Estimating Security Betas Using Prior Information Based on Firm Fundamentals

Document status and date:
Published: 01/04/2016

Document Version:
Publisher's PDF, also known as Version of record

Document license:
Taverne

Estimating Security Betas Using Prior Information Based on Firm Fundamentals

Mathijs Cosemans
Rotterdam School of Management, Erasmus University

Rik Frehen
Tilburg University

Peter C. Schotman
Maastricht University

Rob Bauer
Maastricht University

We propose a hybrid approach for estimating beta that shrinks rolling window estimates toward firm-specific priors motivated by economic theory. Our method yields superior forecasts of beta that have important practical implications. First, unlike standard rolling window betas, hybrid betas carry a significant price of risk in the cross-section even after controlling for characteristics. Second, the hybrid approach offers statistically and economically significant out-of-sample benefits for investors who use factor models to construct optimal portfolios. We show that the hybrid estimator outperforms existing estimators because shrinkage toward a fundamentals-based prior is effective in reducing measurement noise in extreme beta estimates. (*JEL* G11, G12, G14, G17)

Precise estimates of individual stock betas are crucial in many applications of modern finance theory. For instance, portfolio managers need to ensure that their risk exposure stays within predetermined limits and managers need reliable estimates of their company's beta to make capital budgeting decisions. However, as noted by Campbell, Lettau, Malkiel, and Xu (2001), an important practical problem is that "firm-specific betas are difficult to estimate and may well be unstable over time."[1] 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."

The typical approach to reduce measurement error in betas is to group stocks into portfolios, as proposed by Black, Jensen, and Scholes (1972) and Fama and MacBeth (1973). If estimation errors are uncorrelated across stocks, overestimates and underestimates of individual betas will tend to cancel out when stocks are aggregated into portfolios. However, a recent strand of literature stresses the downsides of using portfolios in cross-sectional asset pricing tests. Lewellen, Nagel, and Shanken (2010) demonstrate that the standard tests have low power to reject a model when characteristic-sorted portfolios are used as test assets because of the strong factor structure inherent in such portfolios. Ang, Liu, and Schwarz (2010) show that creating portfolios lowers the precision of risk premium estimates because in doing so valuable information in the cross-section of individual stock betas is destroyed.

We propose a novel approach for estimating individual security betas that incorporates prior information based on firm fundamentals and economic state variables. Our procedure for modeling beta dynamics is a hybrid of the parametric method of Shanken (1990) that relates betas to fundamentals and a rolling sample estimator that is purely data driven. In particular, we shrink rolling beta estimates toward an economically informative prior that is unique to each firm. Our prior specification is motivated by the investment-based asset pricing literature that links a company's beta to its fundamentals.[2] Incorporating prior cross-sectional information about betas can increase the accuracy of beta estimates because a firm's beta likely resembles the betas of firms with similar characteristics. In addition, knowledge about fundamentals can help to improve long-term beta forecasts as we expect a firm's beta to regress over time toward its fundamentals-based prior.

To illustrate the basic idea, consider the following example. Assume that the sample estimate of beta for a utility company is 0.4, and further suppose that it is common knowledge that in the entire universe of stocks, beta is normally distributed around one with a standard deviation of 0.5. Vasicek (1973) argues

---

[1] For evidence of time variation in beta see, among others, Bollerslev, Engle, and Wooldridge (1988), Jagannathan and Wang (1996), Ferson and Harvey (1999), Petkova and Zhang (2005), and Ang and Chen (2007).

[2] See, for example, the theoretical work of Gomes, Kogan, and Zhang (2003), Carlson, Fisher, and Giammarino (2004), Zhang (2005), and Livdan, Sapriza, and Zhang (2009). Empirical evidence that beta is related to firm characteristics is provided by, among others, Ferson and Harvey (1998), Lewellen (1999), and Avramov and Chordia (2006).

that if this prior information is taken into account, the sample estimate of 0.4 is no longer the best estimate of the true beta because it is more likely to be underestimated than overestimated. Therefore, he advocates adjusting the sample estimate toward the cross-sectional mean of one. Karolyi (1992) notes that this common prior ignores relevant firm-specific information that is available prior to sampling. For instance, it is well known that utilities tend to have betas smaller than one. Given this prior knowledge, shrinkage toward the mean is no longer optimal for the utility company because it likely overcorrects the sample estimate. Karolyi (1992) therefore proposes to form industry portfolios and to shrink a firm's sample estimate toward its industry beta.

However, creating portfolios leads to a loss of information in firm-level betas because in practice industry classification is only one of the many potential determinants of beta. For instance, if the utility company is a small, highly levered firm, theory predicts that its beta exceeds the industry average. Although this additional economic information could be incorporated by sorting on multiple characteristics, this would reduce the number of stocks in each portfolio and thereby increase estimation error. We address these issues by specifying a regression-based prior that is firm specific and able to accommodate a large number of characteristics and business-cycle variables.

Our main results are as follows. First, we show that our hybrid estimator leads to significant gains in out-of-sample predictions of beta. Compared to the existing shrinkage estimators of Vasicek (1973) and Karolyi (1992), mean squared errors (MSEs) are more than 15% smaller at the monthly horizon, 25% lower for a one-year forecast period, and up to 40% smaller at the five-year horizon. Our finding that the gains relative to existing methods increase with the horizon highlights the benefits of incorporating fundamentals-based prior information in the estimation of long-term betas. The outperformance over standard rolling window estimators is even larger. For instance, forecast errors produced by the popular five-year rolling estimator based on monthly returns are twice as large as those generated by the hybrid model. Furthermore, we show that assigning portfolio betas to individual stocks, as proposed by Fama and French (1992), also yields inaccurate forecasts of firm-level betas because it ignores the heterogeneity in betas across stocks within each portfolio.

Second, the improved beta forecasts of the hybrid approach offer significant benefits for investors who use factor models to construct optimal portfolios. We illustrate these economic benefits by forming market-neutral portfolios. We find that the portfolio based on covariance forecasts from the hybrid model is the only portfolio that meets the objective of being market neutral ex post. Other beta estimators yield portfolios with significant exposure to market risk because they underestimate the betas of stocks that are bought and overestimate the betas of stocks that are sold short.

Third, we show that by improving the measurement of firm-level betas, the hybrid estimator changes the outcome of cross-sectional asset pricing tests. Specifically, we find that hybrid betas carry a significant price of risk in the

cross-section that is in line with theoretical predictions, even after controlling for a large set of characteristics known to explain variation in returns. In contrast, existing beta estimators used by researchers and practitioners yield risk premium estimates that are insignificantly different from zero. Our findings contradict the view that beta is dead (see, e.g., Fama and French 2004) and provide a rationale for the widespread usage of beta in practice.[3]

By comparing the hybrid model to various simplified alternatives, we identify three key aspects of the approach that drive its superior forecasting ability. First, we show that the hybrid estimator yields better forecasts than rolling window estimators because shrinkage corrects extreme rolling sample estimates of beta that are driven by measurement noise. Shrinkage is effective in the hybrid approach because the prior is unique to each firm and incorporates a broad set of firm characteristics and macroeconomic fundamentals. Conventional Vasicek shrinkage dampens only part of the noise in rolling betas because it employs a common prior that does not make use of cross-sectional information embedded in firm characteristics. The industry-level prior in the Karolyi approach yields little improvement because of the large intra-industry dispersion in betas.

Second, we show that the hybrid model beats other specifications with firm fundamentals because we estimate the prior using a flexible Bayesian method that yields a better bias-variance trade-off than standard frequentist methods. Our Bayesian panel approach increases precision by pooling the loadings on the conditioning variables and mitigates bias by letting other coefficients vary across firms to capture variation in betas unrelated to the characteristics included in the model. The traditional method for estimating conditional betas, that is, running a separate OLS regression for each firm, also allows for firm-level parameter heterogeneity but leads to large measurement errors because it does not exploit cross-sectional information to increase precision. At the other extreme, estimating the prior using a pooled OLS regression leads to precise, but biased, beta estimates because it does not allow for unobserved cross-sectional heterogeneity in betas.

