth ratio. Since the constant term, k, turns out to play no further role, the approximation error is not significant unless there are some time-variations in the higher-order moments of $(c_t - w_t)$.

By definition,

$$\Delta w_{t+1} = \Delta c_{t+1} + (c_t - w_t) - (c_{t+1} - w_{t+1}). \tag{A.5}$$

Therefore, by combining Eqs. (A.4) and (A.5) and iterating forward,

$$k + r_{w,t+1} + [\frac{\rho - 1}{\rho}](c_t - w_t) = \Delta c_{t+1} + (c_t - w_t) - (c_{t+1} - w_{t+1}), \tag{A.6}$$

[26] Campbell's (1993, 1996) model has been extensively studied in the portfolio allocation literature. See Campbell and Viceira (2002) for a review.

$$(c_t - w_t) = \rho k + \rho(c_{t+1} - w_{t+1}) + \rho(r_{w,t+1} - \Delta c_{t+1}), \tag{A.7}$$

$$= \frac{\rho k}{1-\rho} + \sum_{s=1}^{\infty} \rho^s (r_{w,t+s} - \Delta c_{t+s}). \tag{A.8}$$

The expectation of Eq. (A.8) can be substituted back into Eqs. (A.4) and (A.5) to yield

$$\Delta c_{t+1} - E_t[\Delta c_{t+1}] = (E_{t+1} - E_t) \sum_{s=0}^{\infty} \rho^s r_{w,t+s+1} - (E_{t+1} - E_t) \sum_{s=1}^{\infty} \rho^s \Delta c_{t+s}. \tag{A.9}$$

The interpretation of Eq. (A.9) is that to a first-order approximation any unanticipated increase in consumption today must be financed by either higher future returns or lower future consumption.

Intertemporal asset pricing under a general pricing kernel

Suppose that the pricing kernel has only two components: the consumption growth rate and the rate of return on aggregate wealth. In particular, suppose that the pricing kernel is of the form

$$M_{t+1} = \beta^{\theta_1} \left(\frac{C_{t+1}}{C_t} \right)^{\frac{-\theta_2}{\sigma}} (R_{W,t+1})^{\theta_3 - 1} \tag{A.10}$$

for parameters β, θ_1, θ_2, and θ_3. This specification includes Epstein and Zin's (1989) recursive utility function representative agent as a special case. The pricing kernel implied by a recursive utility function takes $\theta_1 \equiv \theta_2 \equiv \theta_3 \equiv \frac{1-\gamma}{1-(\frac{1}{\sigma})}$, where γ is the risk-aversion parameter, σ is the elasticity of intertemporal substitution, and β represents the subjective discount factor. The power utility function is a particularly special case in which $\sigma \equiv \frac{1}{\gamma}$.

By using a second-order Taylor approximation to obtain the log-Euler equations, the following two implications can be obtained:

$$E_t[\Delta c_{t+1}] = \text{constant} + \mu_t + \sigma \frac{\theta_3}{\theta_2} E_t[r_{w,t+1}], \tag{A.11}$$

$$\text{and} \quad E_t[r^e_{i,t+1}] + \frac{V_{ii,t}}{2} = \frac{\theta_2}{\sigma} V_{ic,t} + (1-\theta_3) V_{iw,t}, \tag{A.12}$$

where $\mu_t = \frac{1}{2}\left(\frac{\theta_2}{\sigma}\right) \text{Var}_t[\Delta c_{t+1} - \sigma \frac{\theta_3}{\theta_2} r_{w,t+1}]$, $r^e_{i,t+1}$ is the log-return on an asset i in excess of the risk-free rate, $V_{ii,t} \equiv \text{Var}_t(r^e_{i,t+1})$, $V_{ic,t} \equiv \text{Cov}_t(r^e_{i,t+1}, \Delta c_{t+1})$, and $V_{iw,t} \equiv \text{Cov}_t(r^e_{i,t+1}, r_{w,t+1})$.

Combining Eqs. (A.8) and (A.11) produces the following:

$$c_t - w_t = \text{constant} + (1 - \sigma \frac{\theta_3}{\theta_2}) E_t \sum_{s=1}^{\infty} \rho^s r_{w,t+s} - E_t \sum_{s=1}^{\infty} \rho^s \mu_{t+s}. \tag{A.13}$$

Lettau and Ludvigson (2001a) estimate the left-hand side of Eq. (A.13). In contrast, Campbell (1996), Hodrick, Ng, and Sengmueller (1999), and this paper estimate the right-hand side of Eq. (A.13), in which the consumption variable has been substituted out.

Eq. (A.13) can now be combined with Eq. (A.12) to produce:

$$E_t[r^e_{i,t+1}] + \frac{V_{ii,t}}{2} = \left(1 + \frac{\theta_2}{\sigma} - \theta_3\right) V_{iw,t} + \left(\frac{\theta_2}{\sigma} - \theta_3\right) V_{ih,t} - \left(\frac{\theta_2}{\sigma}\right) V_{i\mu,t}, \tag{A.14}$$

where $r_{h,t+1} \equiv (E_{t+1} - E_t) \sum_{s=1}^{\infty} \rho^s r_{w,t+s+1}$, $V_{ih,t} \equiv \text{Cov}_t(r^e_{i,t+1}, r_{h,t+1})$, $r_{\mu,t+1} \equiv (E_{t+1} - E_t) \sum_{s=1}^{\infty} \rho^s \mu_{t+s+1}$, and $V_{i\mu,t} \equiv \text{Cov}_t(r^e_{i,t+1}, r_{\mu,t+1})$. The terms $r_{h,t+1}$ and $r_{\mu,t+1}$ reflect changes in the expectations of cumulative future market returns and future market volatilities, respectively. These terms are not actually returns on any asset in particular, but the notations are chosen to maintain consistency with prior work. These future expectations are cumulated as exponentially weighted sums, where the weights are the powers of ρ.[27] The first term in Eq. (A.14) reflects an asset's covariation with the current market return. The second term in Eq. (A.14) reflects an asset's covariation with changes in the forecasts of future market returns. The third term in Eq. (A.14) reflects an asset's covariation with changes in the forecasts of future volatilities of the consumption growth rates and the rates of market return. By simplifying the final term, $V_{i\mu,t}$, the asset pricing relation in Eq. (A.14) can be rewritten as,

$$E_t[r^e_{i,t+1}] + \frac{V_{ii,t}}{2} = \left(1 + \frac{\theta_2}{\sigma} - \theta_3\right) V_{iw,t} + \left(\frac{\theta_2}{\sigma} - \theta_3\right) V_{ih,t} - \frac{1}{2}\left(\frac{\theta_2}{\sigma}\right)^2 V_{i\bar{\mu},t}, \quad \text{(A.15)}$$

$$\text{where} \quad r_{\bar{\mu},t+1} = (E_{t+1} - E_t)\sum_{s=1}^{\infty} \rho^s \mathrm{Var}_{t+s}[\Delta c_{t+s+1} - \sigma\frac{\theta_3}{\theta_2} r_{w,t+s+1}]. \quad \text{(A.16)}$$

Using the Law of Iterated Expectations, Eq. (A.11) can be plugged into the final term of Eq. (A.9) to obtain:

$$\Delta c_{t+1} - E_t[\Delta c_{t+1}] = r_{w,t+1} - E_t[r_{w,t+1}] + (1 - \sigma\frac{\theta_3}{\theta_2}) r_{h,t+1} - \frac{1}{2}\frac{\theta_2}{\sigma} r_{\bar{\mu},t+1}. \quad \text{(A.17)}$$

The interpretation of Eq. (A.17) is that an unexpected increase in consumption today must reflect either an unexpectedly high aggregate return today, an improved prospects for higher returns in the future, or lower uncertainty over future consumption and returns. The latter channel reflects reduced precautionary savings due to reduced uncertainty.

Assume that the covariances between shocks to $r_{w,t+1}$ (and powers of these shocks) and the changes in the forecasts of future variances, $(E_{t+1} - E_t)\mathrm{Var}_{t+s+1}[r_{w,t+s+1}]$, are constants for all s. Then by plugging Eq. (A.17) into Eq. (A.16), $r_{\bar{\mu},t+1}$ can be rewritten as:

$$r_{\bar{\mu},t+1} = (1 - \sigma\frac{\theta_3}{\theta_2})^2 (E_{t+1} - E_t)\sum_{s=1}^{\infty} \rho^s \mathrm{Var}_{t+s}[r_{w,t+s+1} + r_{h,t+s+1}]. \quad \text{(A.18)}$$

Let $r_{\nu,t+1} \equiv (E_{t+1} - E_t)\sum_{s=1}^{\infty} \rho^s \mathrm{Var}_{t+s}[r_{w,t+s+1} + r_{h,t+s+1}]$ and $V_{i\nu,t} \equiv \mathrm{Cov}_t(r^e_{i,t+1}, r_{\nu,t+1})$. Then, putting everything together, the asset pricing relationship is now

$$E_t[r^e_{i,t+1}] + \frac{V_{ii,t}}{2} = \gamma V_{iw,t} + (\gamma - 1) V_{ih,t} - \frac{1}{2}(\gamma - 1)^2 V_{i\nu,t}, \quad \text{(A.19)}$$

where $\gamma = \frac{\theta_2}{\sigma} - \theta_3 + 1$. Notice that a single parameter γ determines the risk premium on each of the three sources of risk. If there exists a representative agent with a recursive utility function or a power utility function, the parameter γ corresponds to the risk-aversion coefficient.

