Can an “Estimation Factor” Help Explain Cross-Sectional Returns?

Frederik Lundtofte

First version: September 2004
Current version: February 2006

This research has been carried out within the NCCR FINRISK project on “New Methods in Theoretical and Empirical Asset Pricing”
Can An "Estimation Factor" Help Explain Cross-Sectional Returns?

Frederik Lundtofte
First version: September, 2004
This version: February, 2006

Swiss Institute of Banking and Finance - University of St. Gallen, Rosenbergstrasse 52, CH-9000 St. Gallen, Switzerland. Phone: +41 71 224 7022. Fax: +41 71 224 7088. E-mail: frederik.lundtofte@unisg.ch

I thank David Feldman, Michael Brennan, Yihong Xia, Carsten Sørensen, Fabio Trojani and Björn Hansson for helpful discussions. Financial support from Bankforskningsinstitutet and the Swiss National Science Foundation (NCCR FINRISK and grant 100012-105745/1) is gratefully acknowledged.
Can An ”Estimation Factor” Help Explain Cross-Sectional Returns?

First version: September, 2004
This version: February, 2006

Abstract

We show in a theoretical model that the expected excess return on any asset depends on its covariance not only with the market portfolio, but also with changes in the representative agent’s estimate. We test our model by using GMM and compare it to the Fama-French model. The results suggest that the estimation factor is priced. Moreover, the Hansen-Jagannathan distances show that the conditional and static versions of our derived model perform on a par with the corresponding versions of the Fama-French model.

Keywords: learning, incomplete information, equilibrium, factor pricing models

JEL Classification Codes: C13, G12
1 Introduction

In traditional economic models, it is implicitly assumed that all agents have perfect knowledge of the true parameters involved. However, in making our investments in the stock market, we use estimates of e.g. expected returns, standard deviations, covariances, betas etc., in order to form our portfolio. Our investments are thus sensitive to changes in our estimates of economic fundamentals, and it is intuitive to think that the relative sensitivity of an asset to changes in our estimates would influence its price. We show in a theoretical model how changes in the estimate of the drift in aggregate dividend growth becomes a priced factor.

One purpose of this paper is to investigate whether this ”estimation factor” is important in asset pricing. We will analyze the extent to which the estimation factor affects cross-sectional returns, both theoretically and empirically. Another purpose is to compare our derived model of learning to a benchmark model. Because of its popularity among both academics and practitioners, we have chosen the three-factor Fama-French model as our benchmark.

We show that the expected excess return on any asset depends on its covariance not only with the market portfolio, but also with changes in the representative agent’s estimate. Thus, changes in the representative agent’s estimate enters as a factor in asset pricing through the stochastic discount factor. The representative agent’s estimate is based on realized growth in aggregate dividends and earnings. Hence, observations on realized growth in aggregate dividends and earnings enter into the stochastic discount factor through the representative agent’s estimate (the ”estimation factor”). We find empirical evidence suggesting that the estimation factor is priced, i.e., it comes with a premium. Moreover, empirical tests in connection
with GMM estimations suggest that the estimation factor is important for asset pricing. Further, measures of goodness-of-fit (the Hansen-Jagannathan distances) from our GMM estimations suggest that a conditional version of our derived model performs on a par with a conditional version of the Fama-French model, and that the static version of our derived model performs on a par with the static Fama-French model.

Eugene Fama and Kenneth French caused a sensation when they proposed their three-factor model in Fama and French (1993) and showed that it produced $R^2$s in excess of 90% in time-series regressions. The explanatory factors in their model are the excess return on the market portfolio, joined by a portfolio long in high book-to-market stocks and short in low book-to-market stocks (HML), and a portfolio long in small firms and short in big firms (SMB). Their argument is that stocks with high book-to-market ratios should earn higher returns than stocks with low book-to-market ratios, and that small firms should have higher returns on their stocks than big firms.

The Gordon growth model (Gordon, 1962) implies that there is a relation between an individual stock’s return and the dividends from that particular stock. Campbell and Shiller (1988) estimate a dynamic version of the Gordon growth model, which includes earnings through a VAR specification, i.e., they estimate the relationship between the stock index and the aggregate dividends and earnings. Our contribution is to show how aggregate dividends and earnings enter into the stochastic discount factor through the representative agent’s estimate of the drift in aggregate dividend growth, and especially how the pricing of any asset is separated into one part related to the covariation with the market factor, and another part
related to the covariation with changes in the estimate. We demonstrate that the impact of this estimation factor is statistically and economically significant. One of the advantages of our model over the Gordon growth model is that it is based on aggregate as opposed to asset-specific variables. As such, our model is easier to implement in real-world cross-sectional applications, as it requires fewer observations.

While the previous literature on estimation risk has mainly focused on how estimation risk affects portfolio choice, less has been written about the effect of estimation risk on the aggregate market equity premium, and even fewer contributions are concerned with the effect of estimation risk on cross-sectional stock returns. Klein and Bawa (1976) and Bawa, Brown and Klein (1979) analyze the portfolio choice problem under incomplete information in a single-period setting, while Kandel and Stambaugh (1996) consider a multi-period setting where the investor has a single period horizon. Dothan and Feldman (1986), Detemple (1986), and Gennotte (1986) study general equilibrium under incomplete information, and although these articles have different focuses, they all consider the portfolio choice problem of a partially informed investor in a completely dynamic continuous-time setting. However, Gennotte (1986) studies this issue more extensively. Recent contributions concerning the portfolio choice under estimation risk include Brennan (1998), Brennan and Xia (2001a), Xia (2001), Honda (2003), Knox (2003) and Lundtofte (2006). Notable contributions concerning the equity premium under incomplete information are Veronesi (2000) and Brennan and Xia (2001b). As for the effect of incomplete information on cross-sectional returns, Massa and Simonov (2005) and Zhang (2003) provide two empirical investigations. Both Massa and Simonov (2005) and Zhang (2003) find evidence that estimation risk is in fact priced.
The paper which lies closest to this paper from a technical point of view is perhaps Zhang (2003), who uses survey data of economic forecasters directly to find proxies for e.g. the estimated probability of being in a "good state." In addition to survey data, Zhang (2003) also considers proxying the estimated probability of being in the "good state" by using a probit model with variables that have been shown to have forecasting power for GDP growth in the literature. In this study, we will form estimates using data on aggregate dividends and earnings in an econometric model which is consistent with the estimation procedure of the representative agent. Moreover, in Zhang (2003), the economy has two states, and the unobservability concerns the state of the economy. In contrast, in our model, the unobservability concerns a parameter (the drift in aggregate dividend growth).

Massa and Simonov (2005) first use macro-economic data for estimation according to an econometric model which is consistent with theory, and then use survey data as a check for validity. However, they assume a representative agent with CARA utility, who maximizes the expected utility of final wealth. Finding it more intuitively appealing, and consistent with empirical evidence, we will use a representative agent with CRRA utility instead. Another difference is that we employ a continuous-time framework while Massa and Simonov (2005) use a discrete-time setting.

The starting point of our analysis is a Lucas (1978) economy in which a representative agent does not observe the mean growth in aggregate dividends. We show theoretically that the expected excess return on any asset depends on its covariance not only with the market portfolio, but also with changes in the representative agent’s estimate of the mean growth in dividends. The objectives are twofold: i) to
determine whether the impact of this "estimation factor" is statistically and economically significant, and ii) to compare our derived model to the Fama-French model. Therefore, the model is applied directly to the data and tested cross-sectionally on the excess returns on the 25 Fama-French size and book-to-market portfolios. We estimate the stochastic discount factor using GMM. We find evidence that the impact of the estimation factor is statistically and economically significant. Moreover, the Hansen-Jagannathan distances suggest that a conditional version of the derived model performs on a par with a conditional version of the Fama-French model, and that the static version of our derived model performs on a par with the static Fama-French model. In line with previous studies, we also find that while the conditional models cannot be rejected, the static models can all be rejected using $J$-tests.

In the GMM estimation of the stochastic discount factor including the Fama-French factors, both coefficients associated with the estimation factor are significant at the 10% level. One of these coefficients is also significant at the 5% level. A simple numerical comparison of the effects of the factors in the stochastic discount factor suggests that the estimation factor is economically important. Further, when including the estimation factor into the static Fama-French model, SMB loses its significance. From being highly significant, the $p$-value of the coefficient in front of SMB dramatically rises to about 0.26.

