Efficient Estimation of Firm-Specific Betas and its Benefits for Asset Pricing Tests and Portfolio Choice

Cosemans, M. and Frehen, R.G.P. and Schotman, P.C. and Bauer, R.M.M.J.

CentER, Tilburg University

22 June 2009

Efficient Estimation of Firm-Specific Betas and its Benefits for Asset Pricing Tests and Portfolio Choice

Mathijs Cosemans[a,*], Rik Frehen[a,b], Peter C. Schotman[a,b,c], and Rob Bauer[a,b]

June 22, 2009

Abstract

We improve both the specification and estimation of firm-specific betas. Time variation in betas is modeled by combining a parametric specification based on economic theory with a non-parametric approach based on data-driven filters. We increase the precision of individual beta estimates by setting up a hierarchical Bayesian panel data model that imposes a common structure on parameters. We show that these accurate beta estimates lead to a large increase in the cross-sectional explanatory power of the conditional CAPM. Using the betas to forecast the covariance matrix of returns also results in a significant improvement in the out-of-sample performance of minimum variance portfolios.

Keywords: asset pricing, portfolio choice, time-varying betas, Bayesian econometrics, panel data

JEL classification: C11; C33; G12

Affiliations: [a]Maastricht University, [b]Netspar, [c]CEPR

*An earlier draft of this paper was titled "A Bayesian Panel Data Approach to Explaining Market Beta Dynamics". Corresponding author: Mathijs Cosemans, Maastricht University, P.O. Box 616, 6200 MD Maastricht, Netherlands, e-mail: m.cosemans@finance.unimaas.nl. We are grateful to Dion Bongaerts, Joshua Coval, Martijn Cremers, Joost Driessen, Frank de Jong, Will Goetzmann, Jonathan Ingersoll, Ludovic Phalippou, Jeffrey Pontiff, Stephen Shore, Luis Viceira, and seminar participants at Harvard University, Maastricht University, Stockholm School of Economics, University of Amsterdam, Yale University, Robeco Asset Management, the 2008 North American Summer Meeting of the Econometric Society in Pittsburgh, the 2008 European Meeting of the Econometric Society in Milan, the 2008 Annual Asset Pricing Retreat in Amsterdam, and the Netspar Workshop in The Hague for helpful comments and suggestions. We thank Inquire Europe for financial support. Part of this paper was written while Cosemans was a visiting fellow at Harvard University, Frehen a visiting scholar at Yale University, and Schotman a visiting fellow at SIFR in Stockholm.

Introduction

Precise estimates of firm-specific betas are crucial in many applications of modern finance theory, including asset pricing, corporate cost-of-capital calculations, and risk management. For instance, portfolio managers often have to ensure that their risk exposure stays within predetermined limits and managers need estimates of their company's beta to make capital budgeting decisions. Academics and practitioners have taken two approaches to estimating betas. Under the first one stocks are grouped into portfolios to reduce measurement error, assuming that all stocks within a given portfolio share the same beta (e.g., Fama and MacBeth (1973)). The second method consists of estimating separate time series regressions for each firm to obtain individual betas (e.g, Brennan, Chordia, and Subrahmanyam (1998)).

Apart from this lack of consensus in the literature about the best method to estimate betas, existing studies also fail to provide clear guidance on the best way to model betas. Although a large body of empirical evidence suggests that betas vary over time, existing work uses different specifications to model these changes in betas.[1] Many studies use a parametric approach proposed by Shanken (1990), in which variation in betas is modeled as a linear function of conditioning variables. An alternative, non-parametric approach to model risk dynamics is based on purely data-driven filters, including short-window regressions (Lewellen and Nagel (2006)) and rolling regressions (Fama and French (1997)).[2]

In this paper we improve both the *specification* and *estimation* of firm-specific, time-varying betas. We improve the specification of betas by combining the parametric and non-parametric approaches to modeling time variation in betas. Because the key strengths of each approach are the most important weaknesses of the other, we argue and show that a combination of the two methods leads to more accurate betas than those obtained from each of the two approaches separately. We allow the optimal mix of the two methods to vary across stocks, since individual firms may benefit more or less from either specification, and over time, because the preferred combination during stable market conditions may be different from that in turbulent time periods.

[1] See, for instance, Jagannathan and Wang (1996), Lewellen (1999), Ferson and Harvey (1999), Lettau and Ludvigson (2001), Andersen, Bollerslev, Diebold, and Wu (2005), Avramov and Chordia (2006a), Ang and Chen (2007), and Ang and Kristensen (2009).

[2] An alternative approach has been proposed by Christoffersen, Jacobs, and Vainberg (2007), who calculate forward-looking betas using the information embedded in option data. A drawback of this method is that it requires a cross-section of liquid stock options, which is not available for many small firms.

The parametric specification is appealing from a theoretical perspective because it explicitly links time variation in betas to macroeconomic state variables and firm characteristics (e.g., Gomes, Kogan, and Zhang (2003) and Santos and Veronesi (2004)). However, the main drawback of this approach is that the investor's set of conditioning information is unobservable. Ghysels (1998) shows that misspecifying beta risk may result in serious pricing errors that might even be larger than those produced by an unconditional asset pricing model. In addition, this method can produce excessive variation in betas due to sudden spikes in the macroeconomic variables that are often used as instruments. Finally, many parameters need to be estimated when a large number of conditioning variables is included, which leads to noisy estimates when applied to stocks with a limited number of time series observations. An important advantage of the non-parametric approaches is that they preclude the need to specify conditioning variables, which makes them more robust to misspecification. However, the time series of betas produced by a data-driven approach will always lag the true variation in beta, because using a window of past returns to estimate the beta at a given point in time gives an estimate of the average beta during this time period. Although reducing the length of the window results in timelier betas, the estimation precision of these betas will also decrease.

We improve the estimation of individual stock betas by setting up a Bayesian panel data model that exploits the information in the cross-section of firms to obtain more precise estimates. In particular, we specify hierarchical prior distributions that impose a common structure on parameters while still allowing for cross-sectional heterogeneity. Bayesian methods are especially attractive in settings with individual-level heterogeneity in multiple parameters, because only the parameters of the hierarchical priors where the parameters are assumed to be drawn from have to be estimated. In contrast, methods that estimate every parameter individually without linking it to similar parameters, such as estimating separate time series regressions for every single firm, suffer from poor estimation precision, particularly when the number of time series observations is limited. Intuitively, the Bayes estimator can be interpreted as a weighted average of the least squares estimator for a given cross-section unit and the cross-sectional average coefficient. The Bayes estimator of the firm-specific parameters shrinks the least squares estimator towards the cross-sectional mean. When the number of observations increases, the weight gradually shifts from the prior to the data.

Our panel data approach uses both daily returns and monthly firm-level characteristics to capture the cross-sectional heterogeneity and time series dynamics in monthly betas. Including cross-sectional information increases the accuracy of firm-specific betas because previous studies document a strong cross-sectional relationship between beta and firm characteristics (see, e.g., Fama and French (1992)). Existing work further shows that the use of high-frequency returns yields more precise and timelier estimates of beta than using monthly returns (see, e.g., Bollerslev and Zhang (2003)). We use daily returns instead of intraday returns because market microstructure frictions put an upper limit on the frequency that can be used to estimate betas in practice. We combine the data sampled at different frequencies by implementing the mixed data sampling (MIDAS) approach of Ghysels, Santa-Clara, and Valkanov (2005), which determines the optimal weights given to past data.

We estimate the model using a large panel of individual stocks, which offers several advantages over the alternative of aggregating stocks into portfolios based on characteristics. First, aggregating stocks into portfolios may conceal important information contained in individual stock betas. Ang, Liu, and Schwarz (2008) show that risk premia can be estimated more precisely using individual stocks instead of portfolios, because creating portfolios reduces the cross-sectional variation in betas. A second important drawback is that due to the strong factor structure in the 25 size-B/M sorted portfolios that are often used as test assets in asset pricing studies, traditional cross-sectional tests are flawed and have low power to reject a model, as shown by Lewellen, Nagel, and Shanken (2008). Third, when stocks are grouped into portfolios based on characteristics that have been identified by previous research as determinants of average returns instead of being based on economic theory, the evidence against asset pricing models may be overstated because of data-snooping biases (Lo and MacKinlay (1990)).

Despite the benefits of using individual stocks, most asset pricing studies use characteristics sorted portfolios because it is difficult to estimate firm-level parameters with a reasonable degree of precision when the number of observations is limited. Notable exceptions are Brennan, Chordia, and Subrahmanyam (1998) and Avramov and Chordia (2006a), who use a two stage approach to study the impact of characteristics on risk-adjusted returns. However, both studies estimate separate time series regressions for every single firm, which leads to imprecise beta estimates, particularly for firms with a short return history. Fama and French (2008) even conclude that "given the imprecision of beta estimates for individual stocks, little is lost in omitting them from the cross-section regressions".

Our main empirical findings are as follows. First, we show that modeling time-varying betas as a function of both conditioning variables and past returns dominates traditional specifications in which betas depend on only one of those components. Combining these specifications produces superior beta estimates because they capture different aspects of beta dynamics. We also find that the optimal mix of these specifications varies both over time and across stocks. Second, we show that our panel data approach produces more accurate estimates of firm-specific betas than those obtained from the traditional approach of estimating a separate time series regression for every firm. Specifically, for the average firm the posterior standard deviation of beta is significantly larger in time series regressions than in the panel model. Third, we document strong cross-sectional heterogeneity in firm betas within the 25 size-B/M portfolios that are commonly used to test asset pricing models. This confirms that aggregating stocks into portfolios conceals important information contained in individual stocks and shrinks the cross-sectional variation in betas.

We demonstrate that a more precise estimation and better specification of firm betas has important benefits for asset pricing tests and portfolio choice. In particular, we show that the betas generated by our model have significant explanatory power for the cross-section of returns. Using stocks as test assets and estimating betas in a panel model results in more efficient parameter estimates in cross-sectional asset pricing tests than using portfolios. The estimate of the market premium is positive and statistically significant, even after controlling for firm characteristics. We illustrate the value of our beta specification and estimation method for portfolio choice by using the betas to forecast the covariance matrix of stock returns. We find that the global minimum variance portfolio that is formed using this covariance matrix outperforms minimum variance portfolios based on other strategies, including the naive 1/N rule, the sample covariance matrix, and a static factor model for estimating covariances.

The paper proceeds as follows. In section I we introduce our specification for time-varying betas in a panel data framework. Section II explains the Bayesian approach to inference and section III describes the data. We report our empirical results in section IV and discuss the asset pricing and portfolio choice applications in section V. Section VI concludes.

I The Model

In this section we describe our model for individual betas. For simplicity, we discuss our approach in a conditional CAPM setting but it is straightforward to extend our work to multifactor models. Our goal is to show how to improve the specification and estimation of firm-specific, time-varying betas in any factor model. We start from the following panel data model for excess returns on individual stocks,

$$r_{it} = \alpha_i + \beta_{it-1} r_{Mt} + \epsilon_{it}, \tag{1}$$

where r_{it} is the excess return on stock i in month t, α_i is the risk-adjusted return, β_{it-1} is the conditional market beta, r_{Mt} is the excess market return, and ϵ_{it} is a zero-mean, normally distributed idiosyncratic return shock. Following Avramov and Chordia (2006b), we assume that the covariance matrix of these shocks is diagonal and that idiosyncratic volatility is constant.

Our specification for the conditional beta consists of two components: one part is the realized beta, b_{it}, and the other part is the fundamental beta, β^*_{it},

$$\beta_{it} = \phi_{it} b_{it} + (1 - \phi_{it})\beta^*_{it}, \tag{2}$$

where ϕ_{it} and $(1 - \phi_{it})$ measure the proportion of the beta of firm i that is explained by the realized beta and fundamental beta, respectively. Hereafter we refer to this mixture of realized and fundamental betas as the mixed beta. We allow the optimal combination of fundamental and realized betas to vary not only across firms but also over time. Time variation in ϕ_{it} is modeled as a linear function of market volatility, because the best mix of fundamental and realized betas in turbulent market conditions can be very different from that in stable periods,

$$\phi_{it} = \phi_{0i} + \phi_1 V_{Mt}, \tag{3}$$

where V_{Mt} is the realized market variance, which we calculate by summing the squared daily market returns over the past year. We take the logarithm of the market variance to reduce the impact of outliers and then subtract its time series average and divide by its standard deviation, so that it has mean zero and standard deviation equal to one.

b_{it} is the realized beta that we estimate using daily data according to the Mixed Data Sampling (MIDAS) approach introduced by Ghysels, Santa-Clara, and Valkanov (2005). We choose to estimate realized betas using daily returns because these provide a reasonable balance between efficiency and robustness to microstructure noise (see, Campbell, Lo, and MacKinlay (1997)). However, even at a daily frequency the betas of less liquid stocks might be biased downward. Following Scholes and Williams (1977), we therefore control for nonsynchronous trading effects by adding the covariance of the stock's return with the one-day lagged market return.[3]

The MIDAS approach differs from traditional rolling window estimators of betas by selecting the optimal window for estimating betas using a flexible weighting function. Ghysels, Santa-Clara, and Valkanov (2005) use the MIDAS approach to estimate the market's conditional variance and find that it is superior to traditional GARCH and rolling window methods. In particular, our MIDAS estimator of realized betas is given by:

$$b_{it} = \frac{\sum_{\tau=1}^{\tau^{max}} w_{t-\tau} r_{it-\tau}^{(d)} r_{Mt-\tau}^{(d)}}{\sum_{\tau=1}^{\tau^{max}} w_{t-\tau} r_{Mt-\tau}^{(d)} r_{Mt-\tau}^{(d)}} + \frac{\sum_{\tau=1}^{\tau^{max}} w_{t-\tau} r_{it-\tau}^{(d)} r_{Mt-\tau-1}^{(d)}}{\sum_{\tau=1}^{\tau^{max}} w_{t-\tau} r_{Mt-\tau-1}^{(d)} r_{Mt-\tau-1}^{(d)}}, \tag{4}$$

where t refers to a particular month, τ to a particular trading day, and $w_{t-\tau}$ to the weight given to the product of the return on stock i and the market return, $r_{it-\tau}^{(d)} r_{Mt-\tau}^{(d)}$, and to the squared market return, $r_{Mt-\tau}^{(d)} r_{Mt-\tau}^{(d)}$, on day $t-\tau$. We set the maximum window length τ^{max} equal to 250 days, which is approximately one year of trading days.

We parameterize the weights as a beta function:

$$w_{t-\tau} = \frac{f\left(\frac{\tau}{\tau^{max}}, \kappa_1; \kappa_2\right)}{\sum_{\tau=1}^{\tau_{max}} f\left(\frac{\tau}{\tau^{max}}, \kappa_1; \kappa_2\right)}, \tag{5}$$

where $f(\frac{\tau}{\tau^{max}}, \kappa_1; \kappa_2)$ is the density of a beta distribution. As pointed out by Ghysels, Santa-Clara, and Valkanov (2005), the specification based on the beta function has several advantages. First, it ensures that the weights are positive and sum to one. Second, it is parsimonious because only two parameters need to be estimated. Third, it is flexible as it can take various shapes for different values of the two parameters. We impose a downward sloping pattern on the weights by setting κ_1 equal to 1, which further reduces the number of parameters that need to be

[3] Scholes and Williams (1977) also include a lead term to capture the impact of non-synchronous trading on the market return. We only include a lag term because otherwise the model cannot be used to forecast betas.

estimated. $\kappa_1 = \kappa_2 = 1$ implies equal weights, which corresponds to a rolling window estimator of beta on daily data. $\kappa_1 = 1$ and $\kappa_2 > 1$ correspond to the case of decaying weights. In general, the higher κ_2, the faster the rate of decay and the quicker beta responds to new information.

