CAPM over the Long Run: 1926-2001*

Andrew Ang†
Columbia University and NBER

Joseph Chen‡
University of Southern California

This Version: 15 December 2005

JEL Classifications: C51, G12
Keywords: book-to-market effect, value premium,
conditional CAPM, time-varying beta

*The authors especially thank Geert Bekaert, Mikhail Chernov, Ethan Chiang, Frank Diebold, Ken French, Yufeng Han, Jonathan Lewellen, Stefan Nagel, Zhenyu Wang, and Guofu Zhou for very detailed comments. We also thank Ilan Cooper, Wayne Ferson, Cam Harvey, Bob Hodrick, Michael Johannes, Craig MacKinlay, an anonymous referee, and seminar participants at Columbia University, Duke University, INSEAD, SAC Capital Management, UCLA, UC Riverside, University of Alabama, USC, Wharton, the American Finance Association, the Econometric Society Annual Meeting, the Econometric Society World Congress, a Q-Group meeting, the Simulation-Based and Finite Sample Inference in Finance Conference at Quebec City, and the Western Finance Association. We are grateful to Ken French for making his data available.

†Columbia Business School, 805 Uris Hall, 3022 Broadway, New York, NY 10027; ph: (212) 854-9154; fax: (212) 662-8474; e-mail: aa610@columbia.edu; WWW: http://www.columbia.edu/∼aa610.

‡Marshall School of Business at USC, 701 Exposition Blvd., 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.

CAPM over the Long Run: 1926-2001

A conditional one-factor model can account for the spread in the average returns of portfolios sorted by book-to-market ratios over the long run from 1926-2001. In contrast, earlier studies document strong evidence of a book-to-market effect using OLS regressions in the post-1963 sample. However, the betas of portfolios sorted by book-to-market ratios vary over time and in the presence of time-varying factor loadings, OLS inference produces inconsistent estimates of conditional alphas and betas. We show that under a conditional CAPM with time-varying betas, predictable market risk premia, and stochastic systematic volatility, there is little evidence that the conditional alpha for a book-to-market trading strategy is statistically different from zero.

1 Introduction

Beginning with Basu (1983), many researchers have found significant evidence over the post-1963 period of a book-to-market effect, where stocks with high book-to-market ratios have higher average returns than what the CAPM predicts. This inference is based on conventional OLS with asymptotic standard errors, which relies on the assumptions that factor loadings are constant and that the market risk premium is stable. However, both of these assumptions are violated in data. In particular, betas of book-to-market portfolios vary substantially over time. For example, betas of the highest decile of book-to-market stocks reach over 3.0 prior to 1940 and fall to -0.5 at the end of 2001 (see also Kothari, Shanken and Sloan, 1995; Campbell and Vuolteenaho, 2004; Adrian and Franzoni, 2005).

After taking into account time-varying betas and market risk premia, we find that the conditional alpha of a book-to-market strategy, which goes long the top decile of stocks sorted by book-to-market ratios and shorts the bottom decile of stocks sorted by book-to-market ratios, is statistically insignificant in the long run. Strong evidence of a book-to-market effect can only be found in the post-1963 subsample based on standard OLS inference that assumes betas and market risk premia are constant. Thus, OLS inference is potentially misleading in small samples. Over the long run from 1926 to 2001, there is little evidence of a book-to-market premium and, under a conditional CAPM with time-varying betas, the market factor alone is able to explain the spread between the average returns of portfolios sorted on their book-to-market ratios.

When betas vary over time, standard OLS inference is misspecified and cannot be used to assess the fit of a conditional CAPM. Moreover, when betas vary over time and are correlated with time-varying market risk premia, OLS alphas and betas provide inconsistent estimates of conditional alphas and conditional betas, respectively. We prove that the magnitude of the inconsistency in the unconditional OLS alpha, relative to the true conditional alpha, cannot be determined without direct estimates of the underlying time-varying conditional beta process. This is true even if higher frequency data or short subsamples are used. Moreover, the common practice of employing rolling OLS estimates of betas understates the variance of the true conditional betas. The limiting distribution of the OLS alpha is also distorted from the standard asymptotic distribution which assumes constant betas. This distortion is intensified when shocks are very persistent in small samples. Consequently, a large unconditional OLS alpha may not necessarily imply the rejection of a conditional CAPM.

We estimate a conditional CAPM with time-varying betas, time-varying market risk premia, and stochastic systematic volatility. We directly take into account the time variation of conditional betas in estimating conditional alphas, rather than relying on incorrect OLS inference. Since conditional betas are very persistent, it is not surprising that small samples can generate significant OLS alphas that do not take into account time-varying betas. Thus, our model can explain the appearance of a book-to-market effect inferred from OLS alphas in the post-1963 subsample but not in the pre-1963 subsample, even when the true conditional alpha is constant and close to zero.

Our modelling approach has several advantages. First, Harvey (2001) shows that the estimates of the betas obtained using instrumental variables are very sensitive to the choice of instruments used to proxy for time variation in the conditional betas. Instead of using instrumental variables, we treat the time-varying betas as latent state variables and infer them directly from stock returns. Second, previous estimates of time-varying betas by Campbell and Vuolteenaho (2004), Fama and French (2005), and Lewellen and Nagel (2005), among others, assume discrete changes in betas across subsamples but constant betas within subsamples. That is, they consider the variation across averages of betas in each window, but ignore the variation of the betas within each window. In contrast, we treat betas as endogenous variables that slowly vary over time and directly estimate them.

Third, we capture predictable time variation in both the conditional mean and the conditional volatility of the market excess return. We model time-varying market premia by using a latent state variable for the conditional mean of the excess market return, similar to Merton (1971), Johannes and Polson (2003), Brandt and Kang (2004), among others. We use a stochastic volatility model that provides a better fit to the dynamics of stock returns compared to the GARCH models commonly used in the literature to model time-varying covariances (see comments by Danielsson, 1994, among others). An additional advantage of our framework is that we can take into account prior views on the strength of the book-to-market effect on conditional alphas. Furthermore, we also explicitly examine the finite-sample bias in unconditional OLS alphas and show how their posterior distributions differ from the distributions of conditional alphas.

Over the post-1963 sample, a book-to-market trading strategy that goes long the highest decile portfolio of stocks sorted on book-to-market ratios (value stocks) and goes short the lowest decile portfolio of book-to-market ratio stocks (growth stocks) has an OLS alpha of 0.60% per month with a robust asymptotic p-value, ignoring time variation of betas, of less than 0.01. However, under a one-factor conditional model with time-varying betas, OLS alphas of this magnitude frequently arise in small samples of around forty years. The 0.60% per month point estimate of the OLS alpha lies at the 67%-tile and more than 10% of the left-hand tail lies below zero. In contrast, there is little evidence that the conditional alpha is statistically significant. Using a diffuse prior, more than 10% of the lower-left tail of the posterior distribution of the book-to-market strategy conditional alpha lies below zero. Only an empiricist with an extremely strong prior belief in the existence of the book-to-market premium would conclude that a book-to-market premium exists. Thus, standard OLS inference grossly overstates the statistical significance of the book-to-market premium, even when robust asymptotic t-statistics are employed because it does not take into account time-varying factor loadings.

Our research goals are related to two contemporaneous papers by Lewellen and Nagel (2005) and Petkova and Zhang (2005), who also examine whether a conditional CAPM can explain the book-to-market effect. Lewellen and Nagel (2005) contend that no reasonable degree of covariation in conditional betas and market risk premia can generate the high average returns associated with value stocks in the post-1963 sample. However, they do not address the non-existence of the book-to-market effect in the pre-1963 sample and do not incorporate the larger variation in betas found over the long run from 1921-2001. In addition, Lewellen and Nagel's method of inferring the dynamics of time-varying conditional betas with a series of OLS regressions with constant betas produces inconsistent estimates of both conditional alphas and betas. Petkova and Zhang (2005) also argue that while there is a positive correlation of shocks to the betas of value stocks and shocks to the market risk premium, this correlation is not high enough to explain the book-to-market effect. This correlation is only estimated indirectly, through instrumental proxies for conditional betas and market risk premia. Neither Lewellen and Nagel (2005) nor Petkova and Zhang (2005) examine the distortions induced by time-varying betas on the asymptotic distribution of the OLS alphas, which has as much importance for statistical inference as the size of the bias in the OLS alpha.

Our results question the conventional wisdom that there exists a strong evidence of a book-to-market effect. In particular, we find that a single-factor model performs substantially better than previously believed at explaining the book-to-market premium. Whereas Davis (1994) and Davis, Fama and French (2000) argue for the existence of a book-to-market effect prior to 1963 and advocate the use of an unconditional three-factor model, they neither examine the fit of an unconditional one-factor regression nor estimate a conditional CAPM. We demonstrate that a single conditional one-factor model is sufficient to explain the average returns of book-to-market portfolios both post-1963 and over the long run. We do not claim that a conditional CAPM is the complete model for the cross-section of stock returns. In particular, more powerful tests like the stock characteristic approaches of Daniel and Titman (1997) may be able to reject multi-factor models and their implied conditional CAPM counterparts. Nevertheless, our results show that a simple conditional single-factor model can account for a substantial portion of the book-to-market effect and that the evidence for the book-to-market effect is not as strong as previously believed.

The remainder of the paper is organized as follows. Section 2 discusses various aspects of the book-to-market portfolio returns over the long run from 1926 to 2001. In Section 3, we show that estimating time-varying betas by standard OLS estimators produces biased and inconsistent estimates with distorted asymptotic distributions. We show that the magnitude of the inconsistency and the distortion cannot be corrected without directly estimating the conditional betas. In Section 4, we develop a methodology for consistently estimating time-varying betas in a conditional CAPM. Section 5 presents the estimation results and examines the book-to-market effect under parameter uncertainty, time-varying factor loadings, and small sample biases. Finally, Section 6 concludes.

2 The Book-to-Market Effect Over the Long Run

We focus on the set of decile portfolios sorted on book-to-market ratios constructed by Davis (1994) and Davis, Fama and French (2000).[1] We use the return on a value-weighted portfolio of all stocks listed on the NYSE, AMEX, and NASDAQ as the market return. All returns are calculated in excess of the one-month Treasury bill rate from Ibbotson Associates. Our data differs from other contemporaneous studies in that we focus on the overall book-to-market effect. Loughran (1997) notes that the book-to-market effect is much stronger among smaller stocks. In contrast to our approach that focuses purely on standard book-to-market sorted portfolios, Fama and French (1993, 2005), Lewellen and Nagel (2005) and Petkova and Zhang (2005) enhance the book-to-market effect by placing greater weight on small stocks. These authors construct 2×3 or 5×5 size and book-to-market sorted portfolios. Section 2.1 reexamines the evidence for the book-to-market effect using OLS one-factor regressions. In Section 2.2, we take a first glance at examining the time-varying nature of betas of the book-to-market portfolios.

[1] We obtain data on book-to-market portfolios from Kenneth French's data library, which is at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/.

2.1 Returns on Book-to-Market Portfolios

In Table 1, we report average monthly raw returns and volatilities together with OLS alphas and betas estimated from standard OLS regressions over various samples:

$$r_{i,t} = \hat{\alpha}_T^{OLS} + \hat{\beta}_T^{OLS} r_{m,t} + \varepsilon_{i,t}^{OLS}, \tag{1}$$

where $r_{i,t}$ is the excess stock return, $r_{m,t}$ is the excess market return, and $\varepsilon_{i,t}^{OLS}$ is an orthogonal shock. In equation (1), we denote the estimated alpha of the OLS model as $\hat{\alpha}_T^{OLS}$, with an OLS superscript to emphasize that it is an alpha constructed under the assumptions of OLS. Similarly, we also distinguish the OLS estimate of systematic market risk exposure, $\hat{\beta}_T^{OLS}$, with an OLS superscript. We append $\hat{\alpha}_T^{OLS}$ and $\hat{\beta}_T^{OLS}$ with T subscripts to emphasize that the OLS estimates are computed over a sample size of T. While $\hat{\alpha}_T^{OLS}$ and $\hat{\beta}_T^{OLS}$ in equation (1) should also carry i subscripts to denote that they differ across stocks, we omit them for clarity.

Panel A of Table 1 lists summary statistics for the full sample from July 1926 to December 2001, while Panels B and C cover the subsamples from July 1926 to June 1963 and from July 1963 to December 2001, respectively. For each of these subsamples, we report alphas and betas estimated by OLS, assuming constant alphas and betas over each subsample. We also report statistics on a book-to-market strategy ("BM" portfolio) which is a zero-cost portfolio that goes long the decile 10 book-to-market portfolio (value stocks) and goes short the decile 1 book-to-market portfolio (growth stocks). We compute t-statistics of the OLS alphas using Newey-West (1987) standard errors.

The first surprising result in Table 1 is that the alphas from an unconditional one-factor model are insignificant for book-to-market sorted portfolios over the long run, from 1926 to 2001. In Panel A, which uses the full 75 years of data, there is a weakly increasing relationship between the mean returns and the book-to-market ratios. However, once we control for the market beta, the individual OLS alphas become insignificant and we observe no pattern between the OLS alphas across the book-to-market deciles.[2] In particular, the Newey-West t-statistic for the difference between the OLS alphas of the lowest and highest book-to-market decile portfolios is only 0.97.[3] Much of the lack of a pattern in the alphas can be attributed to the weakly increasing pattern in the betas. Similarly, over the 1926-1963 subsample reported in

[2] Neither Davis (1994) nor Davis, Fama and French (2000) run a simple unconditional CAPM regression, or test for the significance of size or book-to-market factors relative to an unconditional one-factor model.

[3] A Gibbons-Ross-Shaken (1989) (GRS) test for joint significance of the α's across all portfolios fails to reject at the 5% level over 1926-2001. Even from 1963-2001, the GRS test p-value is only borderline significant with a p-value of 0.05.

Panel B, we also fail to find any evidence of a book-to-market effect, as the difference in OLS alphas between value stocks and growth stocks is slightly negative, at -0.16% per month.

In contrast, most prior empirical work examining the book-to-market effect has focused on the period after July 1963, which we report in Panel C. In this post-1963 subsample, the unconditional one-factor model fails. This latter sample has received significantly more attention than the earlier sample because data on firm book values are readily available on COMPUSTAT after this date. The raw average monthly returns of the portfolios over this period exhibit an increasing pattern across the book-to-market decile portfolios. The difference in returns between the value stocks and the growth stocks is 0.53% per month, with a Newey-West t-statistic of 2.16. Once we control for the market factor in an OLS regression, the $\hat{\alpha}_T^{OLS}$ estimates become strictly increasing and the spread in the expected returns widens to 0.60% per month, with a Newey-West t-statistic of 2.51. Unlike the pre-1963 subsample, there is no pattern in the betas across the book-to-market portfolios. This is the familiar result of Fama and French (1992, 1993), who report a strong book-to-market effect in the latter half of the century using OLS alphas.

The main difference across the two subsamples is the presence of a pattern in the OLS estimates of betas in the pre-1963 subsample, but not in the post-1963 subsample. This finding indicates two important facts. First, betas of the book-to-market portfolios appear to vary substantially across time. In the pre-1963 subsample, the OLS beta of the book-to-market strategy is positive at 0.69 and is large enough to explain the performance of the strategy. In the post-1963 subsample, the OLS beta is negative at -0.16 and can no longer explain the performance of the book-to-market strategy. The second fact is that the unconditional OLS regression of equation (1) is misspecified. The OLS specification assumes that betas are constant, but they clearly differ across the two subsamples. We now examine in greater detail the time-varying nature of betas across the long run from 1926 to 2001 and examine the implications of making inference using a misspecified OLS regression described by equation (1).

2.2 Rolling OLS Betas of Book-to-Market Portfolios

We use rolling OLS betas estimated over shorter 60-month windows to provide some evidence which suggests that the true conditional betas vary over time. While the rolling 60-month OLS regression is a common procedure for assessing time-varying betas (since as early as Fama and MacBeth, 1973), we emphasize later in Section 3 that rolling OLS betas do not directly reveal the true betas since OLS estimates of conditional betas are misspecified. Nevertheless, rolling OLS betas can provide some rough characterizations of the true conditional beta process. In particular, the rolling OLS beta estimates provide a glimpse of what the autocorrelation and standard deviation of the true conditional betas are, and can be used to form a prior for our estimates of the true beta data-generating process.

Table 1 shows a remarkable drift in the OLS betas of the book-to-market portfolios over time. For example, in Panel B, from July 1926 to June 1963, stocks with the highest book-to-market ratios have the highest betas. The decile 10 value stock portfolio has a high average return of 1.24% per month and a corresponding high $\hat{\beta}_T^{OLS}$ of 1.66. In contrast, Panel C shows that in the post-1963 subsample, stocks with the highest book-to-market ratios have an OLS beta of $\hat{\beta}_T^{OLS} = 0.95$, but a very high average return of 0.91% per month. To visually illustrate the variation in the OLS betas that we observe in the data, we plot rolling estimates of the market OLS betas over time in Figure 1, similar to Franzoni (2004), Campbell and Vuolteenaho (2004), and Adrian and Franzoni (2005). We compute rolling estimates of the time-varying betas by regressing portfolio returns on the market return using rolling samples of 60 months.

Figure 1 shows that the rolling OLS betas of value stocks are highly persistent, but broadly reflect a downward trend. In particular, the value stock OLS betas reach a high of 2.2 during the 1940s and fall to around 0.5 in December 2001. Figure 1 also shows that the variation in the OLS betas of the growth stock portfolio is much smaller than the variation of the value stock OLS betas. Nevertheless, there is still some evidence that the OLS betas of growth stocks have a slow, mean-reverting component. However, these 60-month rolling OLS betas are, at best, 60-month averages of variation in the true conditional betas. Hence, these plots of time-varying OLS betas strongly suggest that the true conditional betas also vary over time. Since the rolling OLS betas of value stocks in Figure 1 resemble a random walk, we also expect the true conditional betas to be very persistent.

In summary, a one-factor unconditional regression can account for the book-to-market effect over the full sample (1927-2001) and over the pre-1963 sample, but fails over the post-1963 sample. A one-factor unconditional regression produces large $\hat{\alpha}_T^{OLS}$ estimates for the book-to-market strategy only over the post-1963 sample. Moreover, betas of portfolios are not constant as assumed in standard OLS specifications, but vary significantly across time. These results have several implications. First, while the one-factor CAPM regression fails to reject the null that $\hat{\alpha}_T^{OLS} = 0$ in the long run, this does not mean that we can conduct correct inference regarding the true conditional alpha from a data-generating process with time-varying betas since OLS regressions are misspecified. Similarly, while the OLS alpha of the book-to-market strategy is a large $\hat{\alpha}_T^{OLS} =$ 0.60% per month post-1963, this also does not necessarily imply that there exists a positive conditional alpha in the true data-generating process. The fact that the parameters in the OLS regressions change so dramatically across samples suggests that betas, and perhaps other characteristics of the market, vary over time. Furthermore, the instability of the OLS estimates also suggests that the effects of small sample bias and parameter uncertainty may play a role in robust statistical inference. Since the OLS regressions are misspecified, we now develop a framework for making robust inference in a setting with time-varying risk loadings.

3 Theory

The goal of this section is to emphasize the difference between a conditional CAPM and the unconditional one-factor regression estimated by OLS. We show that when conditional betas vary over time, OLS cannot provide consistent estimates of either conditional betas or conditional alphas. Section 3.1 illustrates the differences between a conditional CAPM and an unconditional CAPM. In Section 3.2, we use a highly stylized model to derive closed-form asymptotic distributions for the OLS estimators (but we use a richer specification for our empirical work in Section 4). Sections 3.3 and 3.4 characterize the limiting asymptotic distributions for OLS betas and OLS alphas, respectively.