Our results support papers such as Petkova (2006) and Vassalou (2003), which argue that the Fama-French factors capture information related to present and future investment opportunities, but there may exist other variables that are better at capturing this information. The results of our tests suggest that the estimation factor is able to capture information related to present and future investment.
opportunities.

Interestingly, our paper is also related to Brennan et al. (2004) since, in our formal model, there is a linear relation between changes in the interest rate and changes in the representative agent’s estimate. Brennan et al. (2004) estimate an ICAPM model where changes in the real interest rate and the maximum Sharpe ratio are found to have significant risk premia.

This paper is organized as follows. In section 2, we derive the theoretical results. A description of the data is given in section 3. Section 4 contains our empirical specifications, tests, and results. Here, we estimate the models using GMM, and compare them using the Hansen-Jagannathan distances. Section 5 concludes our results. In the Appendix, we provide a derivation of the representative agent’s estimate.

2 Theoretical Results

Assume that there are \( n \) exogenously given dividend processes \( \{D_i\}_{i=1}^n \) (one for each firm) and hence an aggregate dividend process \( D_t = \sum_{i=1}^n D_{it} \). There is a complete probability space \( (\Omega, F, P) \). Assume further that the dynamics of the aggregate dividend process is given by

\[
\frac{dD_t}{D_t} = \mu_D dt + \sigma_D dZ_t
\]

where both \( \mu_D \) and \( \sigma_D \) are constants, but \( \mu_D \) is unknown. \( Z_t \) is a standard Brownian motion defined over \( (\Omega, F, P) \).

A representative agent has constant relative risk aversion and maximizes ex-
pected utility of consumption, given the set of information $G_t$.

$$U(\{c_s\}_t^T) = E \left[ \int_{s=t}^{T} e^{-\beta_s} \frac{c_s^{1-\gamma}}{1-\gamma} ds \middle| G_t \right],$$

where $\gamma > 0$. The information set that the representative agent uses to form his expectations is small. He will only use observations on two variables; aggregate dividends ($D_t$) and one additional economic variable which we call $X_t$, i.e. $G_t = \sigma((D_s, X_s)_{s \leq t})$. This is consistent with the observation that investors just make use of a small fraction of all the relevant information when they form their expectations, because they are not able to process all the relevant information. The economic variable $X_t$ is assumed to have the following diffusion.

$$dX_t = \mu_X dt + \sigma_X \rho dZ_t + \sigma_X \sqrt{1-\rho^2} dW_t,$$

where $Z_t$ and $W_t$ are independent standard Brownian motions on $(\Omega, F, P)$ and $-1 < \rho < 1$. This is equivalent to $\frac{dX_t}{X_t} = \mu_X dt + \sigma_X dV_t$, where $V_t$ is a standard Brownian motion such that $\langle dV_t, dZ_t \rangle = \rho dt$. This means that $\frac{dX_t}{X_t}$ and $\frac{dD_t}{D_t}$ are correlated with the constant instantaneous coefficient of correlation $\rho$. For simplicity, we will assume that both $\mu_X$ and $\sigma_X$ are known to the representative agent.

As noted by Feldman (2006), we need to know the conditional distribution of the unknown drift in order to re-represent the optimization problem as a Markovian one. The conditional mean $m_t = E [\mu_t | G_t]$, i.e., the expected value of the unknown drift conditional on all available information, can be interpreted as the representative agent’s estimate of the drift term, and, coincidently, it is indeed the optimal estimate if the objective is to minimize the mean squared error. Note that we may write

$$\begin{pmatrix} \frac{dD_t}{D_t} \\ \frac{dX_t}{X_t} \end{pmatrix} = \begin{pmatrix} m_t \\ \mu_X \end{pmatrix} dt + \begin{pmatrix} \sigma_D & 0 \\ \sigma_X \rho & \sigma_X \sqrt{1-\rho^2} \end{pmatrix} \begin{pmatrix} dZ_t \\ dW_t \end{pmatrix}, \quad (2)$$
where
\[
\begin{pmatrix}
  \frac{dZ_t}{dW_t}
\end{pmatrix} = \begin{pmatrix}
  \sigma_D & 0 \\
  \sigma_X \rho & \sigma_X \sqrt{1 - \rho^2}
\end{pmatrix}^{-1} \begin{pmatrix}
  \mu_D - m_t \\
  0
\end{pmatrix} dt + \begin{pmatrix}
  dZ_t \\
  dW_t
\end{pmatrix}.
\]

Further, we see that \(dZ_t\) and \(dW_t\) are the normalized unexpected innovations in \(dD_t/D_t\) and \(dX_t/X_t\), since
\[
\begin{pmatrix}
  dZ_t \\
  dW_t
\end{pmatrix} = \begin{pmatrix}
  \sigma_D & 0 \\
  \sigma_X \rho & \sigma_X \sqrt{1 - \rho^2}
\end{pmatrix}^{-1} \begin{pmatrix}
  \frac{dD_t}{D_t} \\
  \frac{dX_t}{X_t}
\end{pmatrix} - \begin{pmatrix}
  E\left[\frac{dD_t}{D_t} \mid G_t\right] \\
  E\left[\frac{dX_t}{X_t} \mid G_t\right]
\end{pmatrix}.
\]

According to standard filtering theory, \(Z_t\) and \(W_t\) are independent standard Brownian motions with respect to the representative agent’s filtration \(G_t\) (see Liptser and Shiryaev (2001)). Moreover, following Theorem 12.7 in Liptser and Shiryaev (2001), we can determine the representative agent’s estimate \(m_t\).

**Proposition 1** Given a Gaussian prior with mean \(m_0\) and variance \(v_0\), the representative agent’s estimate \(m_t\) satisfies the SDE
\[
dm_t = v_t \frac{1 - \rho^2 \sigma_D}{\sigma_D^2(1 - \rho^2)} dZ_t - v_t \frac{\rho \sigma_D}{\sigma_D \sqrt{1 - \rho^2}} dW_t
\]
(3)

where \(v_t\), the filtering error, is given by
\[
v_t = \frac{v_0 \sigma_D^2(1 - \rho^2)}{\sigma_D^2(1 - \rho^2) + v_0 t}.
\]

The SDE in (3) can be solved as
\[
m_t = \frac{v_t}{v_0} m_0 + \frac{v_t}{v_0} \int_0^t \frac{dD_s}{D_s} - \frac{\rho v_t}{\sigma_D \sigma_X(1 - \rho^2)} \left( \int_0^t \frac{dX_s}{X_s} - \mu_X t \right).
\]
(4)

**Proof.** See Appendix. ■

The consumption good is assumed to be perishable, so all dividends will be consumed in each period \(c_t^* = D_t\), and hence the stochastic discount factor is
given by $\Lambda_t = e^{-\beta t} D_t^{-\gamma}$. Applying Ito’s Lemma, we have

$$\frac{d\Lambda_t}{\Lambda_t} = -\beta dt - \gamma \frac{dD_t}{D_t} + \frac{1}{2} \gamma (\gamma + 1) \left( \frac{dD_t}{D_t} \right)^2.$$  \hspace{1cm} (5)

With the assumed dividend process, the representative agent’s first-order condition implies that the short interest rate is given by

$$r_t = \beta + \gamma m_t - \frac{1}{2} \gamma (\gamma + 1) \sigma^2_D.$$  \hspace{1cm} (6)

Following Cochrane (2001), the cum-dividend expected return on any stock $i$ can be obtained through the following equation.

$$E_t \left[ \frac{dS_i}{S_i} + \frac{D_i}{S_i} dt \right] - r_t dt = -E_t \left[ \frac{d\Lambda_t}{\Lambda_t} \frac{dS_i}{S_i} \right].$$

Inserting the SDE for the stochastic discount factor (equation (5)) into the above equation, we see that the risk premium is still determined by the risk aversion ($\gamma$) and the covariance with aggregate dividends as in the standard Lucas (1978) model.

$$E_t \left[ \frac{dS_i}{S_i} + \frac{D_i}{S_i} dt \right] - r_t dt = \gamma E_t \left[ \frac{dD_i}{D_i} \frac{dS_i}{S_i} \right].$$  \hspace{1cm} (7)

Let $S_M(t)$ denote the price of the claim to the aggregate stream of dividends. In the proposition below, we show that it is separable in the sense that $S_M(t) = D_t f(m_t, t)$, i.e., if today’s dividend increases by a certain percentage, then the stock price will increase by that same percentage.