B Asset pricing relation under a multivariate VAR(1) - GARCH(1,1) distribution

The asset pricing relation of interest is in the form

$$E_t[r^e_{i,t+1}] + \frac{V_{ii,t}}{2} = \gamma_1 V_{iw,t} + \gamma_2 V_{ih,t} + \gamma_3 V_{i\nu,t}, \quad \text{(B.1)}$$

[27] If one uses a second-order Taylor expansion in deriving Eq. (A.4), a new term appears in Eq. (A.14). This term is $\left(\frac{1-\rho}{\rho}\right)\left(\frac{1-\gamma}{1-\sigma}\right) V_{i\xi,t}$, where $\xi = (c_t - w_t)^2$ and $V_{i\xi,t} \equiv \mathrm{Cov}_t(r^e_{i,t+1}, r_{\xi,t+1})$. This term represents an asset's covariation with the conditional variances of the log consumption-to-wealth ratio.

which is in terms of conditional expectations and conditional covariances. This appendix discusses the asset pricing relation in detail when asset returns and state variables follow a joint multivariate VAR(1)-GARCH(1,1) process. A VAR model is fitted to describe the dynamics of the conditional expected returns of the variables in the system while a GARCH specification describes the dynamics of the conditional covariances. In general, there is no need to make additional assumptions regarding the higher-order conditional moments of the distribution. Specifically, there is no need to assume that the distribution is conditionally normal.

Suppose that there is a total of K assets and state variables. Let the first element of the system be the logarithm of the rate of return on aggregate wealth in excess of the risk-free rate. Let the second element of the VAR be the logarithm of the risk-free rate minus the rate of inflation. The other $K-2$ elements are either excess returns of assets to be priced or state variables that forecast future returns or forecast future covariances. Let $\mathbf{f_t}$ denote a K-element vector of these variables.

Consider a first-order K-dimensional VAR system defined by

$$\mathbf{f_{t+1}} = \mu + \mathbf{A}(\mathbf{f_t} - \mu) + \epsilon_{\mathbf{t+1}}, \tag{B.2}$$

where $\mu = E[\mathbf{f}]$, $\mathbf{A}$ is a $K \times K$ parameter matrix, and $\epsilon_{\mathbf{t+1}}$ is a K-dimensional innovation in the vector of state variables. Let $\mathbf{V}_t$ denote the variance-covariance matrix of the next period innovation, ϵ_{t+1}, conditional on information at time t: $\mathbf{V}_t = E_t(\epsilon_{t+1}\epsilon'_{t+1})$. Following Engle and Kroner (1995), let the dynamics of the $\mathbf{V}_t$ be:

$$\mathbf{V}_t = \mathbf{PP}' + \mathbf{\Delta V_{t-1} \Delta}' + \mathbf{\Psi}\epsilon_{\mathbf{t}}\epsilon'_{\mathbf{t}}\mathbf{\Psi}', \tag{B.3}$$

for some lower-triangular K-dimensional matrix, $\mathbf{P}$, and $K \times K$ dimensional matrices $\mathbf{\Delta}$ and $\mathbf{\Psi}$.

The left-hand side of Eq. (B.1), $E_t[r^e_{i,t+1}] + \frac{V_{ii,t}}{2}$, can be derived immediately from this VAR-GARCH specification. Let $\mathbf{ei}$ be a K-dimensional vector with a one in the i-th element and zeros elsewhere. From Eq. (B.2), $E_t[r^e_{i,t+1}] = (\mu + \mathbf{A}(\mathbf{f_t} - \mu))'\mathbf{ei}$. From Eq. (B.3), $V_{ii,t} = \mathbf{ei}'\mathbf{V}_t\mathbf{ei}$.

The first term on the right-hand side of Eq. (B.1) contains the term $V_{iw,t}$. Since $r_{w,t+1}$ is the logarithm of the total return on aggregate wealth, $r_{w,t+1} = (\mathbf{e1} + \mathbf{e2})'\mathbf{f_{t+1}}$. Therefore, $V_{iw,t} = (\mathbf{e1} + \mathbf{e2})'\mathbf{V}_t\mathbf{ei}$.

The second term on the right-hand side of Eq. (B.1) contains the term $V_{ih,t}$, where

$$r_{h,t+1} \equiv (E_{t+1} - E_t)\sum_{s=1}^{\infty}\rho^s r_{w,t+s+1}. \tag{B.4}$$

Under the VAR, the forecast at time t of the s-period ahead state variables is $E_t[\mathbf{f_{t+s}}] = \mu + \mathbf{A}^s(\mathbf{f_t} - \mu)$. Therefore, the change in the forecast of $\mathbf{f_{t+s+1}}$ due to a shock of ϵ_{t+1} is

$$(E_{t+1} - E_t)\mathbf{f_{t+s+1}} = \mathbf{A^s}\epsilon_{\mathbf{t+1}}. \tag{B.5}$$

This function is the *impulse-response* of the s-period ahead variables in response to the innovations at time $t+1$. By using Eq. (B.5),

$$r_{h,t+1} = (\mathbf{e1} + \mathbf{e2})'\sum_{s=1}^{\infty}\rho^s\mathbf{A}^s\epsilon_{\mathbf{t+1}}, \tag{B.6}$$

$$= (\mathbf{e1} + \mathbf{e2})'\rho\mathbf{A}(\mathbf{I} - \rho\mathbf{A})^{-1}\epsilon_{\mathbf{t+1}} = \lambda_{\mathbf{h}}{}'\epsilon_{\mathbf{t+1}}, \tag{B.7}$$

where $\lambda_{\mathbf{h}}{}'$ is defined as $(\mathbf{e1} + \mathbf{e2})'\rho\mathbf{A}(\mathbf{I} - \rho\mathbf{A})^{-1}$. Therefore,

$$V_{ih,t} = \lambda_h{}'\mathbf{V}_t\mathbf{ei}. \tag{B.8}$$

The third term on the right-hand side of Eq. (B.1) contains the term $V_{i\nu,t}$, where

$$r_{\nu,t} = (E_{t+1} - E_t)\sum_{s=1}^{\infty}\rho^s \text{Var}_{t+s}[r_{w,t+s+1} + r_{h,t+s+1}]. \tag{B.9}$$

Note that

$$\text{Var}_{t+s}[r_{w,t+s+1} + r_{h,t+s+1}] = (\mathbf{e1} + \mathbf{e2} + \lambda_{\mathbf{h}})'\mathbf{V}_{t+s}(\mathbf{e1} + \mathbf{e2} + \lambda_{\mathbf{h}}). \tag{B.10}$$

To compute $V_{i\nu,t}$, let $S_{i\mathbf{V},t}$ denote the sum of the covariations of $r^e_{i,t+1}$ with the changes in the forecasts of future covariances of $\mathbf{f_{t+s}}$. That is,

$$S_{i\mathbf{V},t} \equiv E_t\left[\epsilon^e_{i,t+1}\sum_{s=1}^{\infty}\rho^s(E_{t+1} - E_t)\mathbf{V}_{t+s}\right], \tag{B.11}$$

where $\epsilon^e_{i,t+1} = r^e_{i,t+1} - E_t[r^e_{i,t+1}]$. If innovations are conditionally normally distributed, $S_{i\mathbf{V},t} = 0$. However, for a process in which the shocks are not necessarily conditionally normally distributed but the conditional variance follows a multivariate GARCH(1,1) process described by Eq. (B.3), it can be shown that

$$(E_{t+1} - E_t)\mathbf{V}_{t+s} = (\Delta + \Psi)^{s-1}\Psi(\epsilon_{t+1}\epsilon'_{t+1} - \mathbf{V}_t)\Psi'(\Delta' + \Psi')^{s-1}. \tag{B.12}$$

Therefore, $S_{i\mathbf{V},t}$ can be rewritten as

$$\begin{aligned} S_{i\mathbf{V},t} &= E_t\left[\epsilon^e_{i,t+1}\sum_{s=1}^{\infty}\rho^s(\Delta + \Psi)^{s-1}\Psi(\epsilon_{t+1}\epsilon'_{t+1})\Psi'(\Delta' + \Psi')^{s-1}\right], \\ &= \sum_{s=0}^{\infty}\rho^s(\Delta + \Psi)^s\Psi E_t[\epsilon^e_{i,t+1}\epsilon_{\mathbf{t+1}}\epsilon_{\mathbf{t+1}}']\Psi'(\Delta' + \Psi')^s. \end{aligned} \tag{B.13}$$

Note that $S_{i\mathbf{V},t}$ depends only upon the conditional co-skewness of $r^e_{i,t+1}$ with the innovations in $\mathbf{f_{t+1}}$. To further simplify this expression, the following claim can be used.