3.1 The Conditional and Unconditional CAPM

Our model is a reduced-form version of a conditional CAPM:

$$r_{i,t} = \alpha^C + \beta_t r_{m,t} + \bar{\sigma}\varepsilon_{i,t}, \tag{2}$$

where $r_{i,t}$ is the excess stock return, $r_{m,t}$ is the excess market return, $\varepsilon_{i,t}$ is an independent and identically distributed (IID) standard normal shock that is orthogonal to all other shocks, and $\bar{\sigma}$ represents the stock's idiosyncratic volatility. We define the conditional beta of stock i in the standard way as:

$$\beta_t = \frac{\text{cov}_{t-1}(r_{i,t}, r_{m,t})}{\text{var}_{t-1}(r_{m,t})} \tag{3}$$

and define α^C to be the conditional alpha which is the proportion of the conditional expected return that is left unexplained by the stock's systematic exposure. We append the conditional alpha, α^C, with a C superscript to distinguish it from the estimate of alpha obtained from the misspecified OLS, α^{OLS}, from equation (1). While α^C, β_t and $\bar{\sigma}$ should also carry i subscripts to denote that they differ across assets, we omit them for ease of notation.

To complete the model, we specify the dynamics of the market excess return as:

$$r_{m,t} = \mu_t + \sqrt{v_t}\varepsilon_{m,t}, \tag{4}$$

where $\mu_t = \mathrm{E}_{t-1}[r_{m,t}]$ denotes the conditional mean of the market and $v_t = \mathrm{var}_{t-1}[r_{m,t}]$ denotes the conditional market volatility. Under the null of the conditional CAPM, the conditional alpha is zero, $\alpha^C = 0$, and the systematic risk represented by β_t is solely responsible for determining expected returns. If we reject the null hypothesis that $\alpha^C = 0$, we would conclude that the conditional CAPM cannot price the average excess returns of asset i.

The unconditional CAPM used by Black, Jensen and Scholes (1972), Fama and MacBeth (1973), Fama and French (1992, 1993) and others differs from the conditional CAPM in equation (2) by specifying a constant beta over the entire sample period. Many authors, including Fama and French (1993), estimate the regression (1) on portfolios of stocks sorted by book-to-market ratios and reject that the OLS alpha, $\hat{\alpha}_T^{OLS}$, is equal to zero. However, using the unconditional factor model in equation (1) estimated by OLS to make inference regarding the conditional CAPM in equation (2) is treacherous for several reasons.

First, Jagannathan and Wang (1996) show that if time-varying conditional betas are correlated with time-varying market risk premia, then the conditional CAPM in equation (2) is observationally equivalent to an unconditional multifactor model:

$$\mathrm{E}[r_{i,t}] = \alpha^C + \mathrm{cov}(\beta_t, \mu_t) + \bar{\beta}\bar{\mu}_m, \tag{5}$$

where $\bar{\beta} = \mathrm{E}(\beta_t)$ and $\bar{\mu}_m = \mathrm{E}(r_{m,t})$ are the unconditional means of the beta and the market premium process, respectively. Under the null of a conditional CAPM, we would expect that the estimate of the unconditional OLS alpha, $\hat{\alpha}_T^{OLS}$, captures both the conditional alpha, α^C, and the interaction of time-varying factor loadings and market risk premia.

Second, Jagannathan and Wang (1996) show that we need multiple unconditional factors in the OLS regression in equation (1) to capture the same effects as the single-factor conditional model in equation (2), due to the $\mathrm{cov}(\beta_t, \mu_t)$ term in equation (5). Hence, any statement made about the failure of an unconditional CAPM to capture the spread of average returns in the cross-section does not imply that a conditional CAPM cannot explain the cross-sectional spread of average returns. Our main focus is on the ability of the conditional CAPM in equation (2) to explain the cross-section of average returns of stocks sorted by book-to-market ratios, rather than on the unconditional OLS CAPM regression in equation (1).

Third, a conditional CAPM implies an unconditional one-factor model only in the case when β_t is uncorrelated with μ_t. In this special case, equation (5) reduces to $\mathrm{E}[r_{i,t}] = \alpha^C +$

$\bar{\beta}\bar{\mu}_m$. However, when conditional betas and market risk premia are correlated, OLS fails to provide consistent estimates of both the conditional alpha and the conditional betas in equation (2). Moreover, the degree of the inconsistency depends on unknown parameters driving the conditional beta process that are not directly observed. Hence, any inference on the conditional CAPM in equation (2) cannot be made on the basis of OLS estimates of the unconditional one-factor model in equation (1). Finally, the limiting distributions of the OLS alphas and betas ($\hat{\alpha}_T^{OLS}$ and $\hat{\beta}_T^{OLS}$ in equation (1)) are significantly distorted from their conventional OLS asymptotic distributions that assume constant factor loadings. We now illustrate each of these points in the context of a very simple model for which we can analytically derive the limiting distributions of $\hat{\alpha}_T^{OLS}$ and $\hat{\beta}_T^{OLS}$.

3.2 A Simple Model

We first consider the simplest possible setting where stock i's excess return, $r_{i,t}$, is driven by a time-varying beta process. Suppose that in the true conditional model (2), β_t and $r_{m,t}$ are stochastic and correlated with an unconditional correlation of $\rho_{m\beta}$. In this simplest possible setting, we specify that:

$$\begin{aligned} \beta_t &\overset{IID}{\sim} N(\bar{\beta}, \bar{\sigma}_\beta^2) \\ \text{and} \quad r_{m,t} &\overset{IID}{\sim} N(\bar{\mu}_m, \bar{\sigma}_m^2), \end{aligned} \tag{6}$$

where $\bar{\beta}$ is the unconditional beta, $\bar{\sigma}_\beta^2$ is the unconditional variance of beta, $\bar{\mu}_m$ is the unconditional mean of the excess market return, $\bar{\sigma}_m^2$ is the variance of the excess market return, and corr$(r_{m,t}, \beta_t) = \rho_{m\beta}$. In the statistics literature, this is a standard random coefficient model (see, for example, Cooley and Prescott, 1976).

Our goal is to characterize the asymptotic distribution of the OLS estimators:

$$\begin{aligned} \hat{\alpha}_T^{OLS} &= \frac{1}{T}\sum_{t=1}^{T} r_{i,t} - \hat{\beta}_T^{OLS}\frac{1}{T}\sum_{t=1}^{T} r_{m,t} \\ \text{and} \quad \hat{\beta}_T^{OLS} &= \left(\frac{1}{T}\sum_{t=1}^{T}(r_{m,t}-\bar{r}_m)^2\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}(r_{m,t}-\bar{r}_m)r_{i,t}\right), \end{aligned} \tag{7}$$

where $\bar{r}_m = (1/T)\sum r_{m,t}$ represents the sample average of the excess market return, under the data generating process of equation (6). We relegate the full derivation of the asymptotic distribution of $\sqrt{T}(\hat{\alpha}_T^{OLS} - \mathrm{E}[\hat{\alpha}_T^{OLS}])$ and $\sqrt{T}(\hat{\beta}_T^{OLS} - \mathrm{E}[\hat{\beta}_T^{OLS}])$ to Appendix A.[4]

[4] Standard statistics textbooks recognize that in applying OLS to the model of equations (2) and (6), OLS is

We note that the OLS alpha and beta estimates in equation (7) are not pivotal statistics, as their distribution depends explicitly on the parameters of the data-generating process in equation (6). But, the OLS alpha is precisely the statistic most often used by researchers to judge the significance of any CAPM anomaly. Our focus is not to develop a pivotal statistic to estimate time-varying betas, but rather to show how the OLS alpha and beta distributions are affected by time-varying betas. Thus, we point out that inference based on OLS alpha and beta estimates are unreliable in the presence of time-varying factor loadings.

3.3 Asymptotic Distribution of $\sqrt{T}(\hat{\beta}_T^{OLS} - \mathrm{E}[\hat{\beta}_T^{OLS}])$

To understand the distortions that OLS induces on a system with time-varying betas relative to the standard case, it is helpful to write the residual term, $\varepsilon_{i,t}^{OLS}$, of the regression (1) in the form of an omitted variable $(\beta_t - \bar{\beta})r_{m,t}$:

$$\varepsilon_{i,t}^{OLS} = (\beta_t - \bar{\beta})r_{m,t} + \bar{\sigma}\varepsilon_{i,t}. \tag{8}$$

Unlike the usual case of a constant beta, this omitted variable is a product of two normal distributions and can cause at least three distinct problems in the OLS estimates. First, the residuals are heteroskedastic. Second, in practice, β_t is likely to be very persistent (but is assumed to be IID in this simple setting for tractability), which leads to serial correlations in the residuals. Both the problem of heteroskedasticity and serial correlation in $\varepsilon_{i,t}^{OLS}$ can, in principle, be corrected by a heteroskedasticity and autocorrelation consistent (HAC) estimator like Newey and West (1987). Note that this is only an asymptotic correction, so a HAC estimator still ignores the effect of any small sample distortion and bias. The major problem that cannot be corrected by using a HAC estimator is that the OLS residuals, $\varepsilon_{i,t}^{OLS}$, are correlated with the regressor, $r_{m,t}$, which leads to biased and inconsistent OLS estimates:

$$\mathrm{E}[\hat{\beta}_T^{OLS}] = \bar{\beta} + \frac{\rho_{m\beta}\bar{\sigma}_\beta}{\bar{\sigma}_m}\bar{\mu}_m. \tag{9}$$

inconsistent (see, for example, Greene, 2002), but they do not derive the limiting distribution of the OLS estimators. As we show, this derivation is non-trivial as it involves quadratic functions of normals, but this exercise is necessary to interpret both the bias and the sampling dispersion of the OLS estimates. Foster and Nelson (1996) develop a series of rolling weighted OLS regressions, where the optimal weights are a function of the underlying data generating process, that can provide efficient estimates of conditional betas. This case is different to the standard OLS regressions run over the whole sample that are common in the literature. Foster and Nelson also do not consider asymptotic distributions of OLS alphas with time-varying conditional betas.

The magnitude of the inconsistency of $\hat{\beta}_T^{OLS}$ in equation (9) depends on the unknown parameters $\bar{\beta}$, $\rho_{m\beta}$, $\bar{\sigma}_\beta$, and $\bar{\mu}_m$. Lewellen and Nagel (2005) make inferences on the properties the conditional alphas and the conditional betas by estimating a series of high frequency OLS regressions in subsamples. However, note that the inconsistency term in equation (9) depends on the ratio of $\bar{\sigma}_\beta$ to $\bar{\sigma}_m$ which is invariant to the sampling frequency and can be potentially very large. Moreover, even if subsample regressions are used to estimate OLS betas, this does not remove the inconsistency since the conditional betas continue to vary within windows. Hence, there is no way to correct for the inconsistency without knowing $\bar{\beta}$, $\rho_{m\beta}$, and $\bar{\sigma}_\beta$, and these parameters can only be obtained by directly estimating the conditional beta series. In data, since the market risk premia and the variance of the market change over time (see Schwert, 1989), the magnitude of the inconsistency of the OLS estimate, $\hat{\beta}_T^{OLS}$, is also time-varying. The OLS beta provides a consistent estimate of the mean of the true beta process only in the case when the betas are uncorrelated with the market return, $\rho_{m\beta} = 0$.

There is also a distortion of the standard limiting distribution induced by the presence of time-varying betas. The asymptotic distribution of $\sqrt{T}(\hat{\beta}_T^{OLS} - \mathrm{E}[\hat{\beta}_T^{OLS}])$ is given by:

$$\sqrt{T}(\hat{\beta}_T^{OLS} - \mathrm{E}[\hat{\beta}_T^{OLS}]) \xrightarrow{d} N\left(0, (3 + 12\rho_{m\beta}^2)\bar{\sigma}_\beta^2 + (1 - \rho_{m\beta}^2)\bar{\mu}_m^2 \frac{\bar{\sigma}_\beta^2}{\bar{\sigma}_m^2} + \frac{\bar{\sigma}^2}{\bar{\sigma}_m^2}\right). \qquad (10)$$

The last term for the asymptotic variance is $\bar{\sigma}^2/\bar{\sigma}_m^2$, which is the asymptotic variance for the standard OLS case without any time variation in the betas ($\bar{\sigma}_\beta = 0$). The other two terms reflect the contribution of the endogenous regressor β_t that increases the variance of the $\hat{\beta}_T^{OLS}$ estimator relative to the constant beta case.

This increase is not small, even if the betas are uncorrelated with the market return. For example, suppose that $\rho_{m\beta} = 0$, $\bar{\mu}_m = 0.0066$, and $\bar{\sigma}_m = 0.055$, where the excess market parameters are calibrated from the sample mean and sample standard deviation of the monthly excess market returns over 1927-2001. We set the stock idiosyncratic volatility at $\bar{\sigma} = 0.06$ and set the $\bar{\sigma}_\beta$ parameter of the book-to-market portfolios to be 0.468. The last parameter represents the monthly unconditional standard deviation of the betas, which we discuss below in Section 5.1. Then, the ratio of the true asymptotic variance in equation (10) to the standard OLS asymptotic variance is approximately two. This is a conservative estimate because $\rho_{m\beta}$ is unlikely to be zero. Hence, the true limiting variance of $\hat{\beta}_T^{OLS}$ is likely to be larger than the variance implied by standard OLS theory. Therefore, even if we know the correct adjustment for the inconsistency of the $\hat{\beta}_T^{OLS}$ estimator, we cannot at all be confident about the precision of the estimate of the conditional beta.

3.4 Asymptotic Distribution of $\sqrt{T}(\hat{\alpha}_T^{OLS} - \mathrm{E}[\hat{\alpha}_T^{OLS}])$

Jaganathan and Wang (1996) and Lewellen and Nagel (2005), among others, note that the OLS estimate of alpha, $\hat{\alpha}_T^{OLS}$, is a biased estimate of the conditional alpha, α^C, in the conditional CAPM specified in equation (2). In our simple model, $\mathrm{E}[\hat{\alpha}_T^{OLS}]$ is given by:

$$\mathrm{E}[\hat{\alpha}_T^{OLS}] = \alpha^C + \frac{\rho_{m\beta}\bar{\sigma}_\beta}{\bar{\sigma}_m}(\bar{\sigma}_m^2 - \bar{\mu}_m^2). \tag{11}$$

Note that $\hat{\alpha}_T^{OLS}$ provides a consistent estimate of α^C only when the market return process and the conditional betas are uncorrelated. If $\rho_{m\beta} \neq 0$, then direct knowledge of the conditional beta process is required to correct for the inconsistency of $\hat{\alpha}^{OLS}$. As a rough estimate, if we assume that $\rho_{m\beta} = 0.1$ (see below), $\bar{\sigma}_\beta = 0.468$, $\bar{\mu}_m = 0.0066$, and $\bar{\sigma}_m = 0.055$, then equation (11) indicates that $\hat{\alpha}_T^{OLS}$ overstates the true value of α^C by over 0.26% per month.[5]

In our simple model, the asymptotic distribution of the OLS estimate, $\hat{\alpha}_T^{OLS}$, is given by:

$$\begin{aligned}&\sqrt{T}(\hat{\alpha}_T^{OLS} - \mathrm{E}[\hat{\alpha}_T^{OLS}])\\ &\qquad \xrightarrow{d} N\left(0, 10\rho_{m\beta}^2\bar{\mu}_m^2\bar{\sigma}_\beta^2 + (1-\rho_{m\beta}^2)(\bar{\sigma}_m^4 + \bar{\mu}_m^4)\frac{\bar{\sigma}_\beta^2}{\bar{\sigma}_m^2} + (\bar{\sigma}_m^2 + \bar{\mu}_m^2)\frac{\bar{\sigma}^2}{\bar{\sigma}_m^2}\right).\end{aligned} \tag{12}$$

The asymptotic variance of $\hat{\alpha}_T^{OLS}$ in equation (12) has three terms. The third term, $(\bar{\sigma}_m^2 + \bar{\mu}_m^2)\bar{\sigma}^2/\bar{\sigma}_m^2$, is the regular asymptotic variance for the OLS estimate for the intercept term in an environment with non-stochastic betas. In cases where $\bar{\sigma}_\beta$ is large, the asymptotic variance of $\hat{\alpha}_T^{OLS}$ increases substantially when $\bar{\sigma}_\beta$ is not zero, relative to the standard OLS asymptotic variance. Again, we cannot compute the degree of distortion relative to the regular OLS standard error case without knowing $\bar{\sigma}_\beta$ and $\rho_{m\beta}$. Moreover, these parameters cannot be estimated without knowing the conditional, latent beta series. Using asymptotic theory, we can estimate the increase in the asymptotic variance of $\hat{\alpha}_T^{OLS}$ induced by the time-varying regressors by using a HAC estimate of the variance only in the special case when $\rho_{\beta m} = 0$. When $\rho_{\beta m} \neq 0$, HAC estimators are invalid.

In summary, we cannot obtain consistent estimates of conditional betas or alphas by OLS. Neither the adjustments for the magnitude of the inconsistency nor the corrections for the distortions in the asymptotic variances of the OLS estimators can be accomplished without direct

[5] The term we analyze here is not fully present in the empirical analysis of conditional alphas of Lewellen and Nagel (2005) because they (counter-factually) assume that OLS is consistent within each subsample. Our method consistently accounts for the time variation in conditional betas within a given window where Lewellen and Nagel have assumed the OLS betas are consistent.

knowledge of the dynamics of the conditional beta process. We now propose a richer model with time-varying conditional betas, time-varying market risk premia, and time-varying systematic volatility and explain how such a model can be estimated.

4 A Conditional CAPM with Time-Varying Betas

The asymptotic distributions of $\sqrt{T}(\hat{\beta}_T^{OLS} - \mathrm{E}[\hat{\beta}^{OLS}])$ and $\sqrt{T}(\hat{\alpha}_T^{OLS} - \mathrm{E}[\hat{\alpha}^{OLS}])$ in equations (10) and (12) are likely to understate the true variation of $\hat{\beta}_T^{OLS}$ and $\hat{\alpha}_T^{OLS}$ in small samples for two reasons. First, we expect that rather than the conditional betas being drawn from an IID process, conditional betas are more likely to incorporate predictable, slow, mean-reverting components. While we derived the asymptotic distributions using a Central Limit Theorem that can be generalized to incorporate persistence in β_t, a high autocorrelation of the β_t process will cause the asymptotic variance to significantly understate the true variance in small samples. Second, the market process is also empirically not a normal IID process. A more realistic empirical description of the market return is that it also incorporates persistent components both in its conditional mean and conditional volatility. The addition of time-varying components in the market return process further distorts the asymptotic inference of the OLS estimators. For our empirical application, we enrich the simple model of the previous section to incorporate persistent conditional betas, time-varying market risk premia, and stochastic systematic volatility.

4.1 The Model

In our fully specified conditional CAPM, we assume that the latent conditional betas in equation (2) follow an AR(1) process:

$$\begin{aligned} r_{i,t} &= \alpha^C + \beta_t r_{m,t} + \bar{\sigma}\varepsilon_{i,t} \\ \text{and} \quad \beta_t &= \beta_0 + \phi_\beta \beta_{t-1} + \sigma_\beta \varepsilon_{\beta,t}, \end{aligned} \tag{13}$$

where β_t refers to the conditional beta of stock i defined in equation (3). Again, in equation (13), α^C, β_t, $\bar{\sigma}$, β_0, ϕ_β, and σ_β should all have i subscripts, but we omit them for simplicity. We are interested in $r_{i,t}$ representing the returns on the book-to-market strategy. Following the standard set-up of a conditional factor model where the idiosyncratic volatility shocks are uncorrelated with systematic components, we specify $\varepsilon_{i,t}$ to be drawn from an IID normal distribution that is independent of the shocks to the systematic components.