**Proposition 2** The price of the claim to the aggregate stream of dividends is homogeneous of degree one in aggregate dividends, since $S_M(t) = D_t f(m_t, t)$, where

$$f(m_t, t) = \int_t^T \exp \left\{ \left( 1 - \gamma \right) \left( m_t - \frac{\sigma^2_D}{2} \right) - \beta \right\} \left( s - t \right) +$$

$$+ \frac{(1 - \gamma)^2}{2} \int_{u=t}^s \left( \sigma_D + \frac{(s-u)v_u(1-\rho^2\sigma_D)}{\sigma^2_D(1-\rho^2)} \right)^2 du +$$

$$+ \frac{(1 - \gamma)^2}{2} \int_{u=t}^s \frac{(s-u)^2v_u^2\rho^2}{\sigma^2_D(1-\rho^2)} du \right\} ds.$$
Proof. \( S_M(t) = E_t \left[ T_t \int_{t}^{T} D_s ds \right] = E_t \left[ T_t \int_{t}^{T} e^{-\beta(s-t)} \frac{D_s^{1-\gamma}}{D_t^{-\gamma}} ds \right]. \) The solution to the SDE \( \frac{dD_s}{D_s} = m_u du + \sigma_d dW_u \) is given by \( D_s = D_t \exp \left\{ \int_{t}^{s} (m_u - \frac{1}{2} \sigma_d^2) du + \int_{t}^{s} \sigma_d dW_u \right\} \) (for \( \forall s \geq t \)). By applying Ito’s Lemma on \( m_t \) and using the relation in equation (3), we have

\[
\int_{u=t}^{s} m_u du = m_t (s - t) + \int_{u=t}^{s} \frac{(s - u) u (1 - \rho^2 \sigma_d^2)}{\sigma_d^2 (1 - \rho^2)} dW_u - \int_{u=t}^{s} \frac{(s - u) u \rho \sigma_d}{\sigma_d \sqrt{1 - \rho^2}} dW_u.
\]

By Fubini’s Theorem, \( S_M(t) = T_t \int_{t}^{T} E_t \left[ e^{-\beta(s-t)} \frac{D_1^{1-\gamma}}{D_t^{-\gamma}} ds \right] \), and the result follows.

Thus, by Ito’s Lemma, the dynamics of the ”market portfolio” is given by

\[
\frac{dS_M}{S_M} = \frac{dD_t}{D_t} + \frac{f_t}{f} dt + \frac{f_m}{f} dm + \frac{1}{2} \frac{f_{mm}}{f} (dm)^2 + \frac{df}{f} \frac{dD_t}{D_t}.
\] (8)

Rearranging equation (8), we have aggregate dividend growth (\( \frac{dD_t}{D_t} \)) as a function of percentage increase in the value of the market portfolio (\( \frac{dS_M}{S_M} \)) and changes in the representative agent’s estimate (\( dm \)),

\[
\frac{dD_t}{D_t} = \frac{dS_M}{S_M} - \frac{f_t}{f} dt - \frac{f_m}{f} dm - \frac{1}{2} \frac{f_{mm}}{f} (dm)^2 - \frac{df}{f} \frac{dD_t}{D_t}.
\]

Inserting this into the expression for the risk premium for stock \( i \), one has

\[
E_t \left[ \frac{dS_i}{S_i} + \frac{D_t}{S_i} dt \right] - r_t dt = \gamma E_t \left[ \frac{dS_i}{S_i} \frac{dS_M}{S_M} \right] - \gamma \frac{f_m}{f} E_t \left[ \frac{dS_i}{S_i} dm \right].
\] (9)

Reformulating the above expression in terms of covariances, we get

\[
\mu^C_i - r_t = \gamma Cov \left( \frac{dS_i}{S_i} \frac{dS_M}{S_M} \right) - \gamma \frac{f_m}{f} Cov \left( \frac{dS_i}{S_i}, dm \right)
\] (10)

where \( \mu^C_i \) is the expected cum-dividend return. This means that the expected cum-dividend excess return on any stock \( i \) is determined through its covariance not only with the market return (\( \frac{dS_M}{S_M} \)), but also with changes in the estimate (\( dm_i \)).
Thus, estimation risk enters as an important factor in determining cross-sectional stock returns. The model predicts that stocks that *ceteris paribus* have a relatively higher covariance with the market and a relatively lower covariance with upward revisions of the estimate will have a relatively higher risk premium. The interpretation of this result is that stocks that go up when investment opportunities/output technologies are perceived to improve should have a relatively lower risk premium, and that investors should be compensated for carrying market risk. To the extent that returns on growth stocks have a higher covariance with perceived technological improvements (increases in the estimate) than value stocks, the model states that value stocks should have higher expected returns than growth stocks.

Multiplying and dividing by the variances of the market return and the changes in the estimate, we can express equation (10) in terms of time-varying market and estimation betas,

\[ \mu_i^C - r_t = \lambda_{M,t}\beta_{iM,t} + \lambda_m t\beta_{im,t} \] (11)

where \( \beta_{iM,t} = \frac{\text{Cov}(dS_i, dSM)}{\text{Var}[dSM]} \), \( \lambda_{M,t} = \gamma \text{Var}[dSM] \), \( \beta_{im,t} = \frac{\text{Cov}(dS_i, dm)}{\text{Var}[dm]} \) and \( \lambda_m t = -\gamma \frac{E_t[\text{Var}[dM]]}{\text{Var}[dM]} \).

In discrete time, equation (11) can be written as

\[ E_t [r_{i,t+1}] = \lambda_{M,t}\beta_{iM,t} + \lambda_m t\beta_{im,t} \] (12)

where \( \beta_{iM,t} = \frac{\text{Cov}(r_{i,t+1}, r_{M,t+1})}{\text{Var}[r_{M,t+1}]} \), \( \beta_{im,t} = \frac{\text{Cov}(r_{i,t+1}, \Delta m_{t+1})}{\text{Var}[\Delta m_{t+1}]} \). \( r_{i,t+1} \) is the future excess return on stock \( i \), \( r_{M,t+1} \) is the future excess return on the market portfolio, and \( \Delta m_{t+1} = m_{t+1} - m_t \) is the shift in the estimate of the mean growth rate in dividends.

The expression in equation (12) is in terms of conditional expectations, whereas it is useful to test empirically an unconditional model of excess returns. In what follows, we will derive an unconditional model.
The beta formulation in equation (12) implies that the stochastic discount factor is a linear combination of the pricing factors with time-varying coefficients, as shown in the following proposition.

**Proposition 3** The conditional beta pricing model \( E_t \left[ r_{i,t+1}^e \right] = \lambda_{M,t} \beta_{iM,t} + \lambda_{m,t} \beta_{im,t} \) implies that we can find non-trivial \( a_t, b_t, \) and \( c_t \), such that the stochastic discount factor

\[
\pi_{t+1} = a_t + b_t r_{M,t+1}^e + c_t \Delta m_{t+1}
\]

prices any excess return \( r_{i,t+1}^e \), i.e., for any excess return \( r_{i,t+1}^e \), we have

\[
E_t \left[ \pi_{t+1} r_{i,t+1}^e \right] = 0.
\]

**Proof.** Try, for example, \( a_t = 1 + \frac{\lambda_{M,t}}{\text{Var}(r_{M,t+1}^e)} E_t \left[ r_{M,t+1}^e \right] + \frac{\lambda_{m,t}}{\text{Var}(\Delta m_{t+1})} E_t \left[ \Delta m_{t+1} \right] \), \( b_t = -\frac{\lambda_{M,t}}{\text{Var}(r_{M,t+1}^e)} \), and \( c_t = -\frac{\lambda_{m,t}}{\text{Var}(\Delta m_{t+1})} \). For these coefficients, we clearly have

\[
E_t \left[ \pi_{t+1} r_{i,t+1}^e \right] = E_t \left[ \pi_{t+1} \right] E_t \left[ r_{i,t+1}^e \right] + \text{Cov}_t \left( \pi_{t+1}, r_{i,t+1}^e \right) = 0.
\]