Third, the hybrid procedure benefits from combining data sampled at different frequencies, similar to the GARCH-MIDAS approach proposed by Engle, Ghysels, and Sohn (2013) for modeling long- and short-term components of stock market volatility.[4] In particular, we use daily returns to estimate rolling sample betas and monthly data to measure the economic fundamentals in the prior. By using daily data, we obtain more precise rolling

---

[3] Graham and Harvey (2001) report that more than 70% of CFOs use the CAPM to calculate their cost of equity. In addition, Berk and van Binsbergen (2016) find that the CAPM is the dominant model used by mutual fund investors to make their capital allocation decisions.

[4] An appealing feature of both our hybrid model and the GARCH-MIDAS model of Engle, Ghysels, and Sohn (2013) is the direct link between economic fundamentals and measures of financial risk. An important difference between the models is that our approach combines high-frequency time-series data used to estimate rolling betas with prior cross-sectional information provided by low-frequency firm fundamentals. In contrast, in the GARCH-MIDAS model, both the high- and low-frequency components of market volatility are driven by time-series data.

beta estimates than those obtained with the commonly used Fama and MacBeth (1973) procedure that involves monthly returns. We also shorten the five-year estimation window of Fama and MacBeth (1973) to a semiannual window to improve the timeliness of rolling betas. As a result, the rolling sample estimate picks up short-term changes in beta during turbulent periods, while the prior captures long-term movements in beta driven by fundamentals.

1. Methodology

In this section, we develop the framework used to estimate time-varying security betas. Because our goal is to compare different beta estimators rather than different asset pricing models, we focus on a one-factor model. In Section 1.1 we discuss the estimation of rolling sample betas. Section 1.2 describes the specification and estimation of our fundamentals-based prior, and in Section 1.3 we update this prior belief with sample information to obtain hybrid beta estimates. Finally, in Section 1.4 we introduce the existing estimators that we use as benchmark for our hybrid method.

1.1 Rolling sample estimates of beta

We obtain monthly sample estimates of beta from rolling regressions with daily return data. Sampling at a daily frequency provides a reasonable trade-off between efficiency and robustness to microstructure noise. We use a semiannual estimation window to obtain timely estimates that pick up short-term fluctuations in betas, and we later combine them with prior betas that capture long-term information. We estimate rolling sample betas by running the following time-series regression

$$r_{it,s} = \alpha_{it} + \beta_{it} r_{Mt,s} + \epsilon_{it,s}, \tag{1}$$

where $r_{it,s}$ and $r_{Mt,s}$ are daily excess returns on stock i and the market. The subscript $s = (1, 2, \ldots, \tau)$ is used to index the daily returns before the end of month t and τ is the length of the estimation window, that is, $\tau = 125$ trading days. The subscript t is to emphasize that we estimate integrated alphas and betas for each month using a rolling window of daily returns. The intercept α_{it} is the risk-adjusted return. The regression slope β_{it} is our object of interest. The error term $\epsilon_{it,s}$ is a zero-mean, normally distributed idiosyncratic return shock with variance σ_{it}^2.

1.2 Incorporating firm characteristics as prior information

Following Vasicek (1973), we specify uninformative priors on the pricing error α_{it} and idiosyncratic variance σ_{it}^2 and assume that the prior distribution for β_{it} is normal:

$$\beta_{it} \sim \mathcal{N}\left(\bar{\beta}_{it}, \sigma_{\beta_{it}}^2\right). \tag{2}$$

Vasicek (1973) suggests that when no other information is known about a stock except that it comes from a broad universe of stocks, a good choice of the prior density for beta is the cross-sectional distribution of beta in this universe. The prior mean and variance of β_{it} would thus be set equal to the unconditional mean and variance in the cross-section. By assigning the same prior to all stocks, each firm is essentially treated as a random draw from the cross-section. In contrast, we construct a prior for beta that is unique to each firm and economically informative by incorporating observable firm-specific information. Subsequently, we shrink the sample least-squares estimate of a company's beta from Equation (1) toward its fundamentals-based prior beta.

We specify a monthly panel model to elicit a prior for the beta of firm i in month t,[5]

$$R_{it} = \alpha_i^* + \beta_{it|t-1}^* R_{Mt} + \eta_{it}, \tag{3}$$

where R_{it} and R_{Mt} denote monthly excess returns on the stock and the market, respectively, and α_i^* is the risk-adjusted stock return. The idiosyncratic return η_{it} is normally distributed with mean zero and variance $\sigma_{\eta_i}^2$ and is assumed to be independent across stocks. Following Shanken (1990), we parameterize the prior beta as a linear function of conditioning variables

$$\beta_{it|t-1}^* = \delta_{0i} + \delta_{1i} X_{t-1} + \delta_2' Z_{it-1} + \delta_3' Z_{it-1} X_{t-1}, \tag{4}$$

where X_{t-1} is a business-cycle variable and Z_{it-1} is a vector with lagged firm characteristics. We allow the relation between beta and firm characteristics to vary over the business cycle by including the interaction terms $Z_{it-1} X_{t-1}$ in Equation (4) and capture any cyclical pattern in beta unrelated to characteristics by also including X_{t-1} separately. Substitution of (4) into (3) yields

$$R_{it} = \alpha_i^* + (\delta_{0i} + \delta_{1i} X_{t-1} + \delta_2' Z_{it-1} + \delta_3' Z_{it-1} X_{t-1}) R_{Mt} + \eta_{it}. \tag{5}$$

1.2.1 Choosing conditioning variables. Our choice of firm-specific instruments is based on the investment-based asset pricing literature. Gomes, Kogan, and Zhang (2003) derive an explicit relation between market beta and firm size and book-to-market in a general equilibrium setting. They 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. Carlson, Fisher, and Giammarino (2004) argue that value firms are riskier than growth firms because they are more affected by negative aggregate shocks due to higher operating leverage. Zhang (2005) proposes a model in which costly reversibility of capital makes it harder for value firms to reduce the scale of their operations in recessions. Consequently, value firms have countercyclical betas, while those of growth stocks are procyclical. In the

---

[5] We estimate the panel regression using monthly data because some of the conditioning variables in our prior specification are not available at a daily frequency.

model of Livdan, Sapriza, and Zhang (2009), the inflexibility to adjust capital investment to aggregate shocks stems from financial constraints. Specifically, firms with high leverage are more likely to be subject to collateral constraints that limit their ability to smooth dividends by exploiting new investment opportunities. As a result, dividends are more correlated with the business cycle, thereby increasing the risk of these firms.

Motivated by these studies, our set of instruments Z_{it-1} in Equation (4) includes measures of firm size, book-to-market, operating leverage, and financial leverage. We further include momentum, motivated by the empirical finding of Grundy and Martin (2001) that momentum is related to beta dynamics. Since previous work has documented strong variation in betas across industries (Fama and French 1997), we also add a firm's industry classification to our prior model. Existing literature indicates that the relation between firm characteristics and beta varies over the business cycle (e.g., Petkova and Zhang 2005). To capture these time-series dynamics, we follow Jagannathan and Wang (1996) and choose the default spread as indicator of the state of the economy X_{t-1} in Equation (4).[6]

1.2.2 Prior estimation. The parametric specification of beta in Equation (4) is theoretically appealing because it directly links variation in beta to firm-specific and macroeconomic fundamentals. However, in practice two important problems arise when implementing this approach. First, the investor's information set is unobservable, and this is problematic because Ghysels (1998) points out that misspecification of beta risk can result in large pricing errors. Second, including more instruments to mitigate this problem makes estimating the model parameters with precision difficult, particularly for individual stocks.