We expect that the conditional betas in equation (13) vary slowly over time with ϕ_β close to one. This view is suggested both from economic theory and from prior empirical studies. For example, Gomes, Kogan and Zhang (2003), suggest that betas are a function of productivity shocks, which are often calibrated with an autocorrelation of 0.95 at the quarterly horizon. This translates into a monthly autocorrelation of conditional betas above 0.98. In Santos and Veronesi (2004), stock betas change as the ratio of labor income to total consumption changes, which is also a highly persistent variable. Since the firms in the book-to-market portfolios change over time, portfolio reconstitution could also cause time variation in the portfolio betas. Since the OLS rolling betas in Figure 1 have a wide range, we expect that the conditional shocks to the true betas of the book-to-market portfolios can be quite variable, or σ_β could be large. In data, Fama and French (1997) also report substantial variation in factor loadings for industry portfolios, while Ferson and Harvey (1999) show that there is a large variation in the market betas of portfolios sorted by size and book-to-market ratios. Hence, our prior is that β_t should be highly persistent and conditional shocks to β_t can potentially be large.

We further specify that the excess market return, $r_{m,t}$, in equation (4) has a conditional market risk premium, μ_t, and exhibits stochastic systematic volatility, v_t:

$$r_{m,t} = \mu_t + \sqrt{v_t}\varepsilon_{m,t}, \tag{14}$$

where

$$\begin{aligned} \mu_t &= \mu_0 + \phi_\mu \mu_{t-1} + \sigma_\mu \varepsilon_{\mu,t} \\ \text{and} \quad \ln v_t &= v_0 + \phi_v \ln v_{t-1} + \sigma_v \varepsilon_{v,t}. \end{aligned} \tag{15}$$

The shocks, $\varepsilon_{m,t}$, $\varepsilon_{\mu,t}$ and $\varepsilon_{v,t}$, are normally distributed zero mean, unit standard deviation, normally distributed shocks that are potentially correlated. In equation (15), we allow the market risk premium to be a slowly mean-reverting latent process. This is the same specification used in the portfolio allocation literature, beginning with Merton (1971). We model log volatility as a latent AR(1) process, following Jacquier, Polson and Rossi (1994). The log process restricts volatility to be positive and induces fat tails in the distribution for the market return. Since Brandt and Kang (2004) find that the correlation between $\varepsilon_{m,t}$ and $\varepsilon_{\mu,t}$ is insignificant, we set this correlation to be zero. We also specify $\varepsilon_{m,t}$ and $\varepsilon_{v,t}$ to be orthogonal. However, we let $\varepsilon_{\mu,t}$ and $\varepsilon_{v,t}$ have a correlation of $\rho_{\mu v}$. This captures a leverage effect, and allows market conditional expected returns and stochastic volatility to move together. To allow the market risk premia to be correlated with conditional betas, we let $\varepsilon_{\mu,t}$ and $\varepsilon_{\beta,t}$ in equations (13) to (15) have a non-zero correlation of $\rho_{\mu\beta}$.

The OLS alpha, $\hat{\alpha}_T^{OLS}$, estimated from the regression (1) is an implied function of the parameters $\theta = (\mu_0 \;\; \phi_\mu \;\; \sigma_\mu \;\; v_0 \;\; \phi_v \;\; \sigma_v \;\; \rho_{\mu v} \;\; \beta_0 \;\; \phi_\beta \;\; \sigma_\beta \;\; \bar{\sigma} \;\; \alpha^C \;\; \rho_{\mu\beta})$ of the model and the sample size T. Similar to the setting of our simple model in Section 3.2, the limiting distribution of α_T^{OLS} depends on the beta process and the market return process. However, the asymptotic distribution of the OLS alpha in our richer empirical specification (equations (13) to (15)) cannot easily be derived in closed form. In our estimation method, we directly estimate the conditional betas, $\{\beta_t\}$, and the conditional alpha, α^C, and we construct the implied distribution of $\hat{\alpha}_T^{OLS}$ numerically. We stress that our implied distributions of the OLS estimates $\hat{\alpha}_T^{OLS}$ and $\hat{\beta}_T^{OLS}$ are based on the null of the model in equations (13) to (15). However, we show that the model matches the evidence in data on rolling OLS betas and expect that inference under alternative models which allow similar time variation in betas and the market risk premium to also induce large distortions in the distributions of the OLS statistics relative to their standard distributions.

The reduced-form conditional CAPM in equations (13) to (15) falls within the class of conditional CAPM models developed by Harvey (1989), Shanken (1990), Ferson and Harvey (1991, 1993, 1999), Cochrane (1996), and Jagannathan and Wang (1996). Most of these studies use instrumental variables to model the time variation of betas as a linear function of the instruments. Our betas are also time-varying, but instead of relying on instrumental variables, we parameterize the beta itself as an endogenous latent process. This has the advantage of not relying on exogenous predictor variables to capture time-varying betas and avoids any potential omitted variable bias from mis-specifying the set of predictor variables (see Harvey, 2001; Brandt and Kang, 2004). Instead, we infer the betas directly from portfolio returns. Second, we directly model the variation in the betas across time. Campbell and Vuolteenaho (2004), Adrian and Franzoni (2004), Franzoni (2004), and Lewellen and Nagel (2005) document that the betas of book-to-market portfolios change over time, but they do so by estimating constant beta models over different subsamples of data. Section 3.3 shows that this procedure leads to biased and inconsistent estimates with distorted asymptotic distributions.

A special case of our model is an unconditional CAPM, which arises when $\rho_{\mu\beta} = 0$. The model explicitly captures the time variation in market risk premia that previous empirical studies show is important, and when $\rho_{\mu\beta} \neq 0$, the unconditional CAPM does not hold but a conditional CAPM applies. Rather than using GARCH processes to model conditional betas (see, for example, Bekaert and Wu, 2000), our model uses a log volatility model. In GARCH models, betas are time-varying but the variations in the betas are strictly driven by past innovations in returns and do not have independent random components. Danielsson (1994), among others, finds that the GARCH family of models does not remove all non-linear dependencies in stock return data, while autoregressive stochastic volatility models provide better goodness-of-fit for stock return dynamics.

While the model generates heteroskedasticity, one feature of the data that the model is not designed to capture is time-varying idiosyncratic volatility. In the return equation (13), we assume that idiosyncratic volatility is constant at $\bar{\sigma}$. Campbell et al. (2001) show that the idiosyncratic volatility has noticeably trended higher for individual stocks over the 1990s. Incorporating time-varying idiosyncratic volatility would introduce a difficult identification problem between time-varying betas and idiosyncratic risk. We apply the model to stock portfolios, where idiosyncratic risk is lower than at the firm level. Nevertheless, incorporating time-varying idiosyncratic risk would only exacerbate the large variances in OLS alphas that we document, and hence, by ignoring time-varying idiosyncratic risk, our analysis is conservative.

4.2 Estimation

We estimate the model over the full sample, from 1926-2001, to use all available data to pin down the dynamics of the beta process. After estimating the data on the full sample, we examine the small sample distribution of OLS alphas or conditional alphas. In particular, we are especially interested in small samples of the same length as the post-1963 sample, which is the sample where the OLS alpha appears to be significant using conventional t-statistics. Estimating the model only over the short post-1963 sample to infer the distribution of OLS or conditional alphas over the post-1963 period is inefficient and ignores valuable information about the time variation of betas over the long run.

We use a Markov Chain Monte Carlo (MCMC) and Gibbs sampling estimation method that consistently estimates conditional alphas and betas, incorporates the effect of parameter uncertainty, and measures the effect of small sample bias. Appendix B provides a full description of the estimation method.[6] There are three main reasons we use a Bayesian estimation strategy.

First, conditional on the time series of conditional betas ($\{\beta_t\}$), time-varying market risk premia ($\{\mu_t\}$), and time-varying systematic volatility ($\{v_t\}$), the stock return, $r_{i,t}$, in equation (13) is normally distributed. However, the likelihood function for $r_{i,t}$ and $r_{m,t}$ is difficult to

[6] Other similar models are esimated by Liu and Hanssens (1981), Lamoureux and Zhou (1996), Johannes, Polson and Stroud (2002), Johannes and Polson (2003), Jones (2003), Han (2004), and Jostova and Philipov (2005) with Bayesian methods; Harvey, Solnik and Zhou (2002) with GMM; and Brandt and Kang (2004) with simulated maximum likelihood.

derive in closed form because the latent variables $\{\beta_t\}$, $\{\mu_t\}$, and $\{v_t\}$ must be integrated out. This makes maximum likelihood estimation methods difficult to use. Other classical estimation methods, like Generalized Method of Moments (GMM), also entail a potentially difficult optimization problem. In contrast, the Gibbs sampler is fast because it involves drawing from well-defined conditional distributions.

Second, while the asymptotic distribution of the OLS alphas can be derived in closed-form for our simple IID model in Section 3, the asymptotic distribution of the OLS estimators in the conditional CAPM is difficult to derive. MCMC provides posterior distributions whose means can be interpreted as parameter estimates and the inferred estimates of the time series of betas, market risk premia, and systematic volatility are generated as a by-product of the estimation. The estimation method also allows us to extract the exact finite sample distributions of OLS alphas from the posterior distributions of the parameters. We compute the posterior distributions for the OLS alpha, $\hat{\alpha}_T^{OLS}$, for the limiting case where $T = \infty$, and over a finite sample where T corresponds to the post-1963 sample period. Then, we compare these estimates of asymptotic distribution of $\hat{\alpha}_T^{OLS}$ and small sample distribution of $\hat{\alpha}_T^{OLS}$ directly to the estimates of the conditional alpha, α^C.

Finally, we can impose some prior information on some of the parameters, like the parameters that determine the speed of mean reversion of μ_t and β_t that would otherwise lead to identification problems (see the discussions in Brennan, 1998; Johannes, Polson and Stroud, 2002). In particular, the mean-reversion parameter of the betas (ϕ_β) is difficult to pin down. With non-informative priors, the estimate of ϕ_β is almost zero, and the estimates for the betas become degenerate, making $\beta_t \approx r_{i,t}/r_{m,t}$. This makes the likelihood function infinite. We have strong prior beliefs from economic theory that the betas are persistent, so ϕ_β should be close to one, but they must also be bounded above by one to maintain stationarity. Lamoureux and Zhou (1996), Johannes, Polson and Stroud (2002), and Johannes and Polson (2003) all impose informative priors over mean-reversion parameters in related models. We now discuss our choice of prior for α^C, but detail the full specification of all the other priors in Appendix B.

4.3 Priors on α^C

Inference regarding the conditional alpha, α^C, is of critical importance to measuring the economic and statistical significance of the book-to-market premium. We specify informative priors over α^C that range from no prior belief about the value of α^C to a dogmatic belief that an

α^C must exist. Prior beliefs about α^C are represented by the distribution:

$$\alpha^C \sim N\left(\mu^p_{\alpha^C}, (\sigma^p_{\alpha^C})^2\right), \tag{16}$$

where $\mu^p_{\alpha^C}$ is the prior mean and $(\sigma^p_{\alpha^C})^2$ is the prior variance. If $\mu^p_{\alpha^C} = 0$ and $\sigma^p_{\alpha^C}$ is very small, then the researcher believes dogmatically in the conditional CAPM, while a positive $\mu^p_{\alpha^C}$ and a very small $\sigma^p_{\alpha^C}$ represents a researcher with a strong prior that the book-to-market premium is positive. In contrast, setting $\sigma^p_{\alpha^C} = \infty$ or allowing $\sigma^p_{\alpha^C}$ to be sufficient large, represents an effectively diffuse prior that assumes no a priori knowledge about the strength of the value premium.

An alternative specification of priors for the conditional alpha is given by Pastor and Stambaugh (1999) and Pastor (2000), who specify the prior to be directly proportional to idiosyncratic volatility:

$$\alpha^C|\bar{\sigma} \sim N(\mu^p_\alpha, \eta\bar{\sigma}^2), \tag{17}$$

where η is the prior degree of belief in the conditional CAPM. When $\eta = \infty$, mispricing relative to the conditional CAPM is completely unrestricted, while $\eta = 0$ corresponds to dogmatic faith in the conditional CAPM. In the Pastor-Stambaugh prior in equation (17), the prior on α^C is directly linked to the idiosyncratic volatility, which reduces the probability of drawing high Sharpe ratios. Hence, using the Pastor-Stambaugh prior would make us less likely to reject the null of a conditional CAPM. In contrast, our choice of prior on α^C in equation (16) specifies no mechanical link between α^C and $\bar{\sigma}$. With our prior in equation (16), draws of high posterior Sharpe ratios are more likely than under the Pastor-Stambaugh prior and, thus, we bias our results in favor of finding evidence against the conditional CAPM.

4.4 Priors on Time-Varying Betas

Using Figure 1, we can extract some prior information about the latent conditional beta process. Just as the OLS betas are very persistent, we also expect the conditional betas to have slow moving persistent components. We treat the standard deviation of the rolling OLS betas as a sample statistic and compute a similar statistic from our conditional beta estimates to judge the goodness-of-fit of the model. What rolling OLS betas cannot provide, however, are estimates of the true variability of conditional betas, the correlation of conditional betas with market risk premia, or precise estimates of the conditional beta at a particular point in time. Only direct estimates of the conditional betas can accomplish this.

In Table 2, we examine the autocorrelations and standard deviations of the rolling OLS betas of the highest (lowest) book-to-market decile portfolio, which are the value (growth) stocks, along with the book-to-market strategy. We report the 60th order autocorrelation since it is the lowest order autocorrelation that does not use overlapping information. We then compute the first-order autocorrelation implied by an AR(1) process. The implied monthly autocorrelations are highly persistent, with an estimate of 0.993 for the book-to-market strategy. This is a conservative estimate as estimates of autocorrelations are biased downwards in small samples. We compute a tight standard error of 0.003 for the first-order autocorrelation using the delta-method. Although the OLS betas are inconsistent estimates of the conditional betas, we assume that the true conditional betas have a persistence of the same order of magnitude as the persistence of the non-overlapping autocorrelations implied by the rolling OLS betas.

Table 2 also reports the unconditional standard deviations of the rolling 60-month OLS betas. For the growth stock and the value stock portfolios, they are 0.11 and 0.38, respectively. For the book-to-market strategy, the rolling OLS betas exhibit a large degree of time variation, with a volatility of 0.47. Below, we show that rolling averages implied by our estimates of conditional betas closely match this statistic. Armed with this knowledge about the rolling averages of OLS betas, we now directly infer the true conditional betas, $\{\beta_t\}$, by estimating the conditional CAPM in equations (13) to (15).

In our estimation, we are especially careful not to increase the variance of the conditional betas in a manner that implies implausibly large stock return volatility. A model that implies a large stock return variance can potentially produce very disperse posterior distributions with little information. We impose the constraint on our parameter estimates that the total variance of stock returns is kept constant at the level observed in the data. Thus, by construction, systematic and idiosyncratic volatility sum to the observed total volatility of stock returns in data.

5 Empirical Results

We present our estimates of the conditional CAPM with time-varying betas in Section 5.1. Section 5.2 characterizes the posterior distribution of the conditional alphas of the book-to-market strategy and Section 5.3 reports the unconditional OLS alphas implied by our model. In Section 5.4, we consider the additional effects induced by finite sample bias.

5.1 Parameter Estimates of the Conditional CAPM

Table 3 reports the parameter estimates for the conditional CAPM described by equations (13) to (15). We estimate these models using the value stock portfolio, the growth stock portfolio, and the book-to-market strategy. To compute the estimates in Table 3, we use an effectively diffuse prior with $\mu^p_{\alpha^C} = 0$ and $\sigma^p_{\alpha^C} = 1.00\%$ per month in the prior for α^C in equation (16). Changing $\mu^p_{\alpha^C} = 1.00\%$ or using values of $\sigma^p_{\alpha^C}$ larger than 1.00% per month produces virtually identical results. Table 3 reports the mean and standard deviation of the posterior distribution of each parameter. We first characterize the market return process and then investigate the effects of time-varying conditional betas.

The Market Factor

Table 3 shows that the estimated market risk premium process is persistent, with a monthly autocorrelation of $\phi_\mu = 0.956$. Shocks to the conditional mean are not small, with a volatility of $\sigma_\mu = 0.40\%$ per month. These estimates translate to an unconditional volatility of monthly market risk premium of 1.36% and unconditional volatility of annual market risk premium of approximately 2.0% per annum. The log variance, $\ln v_t$, is also a persistent process with an autocorrelation of $\phi_v = 0.941$ and is slightly conditionally negatively correlated with the conditional mean of the market ($\rho_{\mu v} = -0.093$). This is consistent with many studies that find a leverage effect with negative correlations between market volatility and expected returns (see, for example, Campbell and Hentschel, 1992).

In Figure 2, Panel A, we plot estimates of the implied market risk premia and conditional systematic volatility. The estimates of the market risk premia are fairly smooth, but they have moderately large standard error bounds. Pinning down the market risk premia is notoriously difficult. Johannes and Polson (2003) report that for their estimates of NASDAQ expected returns, even a one standard deviation bound often includes zero. Nevertheless, Panel A of Figure 2 shows that market risk premia increase during the late 1930s and the early 1950s, and decline during the 1960s. More recently, market expected returns increase steeply around the time of the OPEC oil shocks in the 1970s. Over the late 1980s and early 1990s, market expected returns are fairly stable but decrease dramatically during the bull market of the late 1990s. During the year 2000, market expected returns start to increase, coinciding with the onset of the last recession. In most of these episodes, volatility moves in opposite directions to expected returns, as shown in Panel B. Our estimate of market volatility reaches a high of close to 17% per month in the early 1930s, and also increases during World War II, the mid-1970s, the 1987 crash, and around the end of the century from 1998-2001.

Time-Varying Beta Estimates

From the estimates of the latent beta process in Table 3, the implied unconditional beta of the value (growth) stock portfolio is 1.21 (1.01). For the book-to-market strategy, the implied unconditional beta is 0.11. Hence over the whole sample, value stocks do have slightly higher betas than growth stocks, but the difference is small. Table 3 also reports that the latent betas, while highly persistent, are fairly volatile. The conditional volatility of the latent betas for value stocks is fairly large at 0.168 per month, compared to 0.132 per month for growth stocks. In comparison, the conditional volatility of the betas for the book-to-market strategy is a modest 0.065 per month.

For the book-to-market portfolios, the correlation between shocks to the conditional betas and shocks to the market risk premium, $\rho_{\mu\beta}$, is large and positive. For value (growth) stocks, $\rho_{\mu\beta}$ is 0.759 (0.882). For the book-to-market strategy, the posterior mean of $\rho_{\mu\beta}$ is 0.408 with a posterior standard deviation of 0.127. Since the unconditional volatility of the market risk premium is 1.36% per month and total market volatility is fixed at 5.5% per month, this implies an unconditional correlation between betas and market returns of 0.1. The large value of $\rho_{\mu\beta}$ has several implications. First, from equation (5), an unconditional one-factor CAPM cannot be the correct specification for risk since $\rho_{\mu\beta} \neq 0$. Second, equations (9) and (11) show that the OLS estimates of betas and alphas are biased and inconsistent. Third, equation (12) shows that the distribution of $\hat{\alpha}_T^{OLS}$ is distorted from its regular OLS asymptotic distribution. We examine the α^C estimates and the implied OLS alpha distributions in detail below.

In Figures 3 and 4, we plot the posterior mean of the time-varying betas produced by the Gibbs sampler. Figure 3 shows the estimates for the value and growth portfolios. The estimated betas of the value stock portfolio exhibit greater variation than the betas of the growth stock portfolio. The conditional betas of value stocks wander from over 3.0 during the late 1930s to below 0.5 in 2001. In contrast, the conditional betas of growth stocks remain in a fairly close neighborhood around 1.0. Figure 4 graphs the estimates for the book-to-market strategy. The conditional betas of the book-to-market strategy reach a high of above 1.5 near-1940, and decline to close to negative 0.5 at the end of 2001. Figure 4 also shows the one posterior standard deviation bound, which is around 0.5 across the whole sample.