We can model the time-varying coefficients in order to obtain an empirically testable model. This can be done by assuming a linear dependence on a vector of state-variables, \( z_t \) (including a constant). Note that the restriction to a linear dependence is not as restrictive as it may sound; it allows for quadratic \((x_t^2)\), cubic \((x_t^3)\) etc. expressions of the original state-variable \( x_t \). As a matter of fact, you can fit any polynomial of the original state-variables. Write the linear dependence as

\[
a_t = a(z_t) = a' z_t
\]

\[
b_t = b(z_t) = b' z_t
\]
\[ c_t = c(z_t) = c'z_t \]

where \( a, b \) and \( c \) are vectors of constants.\(^1\) This means that the stochastic discount factor can be written as

\[ \pi_{t+1} = a'z_t + (b'z_t)r^e_{M,t+1} + (c'z_t)\Delta m_{t+1}. \]

We have now obtained a linear factor model with constant coefficients that can be estimated by using e.g. the GMM. Note that the state-variables enter as factors in the linear factor model with constant coefficients. By taking the unconditional expectation of equation (14), we have the unconditional model

\[ E[\pi_{t+1}r^e_{t,t+1}] = 0. \tag{15} \]

In this paper, we will use the term premium (TERM) to model the time-varying coefficients.\(^2\) The term premium is defined as the difference between the yield of a 10-year Treasury bond and the yield of a one-year Treasury bond. The term premium has been used in the previous literature as a state variable by e.g. Fama (1984) and Campbell (1987). Assuming a linear dependence, the stochastic discount

\(^{1}\)Strictly speaking, \( b_t \) is constant in the model (it is equal to \(-\gamma\)). That is, according to the model, \( b_0 = -\gamma \) and \( b_j = 0 \) for \( j > 0 \).

\(^{2}\)In the previous literature, other state variables have been proposed, such as the default premium and the aggregate dividend yield. The coefficients should be adapted to the representative agent’s information set, so we do not want to use the default premium as a state variable, since there is no default in the theoretical model. Also, since the dividend yield is likely to be non-stationary (we could not reject the hypothesis of a unit root), its inclusion as a state variable is less straightforward. To keep a close connection with the theoretical model and to avoid data mining, we propose a parsimonious model with the term premium as the only state variable.
factor is given by
\[ \pi_{t+1} = a_0 + a_1 T E R M_t + b_0 r^{e}_{M,t+1} + b_1 T E R M_t r^{e}_{M,t+1} + c_0 \Delta m_{t+1} + c_1 T E R M_t \Delta m_{t+1}. \]  

(16)

Note that the converse of Proposition 3 is also true, i.e., the stochastic discount factor in equation (13) implies a conditional beta pricing model, where, to some extent, the time-varying coefficients determine the risk premia. This result is summarized in the proposition below.

**Proposition 4** The stochastic discount factor \( \pi_{t+1} = a_t + b_t r^{e}_{M,t+1} + c_t \Delta m_{t+1} \) implies a conditional beta pricing model

\[ E_t [r^{e}_{i,t+1}] = \lambda_{M,t} \beta_{iM,t} + \lambda_{m,t} \beta_{im,t}, \]

where
\[ \lambda_{M,t} = -b_t \frac{Var_t(r^{e}_{M,t+1})}{E_t[\pi_{t+1}]} \quad \text{and} \quad \lambda_{m,t} = -c_t \frac{Var_t(\Delta m_{t+1})}{E_t[\pi_{t+1}]} \].

**Proof.** This follows from the equation \( 0 = E_t [\pi_{t+1} r^{e}_{i,t+1}] = E_t [\pi_{t+1}] E_t [r^{e}_{i,t+1}] + Cov_t (\pi_{t+1}, r^{e}_{i,t+1}) \).  

In the empirical part of this paper, we will be concerned with testing hypotheses such as \( b_t = 0 \) and \( c_t = 0 \). These are empirical tests of the importance of the factors in the stochastic discount factor (in equation (13)) and, by virtue of Proposition 4 above, they are also tests of their associated risk premia.

Note that, by equation (6), \( dr_t = \gamma dm_t \), i.e., there is a linear relation between changes in the representative agent’s estimate and changes in the interest rate. This relates our paper to Brennan et al. (2004) who estimate an ICAPM model where changes in the real interest rate and the maximum Sharpe ratio are found to have significant risk premia.
3 Description of the Data

Our data have been collected from three sources: the Federal Reserve Bank of St. Louis, and the homepages of Robert Shiller and Kenneth French. We have obtained monthly data on the real dividends ($D_t$) for the S&P Composite Index and real aggregate earnings ($E_t$) from March 1904 to June 2002 from Robert Shiller’s homepage. Monthly data from April 1953 to June 2002 on the returns on the 25 Fama-French portfolios sorted by size and book-to-market, the risk free rate, HML and SMB have been obtained from Kenneth French’s homepage. The Federal Reserve Bank of St. Louis has provided us with monthly data on the term premium (TERM) in the period from April 1953 to June 2002. The term premium (TERM) is defined as the difference between the yield of a 10-year Treasury bond and the yield of a one-year Treasury bond. The data from Robert Shiller on real dividends and earnings prevent us from testing the model on later dates than June 2002, while the data from the Federal Reserve Bank of St. Louis prevent us from testing the model on earlier dates than April 1953.

We use aggregate dividend data instead of aggregate consumption data, because we want to test the exact predictions of the theoretical model by applying it directly to the data. By using dividend data instead of consumption data, Hagiwara and Herce (1997) show that the intertemporal marginal rate of substitution (IMRS) is consistent with the Hansen-Jagannathan bound with a relative risk aversion of 3. Further, the participation rate in the stock market has historically been low (cf. the so-called "stock market participation puzzle"). Basak and Cuoco (1998) demonstrate that, the equity premium is mainly determined by the market participants and can be reconciled with a coefficient of relative risk aversion of around 3.

15
(cf. Hagiwara and Herce (1997)). This gives further support to our choice of using dividend data instead of consumption data.

In the empirical analysis, we assume that the additional economic variable that the representative agent uses in order to form his estimates is aggregate earnings, i.e., \( X_t = E_t \). From a strictly theoretical perspective, aggregate earnings should equal aggregate dividends in the simple Lucas model in the previous section. However, we realize that in the real world, all earnings are not paid out as dividends, but dividends and earnings should clearly be correlated. It is thus reasonable to think that a real world agent would use the information contained in aggregate earnings when he forms his estimates of dividend growth. The assumption that the representative agent knows the expected growth rate in earnings is made only for the purpose of simplifying the analysis. After all, if we are able to obtain strong results by just introducing uncertainty about one parameter (the expected growth in dividends), we can expect to see even stronger results if the uncertainty concerns two parameters. In order to construct a series for the representative agent’s estimate \((m_t)\), we use a discrete-time version of equation (4), where we simply approximate the integrals with sums, according to

\[
m_t = \frac{v_t}{v_0} m_0 + \frac{v_t}{\sigma^2 D (1 - \rho^2)} \sum_{s=1}^t \frac{\Delta D_s}{D_{s-1}} - \frac{\rho v_t}{\sigma D \sigma E (1 - \rho^2)} \left( \sum_{s=1}^t \frac{\Delta E_s}{E_{s-1}} - \mu_E t \right),
\]

where \( \Delta D_s = D_s - D_{s-1} \), and \( \Delta E_s = E_s - E_{s-1} \). The standard deviations of aggregate dividend growth (\( \sigma_D \)) and earnings growth (\( \sigma_E \)), and the correlation between dividend growth and earnings growth (\( \rho \)) are approximated by the sample standard deviations and the sample correlation of the entire period from April 1904 to June 2002. The sample mean of earnings growth from the period April 1904 to April 1953 is taken to be the true drift in earnings (\( \mu_E \)).
The data on TERM start in April 1953, so, in order to make maximum use of the information available, $t = 0$ is taken to be this date. The representative agent forms his prior of the unknown dividend drift ($\mu_D$) by using the arithmetic mean of the dividend growth in the period from April 1904 to April 1953. Thus, $m_0$ is the arithmetic mean of the dividend growth, and the variance of the prior, $v_0$, is the variance of the sample mean, i.e., $v_0 = \sigma_D^2/N$, where $N$ is the number of observations used to form the prior.