We address both issues by estimating the prior model using Bayesian methods. The key advantage of the Bayesian approach is that it allows us to pool some parameters to increase estimation precision, while letting other coefficients vary across firms to capture unobserved heterogeneity in betas. In particular, the firm-specific parameter δ_{0i} in the prior specification mitigates omitted variable bias by picking up the effect on beta of missing conditioning variables that are constant over time but vary across firms.[7] In addition, by pooling the δ_2 and δ_3 loadings on the firm-level conditioning variables, we exploit cross-sectional information to obtain more precise estimates. The pooling of these parameters can be justified by the theoretical work discussed in the previous section that predicts the relation between firm characteristics and beta to be the same across stocks.

The Bayesian approach also uses cross-sectional information to increase the precision of the estimates of δ_{0i} and δ_{1i}. In particular, we specify hierarchical

---

[6] For the sake of parsimony, we do not include interactions between the default spread and industry dummies.

[7] We also let δ_{1i} in (4) be firm specific to allow for heterogeneous exposure to the macroeconomic instrument.

priors that impose a common structure on these parameters, while still allowing for cross-sectional heterogeneity. Intuitively, the OLS estimates of the firm-level parameters are shrunk toward their cross-sectional mean, similar to a random coefficients model. Korteweg and Sorensen (2010) employ a comparable common prior specification for firm-specific parameters. We estimate the parameters of the panel model using Markov chain Monte Carlo methods. A discussion of the prior specification is provided in Appendix A and a detailed description of the estimation procedure is available in the Online Appendix.

1.3 Computing hybrid betas

The posterior moments of $\beta^*_{it|t-1}$ obtained from the panel regression constitute the prior mean and variance for β_{it} in the rolling window regressions. Thus, we set $\bar{\beta}_{it}$ and $\sigma^2_{\beta_{it}}$ in Equation (2) equal to the posterior mean and variance of $\beta^*_{it|t-1}$. Vasicek (1973) derives a formal procedure that combines the sample estimate of beta from (1) with this prior belief to obtain a shrinkage estimate of beta, which is approximately normally distributed with mean and variance given by

$$\tilde{\beta}_{it} = \frac{\bar{\beta}_{it}/\sigma^2_{\beta_{it}} + b_{it}/s^2_{b_{it}}}{1/\sigma^2_{\beta_{it}} + 1/s^2_{b_{it}}} \tag{6}$$

$$\tilde{\sigma}^2_{\beta_{it}} = \frac{1}{1/\sigma^2_{\beta_{it}} + 1/s^2_{b_{it}}}, \tag{7}$$

where b_{it} denotes the sample estimate of β_{it} and $s^2_{b_{it}}$ the OLS sampling variance of b_{it}.[8]

The posterior mean $\tilde{\beta}_{it}$, which we refer to as the hybrid beta, can be expressed as a weighted average of the prior mean and sample estimate of beta

$$\tilde{\beta}_{it} = \phi_{it}\bar{\beta}_{it} + (1 - \phi_{it})b_{it}, \tag{8}$$

with the shrinkage weight ϕ_{it} given by

$$\phi_{it} = \frac{s^2_{b_{it}}}{\sigma^2_{\beta_{it}} + s^2_{b_{it}}}. \tag{9}$$

Equation (9) implies that the degree of shrinkage toward the prior is proportional to the relative precision of the sample estimate and the prior. If the sample

---

[8] Foster and Nelson (1996) develop continuous record asymptotics for rolling betas and demonstrate that the OLS sampling variance overstates the precision of rolling beta estimates because it ignores time variation in betas within the estimation window. We find that computing the asymptotic variance of rolling betas according to the procedure in Foster and Nelson (1996) has little impact on our results because the rolling betas we employ in the hybrid approach are based on a short, semiannual window. From an economic point of view, it is unlikely that a firm's beta moves dramatically in such a short period of time. Because the calculation of the Foster and Nelson (1996) variance is quite involved, we use the standard OLS sampling variance to construct the shrinkage weights in the hybrid approach.

estimate is very imprecise (i.e., $s^2_{b_{it}}$ is large relative to $\sigma^2_{\beta_{it}}$), most weight is given to the prior beta.

1.4 Overview of alternative approaches to estimating betas
In the empirical part of the paper, we compare the performance of our hybrid estimator to that of six alternative beta estimators that are commonly used by researchers and practitioners. In this section, we briefly describe these existing approaches to estimating time-varying security betas, which include a conditional beta model, two rolling window estimators, two shrinkage estimators, and the approach of Fama and French (1992), which assigns portfolio betas to individual stocks.

1.4.1 Conditional beta model. The parametric method of Shanken (1990) models conditional betas as a linear function of a set of instruments. Following Avramov and Chordia (2006), we obtain conditional betas by estimating Equation (5) using a separate time-series regression for each firm. Consequently, all parameters in this specification are treated as firm specific, including the loadings on the conditioning variables.

1.4.2 Rolling window estimators. The simplest approach to modeling time-varying betas is estimating rolling window regressions. A benefit of this method is its robustness to misspecification, since there is no need to specify a set of conditioning variables. An important drawback, however, is that these data-driven filters ignore all time variation in betas within each window. Although shortening the window length results in timelier betas, estimation precision goes down. Because of this balance between timeliness and efficiency, we consider two sets of rolling betas estimated using different window lengths and data frequencies. The first set of betas is obtained by estimating rolling regressions using a five-year window of monthly returns, as proposed by Black, Jensen, and Scholes (1972) and Fama and MacBeth (1973). The other rolling estimator that we consider uses a one-year window of daily returns. Daily, rather than monthly, data are used because in theory the accuracy of beta estimates increases with the sampling frequency (see Andersen et al. 2006).

1.4.3 Shrinkage estimators. We consider two existing approaches that employ shrinkage to improve the accuracy of rolling beta estimates. In the classic Vasicek (1973) procedure, the value-weighted mean and cross-sectional variance of rolling betas are used as prior information. Rolling betas are estimated using a one-year window of daily returns and shrunk toward the common prior depending on the relative precision of the estimates, as in Equation (8). In the approach of Karolyi (1992), stocks are grouped into portfolios based on characteristics and a firm's one-year rolling beta estimate is shrunk toward its one-year rolling portfolio beta. In our implementation of

this approach, prior information is obtained by creating 48 industry portfolios according to the classification of Fama and French (1997).[9]

1.4.4 The Fama-French (1992) approach.   The estimation method of Fama and French (1992) consists of several steps. First, each year at the end of June, stocks are sorted into size deciles based on NYSE breakpoints. Subsequently, each size decile is subdivided into ten portfolios based on the one-year rolling beta estimates of the individual stocks, again using NYSE breakpoints. Equal-weighted daily returns are calculated for each of the resulting 100 size-beta sorted portfolios over the next year. Portfolio betas for month t are then obtained from rolling regressions of daily post-ranking portfolio returns on the daily market return over the one-year window ending in month t. Finally, these portfolio betas are assigned to the individual stocks in each of the 100 size-beta portfolios. In their paper, Fama and French (1992) estimate the pre-ranking and post-ranking betas using monthly returns. We estimate these betas using daily returns because our empirical analysis indicates that daily betas are more accurate predictors of future betas than betas computed from monthly returns.

2. Data

The firm data come from CRSP and Compustat and consist of the daily and monthly return, the book and market value of equity, the book value of total assets, the net sales, and the operating income for all firms listed on the NYSE, AMEX, and NASDAQ. We use the value-weighted portfolio of all stocks as a proxy for the market portfolio. The sample covers the period from August 1964 to December 2011. We include a stock in the analysis for 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 to compute the firm characteristics size, book-to-market, leverage, and momentum. Following Fama and French (1992), we measure firm size by the market value of equity, book-to-market as the ratio of the book and market value of equity, and financial leverage as the ratio of the book value of assets over the market value of equity. Following Gulen, Xing, and Zhang (2009), operating leverage is computed as the three-year moving average of the ratio of the percentage change in operating income before depreciation to the percentage change in sales. Momentum is measured as the cumulative return over the 12 months prior to the current month. We create 48 industry dummies based on the SIC codes of the firms in CRSP using the industry classification of Fama and French (1997). We calculate book-to-market and leverage using accounting data from Compustat as of December

---

[9] We also consider a refinement of this approach in which we use the Hamada (1972) adjustment to account for differences in financial leverage across firms in a given industry. Moreover, in another implementation of the Karolyi (1992) procedure, we form the prior by grouping stocks into 25 size-BE/ME portfolios. Results for these alternative implementations are very similar to those reported in the paper and available upon request.

of the previous year and exclude firms with negative book-to-market equity. Imposing these restrictions leaves a total of 10,889 stocks.