Claim: *Let $\mathbf{B}$ and $\mathbf{C}$ and be matrices such that $\mathbf{BCB}'$ exists and the eigenvalues of $\mathbf{B}\otimes\mathbf{B}$ are all inside the unit circle. Suppose that ρ is a number such that $|\rho| \leq 1$. Then*

$$\sum_{s=0}^{\infty}\rho^s\mathbf{B}^s\mathbf{C}(\mathbf{B}')^s = vec^{-1}\left[(\mathbf{I} - \rho(\mathbf{B}\otimes\mathbf{B}))^{-1}vec(\mathbf{C})\right], \tag{B.14}$$

where $\otimes$ denotes the Kronecker product and vec^{-1} denotes the inverse of the 'vec' operator. □

Proof. Let

$$\mathbf{D} \equiv \sum_{s=0}^{\infty}\rho^s\mathbf{B}^s\mathbf{C}(\mathbf{B}')^s. \tag{B.15}$$

Then,

$$\rho\mathbf{BDB}' = \sum_{s=1}^{\infty}\rho^s\mathbf{B}^s\mathbf{C}(\mathbf{B}')^s, \tag{B.16}$$

$$\mathbf{D} - \rho\mathbf{BDB}' = \mathbf{C}. \tag{B.17}$$

By using Proposition 10.4 of Hamilton (1994),

$$\begin{aligned} \text{vec}(\mathbf{D}) - \rho(\mathbf{B}\otimes\mathbf{B})\text{vec}(\mathbf{D}) &= \text{vec}(\mathbf{C}), \\ \text{vec}(\mathbf{D}) &= (\mathbf{I} - \rho\mathbf{B}\otimes\mathbf{B})^{-1}\text{vec}(\mathbf{C}), \\ \sum_{s=0}^{\infty}\rho^s\mathbf{B}^s\mathbf{C}(\mathbf{B}')^s &= \text{vec}^{-1}\left[(\mathbf{I} - \rho(\mathbf{B}\otimes\mathbf{B}))^{-1}\text{vec}(\mathbf{C})\right]. \end{aligned} \tag{B.18}$$

□

By using this claim, Eq. (B.13) can be rewritten as

$$S_{i\mathbf{V},t} = \text{vec}^{-1}\left[(\mathbf{I} - \rho(\mathbf{\Delta} + \mathbf{\Psi}) \otimes (\mathbf{\Delta} + \mathbf{\Psi}))^{-1}\text{vec}(\mathbf{\Psi} E_t[\epsilon^e_{i,t+1}\epsilon_{\mathbf{t+1}}\epsilon_{\mathbf{t+1}}{}']\mathbf{\Psi}')\right]. \tag{B.19}$$

Finally, by putting everything together, under these distributional assumptions, Eq. (B.1) can be rewritten as

$$\begin{aligned} E_t[r^e_{i,t+1}] + \frac{V_{ii,t}}{2} &= \gamma_1 V_{iw,t} + \gamma_2 V_{ih,t} + \gamma_3 V_{i\nu,t}, \qquad \text{(B.20)} \\ E_t[r^e_{i,t+1}] &= (\mu + \mathbf{A}(\mathbf{f_t} - \mu))'\mathbf{ei}, \\ V_{ii,t} &= \mathbf{ei}'\mathbf{V}_t\mathbf{ei}, \\ V_{iw,t} &= (\mathbf{e1} + \mathbf{e2})'\mathbf{V}_t\mathbf{ei}, \\ V_{ih,t} &= \lambda_h{}'\mathbf{V}_t\mathbf{ei}, \\ V_{i\nu,t} &= (\mathbf{e1} + \mathbf{e2} + \lambda_{\mathbf{h}})' S_{i\mathbf{V},t}(\mathbf{e1} + \mathbf{e2} + \lambda_{\mathbf{h}}), \end{aligned}$$

where $\lambda_{\mathbf{h}}{}' = (\mathbf{e1} + \mathbf{e2})'\rho\mathbf{A}(\mathbf{I} - \rho\mathbf{A})^{-1}$ and $S_{i\mathbf{V},t}$ is defined by Eq. (B.19).

C Analytical calculation of the log-likelihood score of a K-dimensional VAR(1)-GARCH(1,1)

This appendix outlines how one can analytically calculate the score of the log-likelihood function of a K-dimensional VAR(1)-GARCH(1,1) system where the innovations are conditionally normal. Let $\mathbf{f_t}$ denote a K-element vector of variables. The conditional mean of $\mathbf{f_t}$ is given by a first-order K-dimensional VAR system defined by

$$\mathbf{f_{t+1}} - \mu = \mathbf{A}(\mathbf{f_t} - \mu) + \epsilon_{\mathbf{t+1}}, \tag{C.1}$$

where μ is a K-dimensional vector of unconditional means, $\mathbf{A}$ is a $K \times K$ parameter matrix, and $\epsilon_{\mathbf{t+1}}$ is a K-dimensional vector of innovations in the variables. The variance of these innovations is assumed to follow a multivariate GARCH(1,1) process.

Let $\mathbf{V}_t$ denote the conditional variance-covariance matrix of next-period innovation, ϵ_{t+1}, conditional upon information at time t: $\mathbf{V}_t = E_t(\epsilon_{t+1}\epsilon'_{t+1})$.[28] Following the BEKK representation of Engle and Kroner (1995), specify the dynamics of the conditional covariances as:

$$\mathbf{V}_t = \mathbf{P}\mathbf{P}' + \mathbf{\Delta}\mathbf{V_{t-1}}\mathbf{\Delta}' + \mathbf{\Psi}\epsilon_{\mathbf{t-1}}\epsilon'_{\mathbf{t-1}}\mathbf{\Psi}', \tag{C.2}$$

for a lower-triangular K-dimensional matrix, $\mathbf{P}$. For tractability, suppose that matrices $\mathbf{\Delta}$ and $\mathbf{\Psi}$ are also both K-dimensional lower-triangular matrices.[29]

Under the assumption that the innovations are conditionally normally distributed, the log-likelihood function of observations at times (0, $\cdots$, T) from the VAR(1)-GARCH(1,1) model described by Eqs. (C.1) and (C.2) is given by

$$\mathcal{L} = \sum_{t=1}^{T} \frac{-K}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{V}_t| - \frac{1}{2}\epsilon'_{t+1}\mathbf{V}_t^{-1}\epsilon_{t+1}. \tag{C.3}$$

[28] Note that the notation here differs from that found in the GARCH literature, in which $\mathbf{H}_t = E_{t-1}(\epsilon_t\epsilon'_t)$. I use the subscript t to denote the information available at time t rather than the innovation at time t.

[29] It is also common to assume that $\mathbf{\Delta}$ and $\mathbf{\Psi}$ are symmetric rather than lower-triangular. The derivation presented here can be easily modified to that case.

The matrices, $\mathbf{V}_t$, are calculated iteratively according to Eq. (C.2). Suppose that the initial condition $\mathbf{V}_0$ is given exogenously. Note that the parameters of the VAR(1)-GARCH(1,1) are μ, $\mathbf{A}$, $\mathbf{P}$, $\mathbf{\Delta}$, and $\mathbf{\Psi}$. Stack these parameters into a single vector, $\mathbf{\Xi}$, and let ξ denote an element of $\mathbf{\Xi}$.

The t-th score of the log-likelihood function, $\mathbf{s_t}$, is the derivative $\mathbf{s_t} \equiv \frac{\partial \mathcal{L}_t}{\partial \mathbf{\Xi}}$. Since the derivation is more tractable in terms of $\mathbf{V}_t^{-1}$, write the t-th log-likelihood as

$$\mathcal{L}_t = \frac{-K}{2}\log(2\pi) + \frac{1}{2}\log|\mathbf{V}_t^{-1}| - \frac{1}{2}\epsilon_{t+1}'\mathbf{V}_t^{-1}\epsilon_{t+1}. \tag{C.4}$$

Consider the derivative of $\log|\mathbf{V}_t^{-1}|$ with respect to the ij-th element of $\mathbf{V}_t^{-1}$. Note that

$$\frac{\partial}{\partial(V_t^{-1})_{ij}}\log|\mathbf{V}_t^{-1}| = \text{trace}(\mathbf{V}_t^{-1}I_{ij}) = (V_t)_{ji}, \tag{C.5}$$

where I_{ij} is a matrix with a 1 in the ij-th entry and zeros elsewhere. Therefore, the derivative of $\log|\mathbf{V}_t^{-1}|$ with respect to ξ is

$$\frac{\partial}{\partial\xi}\log|\mathbf{V}_t^{-1}| = \sum_{i=1}^{K}\sum_{j=1}^{K}\frac{\partial\log|\mathbf{V}_t^{-1}|}{\partial(V_t^{-1})_{ij}}\frac{\partial(V_t^{-1})_{ij}}{\partial\xi} = \sum_{i=1}^{K}\sum_{j=1}^{K}(V_t)_{ji}\frac{\partial(V_t^{-1})_{ij}}{\partial\xi}. \tag{C.6}$$

For notational simplicity, define an operator Ⓙ as follows. Let $\mathbf{B} = \{b_{ij}\}$ and $\mathbf{C} = \{c_{ij}\}$ be two matrices of the same size. Define the operator Ⓙ as

$$\mathbf{B} \text{Ⓙ} \mathbf{C} \equiv \sum_i\sum_j b_{ij}c_{ij}. \tag{C.7}$$

With this operator, the derivative of $\log|\mathbf{V}_t^{-1}|$ with respect to ξ can now be written as

$$\frac{\partial}{\partial\xi}\log|\mathbf{V}_t^{-1}| = \mathbf{V}_t' \text{Ⓙ} \frac{\partial\mathbf{V}_t^{-1}}{\partial\xi}. \tag{C.8}$$