β^*_{it} is the fundamental beta, parameterized as a function of conditioning variables,

$$\beta^*_{it} = \delta_0 + \delta_1'[Z_{it} \otimes BC_t], \tag{6}$$

where Z_{it} is a vector that contains L firm characteristics and BC_t is a vector that contains a constant and M business cycle variables. This specification allows the relation between beta and firm characteristics to vary over the business cycle. Modeling beta dynamics as a linear function of a set of predetermined instruments goes back to Shanken (1990) and is consistent with the economic motivation for conditional asset pricing models, in which the stochastic discount factor is a function of macroeconomic state variables and factor premia.

We include both firm-specific and macroeconomic variables as instruments for fundamental betas because of their documented predictive power for returns (Fama and French (1989) and Lewellen (1999)). Empirical evidence that systematic risk is related to firm characteristics and business cycle variables is provided by, among others, Jagannathan and Wang (1996), Lettau and Ludvigson (2001), Avramov and Chordia (2006a), and Goetzmann, Watanabe, and Watanabe (2008). The theoretical motivation for choosing firm characteristics as instruments is given by Berk, Green, and Naik (1999) and Gomes, Kogan, and Zhang (2003), who show that the ability of size and book-to-market to explain the cross-section of returns is due to their correlation with the true conditional market beta. They decompose firm value into the value of assets in place and the value of growth options and demonstrate that size captures the component of a firm's systematic risk related to its growth options whereas the book-to-market ratio is a measure of the risk of the firm's assets in place. Zhang (2005) extends this work and argues that because of costly reversibility of capital value firms have countercyclical betas while betas of growth stocks are procyclical. Because the price of risk is also countercyclical his model can explain the value premium within a rational framework. In addition to size and B/M, we also select firm-specific momentum as a conditioning variable to examine whether the momentum effect is related to beta dynamics. Theoretical support for including macroeconomic variables is provided by Santos and Veronesi (2004), who show within a general equilibrium model that market betas vary substantially with the business cycle. Our choice of business cycle variables is motivated by previous work (e.g., Ferson and Harvey (1999)) and includes the default spread, dividend yield, one-month T-bill rate, and term spread.

Substituting equations (2) and (6) into equation (1) leads to the following specification:

$$r_{it} = \alpha_i + \phi_{it} b_{it-1} r_{Mt} + (1 - \phi_{it})(\delta_0 + \delta_1' [Z_{it-1} \otimes BC_{t-1}]) r_{Mt} + \epsilon_{it}. \tag{7}$$

A key objective in this paper is to determine whether the time series dynamics and cross-sectional variation in betas is better described by lagged firm characteristics and macroeconomic state variables, by past realized betas, or by a linear combination of both. Therefore, we are primarily interested in the parameter ϕ_{it} and compare three different specifications: (1) mixed beta (ϕ_{it} unrestricted) (2) fundamental beta ($\phi_{it} = 0$) (3) realized beta ($\phi_{it} = 1$).

II Methodology

A Bayesian Methods

We estimate the model parameters using Bayesian methods.[4] The main advantage of Bayesian inference in our setting is that it allows a very flexible specification for describing the dynamics in beta by imposing a common structure on the model parameters. Updating beliefs according to Bayes' theorem implies that the joint posterior density of the parameters, $p(\theta|y)$, is proportional to the likelihood times the prior density.

$$p(\theta|y) \propto p(y|\theta) p(\theta) \tag{8}$$

where θ is the set of all parameters and y is the full set of data.

The likelihood function for the model in equation (7) is given by

$$p(y|\theta) = \prod_{i=1}^{N} \prod_{t \in T_i} \left(\sigma_{\epsilon_i}^2\right)^{-\frac{1}{2}} \exp\left[-\frac{1}{2\sigma_{\epsilon_i}^2} (r_{it} - \alpha_i - r_{Mt} \beta_{it-1})^2\right], \tag{9}$$

[4]Bayesian methods have been used in a number of asset pricing studies, including Shanken (1987), Harvey and Zhou (1990), Kandel, McCulloch, and Stambaugh (1995), and Cremers (2006). All these studies focus on portfolios and assume that betas are constant. Ang and Chen (2007) and Jostova and Philipov (2005) use Bayesian techniques to obtain time-varying portfolio betas, which they model as latent autoregressive processes.

where $\sigma^2_{\epsilon_i}$ is the idiosyncratic return variance, β_{it-1} is defined as in equation (2), N is the number of stocks, and T_i is the number of monthly return observations for firm i.

B Prior Distributions

We specify conditionally conjugate, hierarchical priors that impose a common structure on the model parameters while still allowing parameters to vary across firms. Thus, our setup combines the benefits of a portfolio approach to estimating betas (e.g., Fama and MacBeth (1973)) and an approach in which separate regressions are estimated for each firm (e.g., Avramov and Chordia (2006a)). Specifically, our choice of prior distributions is as follows:

$$
\begin{aligned}
\alpha_i &\sim N(0, \sigma^2_\alpha) && \text{with} \quad \sigma^2_\alpha \sim IG\,(0.001, 0.001)\,, \\
\phi_{0i} &\sim N(0.5, \sigma^2_{\phi_0}) && \text{with} \quad \sigma^2_{\phi_0} \sim IG\,(0.001, 0.001)\,, \\
\phi_1 &\sim N(0, \sigma^2_{\phi_1}) && \text{with} \quad \sigma^2_{\phi_1} \sim IG\,(0.001, 0.001)\,, \\
\delta_0 &\sim N(0, \sigma^2_{\delta_0}) && \text{with} \quad \sigma^2_{\delta_0} \sim IG\,(0.001, 0.001)\,, \\
\delta_1 &\sim N(0, \Omega_{\delta_1}) && \text{with} \quad \Omega^{-1}_{\delta_1} \sim Wish\left([(L+LM)I]^{-1}, (L+LM)\right), \\
\sigma^2_{\epsilon_i} &\sim IG(0.001, 0.001).
\end{aligned}
$$

We use diffuse priors to minimize their influence on the posterior densities. Following Jostova and Philipov (2005), we specify noninformative prior distributions for the variance parameters σ^2_α, $\sigma^2_{\delta_0}$, $\sigma^2_{\phi_0}$, $\sigma^2_{\phi_1}$ and the idiosyncratic variance $\sigma^2_{\epsilon_i}$, by setting the scale and shape parameters A and B of their inverse gamma (IG) prior distributions equal to 0.001. We set the degrees of freedom parameter ψ of the Wishart prior for $\Omega^{-1}_{\delta_1}$ equal to the dimension of this matrix, $(L + LM)$, because this value gives the lowest possible weight to the prior information (see Gelman, Carlin, Stern, and Rubin (2004)). We set the scale matrix of the Wishart prior equal to $[(L+LM)I]^{-1}$, so that the prior mean of $\Omega^{-1}_{\delta_1}$ is equal to the identity matrix. We give equal prior weight to the fundamental beta and the realized beta by setting the prior mean of ϕ_{0i}, i.e. μ_{ϕ_0}, equal to 0.5 and the prior mean of ϕ_1 equal to 0.[5] We parameterize the MIDAS weights as a beta function and set κ_1 equal to 1. To rule out cases where more recent returns receive less weight than observations in the more distant past, i.e., when $\kappa_2 < 1$, we constrain κ_2 to the

[5] We also considered specifications with μ_{ϕ_0} set equal to 0 or 1. These results are available upon request and show that our findings are robust to the choice of this prior distribution.

interval [1,26]. When $\kappa_2 = 1$ all 250 days receive equal weight in the estimation and when $\kappa_2 = 26$ the cumulative weight given to the 40 most recent days is 99%. We implement this restriction by a change of variable, $\kappa_2 = 1 + 25\kappa_2^*$. For κ_2^* we choose a uniform prior, $\kappa_2^* \sim U[0,1]$.

C Bayesian Inference

We employ Markov Chain Monte Carlo (MCMC) methods to sample from the joint posterior distribution of the parameter vector θ. The main idea is to construct a Markov chain such that the chain converges to a unique stationary distribution that is the posterior density, $p(\theta|y)$. We use the Gibbs sampler, which involves the sequential drawing from the full conditional posterior densities, to obtain draws from the joint posterior density. In particular, first the parameter vector θ is partitioned into B blocks $(\theta^{(1)}, \theta^{(2)}, ..., \theta^{(B)})$. At each iteration of the Gibbs sampler each block is sampled from its posterior distribution conditional on all other blocks and the data. Because the conditional posterior density of κ_2 has a nonstandard form, we cannot directly sample from it. Therefore, we use the Metropolis-Hastings algorithm, in which candidate parameter values are drawn from a proposal density and accepted with a certain probability that is highest in areas of the parameter space where the posterior density is highest (see Chib and Greenberg (1995)). Details on the derivation of the joint posterior density and the conditional posterior distributions are provided in the appendix.

Iterations of the chain converge to draws from the joint posterior. We check convergence by inspecting the standardized cumsum statistics, suggested by Bauwens, Lubrano, and Richard (1999), applying the partial means test based on numerical standard errors, explained by Geweke (2005), and calculating the Gelman-Rubin statistic that compares the variation in output between and within chains, described by Gelman, Carlin, Stern, and Rubin (2004). These diagnostics indicate that the parameter chains have converged after 1,000 iterations. In our empirical analysis we therefore run 5,000 iterations and discard the first 1,000 iterations as burn-in period. The remaining draws are used to summarize the posterior density and to conduct inference.

III Data

The firm data comes from CRSP and Compustat and consist of the monthly return, size, and book-to-market value for a sample of NYSE- and AMEX-listed stocks. To calculate realized betas we further retrieve daily returns from CRSP for these stocks. The sample covers the period from July 1964 to December 2006. Following Avramov and Chordia (2006a), we include a stock in the analysis for a given month t if it satisfies the following criteria. First, its return in the current month t and in the previous 36 months has to be available. Second, data should be available in month t-1 for size as measured by market capitalization and for the book-to-market ratio. We calculate the book-to-market ratio using accounting data from Compustat as of December of the previous year. Finally, in line with Fama and French (1993), we exclude firms with negative book-to-market equity. Imposing these restrictions leaves a total 5,017 stocks over the full sample period and an average of 1,815 stocks per month.

Table I presents summary statistics for the data set. Panel A reports the mean, median, standard deviation and 5th, 25th, 75th, and 95th percentile values of excess stock returns and firm characteristics across all data points. The average monthly excess stock return is 0.69% while the median is -0.16%. The mean (median) firm size is $1.59 (0.16) billion. Because the book-to-market ratio contains some extreme values, we trim all book-to-market outliers to the 0.5th and 99.5th percentile values of the distribution. After trimming, the average (median) book-to-market ratio equals 0.96 (0.75). The cumulative return over the twelve months prior to the current month, which we use as a proxy for momentum, has a mean of 14.65% and a median of 8.60%. Because the distributions of firm size and book-to-market display considerable skewness, we use the logarithmic transformations of these variables in the analysis. Furthermore, we normalize the characteristics by expressing them as deviations from their cross-sectional means to remove any time trend in the average value of the characteristics.

We further retrieve data for the four macroeconomic variables that we use as instruments for the fundamental beta, i.e., the default spread, dividend yield, one-month Treasury bill rate, and term spread. We define the default spread as the yield differential between bonds rated BAA by Moody's and bonds with a Moody's rating of AAA. The dividend yield is calculated as the sum of the dividends paid on the value-weighted CRSP index over the previous 12 months divided by the current level of the index. The term spread is defined as the yield difference between ten-year and one-year Treasury bonds. Panel B shows descriptive statistics for the macroeconomic variables. The average default spread is 1.02%, the mean dividend yield equals 3.01%, the average one-month T-bill rate is 5.69%, and the average term spread is 0.85%.

[Table I about here.]

IV Empirical Results

In section A we study whether betas are driven by lagged conditioning variables or past realized betas. Section B compares the efficiency of the beta estimates produced by the panel model to that of those obtained from time series regressions. In section C we illustrate the loss of information from aggregating stocks into portfolios by showing the cross-sectional variation in firm-level betas within the 25 size-B/M portfolios that are often used to test asset pricing models.

A Beta Specification

A key objective in this paper is to improve the specification of time-varying betas. We investigate whether the time series and cross-sectional variation in betas is best explained by lagged firm characteristics and macroeconomic variables, by past realized betas, or by a linear combination of both. We address this question by estimating the model in equation (7) and examining the distribution of ϕ_{it}, which measures the proportion of beta explained by past realized beta. We first calculate ϕ_{it} based on equation (3) for each draw of the Gibbs sampler. We then calculate for each firm the time series average ϕ_i and its posterior mean.

Figure 1 shows the cross-sectional distribution of these posterior means of ϕ_i. The cross-sectional average is 0.51, which implies that for the average firm the estimate of beta is the average of the fundamental and realized beta estimates. The spread in the distribution shows that for some firms past realized betas are more important determinants of mixed betas while for others lagged fundamental betas have a stronger impact.

[Figure 1 about here.]

Figure 2 plots the evolution of the cross-sectional average of ϕ through time. Interestingly, ϕ increases during periods of high market volatility, such as recessions and the stock market crash in 1987. This implies that more weight should be given to past realized betas and less weight to fundamental betas during turbulent conditions. Because the fundamental beta specification is a function of macroeconomic and firm-specific variables, it captures long-run movements in beta driven by structural changes in the economic environment and in firm- or industry-specific conditions. In contrast, because the realized beta specification is based on high-frequency returns, it picks up short-run fluctuations in beta in periods of high market volatility.[6]

[Figure 2 about here.]

Since we find that conditioning variables motivated by economic theory are important determinants of beta, we now consider the posterior distributions of the parameters underlying the fundamental beta. Table II presents summary statistics of the posterior distribution of the δ_0 and δ_1 parameters in equation (6). The constant term δ_0, which can be interpreted as the average fundamental beta because all conditioning variables are cross-sectionally demeaned, has a posterior mean of 1.01. The results show that all three firm characteristics are important determinants of fundamental betas. Some of the interaction terms between the firm characteristics and macroeconomic variables also capture important variation in market betas, particularly those involving the default spread and one-month T-Bill rate.

[Table II about here.]

As explained in section I, we use the MIDAS approach of Ghysels, Santa-Clara, and Valkanov (2005) to estimate realized betas based on daily return data. This approach incorporates a flexible weighting function that makes it possible to choose the optimal weights given to past data in the estimation. The optimal window strikes a balance between giving equal weight to observations to obtain more precise beta estimates and giving more weight to recent data to obtain betas that are timelier and therefore more relevant. As shown in equation (5), we use a beta weighting function whose shape is determined by two parameters. We set κ_1 equal to 1 and estimate κ_2. We find that in our realized beta specification the posterior mean of κ_2 is equal to 1.16. Figure 3 compares the optimal weighting scheme implied by the posterior mean of κ_2 to the equal weighting scheme used by rolling window estimators. The plot shows that in the optimal scheme the most recent 150 days receive more weight than in the equal weighting scheme because these are most informative for estimating realized betas.