We can compare the conditional variability of the estimated latent betas to the standard deviation of rolling 60-month OLS betas as a specification test of our model estimates. To confirm that the estimates of conditional volatility of the time-varying betas implied by the model match the variability of rolling OLS betas found in the data, we compare the variability of 60-month moving averages of the inferred betas of the book-to-market strategy in Figure 4 to the variability of the 60-month rolling OLS betas reported in Table 2. The match is almost exact. For the book-to-market strategy, the implied rolling 60-month beta average volatility is 0.46, compared with 0.47 in data. Hence, our conditional beta estimates implies rolling betas with a similar degree of variability as the rolling OLS betas from the data.

The large swings of our conditional betas are also consistent with the widely differing point estimates of the OLS alphas across subsamples in Table 1. Taking the posterior mean of the time series of the conditional betas, $\{\hat{\beta}_t\}$, we compute the difference $r_{it} - \hat{\beta}_t r_{mt}$ across subsamples. In the pre-1963 sample, the mean conditional beta is 0.47, the book-to-market strategy returns 0.43% per month, while the excess market return yields 0.85% per month. Thus, in the pre-1963 sample, the mean value of $r_{it} - \hat{\beta}_t r_{mt}$ is 0.03% per month, consistent with an OLS alpha of close to zero. In contrast, over the post-1963 sample, the average conditional beta is -0.11, the book-to-market strategy yields 0.53% per month, and the average market excess return is 0.47% per month. Thus, the mean value of $r_{it} - \hat{\beta}_t r_{mt}$ over the post-1963 sample is 0.58% per month, which is very close to the empirically observed OLS alpha of 0.60% per month reported in Table 1. This suggests that although our model has a constant $\rho_{\mu\beta}$ and constant α^C, the model is capable of generating large differences in OLS alphas in specific sample periods. In particular, the time series of the posterior mean of the conditional betas is consistent with the low (high) OLS alpha in the pre-1963 (post-1963) sample period for the book-to-market strategy.

5.2 Conditional Alphas of the Book-to-Market Strategy

Inference regarding the conditional alpha, α^C, is crucial for judging the fit of the conditional CAPM to explain the value premium. In Table 4, we report the posterior distribution of α^C of the book-to-market strategy. To incorporate various prior views that investors may hold on the strength of the book-to-market effect, we specify several prior distributions, rather than using just one diffuse prior. The priors for α^C in equation (16) range from an effectively uninformative prior with $\mu^p_{\alpha^C} = 0$ and $\sigma^p_{\alpha^C} = 1.00\%$ per month to a highly informative prior with $\mu^p_{\alpha^C} = 0.60\%$ per month and $\sigma^p_{\alpha^C} = 0.10\%$ per month. Since a mean of 0.60% per month corresponds to the $\hat{\alpha}^{OLS}_T$ estimate of the book-to-market strategy over the post-1963 sample (see Table 1), priors with this mean and a low $\sigma^p_{\alpha^C}$ represent a dogmatic belief in the book-to-market effect. For each prior, we report the percentile breakpoints, the mean and the standard deviation of the posterior distribution of α^C.[7]

When we use a prior on α^C with a mean of zero, the value of 0.00% per month lies well above the 10%-tile breakpoint of the posterior, regardless of the standard deviation of the prior. In particular, for the effectively diffuse prior with $\mu^p_{\alpha^C} = 0$ and $\sigma^p_{\alpha^C} = 1.00\%$ per month, the value corresponding to the 10%-tile is -0.01% per month. Hence, an uninformed agent would conclude that the conditional alpha of the book-to-market strategy is insignificantly different from zero. To argue in favor of a strong book-to-market effect, an agent would need to have a strong prior on α^C that has a mean of 0.60% per month and a standard deviation of 0.30% per month or tighter. Under this prior, the posterior distribution of α^C has a lower left-hand tail probability of less than 2.5% for observing a conditional alpha less than zero.

In summary, Table 4 shows that once we account for the time-variation in betas and market risk premia, the evidence against a conditional CAPM is weak using the book-to-market portfolios. In contrast, only the misspecified OLS inference that assumes constant betas based strictly on the post-1963 sample would suggest strong evidence of a book-to-market effect: the OLS alpha has a p-value of 0.006, corresponding to a Newey-West t-statistic of 2.51. However, after accounting for time-varying, conditional betas, only an empiricist with a very strong prior belief in a book-to-market premium would conclude that a conditional CAPM cannot account for the average returns of stocks sorted by book-to-market ratios over the long run.

5.3 OLS Alphas of the Book-to-Market Strategy

We showed in Section 3 that in the simplest IID environment with correlated time-varying betas and market risk premia, the OLS estimate of alpha, $\hat{\alpha}^{OLS}_T$, is inconsistent and the asymptotic distribution of $\sqrt{T}(\hat{\alpha}^{OLS}_T - \mathrm{E}[\hat{\alpha}^{OLS}_T])$ is significantly distorted from its standard OLS distribution with constant betas. We now show that the distortion between α^C and $\hat{\alpha}^{OLS}_T$ are further magnified when we allow for persistent, time-varying betas.

We compute the implied distribution of the OLS alpha, α^{OLS}_T (denoted without a hat to signify it is a random variable), for a sample size of T from the Gibbs sampler. We first compute the limiting distribution of the OLS alpha as $T \to \infty$, which we denote as α^{OLS}. For each observation in the posterior distribution of the model parameters θ, we compute the value of α^{OLS}_T that would result if the data are generated according to equations (13) to (15). We compute

[7] The posterior distributions of α^C are largely unaffected by the estimation of the market process, even if we parameterize the market return to be IID and normally distributed.

this by simulating a time series of $r_{i,t}$ and $r_{m,t}$ of length 100,000 (to proxy for $T = \infty$) at each parameter draw and run an OLS regression on the simulated time series. Since we use a long time series, the constant term from this regression is exactly what the true α_T^{OLS} would be at this particular set of model parameters. We repeat this exercise for every set of parameters in the posterior distribution of the θ, and thus produce the correct posterior distribution of α^{OLS} corresponding to the parameter estimates of the conditional CAPM. The statistic α^{OLS} is a well-defined transformation of the parameters θ, except it is not available in closed form. We report the posterior distribution of α^{OLS} corresponding to the different priors on α^C in Table 5.

Comparing the posterior distributions of α^C in Table 4 and the posterior distributions of α^{OLS} in Table 5 confirms that estimates of alphas obtained by OLS are inconsistent. For an effectively diffuse prior on α^C with a mean of zero and a standard deviation of 1.00% per month, the posterior mean of α^{OLS} is 0.45% per month. In comparison, the posterior mean of α^C is 0.20% per month. The upward bias of α^{OLS} relative to α^C occurs because the estimated correlation between shocks to the market risk premia and shocks to the conditional beta of the book-to-market strategy is positive at 0.41 (see Table 3). For all the priors on α^C, this upward bias is of the order of 20% to 23% per month. Interestingly, in the limit as $T \to \infty$, the posterior standard deviation of α^{OLS} is only slightly larger than the posterior standard deviation of α^C, which are 0.20% and 0.19% per month, respectively.

If we were to base our statistical inference of the book-to-market premium only on α^{OLS}, rather than the correct α^C of the conditional CAPM, Table 5 shows that an investor would conclude that the α^{OLS} is greater than zero regardless of the choice of the prior distribution. In all cases, the value of an alpha of 0.00% per month lies below the 2.5%-tile breakpoint of the posterior distribution of α^{OLS}. Thus, even though we would conclude that the OLS alpha is positive, the true conditional alpha reported in Table 4 is reliably different from 0.00% only for an investor with a very strong prior belief in the book-to-market effect. Hence, the inconsistency of OLS may lead us to conclude that α^{OLS} is positive even if α^C is not.

5.4 Small-Sample Bayesian Analysis of OLS Alphas

A remarkable fact of the simple one-factor OLS regressions of the book-to-market trading strategy in Table 1 is that the $\hat{\alpha}_T^{OLS}$ estimate is 0.60% per month with a Newey-West (1987) t-statistic of 2.51 over the post-1963 sample, but not over the long run. Since the $\hat{\alpha}_T^{OLS}$ estimate is -0.16% per month in the pre-1963 sample, the distribution of $\hat{\alpha}_T^{OLS}$ may be very variable in short samples. In this section, we consider the possible distortions on the posterior distribution of α_T^{OLS}

induced by small samples of length T.[8] Specifically, we show that over small samples, a time-varying beta model with persistent betas but a constant α^C can easily produce one sample in which $\hat{\alpha}_T^{OLS}$ is large, but another sample in which $\hat{\alpha}_T^{OLS}$ is small or negative.

We construct the Bayesian finite sample posterior distribution of α_T^{OLS} in a manner similar to the case of the posterior distribution of α^{OLS}, where $T = \infty$, in Table 5. To compute the finite sample posterior distribution of α_T^{OLS}, we simulate a sample of size T, for each draw θ_i from the posterior distribution of the parameters θ. Since the post-1963 sample corresponds to a time series of length 462 observations, we set $T = 462$. Note that $\alpha_T^{OLS}(\theta)$ is well defined in a small sample as a function of the sample size, T, and the parameters of the model, θ. Hence, for a given sample size, the small sample variable α_T^{OLS} is a valid statistic. To isolate the effect of parameter uncertainty from the effect of small sample bias, we also simulate 10,000 small samples of $T = 462$ holding fixed the parameters of the model at $\bar{\theta}$, the mean of the posterior distribution of θ.

We also consider the effect of various parameters on the model on the small sample distributions of α_T^{OLS} by setting to zero certain parameters of the model in equations (13) to (15). First, we set $\alpha^C = 0$ so that the small sample distribution of α_T^{OLS} is driven only by the correlation between shocks to the beta and shocks to the market risk premium. Second, we set $\rho_{\mu\beta} = 0$, so that an unconditional CAPM holds and OLS alphas are consistent estimates of α^C. Finally, we set both $\alpha^C = 0$ and $\rho_{\mu\beta} = 0$. In each case, we set only the particular parameter in question to zero without re-estimating the model and without changing the other parameters to facilitate comparisons across the specifications.

We report our results in Table 6. As expected, the difference between the α_T^{OLS} and α^C posterior distributions in small samples is even greater than the differences for the limiting case when $T = \infty$, which are reported in Tables 4 and 5. For comparison, the first column in Table 6 lists the posterior distribution of α^C from Table 4 corresponding to the prior of α^C with $\mu^p_{\alpha^C} = 0$ and $\sigma^p_{\alpha^C} = 1.00\%$. The columns under the line "OLS Alpha α_T^{OLS}" report the small-sample posterior distribution of α_T^{OLS}. The column labelled "Full" reports the results based on the full specification of the conditional CAPM, while the other columns set various parameters equal to zero. In all cases, we use an effectively diffuse prior on α^C with $\mu^p_{\alpha^C} = 0$ and $\sigma^p_{\alpha^C} = 1.00\%$ and estimate the full model over the full sample.

[8] Most studies on small sample effects, or sample problems, or "Peso problems" usually focus on term structure (Bekaert, Hodrick and Marshall, 2001) or foreign exchange markets (Evans, 1996), or the aggregate stock market (Rietz, 1988). In contrast, we focus here on small sample inference in the cross-section for the book-to-market trading strategy.

Under the line "With Parameter Uncertainty," we report the posterior distribution of the small sample α_T^{OLS} statistic taking into account parameter uncertainty. The mean of the small sample posterior distribution of α_T^{OLS} at 0.43% per month is largely unchanged from the posterior mean of the population α^{OLS} in Table 5 at 0.45%. However, the small sample posterior distribution of α_T^{OLS} now has a much wider standard deviation of 0.50% per month, compared to the population standard deviation of 0.20% per month in Table 5. The wide tails of the small sample α_T^{OLS} distribution are shown in the percentiles, which range from -0.35% at the 5%-tile to 1.27% at the 95%-tile. The post-1963 $\hat{\alpha}_T^{OLS}$ estimate of 0.60% per month corresponds to the 67%-tile. Clearly, a conditional CAPM can easily generate an $\hat{\alpha}_T^{OLS}$ with a value of 0.60% per month and the posterior 95% confidence bounds of the conditional alpha comfortably encompass zero. In other words, a conditional CAPM can produce outcomes in which $\hat{\alpha}_T^{OLS}$ appears large in one small sample but equals zero in another small sample, particularly when these small samples are only of approximately 40 years in length.

To show that the sampling variation of α^C is not causing the bias and the large tails of the small sample α_T^{OLS} distribution, we next consider setting $\alpha^C = 0$ in the second column under the line "With Parameter Uncertainty." Setting $\alpha^C = 0$ produces a lower mean of the small-sample posterior distribution of α_T^{OLS}, but the variation around the mean is largely unchanged at 0.47% per month. The value of 0.00% per month now falls at the 33%-tile and still makes a small sample draw of $\hat{\alpha}_T^{OLS} = 0.60\%$ per month unsurprising.

In the next column, we set $\rho_{\mu\beta} = 0$. Under this assumption, $\hat{\alpha}_T^{OLS}$ in the unconditional CAPM regression (1) provides a consistent estimate of α^C. While an appropriate HAC standard error may be valid asymptotically, HAC standard errors may vastly understate the tails of the small sample distribution of α_T^{OLS} because the time-varying betas are very persistent. This is indeed the case. The mean of the small-sample posterior distribution of α_T^{OLS} is identical to α^C at 0.23% per month, but the posterior distribution of α_T^{OLS} has a much wider standard deviation than the posterior distribution of α^C (at 0.49% and 0.19% per month, respectively). Finally, if we set both $\alpha^C = 0$ and $\rho_{\mu\beta} = 0$, the small-sample posterior distribution of α_T^{OLS} is centered around zero, but still has a very wide posterior standard deviation of 0.46% per month.

Under the line "Without Parameter Uncertainty" in Table 6, we compute the small sample α_T^{OLS} posterior distribution at the posterior mean of θ, rather than using the entire distribution of θ. Not surprisingly, disregarding parameter uncertainty produces smaller variation of the small sample α_T^{OLS} statistics, but the posterior standard deviations are only slightly smaller than the standard deviations taking into account parameter uncertainty. Thus, disregarding parameter uncertainty does not quantitatively change our results. In all cases, a point estimate of $\hat{\alpha}_T^{OLS} =$ 0.60% per month does not lie anywhere near the upper 10% tail of the small sample posterior distribution of α_T^{OLS}. In summary, a time-varying one-factor model can easily produce what appears to be an anomalous result using OLS alphas with standard asymptotic statistics in small samples, but with correct statistical inference that takes into account time-varying betas, the OLS alphas are statistically insignificant.

Comparing Small Sample and Standard Asymptotic OLS Alpha Distributions

In Figure 5, we compare the posterior small-sample α_T^{OLS} distributions of the book-to-market strategy (in the solid lines) taking into account parameter uncertainty to the asymptotic distribution under the null that $\alpha^{OLS} = 0$ in the regression (1), using a Newey-West (1987) standard error estimate (in the dashed lines). In Panel A, we plot the small-sample α^{OLS} distribution for the full parameter specification corresponding to the column "α_T^{OLS} Full" in Table 6 under the line "With Parameter Uncertainty." Under this specification, we do not impose any parameter constraints. Using the classical asymptotic distribution, we would reject the null that $\alpha^{OLS} = 0$, since the area lying to the right of $\hat{\alpha}_T^{OLS} =$ 0.60% per month is 0.006. In contrast, the exact, small-sample α_T^{OLS} distribution is biased and has much wider tails than the robust asymptotic distribution that assumes constant betas. Under the posterior distribution of α_T^{OLS}, the point estimate of $\hat{\alpha}_T^{OLS} =$ 0.60% per month is no longer reliably different from zero, since 31% of the posterior small-sample α_T^{OLS} distribution lies to the right of the 0.60% line.

In Panel B, we plot the small-sample posterior distribution of α_T^{OLS} imposing the constraint that the conditional alpha, α^C, is zero. This panel clearly illustrates the difference between our results and the conclusions of Lewellen and Nagel (2005). Lewellen and Nagel note that allowing for a correlation between conditionals beta and conditional market risk premia shifts the mean of the distribution of α_T^{OLS} to the right about 10 basis points per month. They argue that the magnitude of the mean shift cannot be large enough to shift the small-sample posterior distribution of α_T^{OLS} all the way to 0.60% per month. We find a slightly larger shift in the posterior mean of 20 basis points per month. However, whereas Lewellen and Nagel (2005) use uncorrected standard, asymptotic OLS theory that assumes constant betas to make their inferences, we find a large posterior standard deviation in the distribution of α_T^{OLS} induced by a conditional CAPM with time-varying betas. Our true, small-sample posterior distribution of α_T^{OLS} has thick tails, so that 15% of the distribution lies to the right of the point estimate of $\hat{\alpha}_T^{OLS} =$ 0.60% per month.

In summary, although the OLS point estimate of $\hat{\alpha}_T^{OLS}$ appears to be large at 0.60% per month over the post-1963 period, the small sample distribution of α_T^{OLS} shows that it is not unusual to observe OLS alphas of this magnitude in small samples of 462 observations. From this point of view, it is not surprising that Cooper, Gutierrez and Marcum (2005) find that book-to-market strategies have difficulty beating the market return in out-of-sample investment strategies in the post-1963 period, despite the conventional OLS evidence of a strong in-sample book-to-market effect over this period. Using statistical inference to take into account time-varying conditional betas, we find little evidence of any book-to-market effect either over the long run, or over the post-1963 sample. Hence, the book-to-market effect may be similar to the size effect, which may be due to incorrect statistical inference (see Chan and Chen, 1988; Knez and Ready, 1997).[9]

6 Conclusion

The book-to-market effect appears to be a strong CAPM anomaly that many researchers consider to be a significant risk factor (see for example, Fama and French, 1993). Over the post-1963 sample, the book-to-market trading strategy generates an OLS alpha of 0.60% per month. Using a Newey-West (1987) estimate of the asymptotic standard error, the post-1963 book-to-market premium appears to be highly statistically significant with a p-value of less than 1%. In contrast, over the pre-1963 sample, the book-to-market strategy generates an OLS alpha of negative 0.16% per month and is not statistically significant. The difference across the two samples can be attributed to time-varying betas in which betas change slowly over time.

Inference of conditional alphas from a conditional CAPM model using unconditional OLS regressions is highly misleading when factor loadings are vary over time. In particular, there is strong evidence that the conditional betas for book-to-market portfolios are time-varying. Conditional betas for the book-to-market strategy, which goes long the highest decile and short the lowest decile of stocks sorted by their book-to-market ratios, range from over 3.0 during the late 1930s to close to negative 0.5 at the end of 2001. When conditional betas are correlated with market risk premia, OLS estimates of alphas are biased, inconsistent, and their asymptotic distributions are severely distorted from standard OLS theory, which assumes constant betas. There is no way to correct the degree of inconsistency or the degree of the distortion without a

[9] Since the size effect was discovered by Banz (1981), the size effect has been negligble. From 1981 to 2001, Fama and French's (1993) SMB size factor has almost a zero premium (-2 basis points per month).

direct knowledge of the time-variation of the conditional beta process.

We propose and directly estimate a conditional CAPM with latent time-varying conditional betas, market risk premia, and stochastic systematic volatility. We find that only an investor with a strong, dogmatic belief in the book-to-market effect would conclude that the conditional alpha of the book-to-market strategy is positive both over the long run, from 1927-2001, and over the post-1963 subsample. Using an effectively uninformative prior, there is little evidence to conclude that the conditional alpha of the book-to-market strategy is different from zero. Since the estimates of betas of book-to-market portfolios are highly correlated with time-varying market risk premia, the distribution of OLS alphas are very dissimilar to the distribution of conditional alphas. In particular, the exact OLS alpha distributions are rather disperse in small samples. Thus, observing a point estimate of an OLS alpha of 0.60% per month over the post-1963 subsample is not at all surprising, even when the true conditional alpha is zero. Indeed, given the time-variation in betas, it is not surprising to observe a high OLS alpha in one small sample, such as the post-1963 sample, but a zero OLS alpha in another small sample, such as the pre-1963 subsample.