We trim outliers in all firm characteristics to the 0.5th and 99.5th percentile values of their cross-sectional distribution. Furthermore, we use the logarithmic transformation of the size, book-to-market, and financial leverage variables because their distributions are skewed. We standardize all characteristics by subtracting the cross-sectional mean and dividing by the cross-sectional standard deviation in each month to remove any time trend in their average value. We measure the default spread as the yield differential between Moody's Baa- and Aaa-rated corporate bonds.

3. Estimation Results

This section presents estimation results for the hybrid beta model. In Section 3.1 we verify whether the fundamentals in the prior model explain variation in firm-level betas. In Section 3.2 we relate the variation in shrinkage weights to firm characteristics and market conditions. Finally, in Section 3.3, we discuss the time-series and cross-sectional properties of hybrid betas.

3.1 Fundamental determinants of betas

Table 1 summarizes the relation between prior betas and firm fundamentals.[10] Consistent with theoretical predictions, we find that all characteristics in the prior model are important determinants of beta. The negative coefficient on size indicates that small firms have higher betas than large firms, and the negative loading on book-to-market implies that value firms are unconditionally less risky than growth firms. As predicted by Carlson, Fisher, and Giammarino (2004), we find that higher operating leverage leads to higher betas. The loadings on the interaction terms between the default spread and the firm characteristics show how betas vary over the business cycle. The positive coefficient on the interaction of the default spread and financial leverage is consistent with the model of Livdan, Sapriza, and Zhang (2009), in which the dividends of firms with higher leverage are more strongly correlated with the business cycle. Finally, we observe a large cross-sectional spread in the firm-specific parameter δ_{0i} in the prior specification in Equation (4), which highlights the importance of also allowing for variation in betas unrelated to fundamentals.

3.2 Variation in shrinkage weights

By construction, the shrinkage parameter in (9) lies between zero and one, with higher values assigning more weight to the prior beta. In Table 2 we relate shrinkage weights to firm characteristics to examine how the shrinkage

---

[10] For brevity, the coefficients on the 48 industry dummies are not reported. We find that most of the industry dummies contain significant prior information about the cross-sectional variation in betas.

Table 1
Fundamental determinants of prior betas

	Panel A: Cross-sectional distribution of firm-specific parameters						
	Mean	SD	5th	25th	Median	75th	95th
Constant	1.004	0.287	0.565	0.856	1.044	1.227	1.507
DEF	0.087	0.111	−0.102	0.037	0.087	0.136	0.269

	Panel B: Posterior moments of pooled parameters	
	Mean	SD
ME	−0.064	0.011
BE/ME	−0.111	0.007
A/ME	0.011	0.008
OPL	0.010	0.003
MOM	0.032	0.003
DEF*ME	0.010	0.006
DEF*BE/ME	−0.005	0.009
DEF*A/ME	0.096	0.005
DEF*OPL	0.002	0.004
DEF*MOM	−0.048	0.004

This table reports coefficient estimates for the determinants of the prior beta, which is parameterized as a linear function of conditioning variables

$$\beta_{it}^* = \delta_{0i} + \delta_{1i} X_{t-1} + \delta_2' Z_{it-1} + \delta_3' Z_{it-1} X_{t-1},$$

where X_{t-1} is the default spread (DEF) and Z_{it-1} is a vector that contains firm size (ME), book-to-market (BE/ME), financial leverage (A/ME), operating leverage (OPL), momentum (MOM), and 48 industry dummies. We use the logarithmic transformations of the size, book-to-market, and financial leverage variables and standardize all firm characteristics by subtracting the cross-sectional mean and dividing by the cross-sectional standard deviation in each month. Furthermore, for each characteristic values smaller than the 0.5th percentile and values greater than the 99.5th percentile of the cross-sectional distribution are set equal to these percentiles. Panel A presents the mean, median, standard deviation, and 5th, 25th, 75th, and 95th percentile values of the cross-sectional distribution of the posterior means of the firm-specific parameters δ_{0i} and δ_{1i}. Panel B presents the posterior means of the pooled parameters δ_2 and δ_3 and the corresponding posterior standard deviations. Coefficients on the industry dummies are not reported for brevity. All results are based on the posterior distribution of the parameters constructed from 1,250 iterations of the Gibbs sampler with a burn-in period of 250 iterations.

intensity varies across stocks. Every month, we form decile portfolios based on shrinkage weights, and for each decile, we compute the cross-sectional mean of various firm characteristics. Results in the first column show that the degree of shrinkage varies widely across firms, ranging from 0.41 for the decile of stocks with the lowest shrinkage weight to 0.86 for those with most shrinkage. Shrinkage intensities are higher for small firms, because their rolling window estimate of beta is less precise due to higher idiosyncratic risk. Stocks with large shrinkage weights also exhibit the widest gap between hybrid betas and rolling betas, suggesting that shrinkage corrects extreme rolling beta estimates that are driven by estimation noise.

The hybrid model not only allows shrinkage weights to vary across stocks but also over time. Figure 1 shows that on average, less shrinkage is applied in volatile markets, such as the stock market crash in 1987 and the credit crisis in 2008. This negative relation between shrinkage intensity and market volatility follows from standard regression theory. When market volatility is high, rolling beta estimates are more precise and therefore less shrinkage is

Table 2
Shrinkage weights and firm characteristics

	Weight	ME	BE/ME	A/ME	OPL	MOM	IVOL	|Δ BETA|
Low	0.41	1.90	0.89	2.27	1.31	14.79	1.44	0.20
2	0.54	1.73	0.83	2.03	1.23	16.37	1.83	0.25
3	0.60	1.58	0.81	1.97	1.23	17.09	2.05	0.27
4	0.64	1.41	0.81	1.94	1.22	17.41	2.24	0.29
5	0.67	1.22	0.82	1.94	1.27	17.48	2.42	0.31
6	0.71	1.04	0.83	1.98	1.27	16.99	2.62	0.33
7	0.74	0.89	0.85	2.03	1.31	16.96	2.84	0.36
8	0.77	0.75	0.89	2.15	1.37	16.64	3.11	0.39
9	0.81	0.56	0.96	2.38	1.37	15.94	3.49	0.44
High	0.86	0.35	1.10	2.89	1.40	12.46	4.34	0.54