[Figure 3 about here.]

[6] Related to this, Engle and Rangel (2008) model low-frequency patterns in market volatility as a function of macroeconomic and financial variables and Hoberg and Welch (2007) compute long- and short-run betas based on different windows of past returns.

We now turn to the mixed betas generated by our model. First, we calculate β_{it} based on equation (2) at each iteration of the Gibbs sampler. Subsequently, we compute for every firm the time series average of the mixed beta and its posterior mean. Figure 4 shows the cross-sectional distribution of these posterior means of β_i. As expected, the distribution is centered around one and has a standard deviation of 0.34. A 95% confidence interval for beta ranges from 0.46 to 1.60, which implies that firms differ substantially in their sensitivity to broad market movements.

[Figure 4 about here.]

In Table III we report summary statistics of the posterior means of all three beta specifications. Because for each t the cross-sectional average of $\bar{\beta}_{it}$ is close to one, the more interesting aspect is the dispersion in betas, both over time and in the cross-section. The left panel in Table III reports properties of the cross-section of betas and the right panel shows time series characteristics of beta. The diagonal elements in these panels show that on average, realized betas display the largest spread, both over time and across firms, while fundamental betas show the least variation. This is consistent with the notion that realized betas capture high-frequency movements in beta and fundamental betas pick up long-run beta fluctuations. Another explanation is that measurement error in the realized beta estimates leads to spurious dispersion or that in addition to firm size, book-to-market, and momentum, other firm characteristics drive variation in beta. The time series and cross-sectional behavior of the mixed betas is a combination of the dynamics of the realized and fundamental betas. Thus, it combines the benefits of both specifications, responding fast to changes in market conditions without producing excessive variation in beta. The off-diagonal elements in Table III are the correlations between the betas generated by the three specifications. Fundamental and realized betas are strongly correlated, both over time and across stocks. Correlation is far from perfect though, as a regression of one on the other has an R^2 of only 0.68. This illustrates that realized and fundamental betas exhibit different cross-sectional characteristics and time series dynamics. Hence, a combination of these two specifications captures different aspects of market beta dynamics.

[Table III about here.]

B Beta Estimation

In this section we compare the precision of beta estimates from the hierarchical Bayesian panel data model to that of estimates from a separate Bayesian time series regression for every firm. We study the estimation efficiency of the two methods for the fundamental beta specification, where ϕ_{it} is fixed at zero. Since this specification requires the estimation of many parameters when a large number of conditioning variables is included, the efficiency gain from using the panel model can be substantial. We measure estimation precision by computing the standard deviation and the 5% and 95% percentile values of the posterior distribution of beta at each point in time.

Figure 5 plots the posterior mean and 5% and 95% percentile values of the posterior distribution of the fundamental beta of IBM from August 1964 through December 2006. The upper graph is based on the estimation output of the panel data model and the lower graph is constructed using the output of a time series regression. The shaded areas in the plot indicate NBER recession periods. The plots show that the confidence interval for beta obtained from the panel regression is much narrower than the interval produced by the time series regression. Noisy estimates of the δ_1 parameters, which measure the influence of the conditioning variables on fundamental betas, lead to wide intervals for beta in the time series model.

[Figure 5 about here.]

The large efficiency gain in the panel model is due to two reasons. First, the δ parameters are pooled across stocks in the panel specification. The panel model therefore exploits the information in the cross-section of stocks to obtain more precise estimates. Second, because we specify hierarchical priors for the firm-specific parameters in the model, only the parameters of the common distribution where the parameters are assumed to be drawn from have to be estimated. The Bayes estimator of the firm-specific parameters in the panel model shrinks the least squares estimator towards the cross-sectional mean. In contrast, in the time series regressions every parameter is estimated individually, which results in poor estimation precision when many parameters need to be estimated and the number of time series observations is small.

Because in our panel approach parameters can be estimated more precisely, it can include more conditioning variables than the traditional approach of estimating a time series regression for every firm used by Avramov and Chordia (2006a). While we include 15 conditioning variables to accurately model beta dynamics, they note that "attention must be restricted to a small number of such variables to ensure some precision in the estimation procedure". To compare the relative efficiency of the panel and time series approaches when a more parsimonious specification for fundamental betas is used, we also estimate the panel model and time series regressions with a set of conditioning variables that is similar to that used by Avramov and Chordia (2006a). In particular, we choose firm size, book-to-market, and two interactions terms between these characteristics and the default spread as instruments.

The confidence intervals for the fundamental beta of IBM based on this reduced set of conditioning variables are displayed in figure 6. As expected, the intervals for beta generated by the panel model and the time series regressions have both narrowed compared to those based on the complete set of conditioning variables. However, the plots show that even when less parameters need to be estimated the panel approach leads to more precise estimates of firm-specific betas than the time series approach.

[Figure 6 about here.]

Because IBM is present in our data set during the entire sample period, many observations are available for beta estimation (509 months). As explained before, we expect the efficiency gain from the hierarchical Bayesian panel data approach to be even larger for firms with a short return history. To summarize the estimation precision for the betas of all firms, we compute the cross-sectional average of the posterior standard deviations of all betas in every month. Figure 7 plots these standard deviations for the panel model and time series regressions. Clearly, the posterior standard deviation of betas estimated using the time series regressions is larger than the standard deviation of betas estimated using the panel regression.

[Figure 7 about here.]

C Portfolio Heterogeneity

The previous section has shown that firm-specific betas are noisy when estimated using time series regressions. To reduce the measurement error in betas, Fama and MacBeth (1973) propose to aggregate stocks into portfolios and run a time series regression for every portfolio to obtain the portfolio's beta. Fama and French (1992) follow this suggestion and assign each stock the beta of the portfolio it belongs to. As pointed out by Ferson and Harvey (1999), such an approach is often used in studies of initial public offerings (IPOs), when no return data is available to estimate beta. They note that this approach only works when the characteristics used for portfolio formation are good proxies for risk, because an important assumption underlying the portfolio approach is that the stocks in a particular portfolio share the same risk characteristics. In case of the widely used 25 portfolios sorted on firm size and book-to-market, it is assumed that firms are homogeneous in their exposure to risk after controlling for size and B/M. When the stocks in a given portfolio have different exposures to other determinants of risk, this method can lead to serious errors. In this section we therefore examine whether firms that are grouped together in a portfolio have similar risk characteristics.

We construct the 25 size-B/M portfolios following the procedure of Fama and French (1993). Subsequently, we calculate for every portfolio j in every month t the cross-sectional average and standard deviation of the excess returns and posterior means of the alphas, betas, and phis of the stocks in that portfolio. The left part of Table IV reports for each portfolio the time series means of these cross-sectional averages. Consistent with prior studies (e.g., Fama and French (1996)), the small-growth portfolio has the lowest average return and a large, negative pricing error. In general, the average portfolio returns display a strong value premium but weak size effect. Importantly, sorting on firm size and B/M does not produce a wide spread in *average* market betas across portfolios, as most portfolio betas are close to one.[7] The table further shows that the phi parameters of large cap portfolios are higher than those of small cap portfolios. This implies that realized betas are the most important determinants of the mixed betas of large firms while fundamental betas have a stronger effect on the mixed betas of small firms.

Table IV also shows the dispersion of the risk and return characteristics across stocks in each portfolio. For all characteristics we observe strong heterogeneity *within* portfolios. In some portfolios the cross-sectional standard deviation of firm-specific alphas is more than 1%. Especially firms that are grouped together in small cap portfolios have significantly different pricing errors. The cross-sectional variation in betas of firms in each portfolio is around 0.30, which implies that the assumption that stocks in the same portfolio have similar risk characteristics is violated. Table IV also reports substantial heterogeneity in phi within portfolios. This

[7]However, unreported results show that value and growth portfolios exhibit very different risk *dynamics*. Confirming the results of Ang and Chen (2007) and Franzoni (2007), we find that the beta of value firms shows a declining trend and is lower than the beta of growth stocks since the 1980s.

means that for some firms in a given portfolio mixed betas are mainly driven by realized betas whereas for others fundamental betas are more important.

[Table IV about here.]

V Applications of Firm-Specific Betas

This section discusses two important applications of the firm-level betas generated by our Bayesian panel data model. In section A we compare the explanatory power of different beta specifications and estimation methods for the cross-section of individual stock returns. Section B uses the beta forecasts to estimate the covariance matrix of stock returns, which we then use to construct minimum variance portfolios.

A Cross-Sectional Tests of the Conditional CAPM

The previous section has shown that aggregating individual stocks into portfolios leads to a substantial loss of information and shrinks the cross-sectional variation in betas. Ang, Liu, and Schwarz (2008) demonstrate that this loss of information can lead to large efficiency losses in cross-sectional tests of asset pricing models. In particular, they show that while creating portfolios reduces estimation error in betas, standard errors of risk premia estimates are higher due to the smaller spread in betas. Consequently, using individual stocks instead of portfolios as base assets allows for more powerful tests of asset pricing models.

Another important reason for using individual stocks in cross-sectional tests of asset pricing models is given by Lewellen, Nagel, and Shanken (2008). They show analytically that due to the strong factor structure in the 25 size-B/M sorted portfolios often used as test assets in asset pricing studies, traditional cross-sectional tests have low power to reject a model. In particular, when theoretical restrictions on cross-sectional slopes are ignored, any factor that is only weakly correlated with the true factors can generate high cross-sectional R^2s and small pricing errors. Lewellen and Nagel (2006) show that the empirical support for several recently proposed asset pricing models weakens considerably when this issue is taken into account. Because individual stock returns do not have a strong factor structure, they are not affected by this problem.

In their analysis, Ang, Liu, and Schwarz (2008) assume constant stock betas, which they estimate by running time series regressions. We extend their work in two directions. First, we improve the specification of betas by allowing for time variation. Second, we use a formal panel data approach to increase the precision of firm-specific beta estimates. We do not claim that the CAPM is the "best" asset pricing model. Our objective is to show the effect of better beta specification and estimation on the pricing ability of the CAPM.

We first estimate betas for all stocks in our sample and for the 25 size-B/M portfolios. We consider four beta specifications (mixed, fundamental, realized, and static) and estimate the models using hierarchical Bayesian panel regressions. These betas are then used as independent variables in second stage monthly cross-sectional regressions of excess returns on betas,

$$r_{it} = \lambda_{0t} + \lambda_{1t}\beta_{it-1} + \lambda_{2t}' x_{it-1} + \upsilon_{it}, \tag{10}$$

where λ_{0t} is the intercept, λ_{1t} the risk premium, and where x_{it-1} is a vector of control variables. We run the cross-sectional regressions for every draw of the Gibbs sampler and calculate the time series average of the cross-sectional coefficients. We then calculate the posterior mean and variance of the Fama-MacBeth estimators. In appendix B we demonstrate that this procedure accounts for measurement error in beta by using the entire posterior distribution of the β_{it} in the estimation.

Columns 1-3 in Table V report the Fama-MacBeth coefficient estimates when individual stocks are the test assets and no control variables are included in the regression ($x_{it-1} = 0$). We find that for the mixed beta specification the intercept is close to zero and insignificant while the risk premium estimate is significantly positive. The λ_{1t} estimate is 0.56%, which is close to the average monthly excess market return during the sample period (0.47%). This implies that the conditional CAPM with mixed betas satisfies the theoretical restriction emphasized by Lewellen and Nagel (2006) that the risk premium should equal the expected excess factor return. For the other three beta specifications the intercepts are significantly different from zero. The risk premium estimates in the realized beta and fundamental beta models are positive and significant but deviate more from the average market return than the premium estimated in the mixed beta model. In terms of explanatory power the mixed beta specification also outperforms the competing approaches to modeling beta. The static CAPM performs worst, because in this model the cross-sectional variation in market betas does not respond to business cycle variations.

Columns 4-6 in Table V show that when portfolios are used as test assets, all beta specifications generate economically large intercepts. Nevertheless, the mixed beta specification again has the highest explanatory power and the static CAPM does worst. We stress that the R^2 should only be compared across beta specifications and should not be used to compare individual stocks and portfolios as test assets, because the dependent variables in the cross-sectional regressions are different. Table V further shows that the standard errors of the parameter estimates are much larger when portfolios are used as test assets than when individual stocks are used, which confirms that sorting stocks into portfolios can lead to large efficiency losses because it reduces the dispersion of betas. In fact, standard errors from using portfolios are more than twice as large as those from using individual stocks.

The last three columns in Table V report estimation results for individual stocks when control variables are added to the cross-sectional regressions. In particular, the vector x_{it-1} contains the firm characteristics size, book-to-market, and momentum. Fama and French (1992) find that the cross-sectional relation between between market beta and average return is flat when tests control for size. We find that while adding these firm characteristics leads to an increase in explanatory power, the risk premium estimate for the mixed beta specification remains significantly positive. Thus, when individual stocks are used as test assets and betas are well-specified and precisely estimated, the positive relation between beta and return no longer disappears when controlling for firm characteristics.

[Table V about here.]

B Beta Forecasts and Minimum Variance Portfolios

An important application of betas is to estimate the covariance matrix of returns, which is used to construct mean-variance efficient portfolios. Traditional implementations of the portfolio theory developed by Markowitz (1952) use sample moments. When the number of assets is large, however, it is difficult to precisely estimate the expected returns and covariances. As a result, asset weights are often extreme and portfolios behave poorly out-of-sample. Many strategies have been proposed to improve the out-of-sample performance of mean-variance portfolios, including shrinkage estimators, imposing short-selling constraints, using asset pricing models to estimate expected returns, and imposing a factor structure on the covariance matrix.[8]

[8]See, e.g., Chan, Karceski, and Lakonishok (1999), Jagannathan and Ma (2003), and Ledoit and Wolf (2003)).

In a recent study, DeMiguel, Garlappi, and Uppal (2007) compare the out-of-sample performance of these approaches to the $1/N$ rule that gives equal weight to all available assets. They conclude that none of the more sophisticated methods consistently outperforms the naive $1/N$ benchmark in terms of Sharpe ratio or certainty equivalent return. Of the alternative models considered, the minimum variance portfolio with short-selling constraints proposed by Jagannathan and Ma (2003) has the highest Sharpe ratio but is not superior to the $1/N$ strategy. The global minimum variance portfolio does well, because it is the only efficient portfolio that does not require estimates of expected returns, which contain large estimation errors. Chan, Karceski, and Lakonishok (1999) compare the out-of-sample performance of minimum variance portfolios based on forecasts of future covariances produced by factor models. They find that there is one major factor, the market, that dominates all other factors. Hence, the one-factor model is adequate for forming the global minimum variance portfolio.

Motivated by these findings, we use the mixed beta forecasts produced by our Bayesian panel data model to forecast the covariance matrix of stock returns and construct the global minimum variance portfolio. We expect our method to outperform competing approaches because it delivers more efficient estimates of firm-specific betas and because it allows for time variation in beta. DeMiguel, Garlappi, and Uppal (2007) admit that their assumption of constant risk is a limitation, but argue that models that allow for time-varying moments are likely to perform poorly out-of-sample because many parameters need to be estimated. However, one of the key advantages of our method is that it can estimate many parameters with high precision. We compare the out-of-sample performance of our approach to that of the traditional sample covariance matrix, the static one-factor structure considered by Chan, Karceski, and Lakonishok (1999), and the $1/N$ rule advocated by DeMiguel, Garlappi, and Uppal (2007).[9] Engle and Colacito (2006) stress the importance of isolating the effect of covariance information from expected returns when the objective is to evaluate different covariance estimators. Because expected returns do not enter the optimization when constructing minimum variance portfolios, differences in the portfolio weights only reflect the effect of different covariance forecasts.