Consider now the derivative of $\epsilon_{t+1}'\mathbf{V}_t^{-1}\epsilon_{t+1}$ with respect to the ij-th element of $\mathbf{V}_t^{-1}$. Since ϵ_{t+1} does not depend upon $(V_t^{-1})_{ij}$,

$$\frac{\partial}{\partial(V_t^{-1})_{ij}}\epsilon_{t+1}'\mathbf{V}_t^{-1}\epsilon_{t+1} = \epsilon_{t+1}'I_{ij}\epsilon_{t+1} = (\epsilon_{t+1}\epsilon_{t+1}')_{ij}. \tag{C.9}$$

Therefore, since $\mathbf{V}_t^{-1}$ is symmetric, the derivative of $\epsilon_{t+1}'\mathbf{V}_t^{-1}\epsilon_{t+1}$ with respect to ξ is :

$$\begin{aligned}\frac{\partial}{\partial\xi}\epsilon_{t+1}'\mathbf{V}_t^{-1}\epsilon_{t+1} &= \sum_{i=1}^{K}\sum_{j=1}^{K}(\epsilon_{t+1}\epsilon_{t+1}')_{ij}\frac{\partial(V_t^{-1})_{ij}}{\partial\xi} + 2\epsilon_{t+1}'\mathbf{V}_t^{-1}\frac{\partial\epsilon_{t+1}}{\partial\xi},\\ &= (\epsilon_{t+1}\epsilon_{t+1}')\text{Ⓙ}\frac{\partial\mathbf{V}_t^{-1}}{\partial\xi} + 2\epsilon_{t+1}'\mathbf{V}_t^{-1}\frac{\partial\epsilon_{t+1}}{\partial\xi}. \end{aligned}\tag{C.10}$$

By using Eqs. (C.8) and (C.10) and the fact that $\frac{\partial\mathbf{B}^{-1}}{\partial\xi} = -\mathbf{B}^{-1}\frac{\partial\mathbf{B}}{\partial\xi}\mathbf{B}^{-1}$, the t-th score with respect to ξ can be written as

$$\frac{\partial\mathcal{L}_t}{\partial\xi} = \frac{1}{2}\left[\epsilon_{t+1}\epsilon_{t+1}' - \mathbf{V}_t'\right] \text{Ⓙ} \left[\mathbf{V}_t^{-1}\frac{\partial\mathbf{V}_t}{\partial\xi}\mathbf{V}_t^{-1}\right] - \epsilon_{t+1}'\mathbf{V}_t^{-1}\frac{\partial\epsilon_{t+1}}{\partial\xi}. \tag{C.11}$$

From Eq. (C.1), the term $\frac{\partial \epsilon_{t+1}}{\partial \xi}$ can be calculated as:

$$\frac{\partial \epsilon_{t+1}}{\partial \Xi} = \Big[\, -(\mathbf{I}_K - \mathbf{A}) | - \mathbf{I}_K \otimes (\mathbf{f_t} - \mu)' | \mathbf{0} \Big] \,, \tag{C.12}$$

where $\mathbf{I}_K$ denotes the K-dimensional identity matrix and $\otimes$ denotes the Kronecker product.

The term $\frac{\partial \mathbf{V}_t}{\partial \xi}$ still needs to be derived. It turns out that $\frac{\partial \mathbf{V}_t}{\partial \xi}$ must be calculated iteratively, since it depends on $\frac{\partial \mathbf{V}_{t-1}}{\partial \xi}$ and $\frac{\partial \epsilon_t}{\partial \xi}$. Note that since V_0 is taken to be exogenous the starting condition for this iteration is $\frac{\partial \mathbf{V}_0}{\partial \xi} = \mathbf{0}$ and $\frac{\partial \epsilon_1}{\partial \xi}$.

Consider first the case in which ξ appears in either μ or $\mathbf{A}$. Note that since

$$\frac{\partial}{\partial \xi} \epsilon_t \epsilon_t' = \epsilon_t (\frac{\partial \epsilon_t}{\partial \xi})' + (\frac{\partial \epsilon_t}{\partial \xi}) \epsilon_t', \tag{C.13}$$

we have

$$\frac{\partial \mathbf{V}_t}{\partial \xi}|_{\{\xi \in (\mu, \mathbf{A})\}} = \Delta \frac{\partial \mathbf{V}_{t-1}}{\partial \xi} \Delta' + \Psi \left[\epsilon_t (\frac{\partial \epsilon_t}{\partial \xi})' + (\frac{\partial \epsilon_t}{\partial \xi}) \epsilon_t' \right] \Psi'. \tag{C.14}$$

Now note that, in general,

$$\frac{\partial}{\partial \mathbf{B}_{ij}} \mathbf{BCB}' = \mathbf{BCI}_{ij}' + \mathbf{I}_{ij} \mathbf{CB}', \tag{C.15}$$

so that

$$\frac{\partial \mathbf{V}_t}{\partial \xi}|_{\{\xi = (\mathbf{P})_{ij}\}} = \Delta \frac{\partial \mathbf{V}_{t-1}}{\partial \xi} \Delta' + \mathbf{PI}_{ij}' + \mathbf{I}_{ij} \mathbf{P}', \tag{C.16}$$

$$\frac{\partial \mathbf{V}_t}{\partial \xi}|_{\{\xi = (\mathbf{\Delta})_{ij}\}} = \Delta \frac{\partial \mathbf{V}_{t-1}}{\partial \xi} \Delta' + \mathbf{\Delta} \mathbf{V}_{t-1} \mathbf{I}_{ij}' + \mathbf{I}_{ij} \mathbf{V}_{t-1} \mathbf{\Delta}', \tag{C.17}$$

$$\frac{\partial \mathbf{V}_t}{\partial \xi}|_{\{\xi = (\mathbf{\Psi})_{ij}\}} = \Delta \frac{\partial \mathbf{V}_{t-1}}{\partial \xi} \Delta' + \mathbf{\Psi} (\epsilon_t \epsilon_t') \mathbf{I}_{ij}' + \mathbf{I}_{ij} (\epsilon_t \epsilon_t') \mathbf{\Psi}'. \tag{C.18}$$

Eqs. (C.11), (C.12), (C.14), (C.16), (C.17), (C.18) and the starting condition $\frac{\partial \mathbf{V}_0}{\partial \xi} = \mathbf{0}$ fully characterize the score of the log-likelihood of the K-dimensional VAR(1)-GARCH(1,1) with conditionally normal innovations.

D Construction of bounds on omitted factors

This appendix derives the bounds on the Sharpe ratio for the factor-mimicking portfolio associated with a factor that may have been omitted from the pricing kernel. One example of an omitted factor is the "surplus consumption ratio" of Campbell and Cochrane (1999). Bekaert and Grenadier (2001) and Barberis, Huang, and Santos (2001) offer other state variables that enter the pricing kernel by affecting preferences. Alternatively, the omitted factor could be interpreted as the "true" changes in the forecasts of future market returns, which could not be observed perfectly in the data.

Suppose that the true pricing kernel has the form

$$M_{t+1} = \beta^{\theta_1} \left(\frac{C_{t+1}}{C_t} \right)^{\frac{-\theta_2}{\sigma}} (R_{W,t+1})^{\theta_3 - 1} (X_{t+1})^{\gamma_x} \,, \tag{D.1}$$

where the variable X_t represents an omitted component of the pricing kernel and γ_x is a new parameter. Under this pricing kernel, the asset pricing relation derived in Appendix A becomes

$$E_t[r_{i,t+1}^e] + \frac{V_{ii,t}}{2} = \gamma V_{iw,t} + (\gamma - 1) V_{ih,t} - \frac{1}{2} (\gamma - 1)^2 V_{i\nu,t} + \gamma_x V_{ix,t}, \tag{D.2}$$

where x_t is the logarithm of X_t and $V_{ix,t} \equiv \text{Cov}_t(r^e_{i,t+1}, x_{t+1})$.

Additionally, suppose that the information in the omitted factor x_t comprises only the portion that is orthogonal to the factors already included in the analysis, except possibly the asset the return of which is to be explained. To formalize this, let i denote any investible factor included in the set of factors, $\mathbf{f_t}$, and suppose that the omitted factor covaries with i but not with any other variable in the VAR. Assume that the additional forecasting information introduced by x_t is independent of the forecasting information already documented in the original set of variables, $\mathbf{f_t}$. This is accomplished by taking the true model for the dynamics of the economy as being given by

$$\begin{bmatrix} \mathbf{f_{t+1}} - \mathbf{E}[\mathbf{f}] \\ x_{t+1} - E[x] \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{A_{fx}} \\ \mathbf{0} & A_{xx} \end{bmatrix} \begin{bmatrix} \epsilon_{\mathbf{f},\mathbf{t+1}} \\ \epsilon_{x,t+1} \end{bmatrix}, \tag{D.3}$$

where $\mathbf{A}$ is the original VAR forecasting parameters under the omitted factor model of Eq. (B.2). Finally, suppose that x_t is not co-skewed with any of the included factors in $\mathbf{f_t}$. Since the omitted variable covaries only with i, the omitted variable bias only affects the VAR estimates of the parameters associated with i. Therefore, since some of the forecasting ability of i may be instrumenting for the forecasting ability of the omitted variable x_t, the bound constructed should be interpreted as the amount of new information x_t must provide in addition to that already captured.