This table reports characteristics for decile portfolios formed on the basis of shrinkage weights. Each month we sort stocks into decile portfolios based on their shrinkage parameter and compute the equal-weighted average of various characteristics of the firms in a given portfolio. ME represents the market value of equity in billions of dollars; BE/ME is the book-to-market ratio; A/ME is the ratio of the book value of assets to the market value of equity; OPL is the three-year moving average of the ratio of the percentage change in operating income before depreciation to the percentage change in sales; MOM is the cumulative return over the twelve months prior to the current month; IVOL is the idiosyncratic volatility obtained from rolling CAPM regressions using semiannual windows of daily returns; and |Δ BETA| is the absolute value of the difference between the hybrid beta and the semiannual rolling beta. The table shows the time-series averages of these monthly characteristics for every decile portfolio.
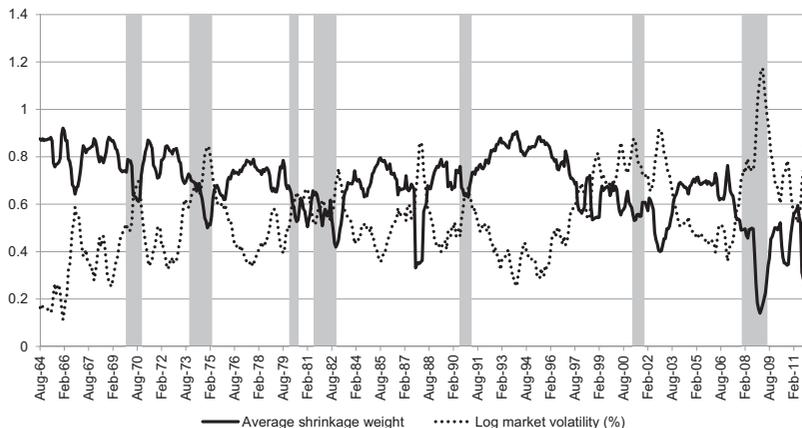
Figure 1
Time-series variation of shrinkage weights
The solid line in this figure plots the evolution through time of the cross-sectional mean of the shrinkage weights defined in Equation (9). The degree of shrinkage depends on the precision of the sample estimate of beta relative to the precision of the prior, with higher values of the shrinkage parameter assigning more weight to the prior beta. The dotted line is the six-month moving average of the logarithm of monthly realized market volatility computed by summing the squared daily returns on the value-weighted portfolio of NYSE/AMEX/NASDAQ stocks in each month. The sample period is August 1964 to December 2011, and the shaded areas indicate NBER recession periods.

needed. In contrast, in stable periods more weight is given to the fundamentals-based prior. Because we allow the combination of prior and rolling betas to vary over time, the hybrid beta estimator can respond quickly to sudden changes in market conditions without producing excessive variation in betas in normal times.

Table 3
Cross-sectional and time-series properties of individual stock betas

	Panel A: Cross-sectional				Panel B: Time series		
	Mean	SD	Implied SD		Mean	SD	Autocorr.
Hybrid	0.99	0.38	0.35		1.01	0.19	0.89
Conditional	0.98	1.14	-		1.02	1.31	0.74
OLS monthly	0.98	0.55	0.41		1.01	0.28	0.87
OLS daily	1.01	0.58	0.50		1.02	0.41	0.93
Vasicek	1.01	0.48	0.43		1.02	0.33	0.94
Karolyi	1.00	0.49	0.43		1.00	0.33	0.94
Fama-French	1.01	0.41	0.40		1.01	0.28	0.88

This table reports cross-sectional and time-series properties of individual stock betas obtained from the estimators discussed in the text. Panel A shows the time-series mean of the value-weighted cross-sectional average of estimated betas and the time-series mean of the cross-sectional standard deviation of betas. It also reports the average implied cross-sectional standard deviation of true betas, $\widehat{Std}(\beta) = [Var(\widehat{\beta}) - \overline{\widehat{Var}}_{\beta_i}]^{1/2}$, that is, the square root of the difference between the sample cross-sectional variance and the average sampling variance of estimated betas. The sampling variance of the shrinkage estimates of beta (Hybrid, Vasicek, and Karolyi) is measured by their posterior variance. The sampling variance of conditional betas, rolling betas (OLS monthly and OLS daily), and Fama-French betas is given by their squared standard error. The implied standard deviation for conditional betas is negative and therefore undefined due to large sampling errors. Panel B presents the value-weighted cross-sectional average of the time-series mean, standard deviation, and autocorrelation of estimated betas. The sample includes all NYSE/AMEX/NASDAQ-listed stocks, and the sample period is August 1964 to December 2011.

3.3 Cross-sectional and time-series properties of betas

Table 3 reports cross-sectional properties of hybrid betas and betas produced by existing estimators. The value-weighted average beta is close to one for all estimators, but the cross-sectional spread in betas varies widely across methods. Conditional betas exhibit the largest cross-sectional dispersion because the parameters in this specification are estimated by running a separate regression for each firm. As a result, a substantial part of the variation in conditional betas is due to estimation error. Following Pastor and Stambaugh (1999), we quantify estimation noise by computing an estimate of the implied cross-sectional variance of true betas as $\widehat{Var}(\beta) = Var(\widehat{\beta}) - \overline{\widehat{Var}}_{\beta_i}$, where the first term on the right is the observed sample cross-sectional variance. The second component is the average sampling variance that reflects measurement error in betas. For conditional betas, this second term is so large that the implied variance is negative. For hybrid betas, the gap between observed and implied variances is small, which indicates that shrinkage of rolling betas toward an economic prior reduces sampling error. The portfolio betas assigned to individual stocks in the Fama and French (1992) approach are also measured with precision. However, in Section 4 we show that this does not necessarily imply that portfolio betas yield accurate forecasts of *firm-level* betas.

Panel B shows that the various estimation procedures also yield different time-series dynamics. Conditional beta estimates are extremely volatile, which again largely reflects estimation noise. Hybrid betas exhibit less time variation than rolling betas because most instruments in the prior are slow moving and included to pick up long-run movements in betas driven by changes in fundamentals.

4. Out-of-Sample Beta Forecasts

In this section we run a horse race between the hybrid estimator and existing beta estimators. Section 4.1 provides direct evidence on the merits of the hybrid approach by comparing its out-of-sample forecasting ability to that of competing estimation techniques. In Section 4.2 we perform a cross-sectional analysis of forecast errors to gain more intuition about the results across different types of stocks. Finally, Section 4.3 examines the performance of various stripped-down versions of the hybrid model to identify the key drivers of its forecasting power.

4.1 Predicting individual stock betas

We generate out-of-sample beta forecasts at the end of every month t using the following procedure. First, we estimate each beta model using only data up to month t and take stock i's beta as forecast for its beta at time $t+k$, which we label $\beta_{it+k|t}^{F}$. We consider monthly, yearly, and five-year forecast horizons, for which k is equal to 1, 12, and 60, respectively. Subsequently, we compare this forecast to the realized beta over the forecast interval that is computed using return data from the start of month $t+1$ to the end of month $t+k$ and denoted by β_{it+k}^{R}. We proceed by reestimating each model using data up to month $t+1$ to produce a forecast for beta at time $t+1+k$. By repeating this procedure every month, we obtain a time series of out-of-sample beta forecasts.

We estimate the realized beta of stock i over the forecast interval k as

$$\beta_{it+k}^{R} \equiv \frac{Cov_{iM,t+k}^{R}}{Var_{M,t+k}^{R}} = \frac{\sum_{h=1}^{k/\Delta} r_{i,t+h\Delta} r_{M,t+h\Delta}}{\sum_{h=1}^{k/\Delta} r_{M,t+h\Delta}^{2}}, \tag{10}$$

where $Cov_{iM,t+k}^{R}$ is the realized covariance between stock i and the market and $Var_{M,t+k}^{R}$ is the realized market variance. These moments are computed using the continuously compounded returns $r_{i,t+h\Delta}$ and $r_{M,t+h\Delta}$ that are defined as the difference in log prices sampled at interval Δ, that is, $p_{i,t+h\Delta} - p_{i,t+(h-1)\Delta}$ and $p_{M,t+h\Delta} - p_{M,t+(h-1)\Delta}$, respectively. We consider one-month, one-year, and five-year window lengths k corresponding to the different forecast horizons that we examine. Andersen et al. (2006) demonstrate that a realized beta measure constructed from high-frequency returns is a consistent estimator of the true integrated beta. In practice, however, market microstructure frictions, such as the bid-ask bounce and nonsynchronous trading effects, put an upper limit on the data frequency that can be used to estimate realized betas.