4 Econometric Estimation and Testing

In this section, we will estimate the stochastic discount factor using the Generalized Method of Moments (GMM) and test our hypotheses regarding the included factors. The excess returns on the 25 Fama-French portfolios, sorted on size and book-to-market, are used as instruments in the estimation of the stochastic discount factor. In addition to the factors already included in the stochastic discount factor in equation (13), we will also include the Fama-French factors HML (“high minus low”) and SMB (“small minus big”). The Fama-French factors are included to see whether the estimation factor has any additional explanatory power over the Fama-French model. Thus, the stochastic discount factor has the following appearance

$$\pi_{t+1} = a_t + b_t r^e_{M,t+1} + c_t \Delta m_{t+1} + d_t HML_{t+1} + e_t SMB_{t+1}, \quad (17)$$

where

$$a_t = a_0 + a_1 TERM_t$$

$$b_t = b_0 + b_1 TERM_t$$

$$c_t = c_0 + c_1 TERM_t$$
\[ d_t = d_0 + d_1 \text{TERM}_t \]
\[ e_t = e_0 + e_1 \text{TERM}_t. \]

Given that \( a_0 \neq 0 \), we can divide \( \pi_{t+1} \) by \( a_0 \), and the scaled stochastic discount factor still prices the excess returns according to \( E_t[\pi_{t+1}r^e_{t+1}] = 0 \). Thus, we can normalize, so that we let \( a_0 = 1 \). Further, we can treat \( a_0 = 1 \) as a "dependent variable" in the regressions, implying that the stochastic discount factor we wish to test has the following appearance.

\[
\pi_{t+1} = 1 - k_1 \text{TERM}_t - k_2 r^e_{M,t+1} - k_3 \text{TERM}_t r^e_{M,t+1} \\
- k_4 \Delta m_{t+1} - k_5 \text{TERM}_t \Delta m_{t+1} \\
- k_6 HML_{t+1} - k_7 \text{TERM}_t HML_{t+1} \\
- k_8 SMB_{t+1} - k_9 \text{TERM}_t SMB_{t+1} \tag{18}
\]

The stochastic discount factor suggested by our previous theoretical analysis thus implies a restriction \( (k_6 = k_7 = k_8 = k_9 = 0) \) on the stochastic discount factor above. Likewise, a conditional version of the Fama-French model is obtained if we apply the restriction \( k_4 = k_5 = 0 \), and the static Fama-French model is obtained through the restriction \( k_1 = k_3 = k_4 = k_5 = k_7 = k_9 = 0 \).

In GMM, the coefficients are chosen so that a quadratic form of pricing errors with a certain weighting matrix \( A \) is minimized. The value of the quadratic form is then used in different tests of mis-specification. The most common choice of the weighting matrix \( A \) is the inverse of the variance-covariance matrix of the model errors. Hansen and Singleton (1982) show that this choice of weighting matrix minimizes the asymptotic variance-covariance matrix of the GMM estimates, and that \( T \) times the minimized sample analog of the quadratic form asymptotically
has a $\chi^2$ distribution with $n - k$ degrees of freedom, where $n$ is the number of orthogonality conditions and $k$ is the number of unknown parameters in the model. This statistic, which is called the $J$-statistic, can thus be used to test whether the pricing errors are zero. However, as pointed out by Jagannathan and Wang (1996), this statistic is unsuitable for model comparisons, because the weighting matrix will be different for different model specifications. Moreover, since the weighting matrix is the inverse of the variance-covariance matrix of the model errors, the $J$-test will reward relatively more "noisy" models.

Hansen and Jagannathan (1997) therefore suggest the use of the inverse of the second moments of the asset returns, i.e. $A = (E[R_t R_t^T])^{-1}$. This choice has the advantage that the weighting matrix remains the same across different model specifications, and hence the minimized quadratic form associated with this weighting matrix is better suited for model comparisons. Hansen and Jagannathan (1997) show that the square-root of the minimized quadratic form is the distance from the candidate stochastic discount factor of a given model to the set of all the discount factors that price the $N$ assets correctly, where the "distance" is the distance in the $L^2$-norm. This distance, between the candidate stochastic discount factor and the set of all the discount factors that price the assets correctly, is referred to as the Hansen-Jagannathan distance, or the HJ-distance for short. The HJ-distance is also the maximum pricing error for a portfolio consisting of the $N$ assets, as shown in Hansen and Jagannathan (1997).

We will use the "optimal" weighting matrix suggested by Hansen and Singleton (1982) to obtain estimates that minimize the asymptotic variance-covariance matrix, and we will report the $J$-statistic. As mentioned earlier, the $J$-statistic should just
be used as a test of mis-specification, not for model comparison. In order to compare
the competing models, we will instead use the HJ-distance.

In Table 1, the coefficient estimates of the stochastic discount factor in equation
(18) are shown together with standard errors, $t$-statistics and $p$-values. The $J$-
statistic and the HJ-distance are also shown. We see that the only factor, except
for the constant, whose associated coefficients are significant is the estimation factor
($\Delta m$). The coefficient in front of $\Delta m_{t+1}$ is significant at the 10\% level, while the
coefficient in front of the cross-term $TERM_i \Delta m_{t+1}$ is significant at the 5\% level.
The $J$-statistic gives a reassuring $p$-value of 0.70, so the model cannot be rejected.
Using Wald-tests, we also test whether the time-varying coefficients in front of the
included factors are zero (see Table 2). We can reject the hypothesis that the time-
varying coefficient in front of the estimation factor ($c_t$) is zero at the 10\% level, i.e.,
we can reject the joint hypothesis that $k_4 = 0$ and $k_5 = 0$ at the 10\% level. The
time-varying coefficients in front of the other factors are found to be insignificant at
the 10\% level.

The fact that the time-varying coefficients in front of the market factor and the
Fama-French factors are insignificant may appear puzzling. One explanation is that
the variable TERM together with changes in the estimate $m$, which is based on div-
idends and earnings, are better at capturing information related to innovations in
the investment opportunity set than the market factor and the Fama-French factors.
Petkova (2006) finds that a model which includes variables that describe investment
opportunities (such as TERM) performs better than the three-factor Fama-French
model, and that the Fama-French factors lose their explanatory power in the pres-
ence of variables such as TERM, which describe the investment opportunity set.
Petkova (2006) further finds that HML is mostly related to the term premium, while SMB is mostly related to the default premium. Moreover, Vassalou (2003) presents evidence that HML and SMB contain news related to future GDP growth, and when news related to GDP growth is present in the asset-pricing model, HML and SMB lose much of their ability to explain the cross-section.

The representative agent’s estimate $m$ is based on past realizations of the growth in aggregate earnings and dividends, which is related to GDP growth. Indeed, when testing the static version of the Fama-French model, containing only the Fama-French factors, then the Fama-French factors are all highly significant (see Table 9). However, the static Fama-French model can be rejected at the 1% level using a $J$-test. Further, as soon as we include the estimation factor, SMB loses its significance (see Table 5). The rise in the $p$-value for the coefficient in front of SMB is dramatic. Before the estimation factor is included, the $p$-value is essentially zero but, as soon as the estimation factor is included, the $p$-value rises to about 0.26. These results suggest that the estimation factor is better at picking up information related to the present and future investment opportunity set than is SMB. TERM further helps to pick up additional information related to present and future investment opportunities and, when this variable is included, the explanatory power of especially the market factor and the Fama-French factors but also the estimation factor is weakened.