Since the introduction of the omitted factor does not change the forecasting power of the included factors, the impulse-response function for the included factors under the omitted factor model is the same as that under the true model. Therefore, the price of risk under the true model is the same as in the omitted factor model. From Eq. (30) under the true model, if i is investible, the expected return on the i-th asset is:

$$\begin{aligned} E_t[r^e_{i,t+1}] + \frac{V_{ii,t}}{2} &= \mathbf{\Lambda}'\mathbf{V}_{\mathbf{if},\mathbf{t}} - \tilde{\mathbf{\Lambda}}'\mathbf{S}_{\mathbf{iV},\mathbf{t}}\tilde{\mathbf{\Lambda}} + \Lambda_x V_{ix,t}, \qquad \text{(D.4)} \\ \text{where} \qquad \mathbf{\Lambda} &= \gamma(\mathbf{e1}+\mathbf{e2}) + (\gamma-1)\lambda_{\mathbf{h}}, \\ \tilde{\mathbf{\Lambda}} &= \frac{1}{\sqrt{2}}(\gamma-1)(\mathbf{e1}+\mathbf{e2}+\lambda_{\mathbf{h}}). \end{aligned}$$

The parameter Λ_x is the sum of γ_x and the x-th element of $\mathbf{\Lambda}$ under the true VAR system. Similarly, if x_t is an investible factor, its expected return can be written as:

$$E_t[x_{t+1}] + \frac{V_{xx,t}}{2} = \mathbf{\Lambda}'\mathbf{V}_{\mathbf{xf},\mathbf{t}} - \tilde{\mathbf{\Lambda}}'\mathbf{S}_{\mathbf{xV},\mathbf{t}}\tilde{\mathbf{\Lambda}} + \Lambda_x V_{xx,t}, \tag{D.5}$$

$$= \Lambda_i V_{ix,t} + \Lambda_x V_{xx,t}, \tag{D.6}$$

where Λ_i denotes the i-th element of $\mathbf{\Lambda}$. Notice that Λ_i in the true model is the same as in the omitted variable model, since the lower-left elements of the true VAR system are zeros. Deriving the last equality uses the assumptions 1) that the omitted factor doesn't covary with the included factors except with i, and 2) that x_t is not co-skewed with the elements of $\mathbf{f_t}$.

Taking the unconditional expectations of Eqs. (D.4) and (D.5), combining the two equations, and solving for Λ_x yields

$$\begin{aligned} E[r^e_i] + \frac{V_{ii}}{2} - (\mathbf{\Lambda}'\mathbf{V_{if}} - \tilde{\mathbf{\Lambda}}'\mathbf{S_{iV}}\tilde{\mathbf{\Lambda}}) &= \rho_{ix}\frac{\sqrt{V_{ii}}}{\sqrt{V_{xx}}}\left(E[x] + \frac{V_{xx}}{2} - \Lambda_i V_{ix}\right) \qquad \text{(D.7)} \\ &= \rho_{ix}\sqrt{V_{ii}}\left(\frac{E[x]+\frac{V_{xx}}{2}}{\sqrt{V_{xx}}}\right) - \rho^2_{ix}V_{ii}\Lambda_i, \end{aligned}$$

where ρ_{ix} is the correlation between r^e_i, and the omitted factor, x. By rearranging and using the fact that the correlation must be less than 1,

$$\left(\frac{E[r^e_i] + \frac{V_{ii}}{2} - (\mathbf{\Lambda}'\mathbf{V_{if}} - \tilde{\mathbf{\Lambda}}'\mathbf{S_{iV}}\tilde{\mathbf{\Lambda}})}{\sqrt{V_{ii}}}\right) + \rho^2_{ix}V_{ii}\Lambda_i \leq \left(\frac{E[x]+\frac{V_{xx}}{2}}{\sqrt{V_{xx}}}\right). \tag{D.8}$$

Suppose Λ_i is non-negative[30], then

$$\left(\frac{E[r_i^e] + \frac{V_{ii}}{2} - (\mathbf{\Lambda}'\mathbf{V_{if}} - \tilde{\mathbf{\Lambda}}'\mathbf{S_{iV}}\tilde{\mathbf{\Lambda}})}{\sqrt{V_{ii}}} \right) \leq \left(\frac{E[x] + \frac{V_{xx}}{2}}{\sqrt{V_{xx}}} \right). \tag{D.9}$$

Eq. (D.9) provide a lower bound for the Sharpe ratio of the omitted factor, which is the expected return of the factor, plus the adjustment for the Jensen's inequality effect, divided by the standard deviation of the factor. The numerator of the lower bound includes the term $E[r_i^e] + \frac{V_{ii}}{2}$ minus $\mathbf{\Lambda}'\mathbf{V_{if}} - \tilde{\mathbf{\Lambda}}'\mathbf{S_{iV}}\tilde{\mathbf{\Lambda}}$, which is the portion of expected returns not explained by the model. The denominator is the standard deviation of i. Hence, the lower bound for the Sharpe ratio of the omitted factor is roughly the Sharpe ratio of the unexplained portion of returns.

[30] The negative Λ_i case can be considered by replacing $r_{i,t+1}^e$ with $-r_{i,t+1}^e$.

References

Andrews, Donald W. K., 1991, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica* 59, 817–858.

Ang, Andrew, and Geert Bekaert, 2001, Stock return predictability: Is it there?, Working paper, Columbia University.

Ang, Andrew, and Joseph S. Chen, 2003, CAPM over the long-run: 1926-2001, Working paper, University of Southern California.

——— , and Yuhang Xing, 2002, Downside correlation and expected stock returns, Working paper, University of Southern California.

Barberis, Nicholas, Ming Huang, and Tano Santos, 2001, Prospect theory and asset prices, *Quarterly Journal of Economics* 116, 1–53.

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny, 1998, A model of investor sentiment, *Journal of Financial Economics* 49, 307–343.

Bekaert, Geert, and Steve R. Grenadier, 2001, Stock and bond pricing in an affine economy, Working paper, Columbia University.

Bekaert, Geert, and Robert J. Hodrick, 2001, Expectations hypotheses tests, *Journal of Finance* 56, 1357–1394.

Bekaert, Geert, and Guojun Wu, 2000, Asymmetric volatility and risk in equity markets, *Review of Financial Studies* 13, 1–42.

Bollerslev, Tim, and Jeffrey M. Wooldridge, 1992, Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances, *Econometric Reviews* 11, 143–172.

Breeden, Douglas T., 1979, An intertemporal asset pricing model with stochastic consumption and investment opportunities, *Journal of Financial Economics* 7, 265–296.

Brennan, Michael J., Ashley W. Wang, and Yihong Xia, 2003, Estimation and test of a simple model of intertemporal capital asset pricing, *Journal of Finance* forthcoming.

Campbell, John Y., 1987, Does saving anticipate declining future labor income?, *Econometrica* 55, 1249–1273.

——— , 1991, A variance decomposition for stock returns, *Economic Journal* 101, 157–179.

——— , 1993, Intertemporal asset pricing without consumption data, *American Economic Review* 83, 487–512.

——— , 1996, Understanding risk and return, *Journal of Political Economy* 104, 298–345.

——— , and John H. Cochrane, 1999, By force of habit: A consumption-based explanation of aggregate stock market behavior, *Journal of Political Economy* 107, 205–251.

Campbell, John Y., and Robert J. Shiller, 1988, Stock prices, earnings, and expected dividends, *Journal of Finance* 43, 661–676.

Campbell, John Y., and Luis M. Viceira, 2002, *Strategic asset allocation: Portfolio choice for long-term investors* (Oxford University Press: Oxford, UK).

Campbell, John Y., and Tuomo Vuolteenaho, 2002, Bad beta, good beta, Working paper, Harvard University—.

Carhart, Mark M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.

Chen, Joseph S., Harrison G. Hong, and Jeremy C. Stein, 2002, Breadth of ownership and stock returns, *Journal of Financial Economics* 2-3, 171–205.

Chen, Nai-Fu, Richard Roll, and Stephen A. Ross, 1986, Economic forces and the stock market, *Journal of Business* 59, 383–403.

Cochrane, John H., 2001, *Asset pricing* (Princeton University Press: Princeton, NJ).

Cumby, Robert E., and John Huizinga, 1992, Testing the autocorrelation structure of disturbances in ordinary least squares and instrumental variables regressions, *Econometrica* 60, 185–195.

Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam, 1998, Investor psychology and security market under- and overreaction, *Journal of Finance* 53, 1839–1885.

Daniel, Kent, and Sheridan Titman, 1997, Evidence on the characteristics of cross-sectional variation in stock returns, *Journal of Finance* 52, 1–33.

Dittmar, Robert F., 2002, Nonlinear pricing kernels, kurtosis preference, and evidence from the cross-section of equity returns, *Journal of Finance* 57, 369–403.

Engle, Robert F., and Kenneth F. Kroner, 1995, Multivariate simultaneous generalized ARCH, *Econometric Theory* 11, 122–150.