Because microstructure noise has a minor impact on liquid stocks, we use intraday data to compute realized betas for the subset of stocks in our sample that are included in the S&P 100 index. Following Andersen et al. (2000), we estimate realized betas using 15-minute returns. We calculate high-frequency returns using transaction prices from the TAQ database. In particular, we construct equally spaced 15-minute returns by computing the logarithmic difference in prices at or immediately before each 15-minute mark. We use the

liquid SPY exchange traded fund (ETF) that tracks the S&P 500 as a proxy for the market index. Our sample of intraday returns extends from January 2, 1996 to December 31, 2011. The out-of-sample period therefore starts in January 1996, and the first beta forecasts for the S&P 100 stocks are obtained by estimating the beta models that we consider using data up to December 1995.

We also study the forecasting ability of the beta estimators for the full cross-section of NYSE/AMEX/NASDAQ-listed stocks and a longer out-of-sample period that starts in August 1984 and ends in December 2011.[11] For this broad universe of stocks, we focus on one- and five-year forecast horizons and compute realized betas using daily returns to mitigate biases due to microstructure issues.[12] However, by using lower frequency (daily) data, the realized beta estimates are less accurate. Andersen, Bollerslev, and Meddahi (2005) analyze this complication in the context of volatility forecasting and demonstrate that because of noise in the realized volatility used as proxy for the true latent volatility, the true predictive accuracy of forecasts is underestimated. Measurement error in realized betas is particularly a concern for small stocks because their returns are more sensitive to microstructure noise. We therefore evaluate the forecast accuracy of beta estimators by computing value-weighted mean squared errors in each out-of-sample period

$$MSE_{t,t+k} = \sum_{i=1}^{N_t} w_{it}(\beta_{it+k}^R - \beta_{it+k|t}^F)^2, \qquad (11)$$

where N_t is the number of stocks in the sample at time t and w_{it} is the weight of each stock.[13]

We use two procedures to evaluate the statistical significance of differences in MSEs generated by the hybrid estimator (MSE^0) and a competing approach (MSE^j). The first method is the Diebold and Mariano (1995) test of equal predictive ability, which is a t-test that takes the form

$$DM_{j,k} = \frac{\bar{d}_{j,k}}{\sqrt{\hat{\sigma}_d^2 / P}}, \qquad (12)$$

where $\bar{d}_{j,k} = \frac{1}{P} \sum_{t=Q}^{T-k} d_{j,t+k}$ with $d_{j,t+k} = MSE_{t,t+k}^j - MSE_{t,t+k}^0$, Q is the length of the in-sample estimation window, and P is the number of out-of-sample observations. $\hat{\sigma}_d^2$ is a consistent estimate of the long-run variance of the loss

---

[11] This start date implies that the first forecasts are formed using the first 20 years of data to estimate the models. The choice of an initial 20-year estimation period is common in the literature (see, e.g., Goyal and Welch 2008).

[12] We omit the monthly horizon because realized betas computed from one month of daily returns are very noisy.

[13] We confirm empirically that the standard errors of realized beta estimates of small stocks are much larger than those of big stocks. Because value weighting mitigates the impact of noise in realized betas of small stocks, it allows for a more reliable evaluation of forecasting performance than equal weighting. Indeed, for all beta estimators we find that value-weighted MSEs are smaller than equal-weighted MSEs. However, the ranking of beta estimators is the same whether the ranking is done using equal-weighted or value-weighted MSEs.

differences $d_{j,t+k}$. Because we make a new prediction every month, forecast errors for the one- and five-year horizons are based on overlapping out-of-sample periods. We use two different HAC estimators of $\hat{\sigma}_d^2$ to account for the autocorrelation in forecast errors caused by the overlapping data. First, we use the Newey and West (1987) estimator with bandwidth set equal to 1.5 times the forecast horizon k. We let the maximal lag length exceed the forecast horizon because the Bartlett kernel underweights higher-order autocorrelations. We also report results based on the HAC estimator of West (1997), which captures the autocorrelation structure by fitting an MA model to the residual series of the forecasting regression. Consequently, higher-order autocorrelations are not downweighted and the number of lags is set equal to $k-1$.[14]

The second approach that we use to evaluate significance is the test of equal finite-sample predictive accuracy proposed by Giacomini and White (2006; denoted GW).[15] This procedure allows us to test for conditional predictive ability and accounts for the effect of parameter estimation uncertainty on relative forecast accuracy.[16] The null hypothesis of the GW test is

$$H_0 : E[MSE_{t,t+k}^j - MSE_{t,t+k}^0 | I_{t-1}] = 0, \tag{13}$$

where I_{t-1} is the information set at time $t-1$. The main idea of testing for conditional predictive ability is to test whether currently available information can be used to predict which forecasting method leads to smaller forecasting errors in the out-of-sample period. In our implementation of the test, we select a q-dimensional vector of elements from the information set that includes a constant and the loss differential in the last period. The GW test statistic is a Wald-type statistic, which follows a χ_q^2 distribution under the null of equal predictive ability. We compute the test statistic using a Newey-West estimate of the variance with bandwidth equal to 1.5 times the forecast horizon.

4.1.1 S&P 100 stocks. Table 4 reports results for the S&P 100 stocks for which we compute realized betas using intraday returns. For each estimator, we present the value-weighted MSE, averaged over time, as well as the ratio of the MSE relative to the MSE of the hybrid model. In addition, we report

---

[14] For additional robustness, we also compute variances using a rectangular kernel with bandwidth $k-1$ and the small-sample adjustment of Harvey, Leybourne, and Newbold (1997). Moreover, we compute p-values for the Diebold and Mariano (1995) test statistic based on a non-parametric bootstrap along the lines of White (2000). Both methods lead to results similar to those based on the Newey and West (1987) and West (1997) HAC estimators.

[15] In theory, the asymptotics of Giacomini and White (2006) no longer apply when a recursive scheme is used to estimate the model parameters. However, Clark and McCracken (2013) present Monte Carlo evidence showing that, in practice, the test works about as well for the recursive scheme as for the rolling scheme.

[16] The object of evaluation in the procedure of Giacomini and White (2006) is the forecasting method, which not only encompasses the forecasting model but also includes the estimation procedure and the data used for estimation. As a result, this test is well suited for MSE comparisons in our setting, which involves models estimated using various methods (Bayesian and frequentist) and different window lengths and data frequencies (daily and monthly).

Table 4
Out-of-sample beta forecasts: S&P 100 stocks

	Hybrid	Conditional	OLS monthly	OLS daily	Vasicek	Karolyi	Fama-French
Panel A: Monthly horizon							
MSE	0.0604	0.6651	0.1624	0.0805	0.0728	0.0730	0.0789
Ratio	1.00	11.02	2.69	1.33	1.21	1.21	1.31
NW *t*-stat		(11.40)	(17.69)	(8.14)	(5.91)	(5.86)	(7.71)
West *t*-stat		[11.43]	[17.73]	[8.17]	[5.93]	[5.87]	[7.72]
GW *p*-value		0.00	0.00	0.00	0.00	0.00	0.00
Panel B: One-year horizon							
MSE	0.0513	0.6827	0.1548	0.0797	0.0711	0.0714	0.0762
Ratio	1.00	13.31	3.02	1.56	1.39	1.39	1.49
NW *t*-stat		(7.89)	(4.34)	(2.86)	(2.20)	(2.22)	(2.26)
West *t*-stat		[8.06]	[5.59]	[3.28]	[2.75]	[2.70]	[2.53]
GW *p*-value		0.00	0.00	0.01	0.03	0.03	0.01
Panel C: Five-year horizon							
MSE	0.0638	0.7587	0.1728	0.1202	0.1078	0.1091	0.1097
Ratio	1.00	11.89	2.71	1.88	1.69	1.71	1.72
NW *t*-stat		(9.28)	(2.86)	(3.68)	(3.74)	(3.76)	(3.26)
West *t*-stat		[9.93]	[3.52]	[3.90]	[3.91]	[3.95]	[3.57]
GW *p*-value		0.00	0.00	0.00	0.01	0.01	0.01