Furthermore, our work shows that in testing for CAPM anomalies, researchers should be very careful using asymptotic normal distributions to conduct statistical inference if the betas of their test portfolios vary over time. In environments with time-varying factor loadings, asymptotic OLS distributions cannot be used for statistical inference because OLS is biased and inconsistent. Furthermore, the distortions from the standard limiting OLS distributions that do not take into account time-varying betas cannot be corrected without directly estimating the conditional betas. Our results emphasize the importance of taking into account time-varying factor loadings before declaring a cross-sectional return pattern anomalous relative to a conditional CAPM.

Appendix

A OLS Estimators Under Stochastic Coefficients

This appendix derives the asymptotic distribution of OLS estimators when the true model has stochastic coefficients and stochastic regressors. Suppose that the true model is:

$$y_t = \alpha + \beta_t x_t + \varepsilon_t, \tag{A-1}$$

where $\varepsilon_t \overset{IID}{\sim} N(0, \bar{\sigma}^2)$ is an independent shock. Moreover suppose that both the coefficient, β_t, and the regressor, x_t, are stochastic with $\beta_t \overset{IID}{\sim} N(\bar{\beta}, \sigma_\beta^2)$ and $x_t \overset{IID}{\sim} N(\mu_x, \sigma_x^2)$, and β_t and x_t have correlation $\rho_{x\beta}$. Suppose that the sample has T observations. If we denote $\beta_t^* = \beta_t - \bar{\beta}$, the model in equation (A-1) can be rewritten in matrix form as:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{Z} + \varepsilon, \tag{A-2}$$

where $\mathbf{Y} = \begin{bmatrix} y_1 \\ \vdots \\ y_T \end{bmatrix}$, $\mathbf{X} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_T \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} \alpha \\ \bar{\beta} \end{bmatrix}$, $\mathbf{Z} = \begin{bmatrix} \beta_1^* x_1 \\ \vdots \\ \beta_T^* x_T \end{bmatrix}$, and $\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_T \end{bmatrix}$.

Suppose that an econometrician obtains misspecified OLS estimates, $\hat{\mathbf{B}} = [\hat{\alpha}, \hat{\beta}]'$ over T observations. Specifically, we estimate:

$$\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}. \tag{A-3}$$

We can write the OLS estimator, $\hat{\mathbf{B}}$, as

$$\hat{\mathbf{B}} = \mathbf{B} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon^*, \tag{A-4}$$

where $\varepsilon^* = \mathbf{Z} + \varepsilon$ is the error term relative to the OLS estimation. Equation (A-4) suggests that the OLS estimator, $\hat{\mathbf{B}}$, is subject to at least three distinct problems:

1. unless $\rho_{x\beta} = 0$, $\hat{\mathbf{B}}$ is subject to an omitted variable bias in $\mathbf{Z}$,
2. ε^* is heteroskedastic in $\mathbf{X}$,
3. and furthermore, unless $\mathbf{Z}$ is IID, OLS residuals are serially correlated.

Therefore, even if $\rho_{x\beta} = 0$, OLS standard errors understate the true variance because of heteroskedasticity and have additional distortions if $\mathbf{Z}$ is not IID (say, for instance, if $\mathbf{Z}$ is positively autocorrelated). When $\rho_{x\beta} = 0$, one can potentially use a HAC estimator of the residual variance. However, this correction is only valid asymptotically. Much more serious problems that cannot be corrected by HAC estimators arise when $\rho_{x\beta} \neq 0$.

A.1 The Inconsistency of $\hat{\mathrm{B}}$

The expectation of the OLS estimator is:

$$\mathrm{E}[\hat{\mathbf{B}}] = \mathbf{B} + \mathrm{E}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}] + \mathrm{E}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon]. \tag{A-5}$$

Since β_t^* and x_t are jointly normally distributed with correlation $\rho_{x\beta}$,

$$\mathrm{E}[\beta_t^* | x_t] = \frac{\rho_{x\beta}\sigma_\beta}{\sigma_x}(x_t - \mu_x).$$

By taking the expectation of equation (A-5) conditional on $\mathbf{X}$, we can write:

$$\begin{aligned}
\mathrm{E}[\hat{\mathbf{B}}|\mathbf{X}] &= \mathbf{B} + \mathrm{E}\left[\left(\frac{1}{T}\mathbf{X}'\mathbf{X}\right)^{-1}\frac{1}{T}\mathbf{X}'\mathbf{Z}\middle|\mathbf{X}\right] + 0, \\
&= \mathbf{B} + \frac{\rho_{x\beta}\sigma_\beta}{\sigma_x}\mathrm{E}\left[\left(\frac{1}{T}\mathbf{X}'\mathbf{X}\right)^{-1}\frac{1}{T}\begin{bmatrix} \sum(x_t - \mu_x)x_t \\ \sum(x_t - \mu_x)x_t^2 \end{bmatrix}\middle|\mathbf{X}\right].
\end{aligned}$$

Since, $\text{plim}(\frac{1}{T}\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{\sigma_x^2}\begin{bmatrix} (\mu_x^2+\sigma_x^2) & -\mu_x \\ -\mu_x & 1 \end{bmatrix}$, $\text{E}[x_t^2] = (\mu_x^2+\sigma_x^2)$, and $\text{E}[x_t^3] = (\mu_x^3 + 3\mu_x\sigma_x^2)$, the unconditional expectation of $\hat{\mathbf{B}}$ is given by:

$$\begin{aligned} \text{E}[\hat{\mathbf{B}}] &= \mathbf{B} + \frac{\rho_{x\beta}\sigma_\beta}{\sigma_x}\frac{1}{\sigma_x^2}\begin{bmatrix} (\mu_x^2+\sigma_x^2) & -\mu_x \\ -\mu_x & 1 \end{bmatrix}\begin{bmatrix} \sigma_x^2 \\ 2\mu_x\sigma_x^2 \end{bmatrix}, \\ &= \mathbf{B} + \frac{\rho_{x\beta}\sigma_\beta}{\sigma_x}\begin{bmatrix} \sigma_x^2-\mu_x^2 \\ \mu_x \end{bmatrix}. \end{aligned} \tag{A-6}$$

Therefore, unless $\rho_{x\beta}=0$, OLS is biased and inconsistent. We denote this inconsistency as:

$$\begin{aligned} \alpha^b &= \hat{\alpha}_T^{OLS} - \text{E}[\hat{\alpha}_T^{OLS}] \overset{\text{plim}}{\rightarrow} \frac{\rho_{x\beta}\sigma_\beta}{\sigma_x}(\sigma_x^2-\mu_x^2) \\ \text{and } \beta^b &= \hat{\beta}_T^{OLS} - \text{E}[\hat{\beta}_T^{OLS}] \overset{\text{plim}}{\rightarrow} \frac{\rho_{x\beta}\sigma_\beta}{\sigma_x}\mu_x. \end{aligned} \tag{A-7}$$

A.2 Asymptotic Distribution of $\sqrt{T}(\hat{\mathbf{B}} - \text{E}[\hat{\mathbf{B}}])$

We begin by writing:

$$\hat{\mathbf{B}} - \text{E}[\hat{\mathbf{B}}] = -\begin{bmatrix} \alpha^b \\ \beta^b \end{bmatrix} + (\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\mathbf{Z} + (\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\varepsilon, \tag{A-8}$$

where α^b and β^b are the asymptotic bias terms in equation (A-7). By the independence of ε, we have:

$$\text{var}(\hat{\mathbf{B}} - \text{E}[\hat{\mathbf{B}}]) = \text{var}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}] + \bar{\sigma}^2\text{E}[(\mathbf{X}'\mathbf{X})^{-1}]. \tag{A-9}$$

The second term of equation (A-9), $\bar{\sigma}^2\text{E}[(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}]$, is the regular standard error obtained by OLS with stochastic regressors. Notice that as $T\to\infty$:

$$\bar{\sigma}^2\text{E}[(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}] \overset{\text{plim}}{\rightarrow} \begin{bmatrix} (\sigma_x^2+\mu_x^2)\frac{\bar{\sigma}^2}{\sigma_x^2} & \mu_x\frac{\bar{\sigma}^2}{\sigma_x^2} \\ \mu_x\frac{\bar{\sigma}^2}{\sigma_x^2} & \frac{\bar{\sigma}^2}{\sigma_x^2} \end{bmatrix}. \tag{A-10}$$

The additional variance term in equation (A-9) is caused by the stochastic coefficient β_t. To analyze this term, we define $\bar{x} = \frac{1}{T}\sum x_t$ and $S_{xx} = (\frac{1}{T}\sum x_t^2) - \bar{x}^2$. This allows us to write:

$$(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\mathbf{Z} = \frac{1}{S_{xx}}\begin{bmatrix} \frac{1}{T}S_{xx}+\bar{x}^2 & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix}\begin{bmatrix} \frac{1}{T}\sum\beta_t^* x_t \\ \frac{1}{T}\sum\beta_t^* x_t^2 \end{bmatrix}. \tag{A-11}$$

If we also define $x_t^* = x_t - \bar{x}$ and $\bar{\beta}^* = \frac{1}{T}\sum\beta_t^*$, we can simplify the expression as follows:

$$\begin{aligned} (\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\mathbf{Z} &= \frac{1}{S_{xx}}\begin{bmatrix} S_{xx}+\bar{x}^2 & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix}\begin{bmatrix} \frac{1}{T}\sum\beta_t^* x_t^* + \frac{1}{T}\sum\beta_t^*\bar{x} \\ \frac{1}{T}\sum\beta_t^* {x_t^*}^2 + \frac{2\bar{x}}{T}\sum\beta_t^* x_t^* + \frac{1}{T}\sum\beta_t^*\bar{x}^2 \end{bmatrix}, \\ &= \frac{1}{S_{xx}}\begin{bmatrix} \frac{-\bar{x}}{T}\sum\beta_t^* {x_t^*}^2 + \frac{S_{xx}-\bar{x}^2}{T}\sum\beta_t^* x_t^* + S_{xx}\frac{1}{T}\sum\beta_t^*\bar{x} \\ \frac{1}{T}\sum\beta_t^* {x_t^*}^2 + \frac{\bar{x}}{T}\sum\beta_t^* x_t^* \end{bmatrix}. \end{aligned} \tag{A-12}$$

By making this expression mean zero and using $S_{xx} = \frac{1}{T}\sum {x_t^*}^2$, we have:

$$(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\mathbf{Z} - \begin{bmatrix} \alpha^b \\ \beta^b \end{bmatrix} = \frac{1}{S_{xx}}\begin{bmatrix} \frac{-\bar{x}}{T}\sum\beta_t^* {x_t^*}^2 + \frac{S_{xx}-\bar{x}^2}{T}\sum\beta_t^* x_t^* - \frac{\alpha^b}{T}\sum {x_t^*}^2 + \frac{S_{xx}\bar{x}}{T}\sum\beta^* \\ \frac{1}{T}\sum\beta_t^* {x_t^*}^2 + \frac{\bar{x}}{T}\sum\beta_t^* x_t^* - \frac{\beta^b}{T}\sum {x_t^*}^2 \end{bmatrix},$$

or equivalently we can write:

$$\sqrt{T}(\hat{\mathbf{B}} - \text{E}[\hat{\mathbf{B}}]) = \frac{1}{\sqrt{T}}\frac{1}{S_{xx}}\begin{bmatrix} -\bar{x}\sum\beta_t^* {x_t^*}^2 + (S_{xx}-\bar{x}^2)\sum\beta_t^* x_t^* - \alpha^b\sum {x_t^*}^2 + S_{xx}\bar{x}\sum\beta^* \\ \sum\beta_t^* {x_t^*}^2 + \bar{x}\sum\beta_t^* x_t^* - \beta^b\sum {x_t^*}^2 \end{bmatrix}. \tag{A-13}$$

Since the expression on the RHS of equation (A-13) has a mean of zero, we can apply a standard Central Limit Theorem to derive the asymptotic distribution of $\sqrt{T}(\hat{\mathbf{B}} - \mathrm{E}[\hat{\mathbf{B}}])$. We compute the asymptotic variance of the RHS of equation (A-13), using the following lemma:

Lemma:

Suppose x and y are two mean-zero normally distributed random variables with variances σ_x^2 and σ_y^2 and correlation ρ. Consider the variables Z^1 and Z^2 defined as:

$$\begin{aligned} Z^1 &= c_1x^2y + c_2xy + c_3x^2 + c_4y \\ \text{and} \quad Z^2 &= d_1x^2y + d_2xy + d_3x^2 + d_4y. \end{aligned} \tag{A-14}$$

Moreover, suppose that $bc_2 + c_3 = 0$ and $bd_2 + d_3 = 0$, where $b = \frac{\rho\sigma_y}{\sigma_x}$. Then, the covariance between Z^1 and Z^2 is given by

$$\text{cov}\ [Z^1, Z^2] = c_1d_1(3 + 12\rho^2)\sigma_y^2\sigma_x^4 + c_2d_2(1-\rho^2)\sigma_y^2\sigma_x^2 + (c_1d_4 + c_4d_1)(1 + 2\rho^2)\sigma_y^2\sigma_x^2 + c_4d_4\sigma_y^2. \tag{A-15}$$

Proof: To compute means and variances of Z^1 and Z^2, we first take conditional expectations given x. The conditional distribution of y given x is normal, with a mean of bx and a variance of $(1-\rho^2)\sigma_y^2$, where $b = \rho\sigma_y/\sigma_x$. Hence, we can derive:

$$\begin{aligned} \mathrm{E}[Z^1|x] &= bc_1x^3 + (bc_2 + c_3)x^2 + bc_4x, \\ \mathrm{E}[Z^1] &= (bc_2 + c_3)\sigma_x^2 = 0. \end{aligned} \tag{A-16}$$

The expression for $\mathrm{E}[Z^2]$ is similar.

To derive the covariance, we first expand:

$$\begin{aligned} \mathrm{E}[Z^1Z^2] = \mathrm{E}[&c_1d_1x^4y^2 + (c_1d_2 + c_2d_1)x^3y^2 + (c_1d_3 + c_3d_1)x^4y + (c_1d_4 + c_4d_1)x^2y^2c_2d_2x^2y^2 \\ &+ (c_2d_3 + c_3d_2)x^3y + (c_2d_4 + c_4d_2)xy^2 + c_3d_3x^4 + (c_3d_4 + c_4d_3)x^2y + c_4d_4y^2]. \end{aligned}$$

Since $\mathrm{E}[y|x] = bx$ and $\mathrm{var}[y|x] = (1-\rho^2)\sigma_y^2$, we have $\mathrm{E}[y^2|x] = (1-\rho^2)\sigma_y^2 + b^2x^2$. We also note that $\mathrm{E}[x^6] = 15\sigma_x^6$ and $\mathrm{E}[x^4] = 3\sigma_x^4$. By first conditioning on x, and noting that the odd moments of a normal distribution are equal to zero, we can compute the expectations of each term:

$$\begin{aligned} \mathrm{E}[x^4y^2] &= \mathrm{E}[((1-\rho^2)\sigma_y^2 + b^2x^2)x^4] = (3 + 12\rho^2)\sigma_y^2\sigma_x^2, \\ \mathrm{E}[x^3y^2] &= 0, \\ \mathrm{E}[x^4y] &= 0, \\ \mathrm{E}[x^2y^2] &= \mathrm{E}[((1-\rho^2)\sigma_y^2 + b^2x^2)x^2] = (1 + 2\rho^2)\sigma_y^2\sigma_x^2, \\ \mathrm{E}[x^3y] &= \mathrm{E}[bx^4] = 3\rho\sigma_y\sigma_x. \end{aligned}$$

Using these expressions, we can derive:

$$\begin{aligned} \mathrm{cov}(Z^1, Z^2) &= \mathrm{E}[Z^1Z^2] - \mathrm{E}[Z^1]\mathrm{E}[Z^2], \\ &= c_1d_1(3 + 12\rho^2)\sigma_y^2\sigma_x^4 + (c_2d_2 + c_1d_4 + c_4d_1)(1 + 2\rho^2)\sigma_y^2\sigma_x^2 \\ &\quad + (c_2d_3 + c_3d_2)(3\rho\sigma_y\sigma_x^3) + 3c_3d_3\sigma_x^4 + c_4d_4\sigma_y^2 \\ &\quad - \left(\frac{\rho\sigma_y}{\sigma_x}c_2 + c_3\right)\left(\frac{\rho\sigma_y}{\sigma_x}d_2 + d_3\right)\sigma_x^4. \end{aligned} \tag{A-17}$$

By imposing $bc_2 + c_3 = 0$ and $bd_2 + d_3 = 0$, we obtain equation (A-15), which completes the proof. ■

We can use the lemma, together with Slutsky's Theorem, to derive the asymptotic variance of $\sqrt{T}(\hat{\mathbf{B}} - \mathrm{E}[\hat{\mathbf{B}}])$. We use Slutsky to take the plims of $\bar{x} \overset{\text{plim}}{\to} \mu_x$ and $S_{xx} \overset{\text{plim}}{\to} \sigma_x^2$. Then, we compute the asymptotic variance of

$\sqrt{T}(\hat{\beta} - \mathrm{E}[\hat{\beta}])$ by using the lemma and setting $c_1 = d_1 = 1$, $c_2 = d_2 = \mu_x$, $c_3 = d_3 = \frac{-\rho_{x\beta}\sigma_\beta}{\sigma_x}\mu_x$, and $c_4 = d_4 = 0$ to obtain:

$$\text{A.Var}\left[\sqrt{T}(\hat{\beta}_T^{OLS} - \mathrm{E}[\hat{\beta}_T^{OLS}])\right] = (3 + 12{\rho_{x\beta}}^2)\sigma_\beta^2 + (1 - {\rho_{x\beta}}^2)\mu_x^2\frac{\sigma_\beta^2}{\sigma_x^2} + \frac{\bar{\sigma}^2}{\sigma_x^2}. \tag{A-18}$$

By setting $c_1 = d_1 = -\mu_x$, $c_2 = d_2 = \sigma_x^2 - \mu_x^2$, $c_3 = d_3 = \frac{-\rho_{x\beta}\sigma_\beta}{\sigma_x}(\sigma_x^2 - \mu_x^2)$, and $c_4 = d_4 = \mu_x\sigma_x^2$, we obtain:

$$\text{A.Var}\left[\sqrt{T}(\hat{\alpha}_T^{OLS} - \mathrm{E}[\hat{\alpha}_T^{OLS}])\right] = (10{\rho_{x\beta}}^2)\mu_x^2\sigma_\beta^2 + (1 - {\rho_{x\beta}}^2)(\sigma_x^4 + \mu_x^4)\frac{\sigma_\beta^2}{\sigma_x^2} + (\sigma_x^2 + \mu_x^2)\frac{\bar{\sigma}^2}{\sigma_x^2}. \tag{A-19}$$

Finally, by setting $c_1 = -\mu_x$, $c_2 = \sigma_x^2 - \mu_x^2$, $c_3 = \frac{-\rho_{x\beta}\sigma_\beta}{\sigma_x}(\sigma_x^2 - \mu_x^2)$, $c_4 = \mu_x\sigma_x^2$, $d_1 = 1$, $d_2 = \mu_x$, $d_3 = \frac{-\rho_{x\beta}\sigma_\beta}{\sigma_x}\mu_x$, and $d_4 = 0$, we can compute the asymptotic covariance between $\sqrt{T}(\hat{\beta} - \mathrm{E}[\hat{\beta}])$ and $\sqrt{T}(\hat{\alpha} - \mathrm{E}[\hat{\alpha}])$, which is given by:

$$-\left[(1 + 11{\rho_{x\beta}}^2)\mu_x\sigma_\beta^2 + (1 - {\rho_{x\beta}}^2)\mu_x^3\frac{\sigma_\beta^2}{\sigma_x^2} + \mu_x\frac{\bar{\sigma}^2}{\sigma_x^2}\right]. \tag{A-20}$$

B Estimating the Model

We estimate the model described by equations (13) to (15) by Gibbs sampling and MCMC. In particular, we estimate the process for the betas and the market risk premium by using the forward-backward algorithm of Carter and Kohn (1994). We estimate the latent stochastic volatility process of the market by adapting the single-state updating algorithm of Jacquier, Polson and Rossi (1994, 2004) to accommodate correlation with the μ_t equation. In general, the individual parameters of equations (13) to (15) can be updated using standard conjugate draws, except we use informative priors for some of the auto-correlation and correlation parameters.