Epstein, Larry G., and Stanley E. Zin, 1989, Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework, *Econometrica* 57, 937–969.

Fama, Eugene F., 1970, Multiperiod consumption-investment decisions, *American Economic Review* 60, 163–174.

———, 1991, Efficient markets II, *Journal of Finance* 46, 1575–1618.

———, and Kenneth R. French, 1988a, Dividend yields and expected stock returns, *Journal of Financial Economics* 22, 3–25.

———, 1988b, Permanent and temporary components of stock prices, *Journal of Political Economy* 96, 246–273.

———, 1989, Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* 25, 23–49.

———, 1992, The cross-section of expected stock returns, *Journal of Finance* 47, 427–465.

———, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

———, 1995, Size and book-to-market factors in earnings and returns, *Journal of Finance* 50, 131–156.

———, 1996, Multifactor explanations of asset pricing anomalies, *Journal of Finance* 51, 55–84.

Fama, Eugene F., and James D. MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 71, 607–636.

Ferson, Wayne E., and Campbell R. Harvey, 1991, The variation of economic risk premiums, *Journal of Political Economy* 99, 385–415.

———, 1999, Conditioning variables and the cross-section of stock returns, *Journal of Finance* 54, 1325–1360.

Ferson, Wayne E., Sergei Sarkissian, and Timothy Simin, 1999, The alpha factor asset pricing model: A parable, *Journal of Financial Markets* 2, 49–68.

Glosten, Lawrence R., Ravi Jagannathan, and David E. Runkle, 1993, On the relation between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance* 48, 1779–1801.

Gomes, Joao, Leonid Kogan, and Lu Zhang, 2002, Equilibrium cross-section of returns, *Journal of Political Economy* forthcoming.

Hamilton, James D., 1994, *Time series analysis* (Princeton University Press: Princeton, New Jersey).

Hansen, Lars P., and Scott F. Richard, 1987, The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models, *Econometrica* 55, 587–613.

Hansen, Lars P., and Kenneth J. Singleton, 1983, Stochastic consumption, risk aversion, and the temporal behavior of asset returns, *Journal of Political Economy* 91, 249–265.

Harvey, Campbell R., 1989, Forecasts of economic growth from the bond and stock markets, *Financial Analysts Journal* 45, 38–45.

———, and Akhtar Siddique, 2000, Conditional skewness in asset pricing tests, *Journal of Finance* 55, 1263–1295.

Hodrick, Robert J., 1992, Dividend yields and expected stock returns: Alternative procedures for inference and measurement, *Review of Financial Studies* 5, 357–386.

———, David T.C. Ng, and Paul Sengmueller, 1999, An international dynamic asset pricing model, *International Tax and Public Finance* 6, 597–620.

Hong, Harrison G., and Jeremy C. Stein, 1999, A unified theory of underreaction, momentum trading, and overreaction in asset markets, *Journal of Finance* 54, 2143–2184.

Jagannathan, Ravi, and Zhenyu Wang, 1996, The conditional CAPM and the cross-section of expected returns, *Journal of Finance* 51, 3–53.

Keim, Donald B., and Robert F. Stambaugh, 1986, Predicting returns in bond and stock markets, *Journal of Financial Economics* 17, 357–390.

Kirby, Chris, 1998, The restrictions on predictability implied by rational asset pricing models, *Review of Financial Studies* 11, 343–382.

Lakonishok, Josef, Andrei Shleifer, and Robert W. Vishny, 1994, Contrarian investment, extrapolation, and risk, *Journal of Finance* 49, 1541–1578.

Lettau, Martin, and Sydney Ludvigson, 2001a, Consumption, aggregate wealth and expected stock returns, *Journal of Finance* 56, 815–849.

———, 2001b, Resurrecting the {C}CAPM: A cross-sectional test when risk premia are time-varying, *Journal of Political Economy* 109, 1238–1287.

Liew, Jimmy, and Maria Vassalou, 2000, Can book-to-market, size and momentum be risk factors that predict economic growth?, *Journal of Financial Economics* 57, 221–245.

MacKinlay, A. Craig, 1995, Multifactor models do not explain deviations from the CAPM, *Journal of Financial Economics* 38, 3–28.

Merton, Robert C., 1973, An intertemporal capital asset pricing model, *Econometrica* 41, 867–887.

Moskowitz, Tobias, 2003, An analysis of (covariance) risk and pricing anomalies, *Review of Financial Studies* 16, 417–457.

Newey, Whitney K., and Kenneth D. West, 1987, A simple positive-definite heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

Ng, David T.C., 2002, The international CAPM when expected returns are time-varying, *Journal of International Money and Finance* forthcoming.

Perez-Quiros, Gabriel, and Allan Timmermann, 2000, Firm size and cyclical variations in stock returns, *Journal of Finance* 55, 1229–1262.

Petkova, Ralitsa, and Lu Zhang, 2003, Is value riskier than growth?, Working paper, University of Rochester.

Poterba, James M., and Lawrence H. Summers, 1988, Mean reversion in stock prices: Evidence and implications, *Journal of Financial Economics* 22, 27–59.

Restoy, Fernando, and Philippe Weil, 1998, Approximate equilibrium asset prices, NBER Working Paper No. 6611.

Roll, Richard, 1977, A critique of the asset pricing theory's tests: Part I, *Journal of Financial Economics* 4, 129–176.

Rosenberg, Barr, Kenneth Reid, and Ronald Lanstein, 1985, Persuasive evidence of market inefficiency, *Journal of Portfolio Management* 11, 9–17.

Ross, Stephen A., 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341–360.

Schwarz, Gideon, 1978, Estimating the dimension of a model, *Annals of Statistics* 48, 1–48.

Vassalou, Maria, 2003, News related to future GDP growth as risk factors in equity returns, *Journal of Financial Economics* 68, 47–73.

Table 1: Summary statistics

Panel A: Variables in the VAR system

Variable	Annualized Mean (%)	Annualized Std. (%)	Autocorrelation Lags 1	2	3	12
emkt	7.93	14.71	0.07	-0.03	0.00	0.04
rrf	1.32	0.90	0.42	0.41	0.35	0.31
bk-mkt	9.44	10.55	0.18	0.06	-0.03	0.20
dyld	3.57	1.01	0.99	0.98	0.96	0.81
term	0.68	0.91	0.96	0.90	0.84	0.51
dflt	0.87	0.37	0.97	0.93	0.90	0.67

Panel B: Contemporaneous correlations of the VAR variables

	emkt	rrf	bk-mkt	dyld	term	dflt
emkt	1.00					
rrf	0.11	1.00				
bk-mkt	-0.37	-0.02	1.00			
dyld	-0.04	-0.08	-0.03	1.00		
term	0.16	0.11	0.11	-0.17	1.00	
dflt	0.08	0.20	0.01	0.48	0.11	1.00

The data set consists of monthly observations from April 1953 to December 1999 (N = 561). The first two variables form a proxy for the real rate of return on the aggregate market portfolio. The variable *emkt* is the rate of return on the value-weighted stock market portfolio in excess of the risk-free rate, and *rrf* is the risk-free rate in excess of the rate of inflation (CPI growth rate). The one-month Treasury bill is used as the risk-free rate. The next variable, *bk-mkt*, is the return on an equal-weighted portfolio that goes long value stocks and shorts growth stocks and represents the book-to-market effect. The final three variables are macroeconomic variables that have been found to possess forecasting ability over future market return: *dyld* is the annualized dividend yield on the S&P 500 stock index, *term* is the annualized yield spread between the 10-year and the 1-year U.S. Treasury bonds, and *dflt* is the annualized yield spread between Baa-rated and Aaa-rated corporate bonds. All variables are expressed as continuously compounded annual returns.

Panel A shows the summary statistics of the variables in the VAR systems. The first column of Panel A shows the annualized means with an adjustment for Jensen's inequality, and the second column of Panel A shows the annualized standard deviations. The last four columns show the autocorrelations at 1-month, 2-month, 3-month, and 12-month lags. Panel B shows the contemporaneous correlations among these variables.

Table 2: Specification tests of the VAR system

Panel A: Selection criteria for the VAR order

Lags	0	1	2	3
Schwarz	-55.6	-65.7	-65.6	-65.4

Panel B: l-tests for residual serial correlations

	emkt	rrf	bk-mkt	dyld	term	dflt
$\chi^2(11)$	8.0	54.3	3.3	14.8	19.3	16.0
[p-val]	[0.71]	[0.00]	[0.99]	[0.19]	[0.06]	[0.14]
$\chi^2(12)$	9.9	54.3	9.8	15.8	19.8	16.2
[p-val]	[0.62]	[0.00]	[0.64]	[0.20]	[0.07]	[0.18]

Panel C Joint forecastability tests

	emkt	rrf	bk-mkt	dyld	term	dflt
$\chi^2(7)$	20.3	68.2	19.8	10182	4006	5106
[p-val]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]

Panel A shows the Schwarz (1978) selection criterion for the lag order of the VAR system. The appropriate lag length minimizes the Schwarz criterion.

Panel B presents the l-tests of Cumby and Huizinga (1992). The l-tests are conducted on the residuals of each variable individually. These tests examine whether the serial correlations of the residuals are jointly zero at lags of order 1 through 11 and at 1 through 12. The asymptotic distributions of the l-statistics are $\chi^2(11)$ and $\chi^2(12)$, respectively.