[9] Brandt, Santa-Clara, and Valkanov (2009) propose a new approach to portfolio optimization, in which portfolio weights are modeled as a function of firm characteristics to exploit cross-sectional patterns in stock returns. They show that this parametric portfolio policy performs well for expected return maximizing portfolios but note that it works less well when the objective is to construct risk minimizing portfolios.

The first estimator of the covariance matrix is the sample covariance matrix,

$$S_t^S = \frac{1}{T-1}\sum_{t=1}^{T}(R_t - \bar{R})(R_t - \bar{R}), \tag{11}$$

where R_t is the vector of monthly stock returns and $\bar{R}$ contains the sample mean returns. Our second covariance estimator is based on the one-factor model. In the first step, we estimate mixed betas using our panel approach and static betas using time series regressions. We use these betas to estimate the covariance matrix in each month according to the one-factor model,

$$S_t^F = s_{Mt}^2 B_t B_t' + D, \tag{12}$$

where B_t is the $N_t \times 1$ vector of betas, s_{Mt}^2 is the sample variance of the market premium, and D is a diagonal matrix that contains the variances of the residuals.

Subsequently, we use the various estimates of the covariance matrix to construct the minimum variance portfolio, by choosing the portfolio weights that solve the following problem:

$$min \quad w_t' S_t w_t, \tag{13}$$

$$s.t. \sum_i w_{it} = 1. \tag{14}$$

The constraint implies that the portfolio is fully invested. Following Chan, Karceski, and Lakonishok (1999) and Jagannathan and Ma (2003), we also consider an extension in which we add a short-selling constraint,

$$w_{it} \geq 0, \quad i = 1, 2, ...N. \tag{15}$$

We form the minimum variance portfolio at the end of each month, based on the forecast of the covariance matrix for the next month. The first portfolio is formed using the first half of the sample period to forecast the covariance matrix of returns according to the methods explained above. Because the sample covariance matrix cannot be positive definite unless the number of return observations per stock is larger than the number of stocks, we only apply this method to a subset of the investment universe. We record the performance of the minimum variance portfolio in the next month and rebalance the portfolio using the new forecast of the covariance matrix. This method produces a time series of monthly returns for global minimum variance portfolios constructed using different covariance estimators. As a benchmark we also form an equally weighted portfolio at the end of each month. Since the objective is to minimize the portfolio variance, we evaluate the performance of the different methods by calculating the realized volatility of the portfolio returns.

Table VI reports annualized risk and return characteristics of the minimum variance portfolios constructed using various forecasts of the covariance matrix. Panel A shows the out-of-sample performance when all stocks in the sample are used in the optimization and short-selling is allowed. The mixed beta specification estimated using the panel data approach outperforms all other methods and produces a portfolio with an annualized standard deviation of 8.12%. The $1/N$ strategy leads to a standard deviation that is almost twice as large (15.40%). The static beta model ranks second and generates an out-of-sample standard deviation of 8.50%. Although our objective is to minimize portfolio variance, we find that the mixed beta approach also leads to the best risk-return tradeoff, as it produces the highest Sharpe ratio of all four approaches, equal to 0.69. Since the minimum variance portfolio constructed using the mixed beta approach does not involve taking short positions, it can also be easily implemented in practice. No short positions are taken because the mixed beta forecasts for all stocks are positive.

In panel B we report the performance for portfolios constructed from a random sample of 250 stocks, which is the same number of stocks considered by Chan, Karceski, and Lakonishok (1999). For this subset of stocks the $1/N$ rule leads to the highest out-of-sample standard deviation (15.93%), followed by the sample covariance matrix, which generates a standard deviation of 15.19% and involves taking large short positions. The mixed beta panel approach again beats all other methods and produces a portfolio with a standard deviation of 8.41%. Thus, the performance of this approach when applied to the smaller sample of stocks is similar to that when all stocks are used in the optimization. An important reason for the relatively poor performance of the naive $1/N$ strategy is that we allocate wealth across individual stocks. In contrast, DeMiguel, Garlappi, and Uppal (2007) apply this policy to allocate wealth across portfolios of stocks. They point out that the loss from naive as opposed to optimal diversification is much larger when allocating wealth across individual assets, because individual stocks have higher idiosyncratic volatility than portfolios.

Jagannathan and Ma (2003) document that the out-of-sample performance of the sample covariance matrix can be improved by imposing no-short-sale constraints, because these reduce sampling error. Panel C reports the risk and return characteristics of the portfolios generated by the four methods when the nonnegativity constraint is imposed on the weights. The random sample of 250 stocks used to form the portfolios is the same as that used in panel B. We find that the standard deviation of the minimum variance portfolio constructed using the sample covariance matrix is indeed lower when no-short-sale restrictions are in place. In fact, this method yields a lower standard deviation than the equally weighted ($1/N$) portfolio or the one-factor model with static betas. We also confirm the finding of Jagannathan and Ma (2003) that imposing no-short-sale constraints reduces the performance of the static factor model. Because the portfolio produced by the mixed beta approach in panel B does not take short positions, it is not affected by the nonnegativity constraint and still has the smallest out-of-sample standard deviation.

[Table VI about here.]

VI Conclusion

Many applications of modern finance theory require precise beta estimates for individual stocks. However, as noted by Campbell, Lettau, Malkiel, and Xu (2001), "firm-specific betas are difficult to estimate and may well be unstable over time". Academics and practitioners have taken two approaches to estimating firm-level betas. The first method sorts stocks into portfolios based on characteristics to reduce measurement error and assigns each firm the beta of the portfolio it belongs to. However, when stocks in the same portfolio have different exposures to other determinants of risk than the characteristics they are sorted on, this approach can lead to serious errors. The second method estimates a separate time series regression for every stock. Although this approach allows each firm to have a different risk exposure, the resulting beta estimates can be very noisy. The literature also uses different specifications to model time variation in betas. Many studies use a parametric approach in which variation in beta is modeled as a linear function of conditioning variables. An alternative, non-parametric approach to model risk dynamics is based on purely data-driven filters. However, both methods have important drawbacks and involve a trade-off between precision and timeliness of beta estimates.

In this paper we therefore improve both the specification and estimation of firm-specific, time-varying betas. We combine the parametric and non-parametric approaches for modeling changes in betas. The precision of firm-level beta estimates is increased by setting up a Bayesian panel data model that exploits the information contained in the cross-section of stocks and imposes a common structure on parameters while still allowing for cross-sectional heterogeneity.

We find that modeling time-varying betas as a function of both conditioning variables and past return data is preferred over traditional beta specifications that are based on only one of these components. Because fundamental and realized betas exhibit different time series dynamics and cross-sectional characteristics, a combination of these specifications captures different aspects of beta. We show that the optimal mixture of these two betas varies across firms and over time. We further demonstrate that our panel data approach yields more precise estimates of firm-level betas than the traditional approach of estimating betas by running a time series regression for every firm. Moreover, we document strong cross-sectional variation in betas of firms that are grouped together in portfolios sorted on size and book-to-market. Consequently, aggregating stocks into portfolios conceals important information contained in individual stock betas and reduces the cross-sectional variation in betas.

We demonstrate that the mixed betas generated by our panel data model lead to a sharpe increase in the pricing ability of the conditional CAPM. The estimate of the market risk premium remains significantly positive when controlling for firm characteristics. The results support the finding of Ang, Liu, and Schwarz (2008) that the use of individual stocks as tests assets instead of portfolios leads to more efficient estimates in cross-sectional tests of asset pricing models. We extend their work by showing that a better specification and more precise estimation of stock-specific betas increases the explanatory power of the CAPM.

Accurate estimates of firm-specific betas are also important for portfolio optimization. Based on the mixed beta estimates produced by our panel model we forecast the covariance matrix of stock returns, which is then used to form minimum variance portfolios. The portfolio constructed using mixed betas from the Bayesian panel approach outperforms portfolios of other strategies, such as the traditional sample covariance matrix and the naive $1/N$ rule, in terms of out-of-sample standard deviation. The mixed beta specification is also superior to competing approaches when short-selling constraints on portfolio weights are imposed.

Since our framework is flexible, it can be readily extended to include multiple risk factors, a different set of conditioning variables for fundamental betas, or another window length for estimating realized betas. In addition, while we have demonstrated the advantages of our approach for asset pricing and portfolio management, it also has important benefits in corporate finance applications. Specifically, because it quickly captures changes in beta and generates precise beta estimates even when little return data is available, our method is well suited for calculating risk-adjusted returns in studies of IPOs and M&As.

A Posterior Distributions

A Joint Posterior Distribution

The joint posterior density is proportional to the product of the likelihood function and the prior distributions of all parameters θ: $p(\theta|y) \propto p(y|\theta)p(\theta)$. Defining β_{it} as in equation (2), stacking the time series observations for every firm i into vectors, and substituting the prior densities specified in section II produces the following joint posterior distribution[10]:

$$
\begin{aligned}
p(\theta|y) =& p(\alpha_i, \sigma_\alpha^2, \phi_{0i}, \sigma_{\phi_0}^2, \phi_1, \sigma_{\phi_1}^2, \kappa_2, \delta_0, \sigma_{\delta_0}^2, \delta_1, \Omega_{\delta_1}^{-1}, \sigma_{\epsilon_i}^2|y) \\
&\propto \prod_{i=1}^{N} \left(\sigma_{\epsilon_i}^2\right)^{-\frac{T_i}{2}} \exp\left[-\frac{1}{2\sigma_{\epsilon_i}^2}(r_i - \alpha_i - r_M\beta_i)'(r_i - \alpha_i - r_M\beta_i)\right] \\
&\times \prod_{i=1}^{N} (\sigma_\alpha^2)^{-\frac{1}{2}} \exp\left[-\frac{\alpha_i^2}{2\sigma_\alpha^2}\right] \times \left(\sigma_\alpha^{-2}\right)^{A_\alpha+1} \exp\left[-\sigma_\alpha^{-2}B_\alpha\right] \\
&\times \prod_{i=1}^{N} (\sigma_{\phi_0}^2)^{-\frac{1}{2}} \exp\left[-\frac{(\phi_{0i}-\mu_{\phi_0})^2}{2\sigma_{\phi_0}^2}\right] \times \left(\sigma_{\phi_0}^{-2}\right)^{A_{\phi_0}+1} \exp\left[-\sigma_{\phi_0}^{-2}B_{\phi_0}\right] \\
&\times (\sigma_{\phi_1}^2)^{-\frac{1}{2}} \exp\left[-\frac{\phi_1^2}{2\sigma_{\phi_1}^2}\right] \times \left(\sigma_{\phi_1}^{-2}\right)^{A_{\phi_1}+1} \exp\left[-\sigma_{\phi_1}^{-2}B_{\phi_1}\right] \\
&\times (\sigma_{\delta_0}^2)^{-\frac{1}{2}} \exp\left[-\frac{\delta_0^2}{2\sigma_{\delta_0}^2}\right] \times \left(\sigma_{\delta_0}^{-2}\right)^{A_{\delta_0}+1} \exp\left[-\sigma_{\delta_0}^{-2}B_{\delta_0}\right] \\
&\times |\Omega_{\delta_1}^{-1}|^{\frac{1}{2}} \exp\left[-\frac{1}{2}\delta_1'\Omega_{\delta_1}^{-1}\delta_1\right] \times |\Omega_{\delta_1}^{-1}|^{\frac{\psi_{\delta_1}-(L+LM)-1}{2}} \exp\left[-\frac{1}{2}tr\left(\left[\psi_{\delta_1}S_{\delta_1}\right]\Omega_{\delta_1}^{-1}\right)\right] \\
&\times \prod_{i=1}^{N} \left(\sigma_{\epsilon_i}^{-2}\right)^{A_\epsilon+1} \exp\left[-\sigma_{\epsilon_i}^{-2}B_\epsilon\right].
\end{aligned}
$$

[10]We use the following parametrization of the inverse gamma distribution,

$$p(y|A,B) = \frac{B^A}{\Gamma(A)}\left(\frac{1}{y^{A+1}}\right)^{A+1} \exp\left(\frac{-B}{y}\right),$$

where $\Gamma(A)$ denotes the Gamma function, A is the shape parameter, and B is the scale parameter. For the Wishart distribution we use the parameterization,

$$p(H|R,\nu) \propto \frac{|H|^{(\nu-k-1)/2}}{|R|^{\nu/2}} \exp\left[\frac{1}{2}tr(R^{-1}H)\right],$$

where k denotes the dimension of the matrix H, ν is the degrees of freedom parameter, and R is the scale matrix.

B Conditional Posterior Distributions

In order to implement the Gibbs sampler we need to derive the full conditional posterior densities for each block of parameters. The conditional densities can be derived from the joint posterior density by ignoring all terms that do not depend on the parameters of interest and then treating the parameters considered to be known as constants. We then obtain the conditional density for the parameters of interest by rearranging the remaining terms into the kernel of a known distribution. We partition the parameter vector θ into the following blocks:

$\theta^{(1)}$: MIDAS weight parameter: (κ_2)

$\theta^{(2)}$: Alpha parameters: (α_i)

$\theta^{(3)}$: Fundamental beta parameters: (δ_0, δ_1)

$\theta^{(4)}$: Firm-specific mixed beta parameters: (ϕ_{0i})

$\theta^{(5)}$: Pooled mixed beta parameter: (ϕ_1)

$\theta^{(6)}$: Variance and covariance parameters: $(\sigma^2_\alpha, \sigma^2_{\delta_0}, \Omega^{-1}_{\delta_1}, \sigma^2_{\phi_0}, \sigma^2_{\phi_1}, \sigma^2_{\epsilon_i})$

To generate samples from the conditional posterior of $\theta^{(1)}$ we use the Metropolis-Hastings algorithm. The conditional posteriors for all other blocks have convenient functional forms. Therefore, we use the Gibbs sampler to iteratively draw from the conditional densities of $\theta^{(2)}$, $\theta^{(3)}$, $\theta^{(4)}$, $\theta^{(5)}$, $\theta^{(6)}$. To simplify notation we rewrite the model in matrix form as

$$r_i = \alpha_i \iota_{T_i} + \phi_{0i} r_M b_i + \phi_1 r_M V_M b_i + r_M W_i \delta - \phi_{0i} r_M W_i \delta - \phi_1 r_M V_M W_i \delta + \epsilon_i, \qquad (16)$$

where r_i is an $T \times 1$ vector of excess returns, r_M an $T \times T$ diagonal matrix of excess market returns, V_M an $T \times T$ diagonal matrix of lagged market volatility, b_i an $T \times 1$ vector of realized betas, and ϵ_i an $T \times 1$ vector of idiosyncratic shocks. Since the δ_0 and δ_1 parameters in block $\theta^{(3)}$ have independent priors, we have simplified the notation further by rewriting $\delta_0 \iota_T + ZBC_i \delta_1$ as $W_i \delta$,where W_i is the $T \times (1 + L + LM)$ matrix of the constant term and conditioning variables. We combine the corresponding precisions $\sigma^{-2}_{\delta_0}$ and $\Omega^{-1}_{\delta_1}$ into the matrix Ω^{-1}_{δ}.