In each of our estimations, we use a burn-in period of 3000 draws and draw 10,000 observations to represent the posterior distributions of parameters and latent variables. Our results are generated using Ox version 3.32 (see Doornik, 2002). Since this model is highly complex, our estimation is probably not the most efficient, but we are confident in its convergence. The autocorrelation of the posterior draws are low, and most importantly, the estimation method works very well on simulated data.

We repeat the conditional CAPM here for ease of reference:

$$\begin{aligned} r_{i,t} &= \alpha^C + \beta_t r_{m,t} + \bar{\sigma}\varepsilon_{i,t}, \\ r_{m,t} &= \mu_t + \sqrt{v_t}\varepsilon_{m,t}, \\ \beta_t &= \beta_0 + \phi_\beta\beta_{t-1} + \sigma_\beta\varepsilon_{\beta,t}, \\ \mu_t &= \mu_0 + \phi_\mu\mu_{t-1} + \sigma_\mu\varepsilon_{\mu,t}, \\ \text{and} \quad \ln v_t &= v_0 + \phi_v \ln v_{t-1} + \sigma_v\varepsilon_{v,t}, \end{aligned} \tag{B-1}$$

where the correlations between all the shock terms are zero except $\mathrm{E}(\varepsilon_{\mu,t}\varepsilon_{v,t}) = \rho_{\mu v}$ and $\mathrm{E}(\varepsilon_{\mu,t}\varepsilon_{\beta,t}) = \rho_{\mu\beta}$. The full set of parameters we draw is $(\theta, \{\mu_t\}, \{\beta_t\}, \{v_t\})$, where

$$\theta = (\mu_0\ \phi_\mu\ \sigma_\mu\ v_0\ \phi_v\ \sigma_v\ \rho_{\mu v}\ \beta_0\ \phi_\beta\ \sigma_\beta\ \bar{\sigma}\ \alpha^C\ \rho_{\mu\beta})$$

is the set of parameters of the model (B-1), $\{\beta_t\}$ is the vector of conditional betas, and $\{\mu_t\}$ ($\{v_t\}$) is the vector of latent conditional means (variances) of the market. Denote the data by $y_t = (r_{m,t}, r_{i,t})$ and the full set of data as $Y = \{y_t\}$. We can break equation (B-1) into several conditional distributions:

1. $p(Y|\theta, \{\beta_t\}, \{\mu_t\}, \{v_t\})$ is the distribution of the data given the conditional betas, conditional means and conditional volatilities of the market,

2. $p(\{\mu_t\}, \{\beta_t\}|\theta, \{v_t\}, Y)$ is the joint distribution of the conditional betas and conditional means of the market, which is an VAR(1) process with correlation between μ_t and β_t,

3. $p(\{v_t\}|\theta, \{\mu_t\}, \{\beta_t\}, Y)$ is the distribution of the conditional market variances, which is a log-normal AR(1) process, and finally,

4. $p(\theta)$ reflects the prior belief about the parameters of the process in (B-1).

The Gibbs sampler involves iterating over the following sets of parameters and states, conditional on other parameters and states:

P1) Latent Conditional Beta and Market Premium States	$\{\mu_t\}$, $\{\beta_t\}$
P2) Latent Conditional Market Variance States	$\{v_t\}$
P3) Market Premium Regression Parameters	μ_0, ϕ_μ
P4) Conditional Beta Regression Parameters	β_0, ϕ_β
P5) Conditional Variance Regression Parameters	v_0, ϕ_v
P6) Conditional Alpha	α^C
P7) Volatility Parameters	$\sigma_\mu, \sigma_v, \sigma_\beta, \bar{\sigma}$
P8) Correlation Parameters	$\rho_{\mu v}, \rho_{\mu\beta}$

Drawing the Conditional Betas and Market Premium Process (P1)

We draw $\{\mu_t, \beta_t\}$ jointly using the multi-move Carter-Kohn (1994) forward-filtering backward-sampling algorithm. This entails running a Kalman filter forward with the state equation:

$$\begin{pmatrix} \mu_t - \bar{\mu} \\ \beta_t - \bar{\beta} \end{pmatrix} = \begin{pmatrix} \phi_\mu & 0 \\ 0 & \phi_\beta \end{pmatrix} \begin{pmatrix} \mu_{t-1} - \bar{\mu} \\ \beta_{t-1} - \bar{\beta} \end{pmatrix} + u_t, \tag{B-2}$$

where u_t is normally distributed bivariate shocks with the covariance matrix $\begin{pmatrix} \sigma_\mu^2 & \rho_{\mu\beta}\sigma_\mu\sigma_\beta \\ \rho_{\mu\beta}\sigma_\mu\sigma_\beta & \sigma_\beta^2 \end{pmatrix}$, and $\bar{\mu} = \mu_0/(1-\phi_\mu)$ and $\bar{\beta} = \beta_0/(1-\phi_\beta)$ are the unconditional means of μ_t and β_t, respectively. We ensure that we match the sample unconditional mean of μ_t in each iteration. There are two observation equations in the Kalman system. First, as Johannes and Polson (2003) note, the observation equation for the market is a heteroskedastic observation equation:

$$r_{m,t} = \bar{\mu} + (\mu_t - \bar{\mu}) + \sqrt{v_t}\varepsilon_{m,t}, \tag{B-3}$$

where v_t is known. The second observation equation for the stock return is:

$$r_{i,t} = \alpha^C + \bar{\beta} r_{m,t} + r_{m,t}(\beta_t - \bar{\beta}) + \bar{\sigma}\varepsilon_{i,t}, \tag{B-4}$$

which is an observation equation with time-varying coefficients since $r_{m,t}$ is known. The time-varying constant term is $\alpha^C + \bar{\beta} r_{m,t}$ and the time-varying factor loading is $r_{m,t}$. Once the Kalman filter is run forward, we backward sample through time following Carter and Kohn (1994).

Drawing the Conditional Market Volatility Process (P2)

Updating the volatility states requires single state updating (see Jacquier, Polson and Rossi, 1994, 2004). For a single state update, the joint posterior for volatility is:

$$p(v_t|v_{t-1}, v_{t+1}, \theta, \{\mu_t\}, Y) \propto p(y_{t+1}|\mu_t, v_t) p(v_t|v_{t-1}, \mu_{t-1}, \mu_t) p(v_{t+1}|v_t, \mu_t, \mu_{t+1}). \tag{B-5}$$

Note that we have set $\rho_{v\beta} = 0$ (from Brandt and Kang, 2004), so the draw of $\{v_t\}$ is unaffected by $\{\beta_t\}$.

Denote $\epsilon_t^\mu = \mu_t - \mu_0 - \phi_\mu \mu_{t-1}$ as the time t residual of the μ_t process. To find the distribution $p(v_t|v_{t-1}, \mu_t, \mu_{t-1}) \equiv p(v_t|v_{t-1}, \epsilon_t^\mu)$, we use the fact that ϵ_t^μ and $\ln v_t$ are jointly normal. Hence, the distribution of $\ln v_t$ conditional on ϵ_t^μ is normally distributed:

$$\ln v_t \sim N\left(v_0 + \phi_v \ln v_{t-1} + \frac{\rho_{\mu v}\sigma_v}{\sigma_\mu}, \sigma_v^2(1-\rho_{\mu v}^2)\right).$$

This implies that we can write:

$$p(v_t|v_{t-1}, \epsilon_t^\mu) \propto \exp\left(-\frac{(\ln v_t - v_0 - \phi_v \ln v_{t-1} - \frac{\rho_{\mu v}\sigma_v}{\sigma_\mu}\epsilon_t^\mu)^2}{2\sigma_v^2(1-\rho_{\mu v}^2)}\right). \tag{B-6}$$

The other two expressions in equation (B-5) are:

$$p(y_{t+1}|\mu_t, v_t) \propto v_t^{-1/2} \exp\left(-\frac{(y_{t+1}-\mu_t)^2}{2v_t}\right) \tag{B-7}$$

and

$$p(v_{t+1}|v_t, \epsilon_{t+1}^{\mu}) \propto \frac{1}{v_t} \exp\left(-\frac{(\ln v_{t+1} - v_0 - \phi_v \ln v_t - \frac{\rho_{\mu v}\sigma_v}{\sigma_\mu}\epsilon_{t+1}^{\mu})^2}{2\sigma_v^2(1-\rho_{\mu v}^2)}\right). \tag{B-8}$$

Substituting equations (B-6) through (B-8) into the joint posterior (B-5), combining the log-normal terms, and completing the square, allows us to write:

$$p(v_t|v_{t-1}, v_{t+1}, \theta, \{\mu_t\}, Y) \propto v_t^{-3/2} \exp\left(-\frac{(y_{t+1}-\mu_t)^2}{2v_t}\right) \exp\left(-\frac{(\ln v_t - \mu_t^*)^2}{2\sigma^{*2}}\right), \tag{B-9}$$

where

$$\mu_t^* = \frac{v_0(1-\phi_v) + \phi_v(\ln v_{t-1} + \ln v_{t+1}) + \frac{\rho_{\mu v}\sigma_v}{\sigma_\mu}(\epsilon_t^{\mu} - \phi_v \epsilon_{t+1}^{\mu})}{(1+\phi_v^2)}$$

and

$$\sigma^{*2} = \frac{\sigma_v^2(1-\rho_{\mu v}^2)}{1+\phi_v^2}.$$

If $\rho_{\mu v} = 0$, then the posterior distribution in equation (B-9) reduces to Jacquier, Polson and Rossi (1994). Since this distribution is not recognizable, we use a Metropolis draw. As suggested by Cogley and Sargent (2005), we use a log-normal density as a proposal:

$$q(v_t) \propto v_t^{-1} \exp\left(-\frac{(\ln v_t - \mu_t^*)^2}{2\sigma^{*2}}\right). \tag{B-10}$$

The acceptance probability for the $(g+1)$th draw is:

$$\left(\frac{v_t^{g+1}}{v_t^g}\right)^{-1/2} \exp\left[-\frac{1}{2}(y_{t+1}-\mu_t)^2 \left(\frac{1}{v_t^{g+1}} - \frac{1}{v_t^g}\right)\right]. \tag{B-11}$$

To draw v_t at the beginning and the end of the sample, we integrate out the initial and end values of v_t by drawing from the log-normal AR(1) process in (B-1), following Jacquier, Polson and Rossi (2004).

Drawing μ_0 and ϕ_μ (P3)

It is hard to pin down ϕ_μ without imposing prior information. In the predictability literature, excess market returns are generally predicted by very persistent variables, such as dividend yields, short rates and term spreads. In theoretical models, expected excess returns vary over business-cycle frequencies and, therefore, are very persistent. Our procedure for drawing ϕ_μ is to use a Random-Walk Metropolis algorithm with a random walk proposal for ϕ_μ bounded from ϕ_L to ϕ_U. Because the random walk is bounded, this is equivalent to drawing from a uniform over ϕ_L to ϕ_U. We set $\phi_L = 0.900$ and $\phi_U = 0.999$.

The acceptance probability for the $(g+1)$th draw is:

$$\exp\left[-\frac{1}{2\sigma_\mu^2}\left(\epsilon_\mu^{g+1'}\epsilon_\mu^{g+1} - \epsilon_\mu^{g'}\epsilon_\mu^{g}\right)\right], \tag{B-12}$$

where ϵ_μ^g is the vector of residuals of the μ_t innovations $\{(\mu_t - \mu_0^g - \phi_\mu^g \mu_{t-1})\}$ from the gth draw. Once ϕ_μ is drawn, we compute μ_0 to match the unconditional market risk premium in data $\bar{\mu}$, by setting $\mu_0 = \bar{\mu}(1-\phi_\mu)$. This is to ensure that the spread of average returns induced by the time-varying betas of the book-to-market portfolios is not being influenced by an implied average excess market return that is far from the data.

Drawing β_0 and ϕ_β (P4)

Given a normal prior, the posterior distribution for β_0 is also normal (see Zellner, 1971). We set the normal prior for β_0 to have zero mean and a variance of 1000.

To draw ϕ_β, we set up an Independence Metropolis draw to use prior information to help identify ϕ_β. We use a normally distributed prior $p(\phi_\beta)$ and draw ϕ_β from a uniform proposal $q(\phi_\beta)$, bounded between ϕ_L and ϕ_U. Our approach to specifying a prior on ϕ_β is as follows. We use our estimates about the mean reversion rate of the 60-month rolling betas in Table 2 to help formulate our prior on ϕ_i. For each portfolio we estimate, we impose a uniform distribution from four standard errors below the implied values of ϕ_i to 0.9999 as our prior on ϕ_β.

The acceptance probability for the $(g+1)$th draw is given by $\pi({\phi_\beta}^{g+1})/\pi({\phi_\beta}^{g})$, where the posterior $\pi(\phi_\beta) = p(Y|\{\beta_t\},\theta)p(\phi_\beta)$ is conjugate normal. This is because the likelihood $p(Y|\{\beta_t\},\theta)$ is normally distributed from the equation for $r_{i,t}$ and β_t in equation (B-1) since the market, $r_{m,t}$ is known:

$$\begin{aligned} r_{i,t} &= \alpha^C + \beta_t r_{m,t} + \bar{\sigma}\varepsilon_{i,t} \\ \text{and} \quad \beta_t &= \beta_0 + \phi_\beta \beta_{t-1} + \sigma_\beta \varepsilon_{\beta,t}. \end{aligned}$$

Drawing v_0 and ϕ_v (P5)

Conditional on volatility, the parameters v_0 and ϕ_v are just regression parameters. These parameters can be updated by a standard conjugate normal draw (see Zellner, 1971).

Drawing α^C (P6)

Given a normal prior, the posterior distribution for α^C is a straightforward regression draw (conjugate normal). The choice of priors for α^C varies as we change the prior mean and prior variance to reflect effective non-informative priors or priors that represent dogmatic belief. For example, for the parameters reported in Table 3, we use a prior normal with zero mean and standard deviation of 1.00% per month.

Drawing σ_μ, σ_v, σ_β, and $\bar{\sigma}$ (P7)

We update the volatility parameters σ_μ and σ_v using standard Inverse Gamma (IG) conjugate draws, assuming IG(ν_0, ν_1) priors (see Zellner, 1971). In all cases, we choose priors with $\nu_0 = \nu_1 = 0$.

Drawing σ_β and $\bar{\sigma}$ is more complicated because we want to constrain the variance of the stock return implied from θ to match the variance of the stock return in data. This ensures that the estimation does not cause the implied variance of the stock return to be greater than that observed in data. The stock variance is given by:

$$\text{var}(r_{i,t}) = \bar{\beta}^2 \text{var}(r_{m,t}) + \text{var}(\beta_t)(\mu_m^2 + \sigma_m^2) + \bar{\sigma}^2, \tag{B-13}$$

where $\text{var}(\beta_t) = \sigma_\beta^2(1-\rho_{\mu\beta}^2)/(1-\phi_\beta^2)$, $\mu_m = \text{E}(r_{m,t})$, and $\sigma_m^2 = \text{var}(r_{m,t})$. We first draw a candidate $\bar{\sigma}$ from the residuals $r_{i,t} - \alpha^C - \beta_t r_{m,t}$ using an IG conjugate draw. Then, we solve equation (B-13) for σ_β, where the values for $\text{var}(r_{i,t})$, $\text{E}(r_{m,t})$, and $\text{var}(r_{m,t})$ are set at their estimates from data. If there is no solution for σ_β, this indicates that the implied idiosyncratic volatility and the volatility of the conditional betas result in a higher total stock variance than what is observed in the data, so we discard and do not update the candidate draw for $\bar{\sigma}$.

Drawing $\rho_{\mu v}$ and $\rho_{\mu\beta}$ (P8)

To impose prior information on the correlation parameters, we use an accept/reject Metropolis algorithm with a normal prior. The resulting posterior $\pi(\rho)$ is derived by Bernardo and Smith (2002, p363), which involves the sample correlation estimate and hypogeometric functions. We draw from a uniform proposal over $\rho_L = -1$ to $\rho_U = +1$. The acceptance probability for the $(g+1)$th draw is given by:

$$\frac{\pi(\rho^{g+1})}{\pi(\rho^g)} \exp\left(-\frac{1}{2\sigma_0^2}[(\rho^g - \mu_0)^2 - (\rho^{g+1} - \mu_0)^2]\right) \tag{B-14}$$

where μ_0 is the prior mean and σ_0^2 is the prior variance.

For $\rho_{\mu v}$, we have strong prior belief that $\rho_{\mu v}$ is negative through the leverage effect. We set $\mu_0 = -0.6$ with $\sigma_0 = 0.2$. Since $\rho_{\mu\beta}$ is a crucial parameter for inferring the OLS alpha, we choose the prior parameters of $\rho_{\mu\beta}$ to be effectively diffuse, with $\mu_0 = 0$ and $\sigma_0 = 0.5$.

References

[1] Adrian T., Franzoni F., 2005. Learning about beta: Time-varying factor loadings, expected returns, and the conditional CAPM. Working paper. Federal Reserve Bank of New York.

[2] Banz, R.W., 1981. The Relationship between Return and Market Value of Common Stocks. Journal of Financial Economics 9, 3-18.

[3] Bekaert, G., Hodrick, R.J., Marshall, D.A., 2001. Peso problem anomalies for term structure anomalies. Journal of Monetary Economics 48, 241-270.

[4] Bekaert, G., Wu, G., 2000. Asymmetric volatility and risk in equity markets. Review of Financial Studies 13, 1-42.

[5] Bernardo, J.M., Smith, A.F.M., 2002. Bayesian Theory. John Wiley & Sons: Chichester.

[6] Black, F., Jensen, M., Scholes, M., 1972. The capital asset pricing model: Some empirical tests. In Jensen, M. (Ed.), Studies in the theory of capital markets. Praeger: New York, 79-121.

[7] Brandt, M., Kang, Q., 2004. On the relationship between the conditional mean and volatility of stock returns: A latent VAR approach. Journal of Financial Economics 72, 217-257.

[8] Brennan, M.J., 1998. The role of learning in dynamic portfolio decisions. European Finance Review 1, 295-306.

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

[10] Campbell, J.Y., Lettau, M., Malkiel, B.G., Xu, Y., 2001. Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk. Journal of Finance 61, 1-43.

[11] Campbell, J.Y., Vuolteenaho, T., 2004. Bad beta, good beta. American Economic Review 94, 1249-1275.

[12] Carter, C.K., Kohn, R., 1994. On gibbs sampling for state space models. Biometrika 81, 541-553.

[13] Chan, K.C., Chen, N.F., 1988. An unconditional asset-pricing test and the role of firm size as an instrumental variable for risk. Journal of Finance 43, 309-325.

[14] Cochrane, J.H., 1996. A cross-sectional test of an investment-based asset pricing model. Journal of Political Economy 104, 572-621.

[15] Cogley, T., Sargent, T.J., 2005. Drifts and volatilities: Monetary policies and outcomes in the post WWII U.S. Review of Economic Dynamics 8, 262-302.