Panel C shows the Wald statistics for the null hypothesis of joint predictability of each variable in the VAR system. The statistics test whether the coefficient estimates on the lagged VAR variables are jointly zero. The test statistics are constructed using a GMM heteroskedasticity- and autocorrelation-consistent estimator and include adjustments for serial correlations using Newey-West (1987). The optimal order of Newey-West lags is determined by the procedure of Andrews (1991) to be three for each system. The asymptotic distribution of the test statistics is $\chi^2(7)$.

Table 3: VAR forecasting equations

Panel A: VAR parameter estimates

μ|$\mathbf{A}$	Mean	emkt$_t$	rrf$_t$	bk-mkt$_t$	dyld$_t$	term$_t$	dflt$_t$	adj. Rsq
emkt$_{t+1}$	0.473*	0.039	0.675	-0.118*	0.284	0.488*	-0.078	2.0%
(s.e.)	(0.135)	(0.056)	(0.713)	(0.058)	(0.203)	(0.206)	(0.643)	
rrf$_{t+1}$	0.111*	-0.002	0.472*	0.004	-0.022*	-0.005	0.084*	18.1%
(s.e.)	(0.022)	(0.003)	(0.053)	(0.003)	(0.009)	(0.010)	(0.027)	
bk-mkt$_{t+1}$	0.679*	0.111*	0.122	0.264*	-0.138	0.221	0.102	5.5%
(s.e.)	(0.201)	(0.038)	(0.513)	(0.050)	(0.159)	(0.155)	(0.461)	
dyld$_{t+1}$	3.103*	-0.015*	-0.033	0.002	0.990*	-0.014*	-0.018	98.8%
(s.e.)	(0.325)	(0.001)	(0.022)	(0.002)	(0.005)	(0.005)	(0.020)	
term$_{t+1}$	0.472*	0.000	0.088	0.005	0.009	0.953*	0.085	92.5%
(s.e.)	(0.165)	(0.003)	(0.057)	(0.005)	(0.012)	(0.019)	(0.054)	
dflt$_{t+1}$	0.794*	-0.004*	0.004	-0.004*	0.015*	-0.013*	0.957*	94.8%
(s.e.)	(0.082)	(0.001)	(0.020)	(0.002)	(0.005)	(0.006)	(0.014)	

Panel B: Variance-covariance parameter estimates

$\mathbf{P}$*100	emkt$_t$	rrf$_t$	bk-mkt$_t$	dyld$_t$	term$_t$	dflt$_t$
emkt$_t$	3.601*					
(s.e.)	(0.474)					
rrf$_t$	-0.048	0.128*				
(s.e.)	(0.026)	(0.047)				
bk-mkt$_t$	-0.645*	0.355	-0.428*			
(s.e.)	(0.258)	(0.199)	(0.209)			
dyld$_t$	-0.086*	-0.043	-0.043*	0.037		
(s.e.)	(0.011)	(0.028)	(0.021)	(0.031)		
term$_t$	0.041*	-0.007	-0.131	-0.198*	0.061	
(s.e.)	(0.020)	(0.022)	(0.096)	(0.062)	(0.149)	
dflt$_t$	0.006	0.012	0.002	0.006	0.084*	-0.002
(s.e.)	(0.005)	(0.010)	(0.028)	(0.045)	(0.008)	(0.011)

Panel C: GARCH parameter estimates

$\mathbf{\Delta}$	emkt$_t$	rrf$_t$	bk-mkt$_t$
emkt$_t$	0.488*		
(s.e.)	(0.183)		
rrf$_t$	0.029*	0.001	
(s.e.)	(0.013)	(0.001)	
bk-mkt$_t$	0.031		0.938*
(s.e.)	(0.069)		(0.022)

$\mathbf{\Psi}$	emkt$_t$	rrf$_t$	bk-mkt$_t$
emkt$_t$	0.142		
(s.e.)	(0.079)		
rrf$_t$	-0.011	0.617*	
(s.e.)	(0.012)	(0.121)	
bk-mkt$_t$	-0.159*		0.047*
(s.e.)	(0.039)		(0.022)

This table shows the parameter estimates of the VAR(1)-GARCH(1,1) model of the multivariate system containing the excess stock market return (*emkt*), the real risk-free rate (*rrf*), the return on the book-to-market portfolio (*bk-mkt*), the dividend yield (*dyld*), the term spread (*term*), and the default spread (*dflt*). The parameters of the model are estimated using QMLE with analytical gradients. The standard errors are calculated by GMM with the scores of the QMLE as the orthogonality conditions using a heteroskedasticity- and autocorrelation-consistent estimator with Newey-West (1987) lags of order three. Parameter estimates marked by an asterisk (*) are asymptotically significant at the 5% level.

Table 4: Estimated coefficients of (co-)skewness

emkt_t	emkt_t	rrf_t	bk-mkt_t	rrf_t	rrf_t	bk-mkt_t	bk-mkt_t	bk-mkt_t
emkt_t	-0.767*							
(s.e.)	(0.186)							
rrf_t	-0.038	-0.066		rrf_t	-0.352*			
(s.e.)	(0.054)	(0.053)		(s.e.)	(0.152)			
bk-mkt_t	0.280*	-0.040	0.263*	bk-mkt_t	0.002	0.034	bk-mkt_t	0.138
(s.e.)	(0.104)	(0.031)	(0.104)	(s.e.)	(0.055)	(0.058)	(s.e.)	(0.187)

This table shows the estimates of the (co-)skewness of the residuals from the VAR-GARCH models. The i, j-th entry of the k-th block indicates the estimate of $s_{ijk} = E[\epsilon_{i,t}\epsilon_{j,t}\epsilon_{k,t}]$. The estimates are expressed as coefficients of (co-)skewness by scaling them by the standard deviations of $\epsilon_{i,t}$, $\epsilon_{j,t}$, and $\epsilon_{k,t}$. Only the (co-)skewness relating to ***emkt***, ***rrf***, and ***bk-mkt*** are shown for brevity. The standard error estimates are obtained using a heteroskedasticity- and autocorrelation-consistent estimator with Newey-West (1987) lags of order three. Parameter estimates marked by an asterisk (*) are asymptotically significant at the 5% level.

Table 5: Weighted sums of impulse-responses, λ_h

	emkt_t	rrf_t	bk-mkt_t	dyld_t	term_t	dflt_t
$1200 * \lambda_h'$	-0.210*	1.352	-0.061	14.943*	2.446	1.412
(s.e.)	(0.072)	(1.033)	(0.078)	(4.895)	(2.328)	(5.506)

This table shows the weighted sums of the impulse-responses of the real market returns across all future horizons, $\lambda_h = (\mathbf{e1} + \mathbf{e2})'\rho\mathbf{A}(\mathbf{I} - \rho\mathbf{A})^{-1}$. The units are in terms of the weighted sums of percentage cumulative future returns (%) in response to a 1% shock in a variable today. The impulse-responses are exponentially weighted at the rate $\rho = 0.9949$.

The standard errors are calculated using a heteroskedasticity- and autocorrelation-consistent estimator with Newey-West (1987) lags of order three. Parameter estimates marked by an asterisk (*) are significant at the 5% level.

Table 6: Average contributions to expected returns

Panel A: From hedge against the current market return

$1200 * E[\mathbf{V_{iw}}]$	emkt$_t$	rrf$_t$	bk-mkt$_t$
$1200 * E[\mathbf{V_{iw}}]$	2.591*	0.017	-1.331
(s.e.)	(0.569)	(0.040)	(0.738)

Panel B: From hedge against changes in the forecasts of expected return

	emkt$_t$	rrf$_t$	bk-mkt$_t$
$1200 * E[\mathbf{V_{ih}}]$	-0.959*	0.005	-0.653
(s.e.)	(0.420)	(0.009)	(1.500)

Panel C: From hedge against changes in the forecasts of volatility

	emkt$_t$	rrf$_t$	bk-mkt$_t$
$1200 * \mathbf{V_{i\nu}}$	-0.002	0.000	0.000
(s.e.)	(0.003)	(0.000)	(0.001)

This table shows the average contributions to annualized percentage expected returns of *emkt*, *rrf*, and bk-mkt$_t$ from various sources of risk. The three sources of risk are the current market return, the changes in forecasts of future expect returns, and the changes in forecasts of future volatilities. The contributions are inferred from the estimates of VAR-GARCH parameters and unconditional skewness. The impulse-responses are exponentially weighted at the rate $\rho = 0.9949$. The standard errors are calculated using a heteroskedasticity- and autocorrelation-consistent estimator with Newey-West (1987) lags of order three. Parameter estimates marked by an asterisk (*) are significant at the 5% level.

Panel A shows the average covariances with the real market returns. These covariances can be interpreted as the average contributions to expected returns from the market beta. The contributions are expressed as annualized percentage returns for a unit increase in γ.

Panel B shows the discounted sums of the impulse-responses scaled by the unconditional covariance matrix. These estimates can be interpreted as the average contributions to returns for a unit increase in $(\gamma - 1)$ from hedging against changes in forecasts of future expected returns.

Panel C shows the cumulative impact on future volatility. These estimates can be interpreted by the asset pricing relation of Eq. (21) as the contributions to expected returns for a unit decrease in $\frac{1}{2}(\gamma - 1)^2$ from hedging against changes in forecasts of future volatility.