We also estimate the model suggested by our previous theoretical model, which can be seen as a reduced version of the model in equation (17) where we have left out the Fama-French factors HML and SMB separately. The results of this estimation are given in Table 3. We cannot reject the hypothesis of a correctly specified model
Table 1 GMM estimates of the unrestricted model together with standard errors, * $t$-statistics and $p$-values. The $J$-statistic and the HJ-distance are also reported. * indicates significance at the 10% level, ** indicates significance at the 5% level, and *** indicates significance at the 1% level.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>$t$-Statistic</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$TERM_t$</td>
<td>0.9732***</td>
<td>0.1375</td>
<td>7.077</td>
<td>0.0000</td>
</tr>
<tr>
<td>$r_{M,t+1}^e$</td>
<td>-0.02979</td>
<td>0.03177</td>
<td>-0.9375</td>
<td>0.3489</td>
</tr>
<tr>
<td>$TERM_t r_{M,t+1}^e$</td>
<td>0.02712</td>
<td>0.04631</td>
<td>0.5856</td>
<td>0.5584</td>
</tr>
<tr>
<td>$\Delta m_{t+1}$</td>
<td>-686.4*</td>
<td>412.2</td>
<td>-1.665</td>
<td>0.0964</td>
</tr>
<tr>
<td>$TERM_t \Delta m_{t+1}$</td>
<td>635.8**</td>
<td>303.9</td>
<td>2.092</td>
<td>0.0369</td>
</tr>
<tr>
<td>$HML_{t+1}$</td>
<td>-0.04671</td>
<td>0.03498</td>
<td>-1.335</td>
<td>0.1823</td>
</tr>
<tr>
<td>$TERM_t HML_{t+1}$</td>
<td>0.04003</td>
<td>0.04749</td>
<td>0.8429</td>
<td>0.3996</td>
</tr>
<tr>
<td>$SMB_{t+1}$</td>
<td>-0.02255</td>
<td>0.03631</td>
<td>-0.6210</td>
<td>0.5349</td>
</tr>
<tr>
<td>$TERM_t SMB_{t+1}$</td>
<td>0.01040</td>
<td>0.05671</td>
<td>0.1834</td>
<td>0.8546</td>
</tr>
</tbody>
</table>

| J-statistic | 12.6 | $p$-value | 0.70 |
| HJ-distance | 0.1277 |          |

on the basis of the $J$-statistic, which produces a reassuring $p$-value of 0.82. The only factor except the term premium with a significant coefficient is the cross-term $TERM_t \Delta m_{t+1}$. Again testing the restriction that the time-varying coefficient in front of the market factor ($h_t$) is zero with a Wald-test, we find that this hypothesis cannot be rejected, as reported in Table 4. We also cannot reject the hypothesis that the time-varying coefficient in front of the estimation factor ($c_t$) is zero in the restricted model using the Wald-test, as seen in Table 4. This can be compared to the results regarding the conditional Fama-French model (cf. Table 7). In the conditional Fama-French model, all the coefficients associated with the market factor and the Fama-French factors are insignificant, and the joint tests produce insignificant
Table 2. Hypothesis tests based on Wald test statistics in the unrestricted model. 

$b_t$, $c_t$, $d_t$, and $e_t$ are the time-varying coefficients in front of the market factor, the estimation factor, HML, and SMB, respectively.

$H_0$: $b_t = 0$, i.e., the joint hypothesis $k_2 = 0$ and $k_3 = 0$.

<table>
<thead>
<tr>
<th>F-statistic</th>
<th>p-value</th>
<th>$\chi^2$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7932</td>
<td>0.4529</td>
<td>1.586</td>
<td>0.4524</td>
</tr>
</tbody>
</table>

$H_0$: $c_t = 0$, i.e., the joint hypothesis $k_4 = 0$ and $k_5 = 0$.

<table>
<thead>
<tr>
<th>F-statistic</th>
<th>p-value</th>
<th>$\chi^2$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.324</td>
<td>0.0988</td>
<td>4.648</td>
<td>0.0979</td>
</tr>
</tbody>
</table>

$H_0$: $d_t = 0$, i.e., the joint hypothesis $k_6 = 0$ and $k_7 = 0$.

<table>
<thead>
<tr>
<th>F-statistic</th>
<th>p-value</th>
<th>$\chi^2$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9552</td>
<td>0.3853</td>
<td>1.910</td>
<td>0.3847</td>
</tr>
</tbody>
</table>

$H_0$: $e_t = 0$, i.e., the joint hypothesis $k_8 = 0$ and $k_9 = 0$.

<table>
<thead>
<tr>
<th>F-statistic</th>
<th>p-value</th>
<th>$\chi^2$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.6003</td>
<td>0.5490</td>
<td>1.201</td>
<td>0.5486</td>
</tr>
</tbody>
</table>

results (see Table 8). The results regarding the conditional models can in turn be compared to the static models (cf. Table 6 and Table 9). In the estimation of the static version of our derived model, the market factor is significant at the 1% level, while the ”estimation factor” is significant at the 5% level, and in the estimation of the static Fama-French model, all the factors are significant at the 1% level.

The HJ-distances of the models with time-varying coefficients are close to each other and are in the range 0.13 to 0.15, indicating that these models have a similar performance. Further, none of the models with time-varying coefficients can be rejected with a J-test, suggesting that they are well-specified. The HJ-distances
of the static models are also close to each other and are in the range 0.32 to 0.39, so these models perform similarly. Thus, the static models have HJ-distances that are about 3 times as large as the HJ-distances of the models with time-varying coefficients, but then again, the static models have fewer parameters. However, the static models can all be rejected using $J$-tests.

*Table 3* GMM estimates of the derived model together with standard errors, $t$-statistics and $p$-values. The $J$-statistic and the HJ-distance are also reported.

* indicates significance at the 10% level, and *** indicates significance at the 1% level.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>$t$-Statistic</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$TERM_t$</td>
<td>0.9304***</td>
<td>0.1104</td>
<td>8.426</td>
<td>0.0000</td>
</tr>
<tr>
<td>$r^c_{M,t+1}$</td>
<td>-0.02180</td>
<td>0.01874</td>
<td>-1.163</td>
<td>0.2452</td>
</tr>
<tr>
<td>$TERM_t r^c_{M,t+1}$</td>
<td>0.02159</td>
<td>0.02944</td>
<td>0.7332</td>
<td>0.4638</td>
</tr>
<tr>
<td>$\Delta m_{t+1}$</td>
<td>-500.0</td>
<td>331.0</td>
<td>-1.511</td>
<td>0.1314</td>
</tr>
<tr>
<td>$TERM_t \Delta m_{t+1}$</td>
<td>447.5*</td>
<td>251.5</td>
<td>1.779</td>
<td>0.0757</td>
</tr>
</tbody>
</table>

$J$-statistic 14.3  $p$-value 0.82

HJ-distance 0.1316
Table 4 Hypothesis tests based on Wald test statistics in the derived model. \( b_t \) and \( c_t \) are the time-varying coefficients in front of the market factor, and the estimation factor, respectively.

\[
H_0: b_t = 0, \text{ i.e., the joint hypothesis } k_2 = 0 \text{ and } k_3 = 0.
\]

\[
F\text{-statistic} \quad 0.9106 \quad p\text{-value} \quad 0.4028
\]

\[
\chi^2 \quad 1.821 \quad p\text{-value} \quad 0.4023
\]

\[
H_0: c_t = 0, \text{ i.e., the joint hypothesis } k_4 = 0 \text{ and } k_5 = 0.
\]

\[
F\text{-statistic} \quad 1.762 \quad p\text{-value} \quad 0.1727
\]

\[
\chi^2 \quad 3.523 \quad p\text{-value} \quad 0.1718
\]