This table reports the mean squared error (MSE) of monthly, yearly, and five-year out-of-sample beta forecasts for stocks included in the S&P 100 index. Beta forecasts for time $t+k$ are formed using data up to month t, with k equal to 1, 12, and 60 for the monthly, one-year, and five-year forecasts, respectively. The out-of-sample period begins in January 1996 and ends in December 2011. The first forecast is based on data from August 1964 to December 1995, and the last forecast uses data up to November 2011. Beta forecasts are compared to realized betas that are computed using intraday returns from the start of month $t+1$ to the end of month $t+k$. Each model is then reestimated using data up to month $t+1$ to produce beta forecasts for time $t+k+1$, which are compared to realized betas at time $t+k+1$. This procedure yields a time series of out-of-sample forecast errors for each stock. For each estimator, the table presents the value-weighted MSE (averaged over time), as well as the ratio of the MSE relative to the MSE of the hybrid model. We further report *t*-statistics for a Diebold and Mariano (1995) test that the MSEs generated by the hybrid estimator and a competing estimator are equal. The *t*-statistics in parentheses are based on Newey-West (1987) standard errors with lag length equal to 1.5 times the number of months in the forecasting period. The *t*-statistics in brackets are based on West (1997) standard errors with the number of lags equal to $k-1$. The last row in each panel shows *p*-values for the test of equal predictive ability of Giacomini and White (2006) based on Newey-West (1987) variances with bandwidth equal to 1.5 times the forecast horizon.

t-statistics for the Diebold and Mariano (1995) test of unconditional predictive ability and *p*-values for the Giacomini and White (2006) test of conditional predictive ability.

We find that the hybrid estimator provides superior out-of-sample forecasts of beta at all horizons, regardless of the procedure used to evaluate significance. MSEs produced by the best existing estimator (Vasicek) are 20% larger at the monthly horizon, 40% larger for the one-year period, and almost 70% larger at the five-year horizon. The outperformance over popular rolling window estimators is even larger. For instance, forecast errors produced by the five-year rolling estimator are more than twice as large as those generated by the hybrid approach. The portfolio approach of Fama and French (1992) yields better forecasts than the daily and monthly rolling estimators but produces MSEs that are 30% (monthly horizon) to 70% (five-year horizon) larger than those obtained with the hybrid estimator. Conditional betas are the worst predictors of future realized betas due to large measurement errors that arise from estimating

Table 5
Out-of-sample beta forecasts: All stocks

	Hybrid	Conditional	OLS monthly	OLS daily	Vasicek	Karolyi	Fama-French
Panel A: One-year horizon							
MSE	0.0970	4.4591	0.1968	0.1153	0.1077	0.1073	0.1136
Ratio	1.00	45.97	2.03	1.19	1.11	1.11	1.17
NW *t*-stat		(4.13)	(5.46)	(3.53)	(2.08)	(2.10)	(2.46)
West *t*-stat		[4.51]	[6.58]	[3.21]	[2.01]	[1.98]	[2.75]
GW *p*-value		0.00	0.00	0.00	0.03	0.03	0.01
Panel B: Five-year horizon							
MSE	0.0946	3.9621	0.1948	0.1431	0.1248	0.1257	0.1259
Ratio	1.00	41.89	2.06	1.51	1.32	1.33	1.33
NW *t*-stat		(6.09)	(2.84)	(6.73)	(4.47)	(4.48)	(3.40)
West *t*-stat		[5.25]	[3.91]	[4.53]	[2.58]	[2.93]	[3.10]
GW *p*-value		0.00	0.01	0.00	0.00	0.00	0.00

This table reports the mean squared error (MSE) of one- and five-year out-of-sample beta forecasts for our sample of NYSE/AMEX/NASDAQ-listed stocks. Beta forecasts for time $t+k$ are formed using data up to month t, with k equal to 12 and 60 for the one- and five-year forecasts, respectively. The out-of-sample period begins in August 1984 and ends in December 2011. The first forecast is based on data from August 1964 to July 1984, and the last forecast uses data up to December 2010. Beta forecasts are compared to realized betas that are computed using daily returns from the start of month $t+1$ to the end of month $t+k$. Each model is then reestimated using data up to month $t+1$ to produce beta forecasts for time $t+k+1$, which are compared to realized betas at time $t+k+1$. This procedure yields a time series of out-of-sample forecast errors for each stock. For each estimator, the table presents the value-weighted MSE (averaged over time), as well as the ratio of the MSE relative to the MSE of the hybrid model. We further report t-statistics for a Diebold and Mariano (1995) test that the MSEs generated by the hybrid estimator and a competing estimator are equal. The t-statistics in parentheses are based on Newey-West (1987) standard errors with lag length equal to 1.5 times the number of months in the forecasting period. The t-statistics in brackets are based on West (1997) standard errors with the number of lags equal to $k-1$. The last row in each panel shows p-values for the test of equal predictive ability of Giacomini and White (2006) based on Newey-West (1987) variances with bandwidth equal to 1.5 times the forecast horizon.

every parameter in the conditional model separately for each firm.[17] Because the differences in MSE between the hybrid approach and existing estimators are not only statistically significant but are also large in economic terms, they can have serious consequences for practical applications, which we explore in Sections 5 and 6.

4.1.2 All stocks. The empirical evidence in Table 5 shows that the superior forecasting ability of the hybrid estimator extends to the full cross-section of NYSE/AMEX/NASDAQ-listed stocks and the longer out-of-sample period starting in August 1984. Prediction errors produced by all specifications are larger for this expanded universe because the daily realized betas used as benchmark are noisier than those constructed from intraday data in Table 4. However, regardless of the way realized betas are measured, the hybrid estimator significantly outperforms all other estimators according to both the Diebold and Mariano (1995) test and the Giacomini and White (2006)

---

[17] The poor performance of the conditional beta specification is not merely a consequence of a few outliers. Unreported results show that even after winsorizing conditional beta forecasts cross-sectionally at the 5th and 95th percentile, MSEs are still much larger than the MSEs produced by all the other beta estimators that we consider.

procedure. The other approaches that involve some form of shrinkage (Vasicek and Karolyi) and the portfolio procedure of Fama and French (1992) yield better predictions of beta than simple rolling estimators, but they do not match the forecasting performance of the hybrid model.

4.2 Cross-sectional analysis of forecast errors

The forecasting results in the previous section raise the question of why the hybrid approach works better than existing beta estimators. To answer this question, we first need to identify the type of stocks for which the hybrid estimator works particularly well. The descriptive results in Table 2 suggest that shrinkage is most beneficial for stocks with extreme sample estimates of beta stemming from large measurement errors. We test this conjecture as follows. At the end of each month, stocks are sorted into decile portfolios based on their predicted beta. Ex ante portfolio betas are measured as the value-weighted average of these beta forecasts. Ex post portfolio betas are estimated by running a regression of daily portfolio returns on a constant and the market return over the next year. Forecast errors are defined as the difference between ex post and ex ante portfolio betas.