B.1 Metropolis-Hastings algorithm to draw κ_2

Since we implement a change of variable, $\kappa_2 = 1 + 25\tilde{\kappa}_2$, we need to draw values for $\tilde{\kappa}_2$. Because the conditional posterior density for $\tilde{\kappa}_2$ does not take a standard form, we cannot use the Gibbs sampler. Instead, we employ the Metropolis-Hastings algorithm, which is a general accept-reject algorithm. In fact, Gelman, Carlin, Stern, and Rubin (2004) show that the Gibbs sampler is a special case of Metropolis-Hastings in which proposed parameter values are accepted with probability one. The M-H algorithm proceeds as follows.

First, a candidate value $\tilde{\kappa}_2^*$ is drawn from a proposal density $q(\tilde{\kappa}_2)$. We apply the Independence Chain M-H algorithm, in which the proposal density is independent across draws. We choose a Beta(1,3) proposal density, which has a mean of 0.25 and standard deviation equal to 0.19. Because the proposal density is not identical to the posterior density, the M-H algorithm does not accept all proposal draws. When a proposal is rejected the parameter value is set equal to the current value. Draws are accepted according to the following probability

$$\pi(\tilde{\kappa}_2^{(g-1)}, \tilde{\kappa}_2^*) = min\left\{1, \frac{p(\tilde{\kappa}_2^*|y)q(\tilde{\kappa}_2^{(g-1)})}{p(\tilde{\kappa}_2^{(g-1)}|y)q(\tilde{\kappa}_2^*)}\right\}. \tag{17}$$

This approach ensures that candidate draws with a high posterior density have a higher probability of being accepted than draws with a low posterior density. Repeating this procedure produces the required sequence of draws from the posterior distribution.

B.2 Conditional posterior α_i

Using Bayes' theorem, we can write:

$$\begin{aligned}
p\left(\alpha_i|y\right) &\propto p\left(y|\alpha_i\right)p\left(\alpha_i\right) \\
&\propto \exp\left[-\frac{1}{2}Q^*\right], \\
\text{where} \quad Q^* &= \left(r_i - \alpha_i - r_M\beta_i\right)'\,\Omega_{\epsilon_i}^{-1}(r_i - \alpha_i - r_M\beta_i) + \frac{\alpha_i^2}{\sigma_\alpha^2} \\
&= \left(X_{\alpha_i} - \alpha_i\iota_{T_i}\right)'\,\Omega_{\epsilon_i}^{-1}\left(X_{\alpha_i} - \alpha_i\iota_{T_i}\right) + \frac{\alpha_i^2}{\sigma_\alpha^2} \\
&= Q_1^* + Q_2^*, \\
\text{with} \quad Q_1^* &= \frac{\left(\alpha_i - \bar{\alpha}_i\right)^2}{\bar{\sigma}_{\alpha_i}^2}, \\
\text{and} \quad Q_2^* &= X_{\alpha_i}'\Omega_{\epsilon_i}^{-1}X_{\alpha_i} - \frac{\bar{\alpha}_i^2}{\bar{\sigma}_{\alpha_i}^2},
\end{aligned}$$

$$\text{and where} \quad X_{\alpha_i} = r_i - \phi_{0i}r_Mb_i - \phi_1 r_M V_M b_i - r_M W_i\delta + \phi_{0i}r_M W_i\delta + \phi_1 r_M V_M W_i\delta.$$

In the derivation of $p\left(\alpha_i|\theta^{-(\alpha_i)}, y\right)$ all parameters in Q_2^* are known, so we can treat Q_2^* as a constant. Thus, $p\left(\alpha_i|\theta^{-(\alpha_i)}, y\right)$ is proportional to $\exp[-\frac{1}{2}Q_1^*]$, which is the kernel of a normal density. Therefore,

$$\begin{aligned}
\alpha_i|\theta^{-(\alpha_i)}, y &\sim N\left(\bar{\alpha}_i, \bar{\sigma}^2_{\alpha_i}\right), \\
\text{with} \quad \bar{\alpha}_i &= \left[\iota'_{T_i}\Omega^{-1}_{\epsilon_i}\iota_{T_i} + \frac{1}{\sigma^2_\alpha}\right]^{-1}\left[\iota'_{T_i}\Omega^{-1}_{\epsilon_i}X_{\alpha_i}\right], \\
\text{and} \quad \bar{\sigma}^2_{\alpha_i} &= \left[\iota'_{T_i}\Omega^{-1}_{\epsilon_i}\iota_{T_i} + \frac{1}{\sigma^2_\alpha}\right]^{-1}.
\end{aligned}$$

B.3 Conditional posterior δ

$$\begin{aligned}
&\delta|\theta^{-(\delta)}, y \sim N\left(\bar{\delta}, \bar{\Omega}_\delta\right), \\
&\text{with} \quad \bar{\delta} = \left[\sum_{i=1}^N \left(\left(1-\phi_{0i}\right) r_M W_i - \phi_1 r_M V_M W_i\right)' \Omega^{-1}_{\epsilon_i} \left(\left(1-\phi_{0i}\right) r_M W_i - \phi_1 r_M V_M W_i\right) + \Omega^{-1}_\delta\right]^{-1} \\
&\qquad\quad \times \left[\sum_{i=1}^N X'_{\delta i}\Omega^{-1}_{\epsilon_i}\left(\left(1-\phi_{0i}\right) r_M W_i - \phi_1 r_M V_M W_i\right)\right], \\
&\text{and} \quad \bar{\Omega}_\delta = \left[\sum_{i=1}^N \left(\left(1-\phi_{0i}\right) r_M W_i - \phi_1 r_M V_M W_i\right)' \Omega^{-1}_{\epsilon_i} \left(\left(1-\phi_{0i}\right) r_M W_i - \phi_1 r_M V_M W_i\right) + \Omega^{-1}_\delta\right]^{-1},
\end{aligned}$$

and where $\quad X_{\delta i} = r_i - \alpha_i\iota_{T_i} - \phi_{0i} r_M b_i - \phi_1 r_M V_M b_i$.

B.4 Conditional posterior ϕ_{0i}

$$\begin{aligned}
&\phi_{0i}|\theta^{-(\phi_{0i})}, y \sim N\left(\bar{\phi}_{0i}, \bar{\sigma}^2_{\phi_{0i}}\right), \\
&\text{with} \quad \bar{\phi}_{0i} = \left[(b_i - W_i\delta)' r_M \Omega^{-1}_{\epsilon_i} r_M (b_i - W_i\delta) + \frac{1}{\sigma^2_{\phi_0}}\right]^{-1}\left[(b_i - W_i\delta)' r_M \Omega^{-1}_{\epsilon_i} X_{\phi_{0i}} + \frac{\mu_{\phi_0}}{\sigma^2_{\phi_0}}\right], \\
&\text{and} \quad \bar{\sigma}^2_{\phi_{0i}} = \left[(b_i - W_i\delta)' r_M \Omega^{-1}_{\epsilon_i} r_M (b_i - W_i\delta) + \frac{1}{\sigma^2_{\phi_0}}\right]^{-1},
\end{aligned}$$

and where $\quad X_{\phi_{0i}} = r_i - \alpha_i\iota_{T_i} - r_M W_i\delta - \phi_1 r_M V_M (b_i - W_i\delta)$.

B.5 Conditional posterior ϕ_1

$$
\begin{aligned}
&\phi_1|\theta^{-(\phi_1)}, y \sim N\left(\bar{\phi}_1, \bar{\sigma}^2_{\phi_1}\right),\\
&\quad \text{with} \quad \bar{\phi}_1 = \left[\sum_{i=1}^{N} (b_i - W_i\delta)' \, r_M V_M \Omega_{\epsilon_i}^{-1} r_M V_M \, (b_i - W_i\delta) + \frac{1}{\sigma^2_{\phi_1}}\right]^{-1} \left[\sum_{i=1}^{N} X'_{\phi_{1i}} \Omega_{\epsilon_i}^{-1} r_M V_M \, (b_i - W_i\delta)\right],\\
&\quad \text{and} \quad \bar{\sigma}^2_{\phi_1} = \left[\sum_{i=1}^{N} (b_i - W_i\delta)' \, r_M V_M \Omega_{\epsilon_i}^{-1} r_M V_M \, (b_i - W_i\delta) + \frac{1}{\sigma^2_{\phi_1}}\right]^{-1},
\end{aligned}
$$

and where $\quad X_{\phi_{1i}} = r_i - \alpha_i \iota_{T_i} - \phi_{0i} r_M (b_i - W_i\delta) - r_M W_i \delta.$

B.6 Conditional posteriors $\sigma^2_\alpha, \Omega_\delta^{-1}, \sigma^2_{\phi_0}, \sigma^2_{\phi_1}, \sigma^2_{\epsilon_i}$

$$
\sigma^2_\alpha|\theta^{-(6)}, y \sim IG\left(\frac{N + 2A_\alpha}{2}, \frac{\sum_{i=1}^N \alpha_i^2 + 2B_\alpha}{2}\right),
$$

$$
\Omega_\delta^{-1}|\theta^{-(6)}, y \sim Wish\left(\left[\delta\delta' + (\psi_\delta S_\delta)\right]^{-1}, \psi_\delta + 1\right),
$$

$$
\sigma^2_{\phi_0}|\theta^{-(6)}, y \sim IG\left(\frac{N + 2A_{\phi_0}}{2}, \frac{\sum_{i=1}^N (\phi_{0i} - \mu_{\phi_0})^2 + 2B_{\phi_0}}{2}\right),
$$

$$
\sigma^2_{\phi_1}|\theta^{-(6)}, y \sim IG\left(\frac{1 + 2A_{\phi_1}}{2}, \frac{\phi_1^2 + 2B_{\phi_1}}{2}\right),
$$

$$
\sigma^2_{\epsilon_i}|\theta^{-(6)}, y \sim IG\left(\frac{T_i + 2A_\epsilon}{2}, \frac{(r_i - \alpha_i - r_M\beta_i)'(r_i - \alpha_i - r_M\beta_i) + 2B_\epsilon}{2}\right).
$$

B Cross-Sectional Asset Pricing Tests

In this appendix we show how we account for measurement error in betas in the cross-sectional asset pricing tests. We consider the cross-sectional regression model described in section V,

$$r_{it} = \lambda_{0t} + \lambda_{1t}\beta_{it-1} + \lambda_{2t}' x_{it-1} + \upsilon_{it}. \tag{18}$$

Conditional on β_{it-1} we can estimate the parameters λ_t, either Bayesian or classical, using the Fama-MacBeth approach. Let $W_{it} = (1 \ \beta_{it} \ x_{it}')'$ and let $\hat{\lambda}_t$ denote the cross-sectional OLS estimator of λ_t. The Fama-MacBeth estimator of the average λ is

$$\hat{\lambda} = \frac{1}{T}\sum_t \hat{\lambda}_t = \frac{1}{T}(W_{t-1}' W_{t-1})^{-1} W_{t-1}' r_t, \tag{19}$$

with covariance matrix

$$S \equiv Var\left(\sqrt{T}(\hat{\lambda} - \lambda)\right) = \frac{1}{T}\sum_t (\hat{\lambda}_t - \hat{\lambda})(\hat{\lambda}_t - \hat{\lambda})'. \tag{20}$$

The Fama-MacBeth procedure overstates the precision of parameter estimates in the cross-sectional regressions because it ignores estimation errors in the β_{it}.

As explained in section II, the Gibbs sampler has produced a series of L draws from the posterior density $p(\beta|y)$, where β contains the entire collection of all β_{it} and y is a shorthand for all data used in estimating the betas. Given the $\beta_{it}^{(\ell)}$ from the ℓ^{th} iteration of the Gibbs sampler we can form the regressor matrix $W_t^{(\ell)}$ and using $W_t^{(\ell)}$ construct the conditional mean $\hat{\lambda}^{(\ell)}$ and covariance matrix $S^{(\ell)}$. From these we form the unconditional estimators

$$\bar{\lambda} = \frac{1}{L}\sum_\ell \hat{\lambda}^{(\ell)} \tag{21}$$

and

$$\bar{S} = \frac{1}{L}\sum_\ell S^{(\ell)} + \frac{1}{L}\sum_\ell (\hat{\lambda}^{(\ell)} - \bar{\lambda})(\hat{\lambda}^{(\ell)} - \bar{\lambda})'. \tag{22}$$

The estimates $\bar{\lambda}$ and $\bar{S}$ can be interpreted as the posterior mean and variance of the average λ_t if we assume that the prior on λ_t is uniform and λ does not affect the posterior density of β_{it}.

References

Andersen, T., T. Bollerslev, F. Diebold, and J. Wu, 2005, A framework for exploring the macroeconomic determinants of systematic risk, *American Economic Review* 95, 398–404.

Ang, A., and J. Chen, 2007, CAPM over the long run: 1926-2001, *Journal of Empirical Finance* 14, 1–40.

Ang, A., and D. Kristensen, 2009, Testing conditional factor models, Working paper, Columbia University.

Ang, A., J. Liu, and K. Schwarz, 2008, Using stocks or portfolios in tests of factor models, Working paper, Columbia University.

Avramov, D., and T. Chordia, 2006a, Asset pricing models and financial market anomalies, *Review of Financial Studies* 19, 1001–1040.

Avramov, D., and T. Chordia, 2006b, Predicting stock returns, *Journal of Financial Economics* 82, 387–415.

Bauwens, L., M. Lubrano, and J. Richard, 1999, *Bayesian Inference in Dynamic Econometric Models* (Oxford University Press: Oxford).

Berk, J., R. Green, and V. Naik, 1999, Optimal investment, growth options, and security returns, *Journal of Finance* 54, 1553–1607.

Bollerslev, T., and B. Zhang, 2003, Measuring and modeling systematic risk in factor pricing models using high-frequency data, *Journal of Empirical Finance* 10, 533–558.

Brandt, M., P. Santa-Clara, and R. Valkanov, 2009, Parameteric portfolio policies: Exploiting characteristics in the cross section of equity returns, *Review of Financial Studies*, Forthcoming.

Brennan, M., T. Chordia, and A. Subrahmanyam, 1998, Alternative factor specifications, security characteristics, and the cross-section of expected stock returns, *Journal of Financial Economics* 49, 345–373.

Campbell, J., M. Lettau, B. Malkiel, and Y. Xu, 2001, Have individual stocks become more volatile? an empirical exploration of idiosyncratic risk, *Journal of Finance* 56, 1–43.

Campbell, J., A. Lo, and A. MacKinlay, 1997, *The Econometrics of Financial Markets* (Princeton University Press: Princeton).

Chan, L., J. Karceski, and J. Lakonishok, 1999, On portfolio optimization: Forecasting covariances and choosing the risk model, *Review of Financial Studies* 12, 937–974.

Chib, S., and E. Greenberg, 1995, Understanding the Metropolis-Hastings algorithm, *The American Statistician* 49, 327–335.

Christoffersen, P., K. Jacobs, and G. Vainberg, 2007, Forward-looking betas, Working Paper, McGill University.

Cremers, M., 2006, Multifactor efficiency and Bayesian inference, *Journal of Business* 79, 2951–2998.

DeMiguel, V., L. Garlappi, and R. Uppal, 2007, Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?, *Review of Financial Studies*, Forthcoming.

Engle, R., and R. Colacito, 2006, Testing and valuing dynamic correlations for asset allocation, *Journal of Business and Economic Statistics* 24, 238–253.