Table 5 GMM estimates of a static version of the model estimated in Table 1, together with standard errors, \( t \)-statistics and \( p \)-values. The \( J \)-statistic and the HJ-distance are also reported. *** indicates significance at the 1% level.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>( t )-Statistic</th>
<th>( p )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r^S_{M,t+1} )</td>
<td>0.06427***</td>
<td>0.01163</td>
<td>5.525</td>
<td>0.0000</td>
</tr>
<tr>
<td>( \Delta m_{t+1} )</td>
<td>1121***</td>
<td>384.8</td>
<td>2.913</td>
<td>0.0037</td>
</tr>
<tr>
<td>( HML_{t+1} )</td>
<td>0.1174***</td>
<td>0.02178</td>
<td>5.391</td>
<td>0.0000</td>
</tr>
<tr>
<td>( SMB_{t+1} )</td>
<td>0.01859</td>
<td>0.01646</td>
<td>1.130</td>
<td>0.2591</td>
</tr>
<tr>
<td>( J )-statistic</td>
<td>46.7</td>
<td></td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td>HJ-distance</td>
<td>0.3222</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6 GMM estimates of a static version of the derived model, together with standard errors, \( t \)-statistics and \( p \)-values. The \( J \)-statistic and the HJ-distance are also reported. ** indicates significance at the 5% level, and *** indicates significance at the 1% level.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>( t )-Statistic</th>
<th>( p )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r^S_{M,t+1} )</td>
<td>0.04995***</td>
<td>0.01029</td>
<td>4.854</td>
<td>0.0000</td>
</tr>
<tr>
<td>( \Delta m_{t+1} )</td>
<td>615.0**</td>
<td>308.8</td>
<td>1.991</td>
<td>0.0469</td>
</tr>
<tr>
<td>( J )-statistic</td>
<td>49.3</td>
<td></td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td>HJ-distance</td>
<td>0.3887</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 7 GMM estimates of the conditional Fama-French model together with standard errors, $t$-statistics and $p$-values. The $J$-statistic and the HJ-distance are also reported. *** indicates significance at the 1% level.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>$t$-Statistic</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$TERM_t$</td>
<td>0.8610***</td>
<td>0.1142</td>
<td>7.540</td>
<td>0.0000</td>
</tr>
<tr>
<td>$r_{M,t+1}^e$</td>
<td>-0.02381</td>
<td>0.02758</td>
<td>-0.8633</td>
<td>0.3883</td>
</tr>
<tr>
<td>$TERM_t r_{M,t+1}^e$</td>
<td>0.02906</td>
<td>0.03835</td>
<td>0.7579</td>
<td>0.4488</td>
</tr>
<tr>
<td>$HML_{t+1}$</td>
<td>0.01040</td>
<td>0.03146</td>
<td>0.3305</td>
<td>0.7411</td>
</tr>
<tr>
<td>$TERM_t HML_{t+1}$</td>
<td>0.003581</td>
<td>0.03927</td>
<td>0.09121</td>
<td>0.9274</td>
</tr>
<tr>
<td>$SMB_{t+1}$</td>
<td>0.007345</td>
<td>0.03022</td>
<td>0.2430</td>
<td>0.8081</td>
</tr>
<tr>
<td>$TERM_t SMB_{t+1}$</td>
<td>-0.01420</td>
<td>0.04590</td>
<td>-0.3094</td>
<td>0.7572</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$J$-statistic</th>
<th>$p$-value</th>
<th>HJ-distance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15.8</td>
<td>0.60</td>
<td>0.1480</td>
</tr>
</tbody>
</table>

Table 8 Hypothesis tests based on Wald test statistics in the conditional Fama-French model. $b_t$, $d_t$, and $e_t$ are the time-varying coefficients in front of the market factor, HML, and SMB, respectively.

$H_0$: $b_t = 0$, i.e., the joint hypothesis $k_2 = 0$ and $k_3 = 0$.

- $F$-statistic: 0.3786, $p$-value: 0.6850
- $\chi^2$: 0.7572, $p$-value: 0.6848

$H_0$: $d_t = 0$, i.e., the joint hypothesis $k_6 = 0$ and $k_7 = 0$.

- $F$-statistic: 0.2482, $p$-value: 0.7803
- $\chi^2$: 0.4965, $p$-value: 0.7802

$H_0$: $e_t = 0$, i.e., the joint hypothesis $k_8 = 0$ and $k_9 = 0$.

- $F$-statistic: 0.05158, $p$-value: 0.9497
- $\chi^2$: 0.1032, $p$-value: 0.9497
Table 9 GMM estimates of the static Fama-French model together with standard errors, \( t \)-statistics and \( p \)-values. The \( J \)-statistic and the HJ-distance are also reported. *** indicates significance at the 1% level.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>( t )-Statistic</th>
<th>( p )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r_{M,t+1} )</td>
<td>0.2341***</td>
<td>0.01564</td>
<td>14.97</td>
<td>0.0000</td>
</tr>
<tr>
<td>( HML_{t+1} )</td>
<td>0.1980***</td>
<td>0.02492</td>
<td>7.945</td>
<td>0.0000</td>
</tr>
<tr>
<td>( SMB_{t+1} )</td>
<td>0.1268***</td>
<td>0.01965</td>
<td>6.453</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
\hat{a}_{t} &= 1 - \hat{k}_1TERM_t, \\
\hat{b}_{t} &= -\hat{k}_2 - \hat{k}_3TERM_t, \\
\hat{c}_{t} &= -\hat{k}_4 - \hat{k}_5TERM_t, \\
\hat{d}_{t} &= -\hat{k}_6 - \hat{k}_7TERM_t, \\
\hat{e}_{t} &= -\hat{k}_8 - \hat{k}_9TERM_t
\end{align*}
\]

Given the coefficient estimates in Table 1, we can estimate the time-varying coefficients of the unrestricted conditional model, using the following relations.

Note that \( a_0 \) cannot be determined. To make a simple comparison of the economic importance of the various factors in the stochastic discount factor, we can compare the magnitude of \( \left| \frac{\hat{b}_{t}r_{M,t+1}}{a_0} \right| \) (i.e. the average absolute value of \( \frac{\hat{b}_{t}r_{M,t+1}}{a_0} \)) to the magnitude of \( \left| \frac{\hat{c}_{t}m_{t+1}}{a_0} \right| \) (i.e. the average absolute value of \( \frac{\hat{c}_{t}m_{t+1}}{a_0} \)), and so on.

Table 10 gives a summary of the economic importance of the factors. Remarkably, \( \left| \frac{\hat{c}_{t}m_{t+1}}{a_0} \right| \) is more than eight times as large as \( \left| \frac{e_{t}SMB_{t+1}}{a_0} \right| \), and more than three times as large as \( \left| \frac{\hat{b}_{t}r_{M,t+1}}{a_0} \right| \), and \( \left| \frac{\hat{d}_{t}HML_{t+1}}{a_0} \right| \), respectively. This is further evidence suggesting that the estimation factor is important in determining the price of an asset.
Table 10  Comparison of the economic importance of the factors in the unrestricted model.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Value</th>
<th>Relative value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{a}_t/a_0$</td>
<td>0.7994</td>
<td>2.70</td>
</tr>
<tr>
<td>$\hat{b}<em>t r</em>{M,t+1}/a_0$</td>
<td>0.08047</td>
<td>0.271</td>
</tr>
<tr>
<td>$\hat{c}<em>tm</em>{t+1}/a_0$</td>
<td>0.2965</td>
<td>1.00</td>
</tr>
<tr>
<td>$\hat{d}<em>t HML</em>{t+1}/a_0$</td>
<td>0.07815</td>
<td>0.264</td>
</tr>
<tr>
<td>$\hat{e}<em>t SMB</em>{t+1}/a_0$</td>
<td>0.03543</td>
<td>0.119</td>
</tr>
</tbody>
</table>

5 Conclusions

We analyze the importance of a derived "estimation factor" in the pricing of cross-sectional returns. Our starting point is a Lucas (1978) economy in which a representative agent does not observe the mean growth in aggregate dividends. We show that the expected excess return on any asset depends on its covariance not only with the market portfolio, but also with changes in the representative agent’s estimate of the mean growth in aggregate dividends. We test the model on cross-sectional excess returns on the 25 Fama-French size and book-to-market portfolios by estimating the stochastic discount factor with the use of GMM. We find evidence suggesting that the impact of the estimation factor is statistically and economically significant. The results also show that a conditional version of our model performs on a par with a conditional version of the Fama-French model, and that the static version of our model performs on a par with the static Fama-French model. However, the static models can be rejected at the 1% level using $J$-tests, whereas the conditional models cannot be rejected on the basis of $J$-tests. The results of our tests suggest that
the estimation factor is able to capture information related to present and future investment opportunities, and that it thus emerges as an important factor in asset pricing.

The results in this paper call for further studies of cross-sectional models of learning. The analysis contained in our paper could be developed further in at least two areas. Firstly, the theoretical model could be refined: e.g., the drift in the aggregate dividend growth could be made stochastic and mean-reverting; we could introduce labor income, and we could let the preferences allow for habit persistence. Secondly, the model could be tested on international data to see whether the results differ from the ones obtained using data from the US market.

Appendix: Proof of Proposition 1

Here, we will give the proof of Proposition 1, i.e., we will solve the filtering problem of the representative agent.