[16] Cooley, T.F., Prescott, E.C., 1976. Estimation in the presence of stochastic parameter variation. Econometrica 44, 167-183.

[17] Cooper, M., Gutierrez, R.C., Marcum, W., 2005. On the predictability of stock returns in real time. Journal of Business 78, 469-500.

[18] Daniel, K., Titman, S., 1997. Evidence on the characteristics of cross sectional variation in stock returns. Journal of Finance 52, 1-33.

[19] Danielsson, J., 1994. Stochastic volatility in asset prices: Estimation with simulated maximum likelihood. Journal of Econometrics 64, 375-400.

[20] Davis, J.L., 1994. The cross-section of realized stock returns: The pre-COMPUSTAT evidence. Journal of Finance 49, 1579-1593.

[21] Davis, J.L., Fama, E.F., French, K.R., 2000. Characteristics, covariances, and average returns: 1929-1997. Journal of Finance 55, 389-406.

[22] Doornik, J.A., 2002. Object-oriented matrix programming using Ox. 3rd ed. Timberlake Consultants Press: London and Oxford University: www.nuff.ox.ac.uk/Users/Doornik.

[23] Evans, M.D.D., 1996. Peso problems: Their theoretical and empirical implications. In Maddala, G.S., Rao, C.R. (Eds.), Handbook of Statistics: Statistical Methods in Finance. Elsevier: North-Holland.

[24] Fama, E.F., French, K.R., 1992. The cross-section of expected stock returns. Journal of Finance 47, 427-465.

[25] Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3-56.

[26] Fama, E.F., French, K.R., 1997. Industry costs of equity. Journal of Financial Economics 43, 153-193.

[27] Fama, E.F., French, K.R., 2005. The value premium and the CAPM. Journal of Finance (to appear).

[28] Fama, E.F., MacBeth, J.D., 1973. Risk, return, and equilibrium: Empirical tests. Journal of Political Economy 71, 607-636.

[29] Ferson, W.E., Harvey, C.R., 1991. The variation of economic risk premiums. Journal of Political Economy 99, 385-415.

[30] Ferson, W.E., Harvey, C.R., 1993. The risk and predictability of international equity returns. Review of Financial Studies 6, 527-566.

[31] Ferson, W.E., Harvey, C.R., 1999. Conditioning variables and the cross-section of stock returns. Journal of Finance 54, 1325-1360.

[32] Foster, D.P., Nelson, D.B., 1996. Continuous record asymptotics for rolling sample variance estimators. Econometrica 64, 139-174.

[33] Franzoni, F., 2004. Where is beta going? The riskiness of value and small stocks. Working paper. HEC.

[34] Gibbons, M.R., Ross, S.A., Shanken, J., 1989. A test of the efficiency of a given portfolio. Econometrica 57, 1121-1152.

[35] Gomes, J.F., Kogan, L., Zhang, L., 2003. Equilibrium cross-section of returns. Journal of Political Economy 111, 693-732.

[36] Greene, W.H., 2002. Econometric Analysis. 5th ed., Prentice Hall: New York.

[37] Han, Y., 2004. Asset allocation with a high dimensional latent factor stochastic volatility model. Review of Financial Studies (to appear).

[38] Harvey, C.R., 1989. Time-varying conditional covariances in tests of asset pricing models. Journal of Financial Economics 24, 289-317.

[39] Harvey, C.R., 2001. The specification of conditional expectations. Journal of Empirical Finance 8, 573-638.

[40] Harvey, C.R., Solnik, B., Zhou, G., 2002. What determines expected international asset returns. Annals of Economics and Finance 3, 249-298.

[41] Jacquier, E., Polson, N.G., Rossi, P.E., 1994. Bayesian analysis of stochastic volatility models. Journal of Business and Economic Statistics 12, 371-417.

[42] Jacquier, E., Polson, N.G., Rossi, P.E., 2004. Bayesian analysis of fat-tailed stochastic volatility models with correlated errors. Journal of Econometrics 122, 185-212.

[43] Jagannathan, R., Wang, Z., 1996. The conditional CAPM and the cross-section of expected returns. Journal of Finance 51, 3-53.

[44] Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. Journal of Finance 48, 65-91.

[45] Johannes, M., Polson, N., 2003. MCMC methods for financial econometrics. In Aït-Sahalia, Y., Hansen, L.P. (Eds.), Handbook of Financial Econometrics (to appear).

[46] Johannes, M., Polson, N., Stroud, J., 2002. Sequential optimal portfolio performance: Market and volatility timing. Working Paper. Columbia University.

[47] Jones, C.S., 2003. The dynamics of stochastic volatility: Evidence from underlying and options markets. Journal of Econometrics 116, 181-224.

[48] Jostova, G., Philipov, A., 2005. Bayesian analysis of stochastic betas. Journal of Financial and Quantitative Analysis 40, 747-778.

[49] Knez, P.J., Ready, M.J., 1997. On the robustness of size and book-to-market in cross-sectional regressions. Journal of Finance 52, 1355-1382.

[50] Kothari, S.P., Shanken, J., Sloan, R., 1995. Another look at the cross-section of expected stock returns. Journal of Finance 50, 185-224.

[51] Lamoureux, C.G., Zhou, G., 1996. Temporary components of stock returns: What do the data tell us? Review of Financial Studies 9, 1033-1059.

[52] Lewellen, J., Nagel, S., 2005. The conditional CAPM does not explain asset-pricing anomalies. Journal of Financial Economics (to appear).

[53] Loughran, T., 1997. Book-to-market across firm size, exchange, and seasonality: Is there an effect? Journal of Financial and Quantitative Analysis 32, 249-268.

[54] Liu, L.M., Hanssens, D.M., 1981. A Bayesian approach to time-varying cross-sectional regression models. Journal of Econometrics 15, 341-356.

[55] Merton, R.C., 1971. Optimum consumption and portfolio rules in a continuous-time model. Journal of Economic Theory 3, 373-413.

[56] Newey, W.K., West, K.D., 1987. A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703-8.

[57] Pástor, L., 2000. Portfolio selection and asset pricing models. Journal of Finance 55, 179-223.

[58] Pástor, L., Stambaugh, R., 1999. Costs of equity capital and model mispricing. Journal of Finance 54, 67-121.

[59] Petkova, R., Zhang, L., 2005. Is value riskier than growth? Journal of Financial Economics 78, 187-202.

[60] Rietz, T.A., 1988. The equity risk premium: A solution. Journal of Monetary Economics 22, 117-131.

[61] Santos, T., Veronesi, P., 2004. Conditional betas. Working paper. Columbia University.

[62] Schwert, G.W., 1989. Why does stock market volatility change over time? Journal of Finance 44, 1115-1154.

[63] Shanken, J., 1990. Intertemporal asset pricing: An empirical investigation. Journal of Econometrics 45, 99-120.

[64] Zellner, A., 1971. An introduction to Bayesian inference in econometrics. John Wiley & Sons: Chichester.

Table 1: Summary Statistics of Book-to-Market Portfolios

Decile	1	2	3	4	5	6	7	8	9	10	BM	Market
Panel A: Full Sample (1926:07-2001:12)												
Monthly Return (%)	0.59	0.69	0.68	0.63	0.76	0.76	0.84	0.96	1.08	1.07	0.48	0.66
Monthly Volatility (%)	5.89	5.60	5.47	6.17	5.73	6.32	6.82	6.93	8.21	9.39	6.71	5.52
$\hat{\beta}_T^{OLS}$	1.01	0.98	0.95	1.06	0.97	1.07	1.13	1.14	1.31	1.42	0.41	
$\hat{\alpha}_T^{OLS}$	-0.08	0.05	0.06	-0.06	0.11	0.06	0.09	0.21	0.21	0.14	0.21	
$\hat{\alpha}_T^{OLS}$ N-W t-stat	[1.16]	[0.84]	[1.08]	[0.92]	[1.61]	[0.72]	[0.88]	[1.90]	[1.50]	[0.76]	[0.97]	
Panel B: Subsample 1926:07-1963:06												
Monthly Return (%)	0.80	0.88	0.87	0.79	1.02	0.90	0.98	1.17	1.35	1.24	0.43	0.85
Monthly Volatility (%)	6.41	6.34	6.09	7.40	6.84	7.83	8.68	8.86	10.74	12.25	8.33	6.43
$\hat{\beta}_T^{OLS}$	0.96	0.96	0.91	1.11	1.02	1.16	1.27	1.29	1.53	1.66	0.69	
$\hat{\alpha}_T^{OLS}$	-0.02	0.07	0.10	-0.15	0.15	-0.08	-0.11	0.08	0.04	-0.17	-0.16	
$\hat{\alpha}_T^{OLS}$ N-W t-stat	[0.22]	[0.78]	[1.17]	[1.57]	[1.48]	[0.66]	[0.66]	[0.45]	[0.19]	[0.61]	[0.47]	
Panel C: Subsample 1963:07-2001:12												
Monthly Return (%)	0.38	0.50	0.50	0.48	0.51	0.62	0.70	0.75	0.82	0.91	0.53	0.47
Monthly Volatility (%)	5.34	4.78	4.80	4.70	4.40	4.39	4.35	4.33	4.61	5.34	4.66	4.47
$\hat{\beta}_T^{OLS}$	1.11	1.02	1.02	0.97	0.89	0.90	0.84	0.84	0.88	0.95	-0.16	
$\hat{\alpha}_T^{OLS}$	-0.14	0.02	0.02	0.03	0.09	0.20	0.30	0.36	0.40	0.46	0.60	
$\hat{\alpha}_T^{OLS}$ N-W t-stat	[1.40]	[0.35]	[0.27]	[0.28]	[0.91]	[2.36]	[2.65]	[3.24]	[3.44]	[2.74]	[2.51]	

This table reports the summary statistics for the value-weighted book-to-market portfolios. Each month, stocks listed on NYSE, AMEX and Nasdaq are sorted according to their past book-to-market ratios into deciles of lowest book-to-market stocks (decile 1: growth stocks) to highest book-to-market stocks (decile 10: value stocks). Panels A, B, and C report the average monthly raw excess returns and monthly volatilities over various sample periods. The panels also report the unconditional OLS betas ($\hat{\beta}_T^{OLS}$), unconditional OLS alphas ($\hat{\alpha}_T^{OLS}$), and the t-statistics for the $\hat{\alpha}_T^{OLS}$ estimates computed using Newey-West (1987) (N-W) standard errors with three lags (in square brackets). Panel A covers the sample period from July 1926 to December 2001. Panels B and C cover the sample period from July 1926 to June 1963 and from July 1963 to December 2001, respectively. The column labelled "BM" lists the summary statistics for the strategy that goes long the decile 10 book-to-market portfolio and shorts the decile 1 book-to-market portfolio (book-to-market strategy). The column labelled "Market" reports the statistics for the value-weighted market portfolio.

Table 2: Summary Statistics of the 60-month Rolling Betas

	Value	Growth	Book-to-Market Strategy
60^{th} Autocorrelation	0.629	0.509	0.652
	(0.129)	(0.153)	(0.134)
Implied Monthly Autocorrelation	0.992	0.989	0.993
	(0.003)	(0.005)	(0.003)
Std. Deviation	0.377	0.108	0.468

This table reports the monthly mean-reversion parameter of the OLS betas, which are estimated by rolling 60-month OLS regressions. For each portfolio, we estimate the market beta over each 60-month subsample over the period July 1926 to December 2001. We compute their 60th autocorrelations and take their 60th roots as measures of the implied monthly mean-reversion parameters under the null of an AR(1) process. We also report the unconditional standard deviation of the 60-month rolling OLS betas. The "value" stock portfolio is the highest book-to-market (decile 10) portfolio, while the "growth" stock portfolio is the lowest book-to-market (decile 1) portfolio. The "book-to-market strategy" represents returns on a strategy that goes long value stocks and goes short growth stocks.

Table 3: Model Parameter Estimates

Market Parameters		Stock Return and Beta Parameters			
			Value	Growth	Bk-Mkt
μ_0	0.001	β_0	0.014	0.021	0.001
	(0.001)		(0.007)	(0.007)	(0.002)
ϕ_μ	0.956	ϕ_β	0.988	0.981	0.992
	(0.005)		(0.003)	(0.005)	(0.003)
σ_μ	0.004	σ_β	0.168	0.132	0.065
	(0.001)		(0.026)	(0.029)	(0.016)
v_0	-0.361	$\bar{\sigma}$	0.063	0.034	0.052
	(0.638)		(0.022)	(0.007)	(0.002)
ϕ_v	0.941	α^C	0.000	-0.002	0.002
	(0.110)		(0.002)	(0.001)	(0.002)
σ_v	0.274	$\rho_{\mu\beta}$	0.759	0.882	0.408
	(0.177)		(0.086)	(0.015)	(0.127)
$\rho_{\mu v}$	-0.093				
	(0.083)				

We report posterior means and standard deviations of parameters for the model:

$$\begin{aligned} r_{i,t} &= \alpha^C + \beta_t r_{m,t} + \bar{\sigma}\varepsilon_{i,t}, \\ r_{m,t} &= \mu_t + \sqrt{v_t}\varepsilon_{m,t}, \\ \beta_t &= \beta_0 + \phi_\beta \beta_{t-1} + \sigma_\beta \varepsilon_{\beta,t}, \\ \mu_t &= \mu_0 + \phi_\mu \mu_{t-1} + \sigma_\mu \varepsilon_{\mu,t} \\ \text{and} \quad \ln v_t &= v_0 + \phi_v \ln v_{t-1} + \sigma_v \varepsilon_{v,t}, \end{aligned}$$

where $r_{i,t}$ is an excess rate of return on a portfolio and $r_{m,t}$ is the excess rate of return on the market portfolio. The return shocks, $\varepsilon_{i,t}$ and $\varepsilon_{m,t}$, are independent standard normals. The conditional moments, β_t, μ_t, and $\ln v_t$, follow latent AR(1) processes, where the shocks, $\varepsilon_{m,t}$, $\varepsilon_{\mu,t}$, and $\varepsilon_{v,t}$, are standard normals. The correlation between $\varepsilon_{m,t}$ and $\varepsilon_{\mu,t}$ is $\rho_{\mu\beta}$, the correlation between $\varepsilon_{\mu,t}$ and $\varepsilon_{v,t}$ is $\rho_{\mu v}$, and the correlations between other error terms are zero. We separately estimate each portfolio with the market as a pair, but the estimates of the market are almost identical across all three portfolios. Value (growth) stocks refer to the highest (lowest) book-to-market decile portfolio. The column labelled "Bk-Mkt" refers to the return on a strategy of going long the value stock portfolio and going short the growth stock portfolio. All models are estimated over the full sample from July 1926 to December 2001, and we use a normal prior on α^C with zero mean and a standard deviation of 1% per month. We list the priors of other parameters in Appendix B.

Table 4: Conditional Alpha, α^C, of the Book-to-Market Strategy

	Prior Distribution of α^C							
$\mu^p_{\alpha^C}$	0.00	0.00	0.00	0.00	0.60	0.60	0.60	0.60
$\sigma^p_{\alpha^C}$	0.10	0.30	0.50	1.00	0.10	0.30	0.50	1.00
Mean	0.06	0.22	0.20	0.23	0.50	0.33	0.30	0.24
Std Dev	0.09	0.17	0.18	0.19	0.11	0.17	0.18	0.19
Percentiles								
0.010	-0.16	-0.20	-0.23	-0.23	0.20	-0.03	-0.13	-0.21
0.025	-0.12	-0.13	-0.15	-0.14	0.26	0.03	-0.05	-0.12
0.050	-0.09	-0.07	-0.09	-0.08	0.31	0.08	0.01	-0.06
0.100	-0.06	-0.01	-0.02	-0.01	0.37	0.13	0.06	0.01
0.250	0.00	0.10	0.09	0.11	0.44	0.22	0.17	0.12
0.500	0.06	0.23	0.20	0.23	0.51	0.32	0.29	0.24
0.750	0.11	0.38	0.32	0.35	0.58	0.44	0.43	0.36
0.900	0.17	0.38	0.43	0.47	0.63	0.55	0.53	0.49
0.950	0.20	0.41	0.50	0.54	0.67	0.62	0.58	0.56
0.975	0.24	0.47	0.56	0.61	0.70	0.68	0.64	0.63
0.990	0.27	0.53	0.64	0.70	0.74	0.75	0.72	0.72

This table reports the posterior distribution of the conditional alpha, α^C, from the conditional CAPM described by equations (13) to (15) for the book-to-market strategy, which goes long the decile 10 book-to-market portfolio (value stocks) and goes short the decile 1 book-to-market portfolio (growth stocks). The table reports the posterior distribution of the conditional alpha, α^C, in equation (13). We vary the mean, $\mu^p_{\alpha^C}$, and standard deviation, $\sigma^p_{\alpha^C}$, of the normal prior distribution on α^C as we move across the columns. We report various percentile points of the posterior distribution, in addition to posterior means and standard deviations. The table entries are expressed in terms of percentage returns per month. The models are estimated over July 1926 to December 2001.

Table 5: OLS Alpha, α^{OLS}, of the Book-to-Market Strategy

	Prior Distribution of α^C							
$\mu^p_{\alpha^C}$	0.00	0.00	0.00	0.00	0.60	0.60	0.60	0.60
$\sigma^p_{\alpha^C}$	0.10	0.30	0.50	1.00	0.10	0.30	0.50	1.00
Mean	0.29	0.43	0.43	0.45	0.70	0.55	0.52	0.47
Std Dev	0.14	0.17	0.20	0.20	0.14	0.19	0.21	0.21
Percentiles								
0.010	0.00	-0.01	-0.04	-0.04	0.35	0.14	0.06	-0.01
0.025	0.04	0.06	0.04	0.05	0.43	0.21	0.13	0.07
0.050	0.08	0.12	0.11	0.12	0.47	0.25	0.19	0.14
0.100	0.12	0.20	0.18	0.20	0.53	0.32	0.26	0.22
0.250	0.19	0.32	0.30	0.32	0.61	0.42	0.37	0.34
0.500	0.28	0.46	0.43	0.44	0.71	0.54	0.51	0.46
0.750	0.37	0.56	0.55	0.57	0.80	0.67	0.66	0.60
0.900	0.45	0.61	0.67	0.70	0.88	0.78	0.77	0.73
0.950	0.51	0.67	0.76	0.79	0.94	0.87	0.83	0.82
0.975	0.56	0.74	0.84	0.87	0.99	0.94	0.91	0.90
0.990	0.64	0.82	0.93	0.96	1.05	1.04	1.01	1.00

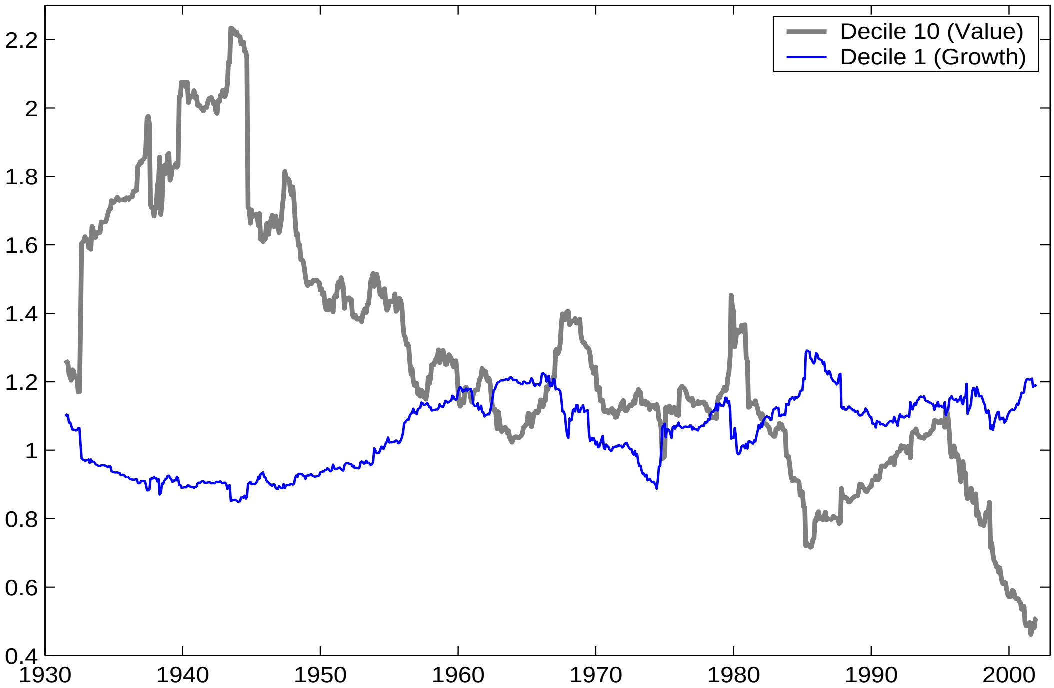
This table reports the posterior distribution of the OLS alpha, α^{OLS}, corresponding to a sample size of $T = \infty$ from the conditional CAPM described by equations (13) to (15) for the book-to-market strategy, which goes long the decile 10 book-to-market portfolio (value stocks) and goes short the decile 1 book-to-market portfolio (growth stocks). We compute the posterior distribution of α^{OLS} by simulating a time-series of 100,000 observations for each observation in the posterior distribution of the model parameters θ. For each simulated time-series, we estimate equation (1) and record the estimated α^{OLS}. We vary the mean, $\mu^p_{\alpha^C}$, and standard deviation, $\sigma^p_{\alpha^C}$, of the normal prior distribution on α^C as we move across the columns. We report various percentile points of the posterior distribution, in addition to posterior means and standard deviations. The table entries are expressed in terms of percentage returns per month. The models are estimated over July 1926 to December 2001.