Engle, R., and J. Rangel, 2008, The spline-GARCH model for low-frequency volatility and its global macroeconomic causes, *Review of Financial Studies* 21, 1187–1222.

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

Fama, E., and K. French, 1992, The cross-section of expected stock returns, *Journal of Finance* 47, 427–465.

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

Fama, E., and K. French, 1996, Multifactor explanations of asset pricing anomalies, *Journal of Finance* 51, 55–84.

Fama, E., and K. French, 1997, Industry costs of equity, *Journal of Financial Economics* 43, 153–193.

Fama, E., and K. French, 2008, Dissecting anomalies, *Journal of Finance* 63, 1653–1678.

Fama, E., and J. MacBeth, 1973, Risk, return and equilibrium: Empirical tests, *Journal of Political Economy* 71, 607–636.

Ferson, W., and C. Harvey, 1999, Conditioning variables and the cross section of stock returns, *Journal of Finance* 54, 1325–1360.

Gelman, A., J. Carlin, H. Stern, and D. Rubin, 2004, *Bayesian Data Analysis* (Chapman and Hall/CRC: Boca Raton, Florida) 2nd edn.

Geweke, J., 2005, *Contemporary Bayesian Econometrics and Statistics* (Wiley: New York).

Ghysels, E., 1998, On stable factor structures in the pricing of risk: Do time-varying betas help or hurt?, *Journal of Finance* 53, 549–573.

Ghysels, E., P. Santa-Clara, and R. Valkanov, 2005, There is a risk-return trade-off after all, *Journal of Financial Economics* 76, 509–548.

Goetzmann, W., A. Watanabe, and M. Watanabe, 2008, Investor expectations, business conditions, and the pricing of beta-instability risk, Working Paper, Yale School of Management.

Gomes, J., L. Kogan, and L. Zhang, 2003, Equilibrium cross-section of returns, *Journal of Political Economy* 111, 693–732.

Harvey, C., and G. Zhou, 1990, Bayesian inference in asset pricing tests, *Journal of Financial Economics* 26, 221–254.

Hoberg, G., and I. Welch, 2007, Aged and recent market betas in securities prices, Working Paper, University of Maryland and Brown University.

Jagannathan, R., and T. Ma, 2003, Risk reduction in large portfolios: Why imposing the wrong constraints helps, *Journal of Finance* 58, 1651–1683.

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

Jostova, G., and A. Philipov, 2005, Bayesian analysis of stochastic betas, *Journal of Financial and Quantitative Analysis* 40.

Kandel, S., R. McCulloch, and R. Stambaugh, 1995, Bayesian inference and portfolio efficiency, *Review of Financial Studies* 8, 1–53.

Ledoit, O., and M. Wolf, 2003, Improved estimation of the covariance matrix of stock returns with an application to portfolios selection, *Journal of Empirical Finance* 10, 603–621.

Lettau, M., and S. Ludvigson, 2001, Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying, *Journal of Political Economy* 109, 1238–1287.

Lewellen, J., 1999, The time-series relations among expected return, risk, and book-to-market, *Journal of Financial Economics* 54, 5–43.

Lewellen, J., and S. Nagel, 2006, The conditional CAPM does not explain asset-pricing anomalies, *Journal of Financial Economics* 82, 289–314.

Lewellen, J., S. Nagel, and J. Shanken, 2008, A skeptical appraisal of asset-pricing tests, *Journal of Financial Economics*, Forthcoming.

Lo, A., and A. MacKinlay, 1990, Data-snooping biases in tests of financial asset pricing models, *Review of Financial Studies* 3, 431–467.

Markowitz, H., 1952, Portfolio selection, *Journal of Finance* 7, 77–91.

Santos, T., and P. Veronesi, 2004, Conditional betas, Working paper, Columbia University and University of Chicago.

Scholes, M., and J. Williams, 1977, Estimating betas from nonsynchronous data, *Journal of Financial Economics* 5, 309–327.

Shanken, J., 1987, A Bayesian approach to testing portfolio efficiency, *Journal of Financial Economics* 19, 195–215.

Shanken, J., 1990, Intertemporal asset pricing: An empirical investigation, *Journal of Econometrics* 45, 99–120.

Zhang, L., 2005, The value premium, *Journal of Finance* 60, 67–103.

Table I: **Summary Statistics for Firm Characteristics and Macroeconomic Variables**

This table presents descriptive statistics for stock returns, firm characteristics, and macroeconomic variables for 510 months from July 1964 through December 2006. Panel A reports the mean, median, standard deviation and 5th, 25th, 75th, and 95th percentile values of firm characteristics for a total of 5,017 stocks over the full sample period and an average of 1,815 stocks per month. We include a stock in the sample for a given month t if it satisfies the following criteria. First, its return in the current month, t, and over the past 36 months has to be available. Second, data should be available in month *t-1* for size as measured by market capitalization and for the book-to-market ratio. We exclude firms with negative book-to-market equity. XRET is the return in excess of the risk-free rate, MV represents the market capitalization in billions of dollars, and BM is the book-to-market ratio, for which values smaller than the 0.5th percentile and values greater than the 99.5th percentile are set equal to the 0.5th percentile and 99.5th percentile values, respectively. MOM is the cumulative return over the twelve months prior to the current month. Panel B shows the mean, median, standard deviation and 5th, 25th, 75th, and 95th percentile values of macroeconomic variables. DEF is the default spread, defined as the yield differential between bonds rated BAA by Moody's and bonds with a Moody's rating of AAA. DY is the dividend yield on the value-weighted CRSP index. The dividend yield is calculated as the sum of the dividends paid on the index in the previous year divided by the current level of the index. TBILL is the one-month Treasury bill rate. TERM is the term spread, defined as the yield difference between ten-year and one-year Treasury bonds.

	Mean	Std. dev.	5th	25th	Median	75th	95th
Panel A: Firm characteristics							
XRET (%)	0.69	12.31	-17.48	-5.99	-0.16	6.42	21.29
MV ($ billions)	1.59	5.60	0.01	0.03	0.16	0.81	6.67
BM	0.96	0.82	0.18	0.44	0.75	1.22	2.45
MOM (%)	14.65	49.29	-49.12	-14.24	8.60	34.46	96.30
Panel B: Macroeconomic variables							
DEF (%)	1.02	0.43	0.55	0.73	0.90	1.21	1.92
DY (%)	3.01	1.10	1.30	2.02	2.96	3.77	4.84
TBILL (%)	5.69	2.70	1.56	4.08	5.16	6.96	10.57
TERM (%)	0.85	1.14	-1.14	0.08	0.78	1.69	2.83

Table II: **Posterior Distribution of Fundamental Beta Parameters**

This table reports the Bayesian posterior distribution of the determinants of the fundamental beta, which is parameterized as a linear function of firm characteristics and business cycle variables,

$$\beta^*_{it} = \delta_0 + \delta_1'[Z_{it} \otimes BC_t],$$

where Z_{it} is a vector that contains L firm characteristics and BC_t is a vector that contains a constant and M business cycle variables. MV is the log of firm size, BM is the log of the book-to-market ratio, and MOM is the cumulative return over the twelve months prior to the current month. These firm characteristics are expressed as deviations from their cross-sectional mean in every period. DEF is the default spread, DY is the dividend yield, TBILL is the one-month Treasury bill rate, and TERM is the term spread. The table presents the mean, median, standard deviation and 5th, 25th, 75th, and 95th percentile values of the posterior distribution of the delta parameters, based on 5,000 iterations of the Gibbs sampler and a burn-in period of 1,000 iterations. All δ_1 parameters are multiplied by 100.

	Mean	Std. dev.	5th	25th	Median	75th	95th
Constant (δ_0)	1.01	0.01	0.98	0.99	1.00	1.01	1.02
MV	-3.62	0.68	-4.77	-4.05	-3.66	-3.20	-2.47
BM	-5.30	1.32	-7.38	-6.18	-5.44	-4.34	-3.10
MOM	-10.02	2.99	-14.97	-12.05	-9.99	-8.18	-5.28
MV*TBILL	0.65	0.29	0.16	0.45	0.65	0.84	1.14
MV*TERM	1.57	0.28	1.08	1.40	1.58	1.74	2.03
MV*DEF	0.55	0.52	-0.33	0.20	0.53	0.93	1.38
MV*DY	-1.41	0.08	-1.55	-1.47	-1.40	-1.35	-1.27
BM*TBILL	-3.11	0.58	-3.98	-3.52	-3.09	-2.74	-2.20
BM*TERM	0.35	0.73	-0.84	-0.13	0.39	0.83	1.49
BM*DEF	13.15	0.95	11.62	12.53	13.18	13.74	14.64
BM*DY	-0.05	0.14	-0.30	-0.14	-0.04	0.05	0.17
MOM*TBILL	3.05	1.02	1.43	2.29	3.06	3.81	4.67
MOM*TERM	-1.11	0.89	-2.55	-1.67	-1.08	-0.54	0.34
MOM*DEF	-28.91	1.70	-31.65	-30.02	-29.00	-27.76	-26.02
MOM*DY	8.18	0.23	7.81	8.03	8.18	8.32	8.56

Table III: **Beta Summary Statistics**

This table reports summary statistics on the dispersion of betas and the correlation between mixed, fundamental, and realized betas. The left panel reports the properties of the cross-section of betas based on the time series average of the cross-sectional covariances

$$S_{cross,t} = \frac{1}{N_t} \sum_i \left(\bar{\beta}_{it}^{(j)} - \bar{\bar{\beta}}_t^{(j)} \right) \left(\bar{\beta}_{it}^{(k)} - \bar{\bar{\beta}}_t^{(k)} \right),$$

where the indices j and k refer to the model (Mixed, Realized, Fundamental), betas are evaluated at their posterior means $\bar{\beta}_{it}$, and where $\bar{\bar{\beta}}_t$ is the average beta at time t. The right panel considers the cross-sectional average of the time series covariances

$$S_{time,i} = \frac{1}{T_i} \sum_t \left(\bar{\beta}_{it}^{(j)} - \bar{\bar{\beta}}_i^{(j)} \right) \left(\bar{\beta}_{it}^{(k)} - \bar{\bar{\beta}}_i^{(k)} \right),$$

where $\bar{\bar{\beta}}_i$ is the average beta of firm i. The diagonal elements in both panels have been transformed into standard deviations. The off-diagonal elements of both the cross-sectional and time series covariance matrices have been rescaled to correlations (*italics*).

	Cross-sectional (S_{cross})			Time Series (S_{time})		
	Mixed	Realized	Fund'l	Mixed	Realized	Fund'l
Mixed beta	0.33			0.24		
Realized beta	*0.95*	0.61		*0.94*	0.46	
Fundamental beta	*0.94*	*0.83*	0.10	*0.93*	*0.82*	0.09

Table IV: **Heterogeneity within 25 Size-B/M Portfolios**

This table presents the time series average of characteristics of 25 size-B/M sorted portfolios and the cross-sectional spread in these characteristics. The portfolios are constructed annually by sorting stocks independently into size and B/M quintiles at the end of June. The 25 portfolios are then formed as the intersections of these five size and B/M quintiles. Subsequently, we calculate for every portfolio j at every time t the equally weighted cross-sectional average of the posterior means of the firm-specific alphas, betas, and phis, and of the excess returns of the stocks in the portfolio. We also calculate for every portfolio j at every time t the cross-sectional standard deviation of the posterior means of the firm-specific alphas, betas, and phis, and of the excess returns of the stocks in the portfolio. The table shows for every portfolio the time series means of these cross-sectional averages and standard deviations.

Size	Book-to-Market equity (B/M) quintiles									
quintiles	Low	2	3	4	High	Low	2	3	4	High
	Average Return					Return Variation				
Small	0.18	0.54	0.67	0.97	0.93	15.56	13.84	13.11	12.50	13.83
2	0.72	0.83	0.80	0.83	0.85	12.98	11.16	10.40	10.13	11.34
3	0.67	0.69	0.76	0.95	0.87	11.03	9.83	9.04	8.89	9.82
4	0.65	0.62	0.72	0.84	0.94	9.26	8.26	7.83	7.51	8.33
Big	0.48	0.46	0.59	0.72	0.84	7.27	6.93	6.57	6.35	6.56
	Average Alpha					Alpha Variation				
Small	-0.40	-0.06	0.03	0.03	-0.24	1.50	1.27	1.21	1.12	1.28
2	0.21	0.32	0.29	0.23	-0.04	1.32	1.07	0.95	0.95	1.13
3	0.43	0.41	0.37	0.31	0.10	1.09	0.88	0.77	0.76	0.91
4	0.48	0.41	0.32	0.23	0.14	0.79	0.61	0.56	0.53	0.59
Big	0.45	0.35	0.29	0.21	0.12	0.49	0.40	0.37	0.36	0.40
	Average Beta					Beta Variation				
Small	1.21	1.15	1.10	1.04	1.03	0.35	0.31	0.31	0.31	0.32
2	1.24	1.14	1.08	1.04	1.06	0.32	0.30	0.29	0.29	0.29
3	1.23	1.11	1.03	1.00	1.06	0.33	0.30	0.29	0.29	0.29
4	1.13	1.05	0.99	0.95	1.00	0.31	0.29	0.29	0.29	0.29
Big	1.04	1.00	0.95	0.91	0.94	0.27	0.25	0.26	0.25	0.24
	Average Phi					Phi Variation				
Small	0.45	0.45	0.46	0.47	0.47	0.09	0.10	0.11	0.11	0.11
2	0.47	0.48	0.49	0.49	0.49	0.10	0.11	0.12	0.12	0.11
3	0.49	0.49	0.50	0.52	0.53	0.10	0.11	0.13	0.14	0.12
4	0.50	0.51	0.53	0.55	0.55	0.11	0.12	0.13	0.13	0.11
Big	0.51	0.53	0.54	0.54	0.56	0.13	0.12	0.11	0.11	0.10

Table V: **Cross-Sectional Asset Pricing Tests**

This table reports Fama-MacBeth coefficient estimates for 5,017 NYSE-AMEX stocks (columns 1-3 and 7-9) and for 25 size-B/M portfolios (columns 4-6). The sample period is August 1964 through December 2006 (509 months). In the first stage betas are estimated using the Bayesian panel data approach outlined in section I. These betas are then used as independent variables in second stage cross-sectional regressions of returns on betas,

$$r_{it} = \lambda_{0t} + \lambda_{1t}\beta_{it-1} + \lambda_{2t}' x_{it-1} + \upsilon_{it}.$$

In columns 1-6 no control variables are included ($x_{it-1} = 0$). In columns 7-9 the vector x_{it-1} includes the firm characteristics size, book-to-market, and momentum. We use the entire posterior distribution of beta in the estimation to account for measurement error in beta. The table shows the estimated intercept (λ_0) and risk premium (λ_1) for four different beta specifications. Standard errors are in parentheses and corresponding t-stats are reported in brackets.