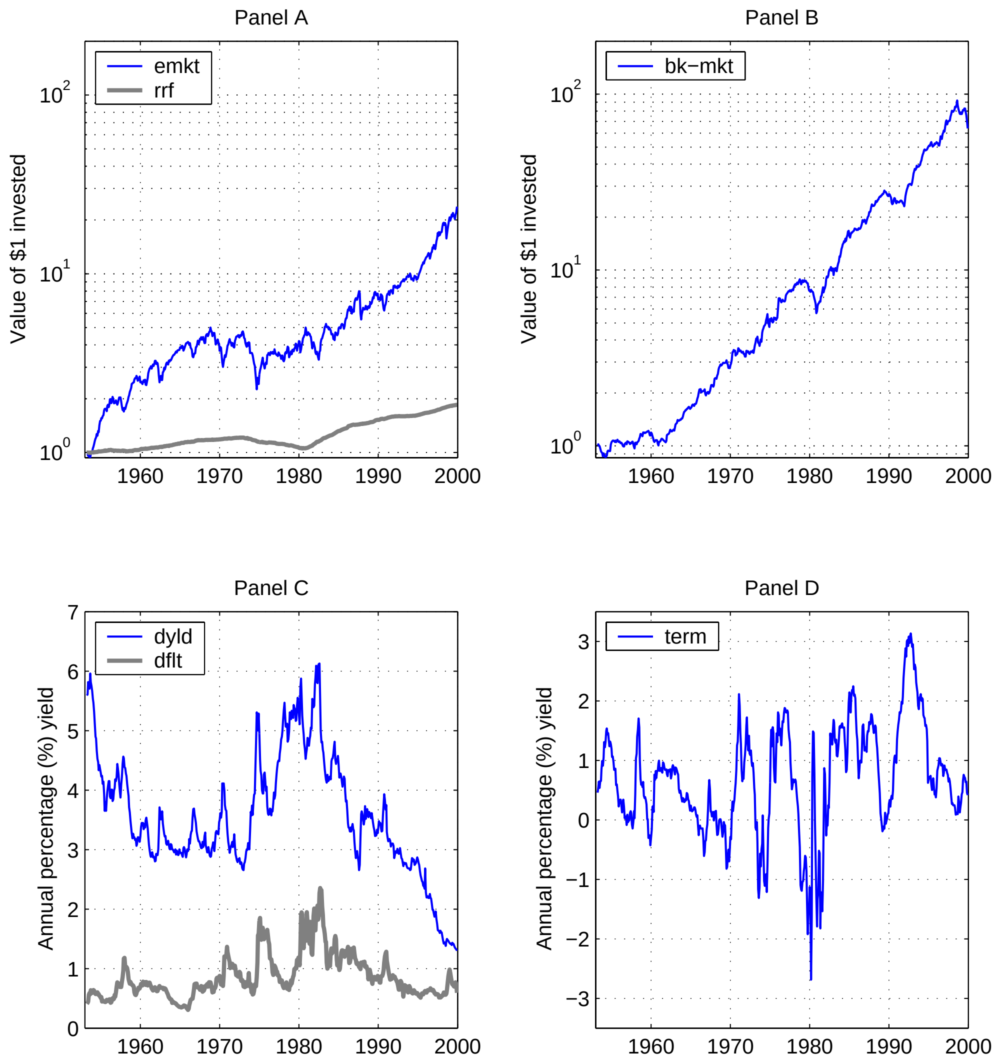
Table 7: Returns left unexplained by the models

	No model	Conditional CAPM	Intertemporal CAPM
bk-mkt effect	8.88%	12.68%	12.71%
(s.e.)	(1.54%)	(1.82%)	(1.48%)
Wald	33.24	48.79	73.55
[p-val]	[0.00]	[0.00]	[0.00]
Lower bound			1.27
(s.e.)			(0.14)

This table shows the returns of the book-to-market effect left unexplained by various models. The first column, labeled "No model," shows the time-series averages of $E_t[r^e_{i,t+1}] + \frac{V_{ii,t}}{2}$, implied by the VAR-GARCH models. The second column, labeled "Conditional CAPM," is the average unexplained returns after accounting for the conditional market beta. The third column is the average unexplained returns after accounting for the market beta and hedges against changes in forecasts of future market returns and future market volatilities. The returns are expressed as annualized percentages.

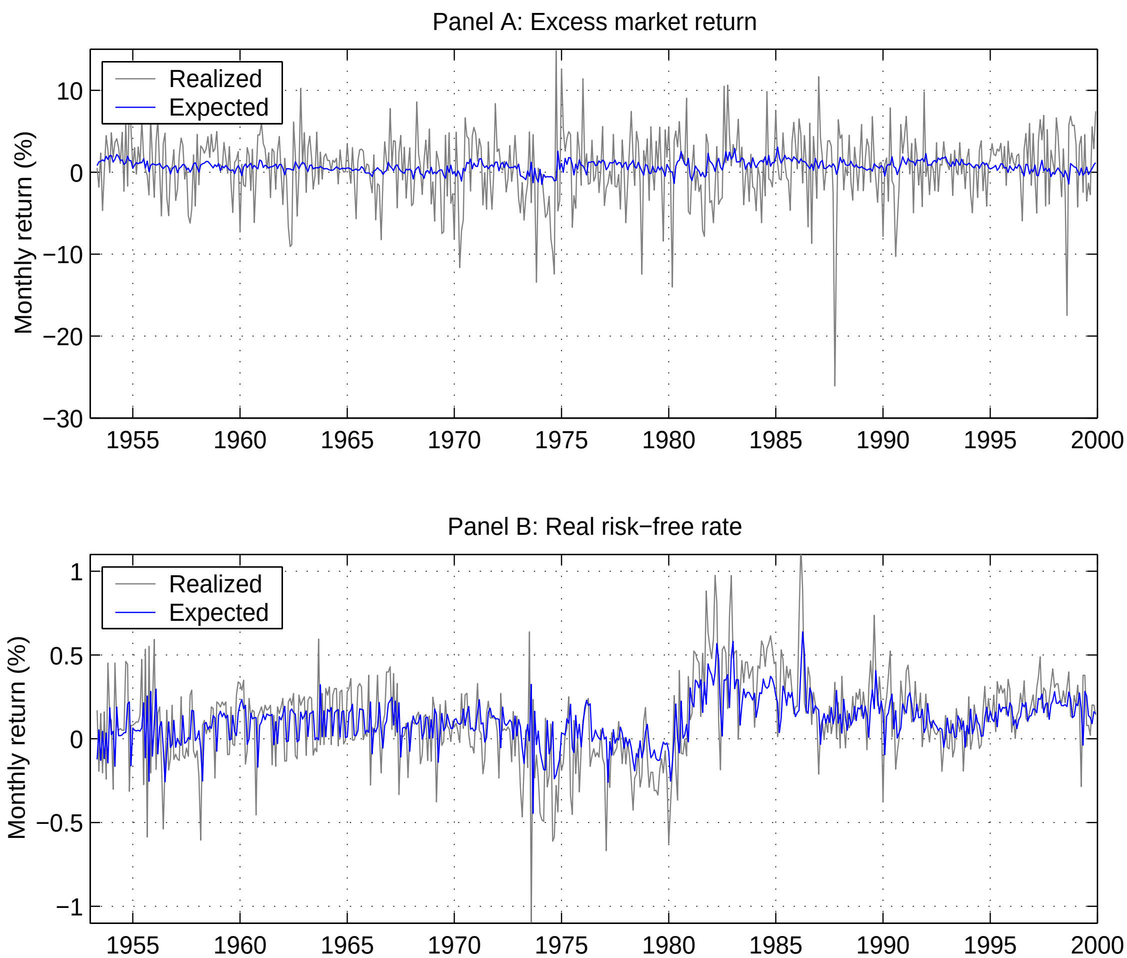
The Wald statistics are for the tests of the hypothesis that the returns left unexplained by the models are zero. The "Lower bound" is the lower bound of the Sharpe ratio for an omitted factor that would account for the unexplained portion of each return. The Sharpe ratios are expressed as annualized values. The standard errors and p-values are calculated using a heteroskedasticity- and autocorrelation-consistent estimator with Newey-West (1987) lags of order three.

Figure 1: Plots of the variables in the VAR systems



This figure shows the variables in the VAR systems. Panel A shows the cumulative excess return of the market, *emkt*, and the real risk-free rate, *rrf*. The plot graphs the cumulative value of a $1 invested ($1 long and $1 short) in April 1953 and left in the asset. The y-axis is a logarithmic scale. Panel B shows the cumulative returns on the book-to-market portfolio, *bk-mkt*. Panel C shows the dividend yield, *dyld*, and the default spread, *dflt*, in annualized, continuously compounded terms. Panel D shows the term spread, *term*, in annualized, continuously compounded terms.

Figure 2: Estimates of conditional expected returns from VAR(1)



This figure shows the estimated conditional expected returns of *emkt* and *rrf* implied by the VAR estimates of the *bk-mkt* system. The results from the *size* and *mom* systems are similar. The plots also show the realized returns on *emkt* and *rrf*. The returns are expressed as continuously compounded monthly returns.