In matrix form, the dynamics of the representative agent’s observations on aggregate dividend growth and the economic variable $X_t$ are given by

$$
\begin{pmatrix}
\frac{dD_t}{D_t} \\
\frac{dX_t}{X_t}
\end{pmatrix} = 
\begin{pmatrix}
\mu_D \\
\mu_X
\end{pmatrix} dt + 
\begin{pmatrix}
\sigma_D & 0 \\
\sigma_X \rho & \sigma_X \sqrt{1-\rho^2}
\end{pmatrix} 
\begin{pmatrix}
dZ_t \\
dW_t
\end{pmatrix}.
$$

This can also be written as

$$
\begin{pmatrix}
\frac{dD_t}{D_t} \\
\frac{dX_t}{X_t}
\end{pmatrix} = 
\begin{pmatrix}
m_t \\
\mu_X
\end{pmatrix} dt + 
\begin{pmatrix}
\sigma_D & 0 \\
\sigma_X \rho & \sigma_X \sqrt{1-\rho^2}
\end{pmatrix} 
\begin{pmatrix}
d\bar{Z}_t \\
d\bar{W}_t
\end{pmatrix},
$$

where

$$
\begin{pmatrix}
d\bar{Z}_t \\
d\bar{W}_t
\end{pmatrix} = 
\begin{pmatrix}
\sigma_D & 0 \\
\sigma_X \rho & \sigma_X \sqrt{1-\rho^2}
\end{pmatrix}^{-1} 
\begin{pmatrix}
\mu_D - m_t \\
0
\end{pmatrix} dt + 
\begin{pmatrix}
d\bar{Z}_t \\
d\bar{W}_t
\end{pmatrix}.
$$
To fit into the Liptser and Shiryaev (2001) framework, we can write our observations as
\[
\begin{pmatrix}
\frac{dD_t}{dt} \\
\frac{dX_t}{dt}
\end{pmatrix} = \begin{pmatrix} 0 & 1 \\ \mu_X & 0 \end{pmatrix} \mu_d + \begin{pmatrix} \sigma_D & 0 \\ \sigma_X \rho & \sigma_X \sqrt{1-\rho^2} \end{pmatrix} \begin{pmatrix} dZ_t \\ dW_t \end{pmatrix}.
\]

Thus, by Theorem 12.7 in Liptser and Shiryaev (2001), the representative agent’s estimate \((m_t)\) satisfies the SDE
\[
dm_t = v_t \begin{pmatrix} 1 & 0 \end{pmatrix} \left( \begin{pmatrix} \sigma_D^2 & \rho \sigma_D \sigma_X \\ \rho \sigma_D \sigma_X & \sigma_X^2 \end{pmatrix} \right)^{-1} \begin{pmatrix} \left( \frac{dD_t}{dt} \right) \\ \left( \frac{dX_t}{dt} \right) \end{pmatrix} - \begin{pmatrix} m_t \\ \mu_X \end{pmatrix} dt,
\]
where the filtering error \((v_t)\) evolves according to
\[
\frac{dv_t}{dt} = -v_t^2 \begin{pmatrix} 1 & 0 \end{pmatrix} \left( \begin{pmatrix} \sigma_D^2 & \rho \sigma_D \sigma_X \\ \rho \sigma_D \sigma_X & \sigma_X^2 \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix}.
\]

Simple matrix algebra gives:
\[
\left( \begin{pmatrix} \sigma_D^2 & \rho \sigma_D \sigma_X \\ \rho \sigma_D \sigma_X & \sigma_X^2 \end{pmatrix} \right)^{-1} = \frac{1}{\sigma_D^2 \sigma_X^2 (1-\rho^2)} \begin{pmatrix} \sigma_X^2 & -\rho \sigma_D \sigma_X \\ -\rho \sigma_D \sigma_X & \sigma_D^2 \end{pmatrix},
\]
\[
\left( \begin{pmatrix} \sigma_D^2 & \rho \sigma_D \sigma_X \\ \rho \sigma_D \sigma_X & \sigma_X^2 \end{pmatrix} \right)^{-1} = \frac{1}{\sigma_D^2 \sigma_X^2 (1-\rho^2)} \begin{pmatrix} \sigma_X^2 & -\rho \sigma_D \sigma_X \\ -\rho \sigma_D \sigma_X & \sigma_D^2 \end{pmatrix},
\]
\[
\left( \begin{pmatrix} \sigma_D^2 & \rho \sigma_D \sigma_X \\ \rho \sigma_D \sigma_X & \sigma_X^2 \end{pmatrix} \right)^{-1} = \frac{1}{\sigma_D^2 \sigma_X^2 (1-\rho^2)} \begin{pmatrix} \sigma_X^2 & -\rho \sigma_D \sigma_X \\ -\rho \sigma_D \sigma_X & \sigma_D^2 \end{pmatrix},
\]
By equation (24), we have
\[
dm_t = v_t \begin{pmatrix} 1 & 0 \end{pmatrix} \left( \begin{pmatrix} \sigma_X^2 & -\rho \sigma_D \sigma_X \\ -\rho \sigma_D \sigma_X & \sigma_D^2 \end{pmatrix} \right) \begin{pmatrix} \sigma_D^2 & 0 \\ \sigma_X \rho & \sigma_X \sqrt{1-\rho^2} \end{pmatrix} \begin{pmatrix} dZ_t \\ dW_t \end{pmatrix} = v_t \begin{pmatrix} 1 & 0 \end{pmatrix} \left( \begin{pmatrix} \sigma_X^2 & -\rho \sigma_D \sigma_X \\ -\rho \sigma_D \sigma_X & \sigma_D^2 \end{pmatrix} \right) \begin{pmatrix} dZ_t \\ dW_t \end{pmatrix} =
\]

30
\[ v_t = \frac{1 - \rho^2 \sigma_D}{\sigma_D^2 (1 - \rho^2)} \sigma_D Z_t - v_t \frac{\rho}{\sigma_D \sqrt{1 - \rho^2}} dW_t. \]

Equation (25) reads

\[ \frac{dv_t}{dt} = -\frac{v_t^2}{\sigma_D^2 (1 - \rho^2)}. \tag{26} \]

Equation (24) can also be written as

\[ dm_t = \frac{v_t}{\sigma_D \sigma_X (1 - \rho^2)} \left( \sigma_X^2 \left( \frac{dD_t}{D_t} - m_t dt \right) - \rho \sigma_D \sigma_X \left( \frac{dX_t}{X_t} - \mu_X dt \right) \right), \]

or

\[ \frac{\sigma_D^2 \sigma_X^2 (1 - \rho^2)}{v_t} dm_t + \sigma_X^2 m_t dt = \sigma_X^2 \frac{dD_t}{D_t} - \rho \sigma_D \sigma_X \left( \frac{dX_t}{X_t} - \mu_X dt \right). \tag{27} \]

By Ito’s Lemma, we have

\[ d\left( \frac{m_t}{v_t} \right) = \frac{1}{v_t} dm - \frac{1}{v_t^2} dv_t dt = \frac{1}{v_t} dm_t + \frac{1}{\sigma_D^2 (1 - \rho^2)} m_t dt. \]

Thus, equation (27) can be written as

\[ \sigma_D^2 \sigma_X^2 (1 - \rho^2) d\left( \frac{m_t}{v_t} \right) = \sigma_X^2 \frac{dD_t}{D_t} - \rho \sigma_D \sigma_X \left( \frac{dX_t}{X_t} - \mu_X dt \right). \]

Solving this SDE, we have

\[ \frac{m_t}{v_t} = \frac{m_0}{v_0} + \frac{1}{\sigma_D^2 (1 - \rho^2)} \int_{s=0}^{t} \frac{dD_s}{D_s} - \frac{\rho}{\sigma_D \sigma_X (1 - \rho^2)} \left( \int_{s=0}^{t} \frac{dX_s}{X_s} - \mu_X t \right), \]

from which we can solve for the representative agent’s estimate,

\[ \frac{m_t}{v_0} = \frac{v_t}{v_0} m_0 + \frac{v_t}{\sigma_D^2 (1 - \rho^2)} \int_{s=0}^{t} \frac{dD_s}{D_s} - \frac{\rho v_t}{\sigma_D \sigma_X (1 - \rho^2)} \left( \int_{s=0}^{t} \frac{dX_s}{X_s} - \mu_X t \right). \]

Moreover, equation (26) is a Riccati equation, which can be solved as

\[ v_t = \frac{v_0^2 \sigma_D^2 (1 - \rho^2)}{\sigma_D^2 (1 - \rho^2) + v_0 t}. \]
References