	Stocks - No controls			Portfolios - No controls			Stocks - With controls		
	λ_0	λ_1	Adj. R^2 (%)	λ_0	λ_1	Adj. R^2 (%)	λ_0	λ_1	Adj. R^2 (%)
Panel A: Mixed Beta									
Estimate	-0.04	0.56	4.21	0.61	0.08	18.30	0.04	0.63	6.70
SE	(0.17)	(0.26)		(0.42)	(0.48)		(0.17)	(0.30)	
t-stat	[-0.24]	[2.15]		[1.45]	[0.16]		[0.22]	[2.09]	
Panel B: Realized Beta									
Estimate	0.35	0.35	3.47	0.61	0.13	16.66	0.31	0.41	6.26
SE	(0.13)	(0.16)		(0.29)	(0.29)		(0.12)	(0.16)	
t-stat	[2.73]	[2.24]		[2.07]	[0.44]		[2.51]	[2.48]	
Panel C: Fundamental Beta									
Estimate	0.64	0.67	2.05	1.14	-0.42	16.24	0.72	0.72	4.55
SE	(0.30)	(0.34)		(0.76)	(0.78)		(0.24)	(0.40)	
t-stat	[2.13]	[1.97]		[1.50]	[-0.54]		[3.04]	[1.80]	
Panel D: Static Beta									
Estimate	0.63	0.09	0.65	1.22	-0.45	9.39	0.87	-0.15	3.78
SE	(0.28)	(0.20)		(0.47)	(0.49)		(0.29)	(0.22)	
t-stat	[2.25]	[0.45]		[2.60]	[-0.92]		[2.98]	[-0.70]	

Table VI: **Risk and Return Characteristics of Global Minimum Variance Portfolios**

This table reports the out-of-sample performance of global minimum variance portfolios that are formed at the end of each month from December 1985 through December 2006 out of a universe of NYSE-AMEX stocks. The optimization procedure uses forecasts of the covariance matrix of returns produced by different models. Panel A reports the out-of-sample performance of these minimum variance portfolios when all stocks in the sample are used in the optimization and without any constraints imposed on the weights. Panel B reports results for this unconstrained optimization when a random sample of 250 stocks is used to construct the portfolios. Panel C reports results for this reduced investment universe when a nonnegativity constraint is imposed on the portfolio weights (no short-selling). The mean return and standard deviation are those of excess returns. Mean return, standard deviation, and Sharpe ratio are annualized. Short interest is in percentages.

Model	Mean	Std. Dev.	Sharpe Ratio	Short Interest
Panel A: Unconstrained (all stocks)				
Equally weighted ($1/N$)	8.57	15.40	0.56	0.00
Static beta (TS model)	5.24	8.50	0.61	-64.64
Mixed beta (Panel model)	5.60	8.12	0.69	0.00
Panel B: Unconstrained (250 stocks)				
Sample covariance matrix	4.30	15.19	0.28	-144.79
Equally weighted ($1/N$)	9.46	15.93	0.59	0.00
Static beta (TS model)	7.81	13.32	0.59	-60.35
Mixed beta (Panel model)	5.70	8.41	0.68	0.00
Panel C: Nonnegativity Constrained (250 stocks)				
Sample covariance matrix	3.02	11.74	0.26	0.00
Equally weighted ($1/N$)	9.46	15.93	0.59	0.00
Static beta (TS model)	7.55	15.60	0.48	0.00
Mixed beta (Panel model)	5.70	8.41	0.68	0.00

Figure 1: **Cross-Sectional Distribution of Phi**

This figure shows the cross-sectional distribution of the time series average of the parameter ϕ_i, which measures the proportion of beta explained by past realized beta,

$$\beta_{it} = \phi_{it} b_{it} + (1 - \phi_{it})\beta^*_{it},$$

where b_{it} is the realized beta of firm i, β^*_{it} is the fundamental beta, and where ϕ_{it} is given by

$$\phi_{it} = \phi_{0i} + \phi_1 V_{Mt},$$

where V_{Mt} is the realized market variance. We first calculate ϕ_{it} based on equation (3) for each draw of the Gibbs sampler. We then calculate for each firm the time series average ϕ_i and its posterior mean. This figure shows the cross-sectional distribution of these posterior means.

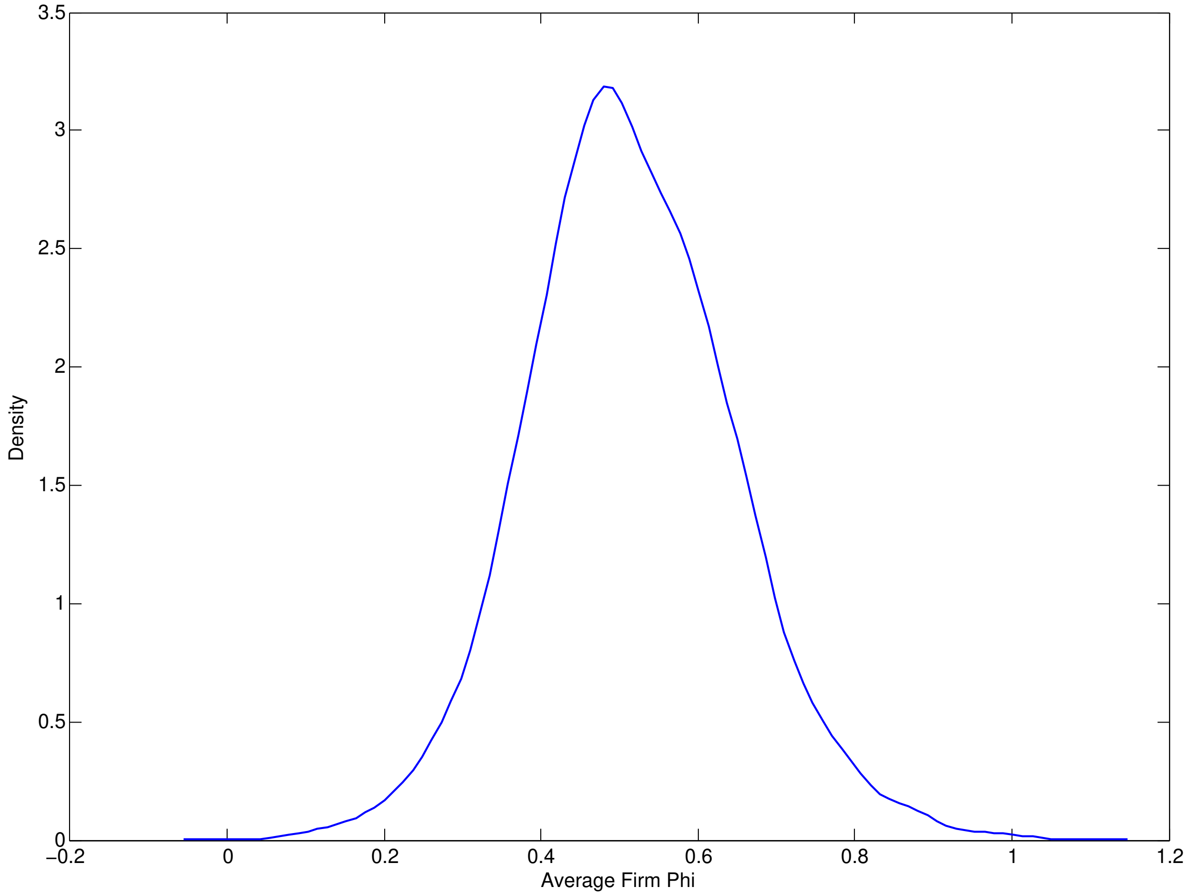


Figure 2: **Evolution of Phi through Time**

This figure plots the evolution through time of the cross-sectional average of ϕ_{it}. We first calculate at each iteration of the Gibbs sampler ϕ_{it} based on equation (3). We then compute its posterior mean and the cross-sectional average of these posterior means in each month from July 1964 through December 2006. Shaded areas indicate NBER recession periods.

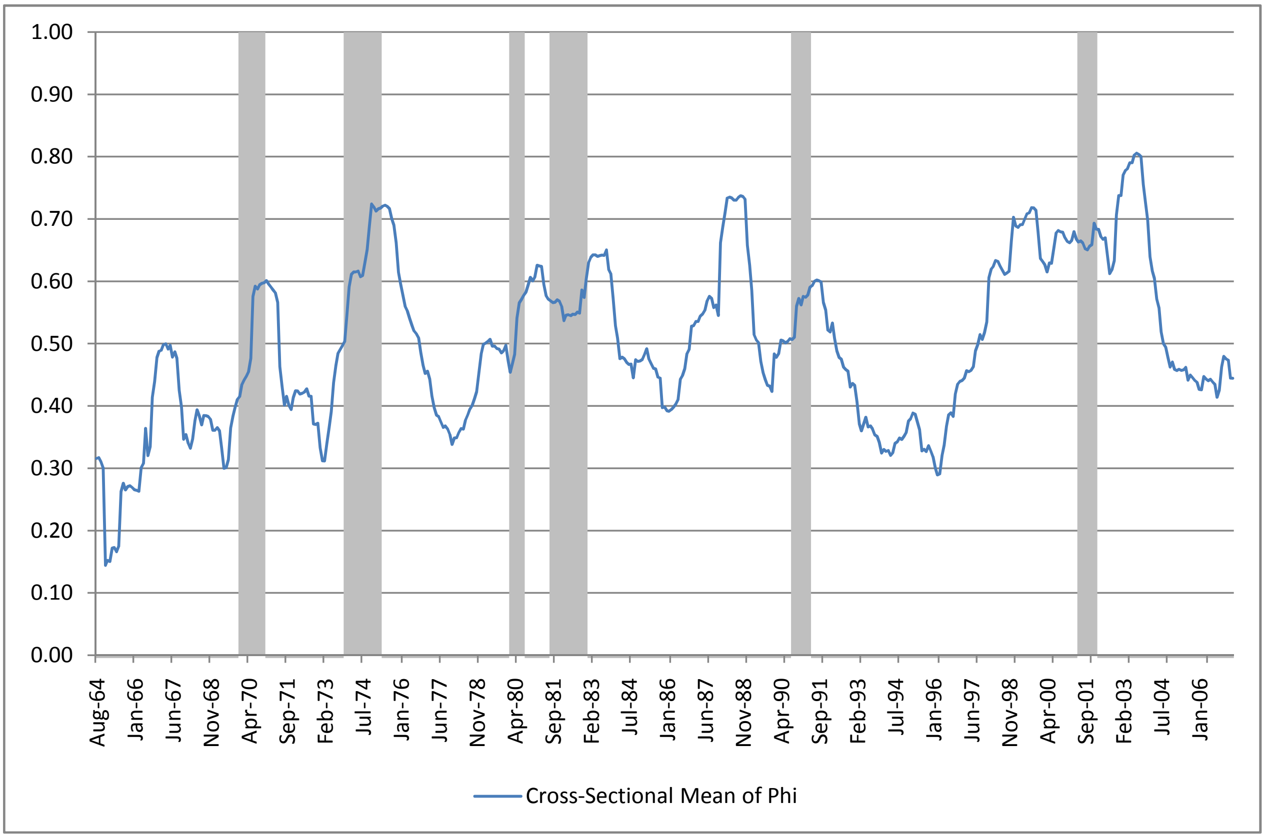


Figure 3: **Optimal versus Equal Weighting Scheme for Estimating Realized Beta**

This figure compares the equal weights in the traditional rolling window estimator of realized betas to the weights implied by the MIDAS weighting function in equation (5) for the realized beta estimator in equation (4). We set the maximum window length equal to 250 trading days.

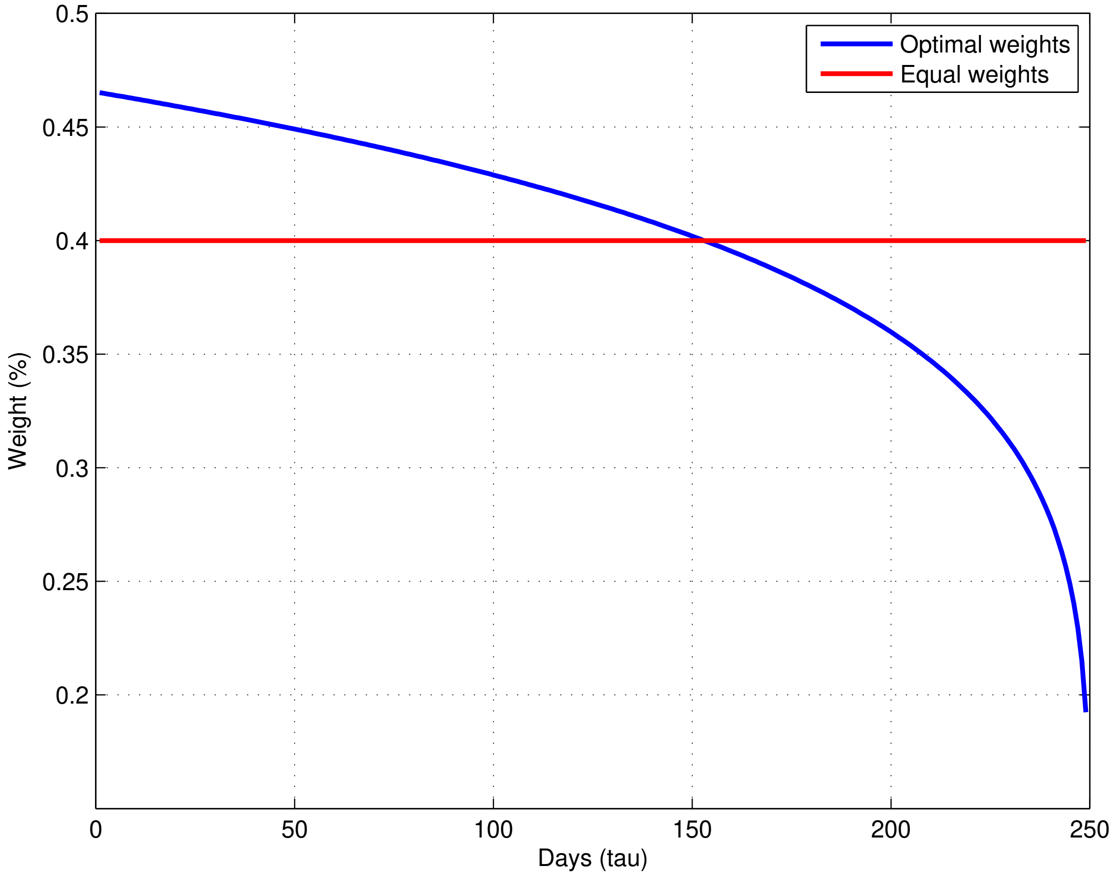


Figure 4: **Cross-Sectional Distribution of Firm Betas**

This figure shows the cross-sectional distribution of average firm betas. We first calculate at each iteration of the Gibbs sampler the beta for firm i at time t based on the model in equation (7). Subsequently, we compute the time series averages of these conditional betas. We then calculate for each firm the posterior mean of its time series average beta. This figure shows the cross-sectional distribution of these posterior means.