Table 6: Finite-Sample OLS Alphas

	Conditional Alpha α^C	OLS Alpha α_T^{OLS}							
		With Parameter Uncertainty				Without Parameter Uncertainty			
	Full	Full	$\alpha^C = 0$	$\rho_{\mu\beta} = 0$	$\alpha^C = 0$ $\rho_{\mu\beta} = 0$	Full	$\alpha^C = 0$	$\rho_{\mu\beta} = 0$	$\alpha^C = 0$ $\rho_{\mu\beta} = 0$
Mean	0.23	0.43	0.20	0.23	0.01	0.42	0.20	0.24	0.01
Std Dev	0.19	0.50	0.47	0.49	0.46	0.42	0.42	0.42	0.42
Percentiles									
0.010	-0.23	-0.75	-0.88	-1.07	-1.19	-0.57	-0.79	-0.83	-1.06
0.025	-0.14	-0.53	-0.68	-0.77	-0.90	-0.39	-0.62	-0.62	-0.85
0.050	-0.08	-0.35	-0.51	-0.56	-0.73	-0.23	-0.46	-0.45	-0.68
0.100	-0.01	-0.16	-0.33	-0.36	-0.54	-0.07	-0.30	-0.27	-0.50
0.250	0.01	0.12	-0.09	-0.06	-0.26	0.16	-0.07	-0.02	-0.25
0.500	0.23	0.41	0.17	0.24	0.00	0.41	0.18	0.24	0.01
0.750	0.35	0.71	0.46	0.53	0.27	0.67	0.44	0.49	0.27
0.900	0.47	1.03	0.77	0.83	0.55	0.94	0.71	0.74	0.51
0.950	0.54	1.27	0.98	1.03	0.73	1.13	0.90	0.91	0.68
0.975	0.61	1.48	1.21	1.23	0.92	1.32	1.09	1.07	0.84
0.990	0.70	1.78	1.49	1.47	1.17	1.56	1.33	1.28	1.05

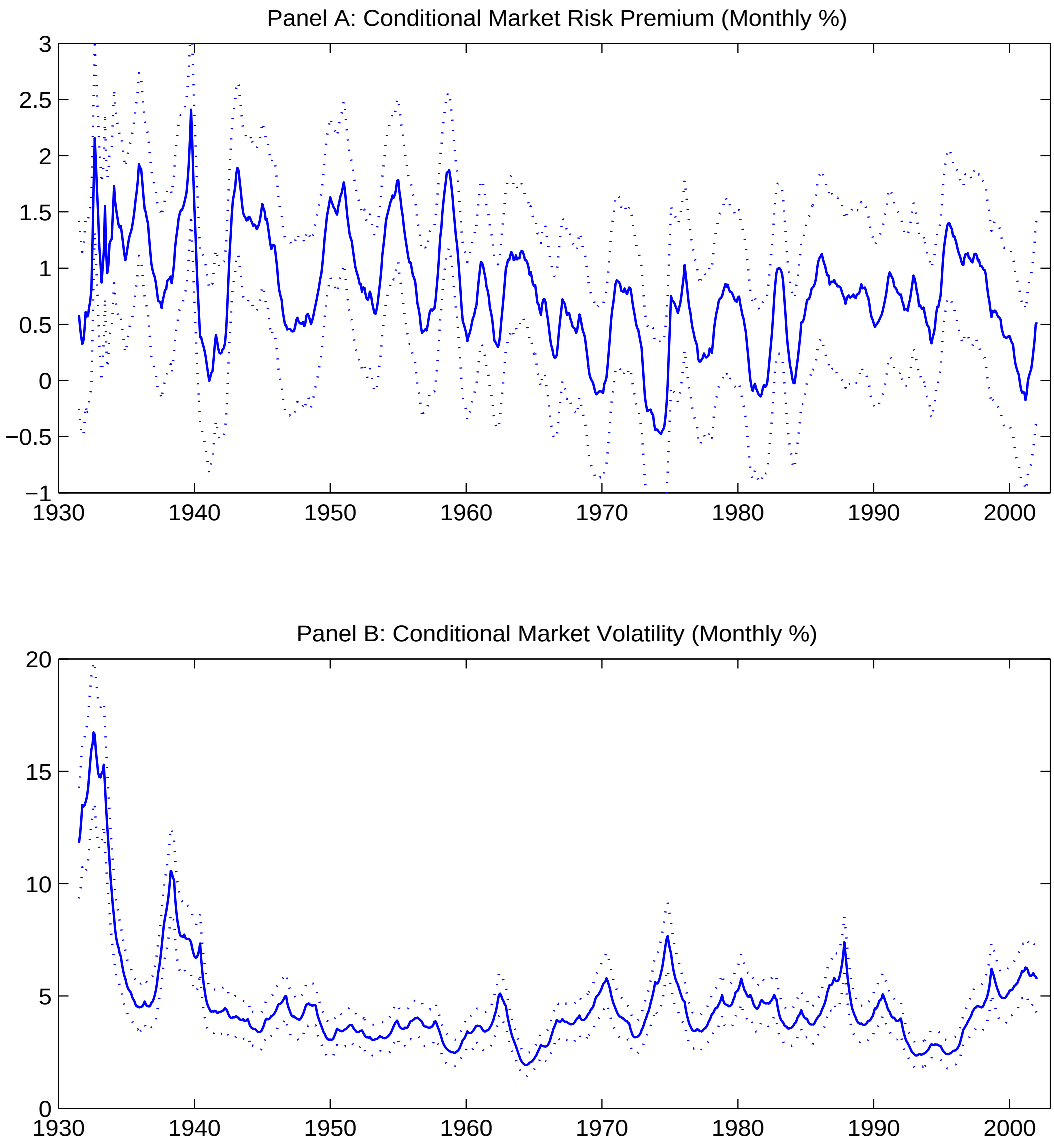
This table reports the small-sample posterior distribution of α_T^{OLS} from the conditional CAPM described by equations (13) to (15) for the book-to-market strategy for a small sample of length $T = 462$, which corresponds to the length of the post-1963 sample. In the columns under the line "With Parameter Uncertainty," we obtain the small-sample posterior distribution of α_T^{OLS} by simulating time-series of 462 observations for each observation in the posterior distribution of the model parameters θ. For each simulated time-series, we estimate the OLS alpha. In the columns under the line "Without Parameter Uncertainty," we disregard the effect of parameter uncertainty on the small-sample posterior distribution of α_T^{OLS} by simulating only from the posterior mean of the posterior parameter distributions. The first column repeats the posterior distribution of α^C from Table 4 for comparison. The columns under the line "OLS Alpha α_T^{OLS}" report the small-sample posterior distribution of α_T^{OLS}. The columns labelled "Full" report the results based on the full specification of the conditional CAPM, while the other columns set various parameters equal to zero. In all cases, we use a normal prior on α^C with zero mean and a standard deviation of 1% per month and estimate the full model over July 1926 to December 2001. All table entries are expressed in terms of percentage returns per month.

Figure 1: Estimates of 60-month Rolling OLS Betas for Book-to-Market Portfolios



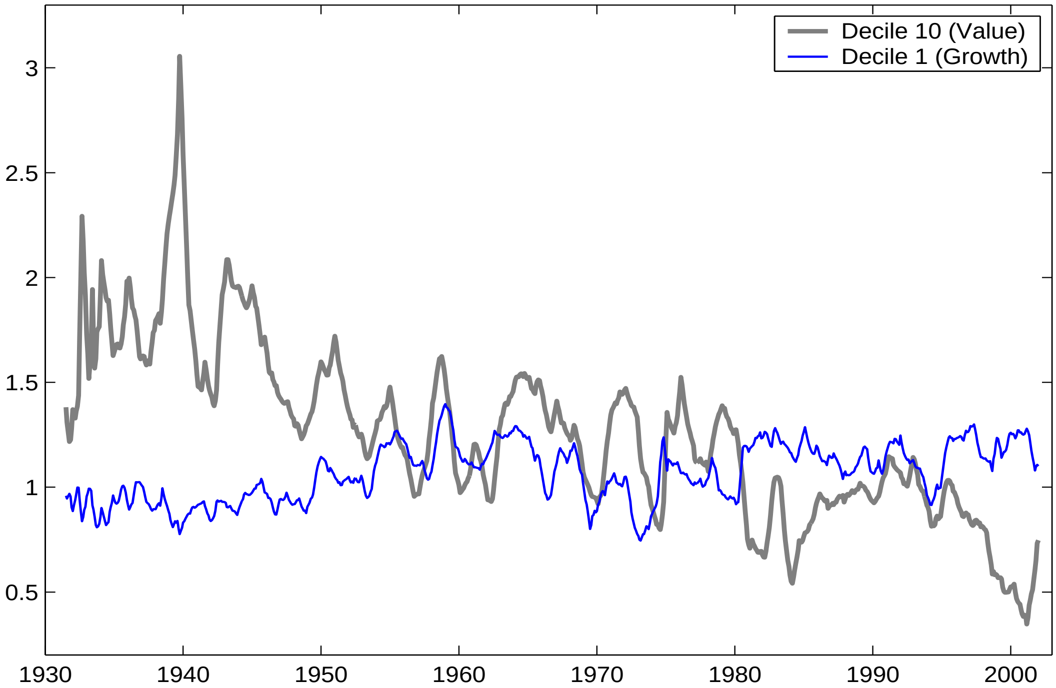
This figure shows the 60-month rolling OLS betas of the book-to-market decile portfolios from July 1931 to December 2001 for the decile 10 book-to-market portfolio (value stocks) and the decile 1 book-to-market portfolio (growth stocks). For each portfolio in each month, we estimate OLS beta using the past 60 months of observations using the regression in equation (1).

Figure 2: Time-Varying Market Risk Premia and Market Volatility



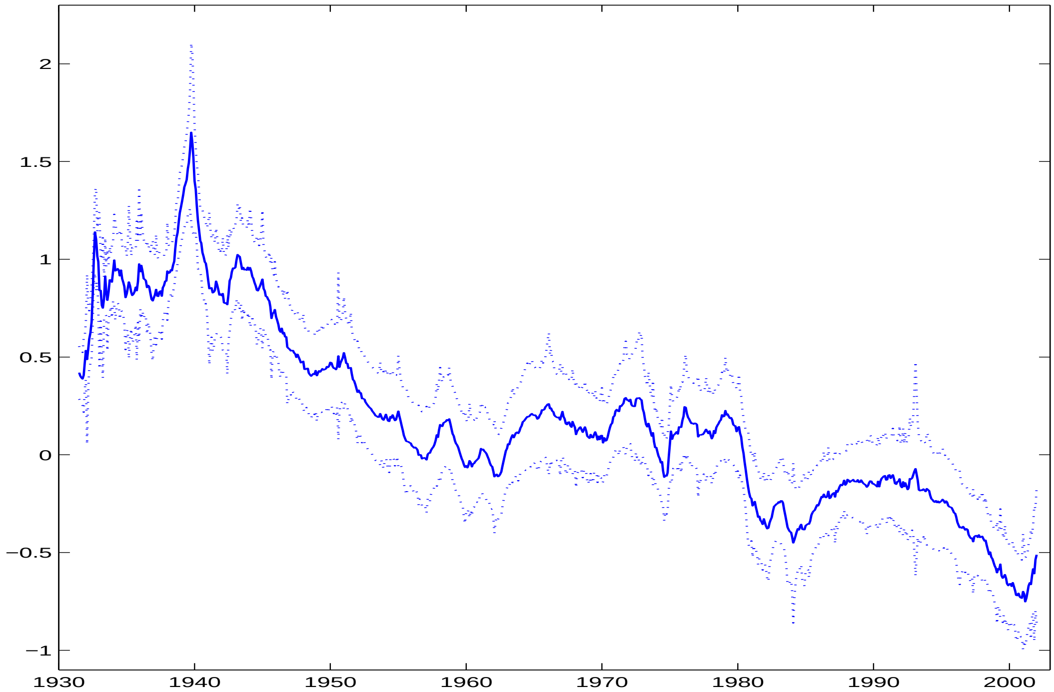
We plot the estimates of time-varying market risk premia (Panel A) and market volatility (Panel B) obtained by a Gibbs sampling estimation of the conditional CAPM described by equations (13) to (15). The dotted lines show a one posterior standard deviation bound.

Figure 3: Estimates of Time-Varying Betas for Book-to-Market Portfolios



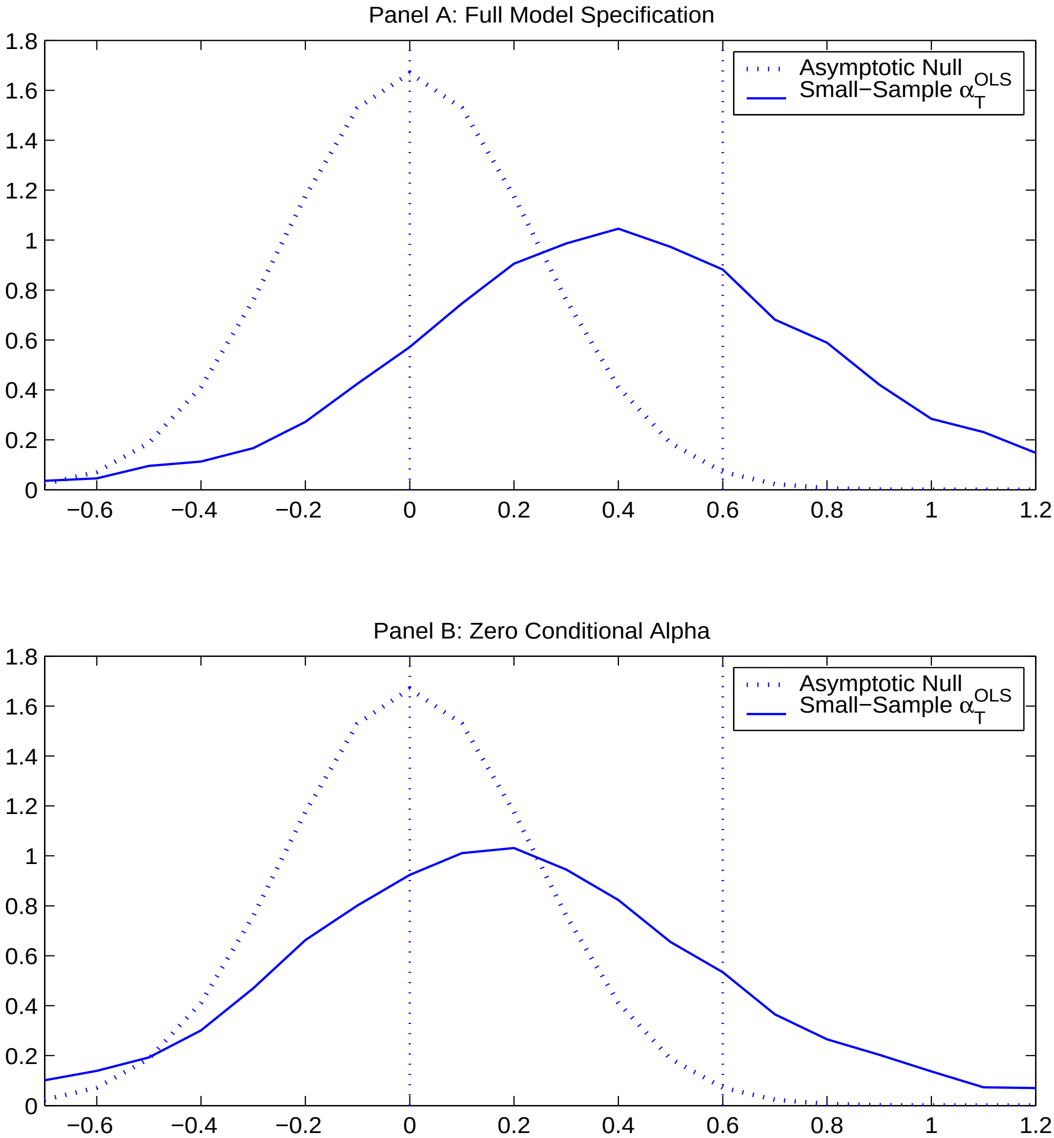
These plots show the inferred estimates of time-varying betas obtained by the Gibbs sampling estimation of the conditional CAPM described by equations (13) to (15) for the decile 10 book-to-market portfolio (value stocks) and the decile 1 book-to-market portfolio (growth stocks).

Figure 4: Estimates of Time-Varying Betas for the Book-to-Market Strategy



This plot shows the inferred estimates of time-varying betas obtained by the Gibbs sampling estimation of the conditional CAPM described by equations (13) to (15) for the strategy of going long the decile 10 book-to-market portfolio and going short the decile 1 book-to-market portfolio (the book-to-market strategy). The dotted lines show a one posterior standard deviation bound.

Figure 5: The Small-Sample Posterior Distribution of the OLS Alpha



The solid lines show the inferred small-sample posterior distribution of α_T^{OLS} from the conditional CAPM described by equations (13) to (15) for the book-to-market strategy for a sample size of $T = 462$, which corresponds to the post-1963 sample. Panel A graphs the distribution for the full model specification (without any parameter restrictions) and Panel B graphs the distribution where $\alpha^C = 0$. In both panels, we account for parameter uncertainty by simulating a time-series of 462 months and computing the OLS alpha for each point θ_i from the posterior distribution (consisting of 10,000 points) of the model parameters. For each simulated time series, we run an OLS regression and record its estimated α_T^{OLS}. The plots show the probability density function of the posterior small-sample α_T^{OLS} in solid lines. We also plot the probability density function of the asymptotic distribution of α^{OLS} assuming constant betas under the null that $\alpha^{OLS} = 0$ with a robust Newey-West (1987) standard error estimate over the post-1963 sample in dashed lines. The figures also indicate the location of the null hypothesis of $\alpha^{OLS} = 0$ as well as the location of the empirically observed $\hat{\alpha}_T^{OLS}$ of 0.60% per month with vertical dashed lines. The numbers on the x-axis in each panel are in percentage returns per month.