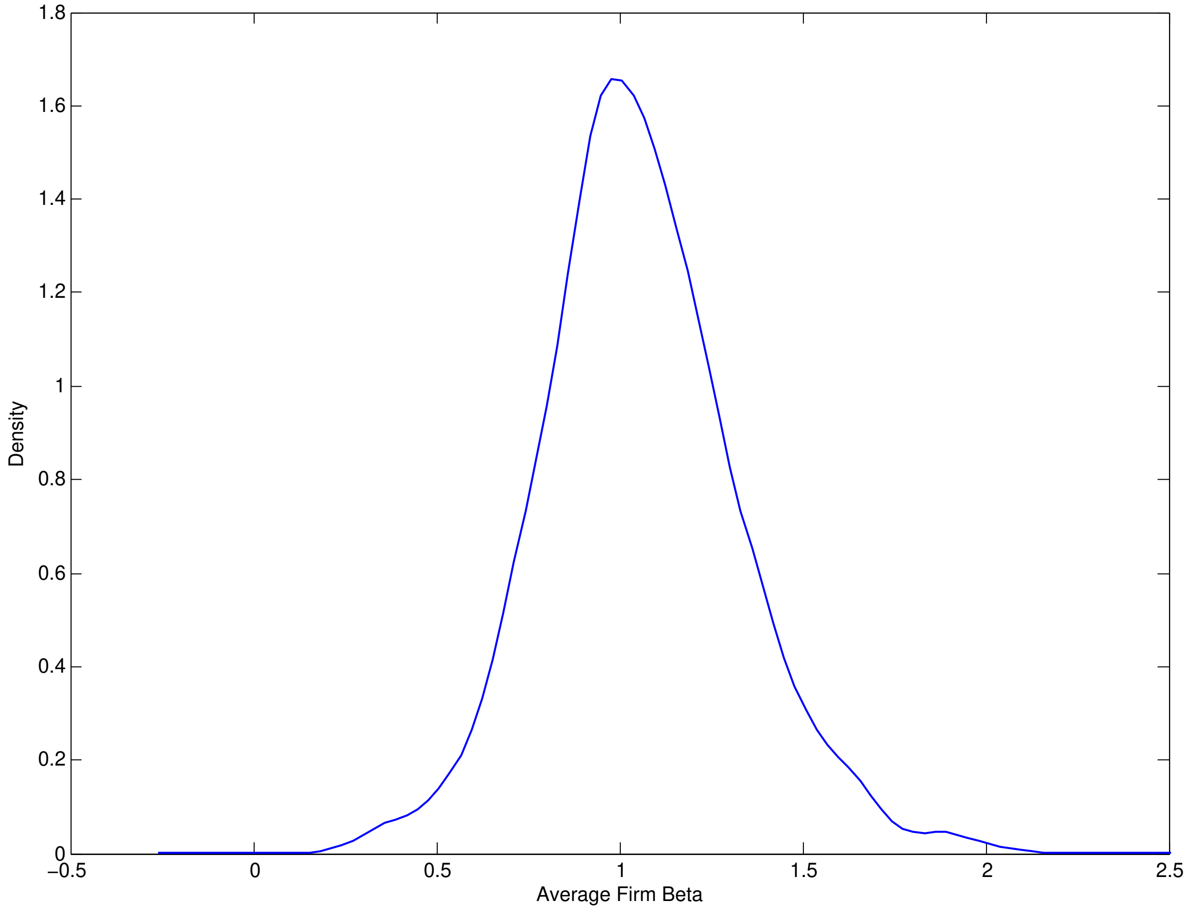
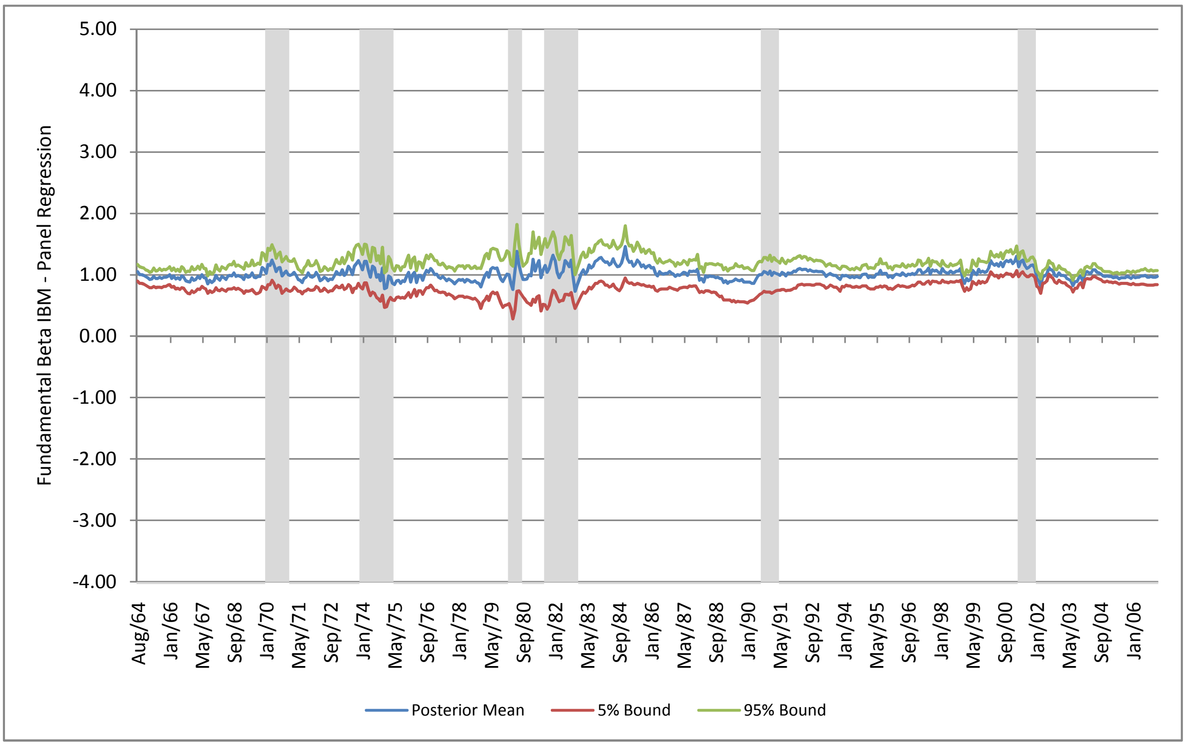


Figure 5: **Confidence Interval for IBM Beta: Panel versus Time Series Regression**

This figure plots the mean and 5% and 95% percentile values of the posterior distribution of the fundamental beta of IBM in each month from August 1964 through December 2006. The fundamental beta is modeled as a linear function of firm characteristics and macroeconomic state variables. The upper graph is based on the estimation output of the hierarchical panel data model presented in section I of the paper and the lower graph is constructed using the output of a time series regression. Shaded areas indicate NBER recession periods.



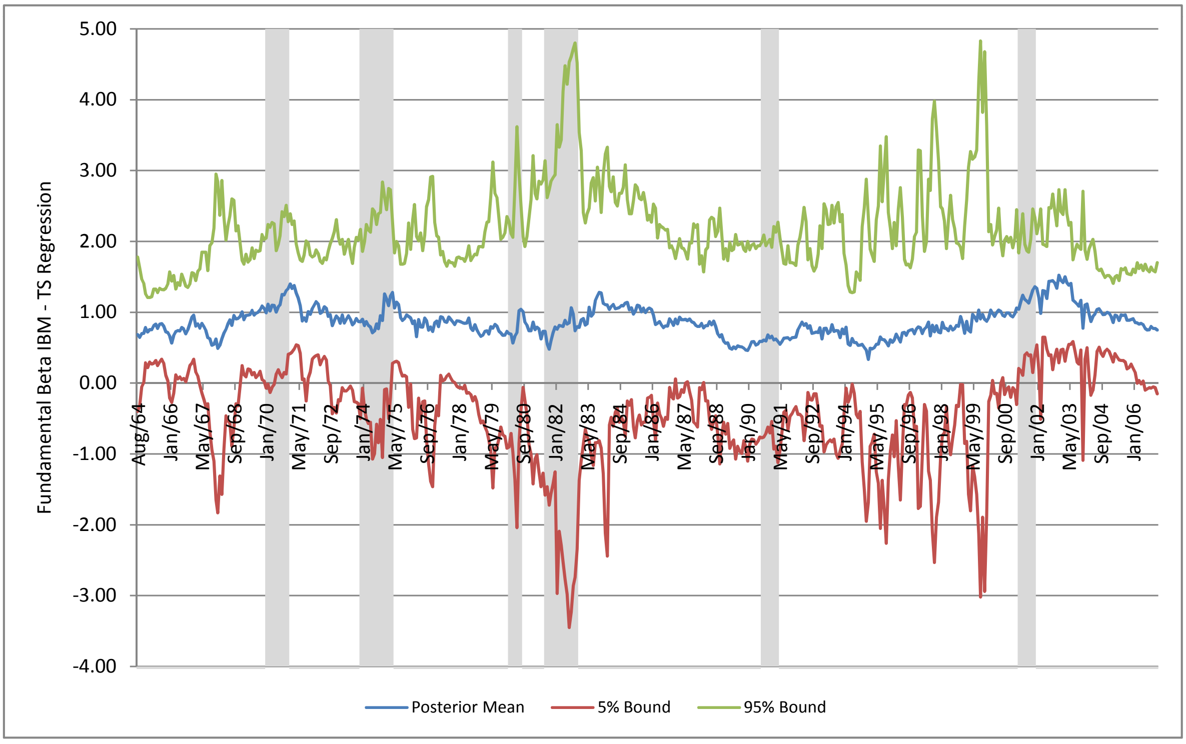
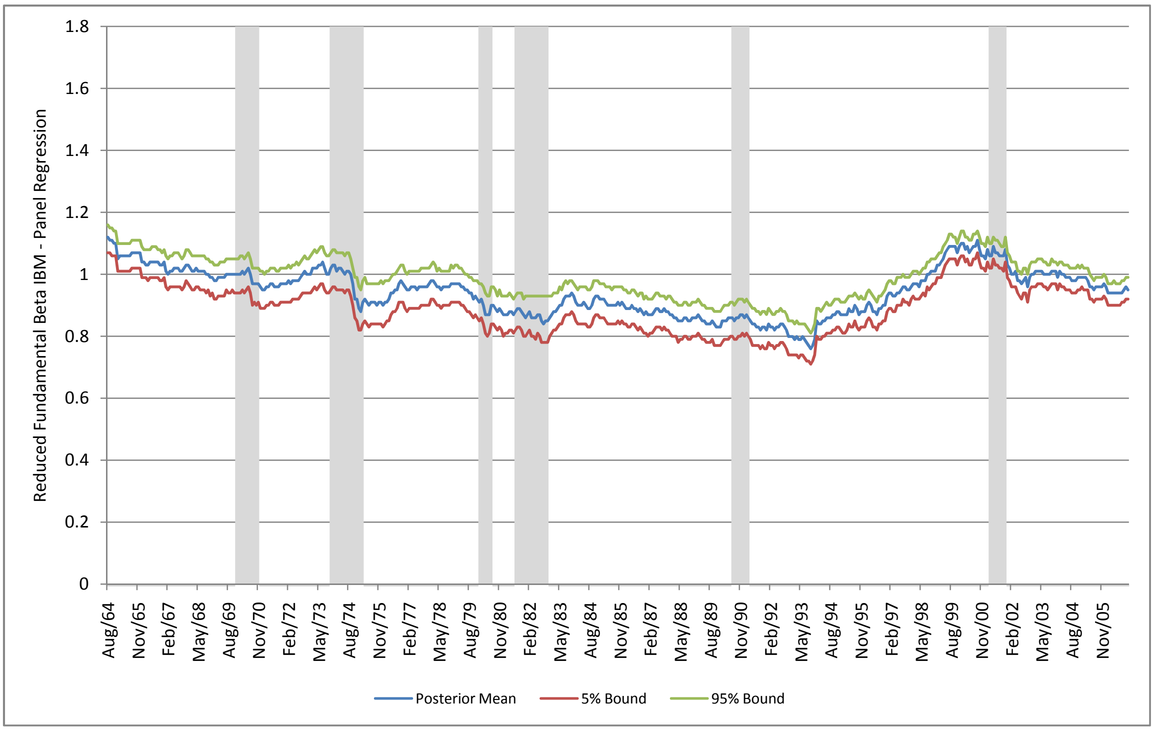


Figure 6: **Confidence Interval for Reduced Fundamental Beta of IBM**

This figure plots the mean and 5% and 95% percentile values of the posterior distribution of the fundamental beta of IBM in each month from August 1964 through December 2006. The fundamental beta is modeled as a linear function of a reduced set of firm characteristics and macroeconomic state variables. The upper graph is based on the estimation output of the hierarchical panel data model presented in section I of the paper and the lower graph is constructed using the output of a time series regression. Shaded areas indicate NBER recession periods.



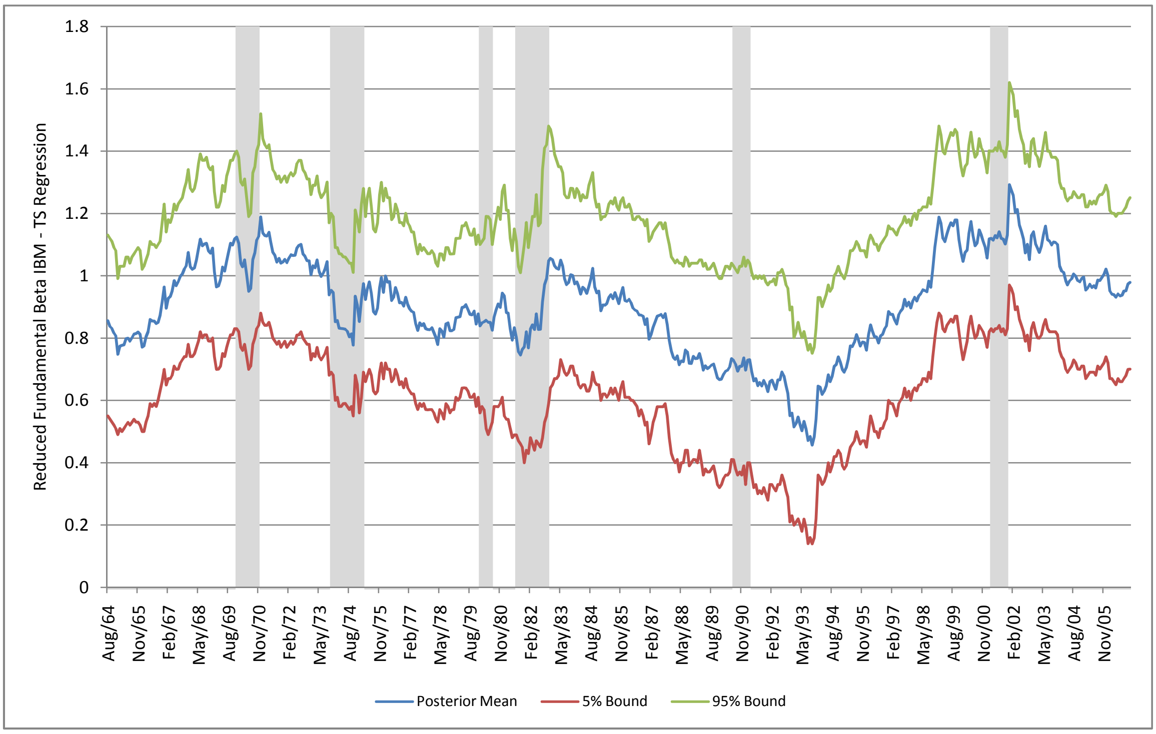


Figure 7: **Average Posterior Standard Deviation of Fundamental Betas**

This figure plots the cross-sectional average of the posterior standard deviations of the fundamental betas of all firms in the sample from August 1964 through December 2006. In the upper graph the fundamental beta is modeled as a linear function of firm characteristics and macroeconomic state variables and in the lower graph fundamental betas depend on a reduced set of conditioning variables. Posterior standard deviations are based on the estimation output of the hierarchical panel data model and the output of time series regressions estimated for every firm. Shaded areas indicate NBER recession periods.