Figure 2 plots the average forecast error for each beta decile.[18] The black bars represent the deciles for which the hybrid estimator produces significantly lower MSEs than a competing approach according to the Diebold and Mariano (1995) test. A clear pattern emerges: all existing estimators significantly underestimate low betas and overestimate high betas. Average forecast errors for the daily rolling window estimator range from 0.32 for the low-beta decile to -0.25 for the decile of high-beta stocks. The five-year rolling estimator based on monthly returns performs even worse as it underestimates the beta of stocks in the bottom decile by 0.37 and overestimates the beta of the top portfolio by 0.47. The Vasicek and Karolyi estimators dampen only part of the noise in rolling beta estimates. Forecast errors for the low-beta deciles remain significantly positive, while the high-beta portfolios still exhibit substantial negative forecast errors. The estimation approach of Fama and French (1992) does better than these conventional shrinkage estimators. An important downside of this procedure is that it ignores the cross-sectional heterogeneity in betas within each of the size-beta portfolios stocks are sorted into. Consequently, even if the betas of the size-beta sorted portfolios themselves are unbiased, assigning these portfolio betas to individual stocks induces a bias in firm-level betas that leads to higher MSEs in the firm-level forecasts in Tables 4 and 5. However, this problem is less severe in the portfolio-level test in Figure 2 because upward and downward biases in firm-level betas tend to cancel out at the portfolio level. Nevertheless, the Fama and French (1992) method still yields sizeable forecast errors for high- and low-beta deciles. In contrast, prediction errors produced by the hybrid

---

[18] We omit the Karolyi approach to save space. The results are very similar to those for the Vasicek estimator.
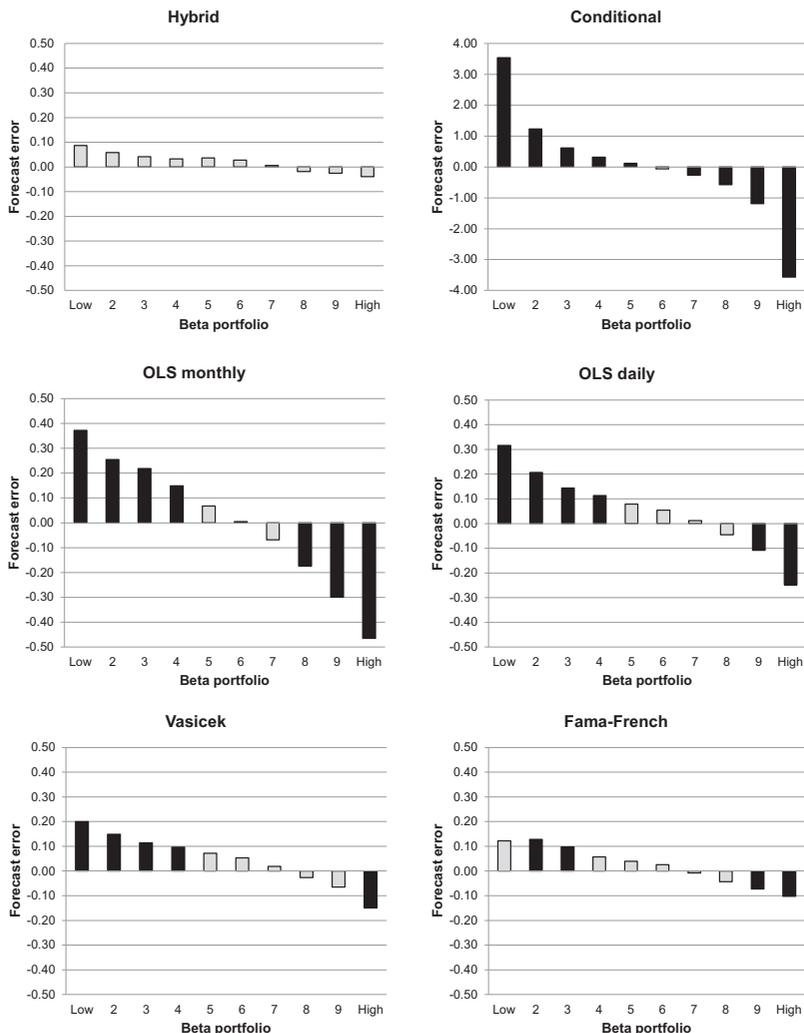
Figure 2
Average forecast errors for beta portfolios
This figure shows average forecast errors for decile portfolios formed by sorting all NYSE/AMEX/NASDAQ-listed stocks on their predicted beta. Beta forecasts are obtained using the procedure described in the text. The first forecast is based on data from August 1964 to July 1984, and the last forecast uses data up to December 2010. Ex ante portfolio betas are measured as the value-weighted average of the beta forecasts. The beta portfolios are held for one month and rebalanced at the end of the month. Ex post portfolio betas are estimated by running a regression of daily value-weighted portfolio returns on a constant and the daily market return over the next year. Forecast errors are defined as the difference between ex post and ex ante portfolio betas and are averaged over time. Black bars represent the portfolios for which the difference in the MSEs produced by the hybrid estimator and an alternative beta approach is statistically significant at the 5% level according to the test of Diebold and Mariano (1995).

estimator are smaller and do not exhibit such a pronounced pattern across deciles. The Diebold and Mariano (1995) test confirms that existing methods generate significantly larger MSEs for the extreme beta portfolios than the hybrid model.

Why does the hybrid estimator perform better for stocks with extreme beta estimates than other methods? We answer this question by relating the cross-sectional spread in betas to measurement error. More specifically, we regress the squared deviation of beta forecasts from their cross-sectional mean on a set of firm characteristics and on a firm's idiosyncratic volatility. Our motivation for doing so is that from a theoretical point of view, we do not expect a relation between idiosyncratic risk and the spread in betas after controlling for firm characteristics that are known to drive variation in betas. Empirically, however, a positive relation may exist because higher idiosyncratic risk increases the standard error of beta estimates (all else equal), leading to more extreme sample estimates of beta. However, with shrinkage, higher idiosyncratic risk and therefore noisier sample betas imply that less weight is given to the sample estimate of beta and more weight is assigned to the prior. Thus, if shrinkage reduces measurement error in betas, we expect the positive relation between idiosyncratic risk and dispersion in betas to be weaker for shrinkage estimates of beta than for sample estimates of beta. The more precise the prior is, the more effective the shrinkage, and the weaker the relation between idiosyncratic risk and cross-sectional variation in betas will be.

The results reported in Table 6 confirm that existing estimators generate more extreme betas for stocks with higher idiosyncratic volatility, even after controlling for various firm characteristics. The standardized coefficient on idiosyncratic risk is largest for standard rolling window betas and for conditional betas estimated using time-series regressions. The Vasicek and Karolyi shrinkage methods and the Fama-French approach do a better job but still produce extreme beta estimates driven by sampling error. In contrast, for the hybrid estimator, we do not find a significant relation between idiosyncratic risk and cross-sectional spread in betas, which indicates that the cross-sectional dispersion in hybrid betas is largely unrelated to measurement noise.

In sum, we find that the hybrid model produces more accurate beta forecasts than alternative approaches because shrinkage toward a fundamentals-based prior corrects the tendency of rolling sample estimators to overpredict at high beta estimates and underpredict at low estimates. Existing shrinkage estimators and methods that group stocks into portfolios offer only limited improvement over rolling estimators and yield significantly larger prediction errors than the hybrid method.

4.3 Decomposition of hybrid beta forecasting performance

The analysis in the preceding section leaves us with one final question: why is shrinkage in the hybrid model more effective than in traditional methods? The main features of the hybrid model that set it apart from other shrinkage

Table 6
Determinants of cross-sectional spread in beta forecasts

	Constant	ME	BE/ME	A/ME	OPL	MOM	IVOL
Hybrid	0.16	−0.01	−0.04	0.01	0.00	0.09	0.01
	(75.38)	(−1.43)	(−5.09)	(0.81)	(0.13)	(7.54)	(0.98)
Conditional	4.86	−0.40	−2.64	0.93	−0.03	1.44	5.24
	(45.80)	(−6.24)	(−19.29)	(12.12)	(−7.22)	(5.70)	(18.95)
OLS monthly	0.39	0.00	−0.01	−0.05	−0.00	0.00	0.22
	(35.27)	(0.46)	(−0.77)	(−2.79)	(−1.38)	(0.18)	(7.78)
OLS daily	0.33	0.02	−0.02	−0.04	0.00	0.06	0.24
	(62.20)	(3.11)	(−3.13)	(−7.52)	(2.15)	(2.93)	(14.66)
Vasicek	0.22	0.01	−0.02	−0.03	0.00	0.03	0.09
	(45.29)	(2.65)	(−2.28)	(−5.89)	(2.38)	(3.14)	(8.95)
Karolyi	0.23	0.02	−0.01	−0.03	0.00	0.04	0.10
	(46.60)	(3.30)	(−1.91)	(−7.57)	(1.71)	(3.02)	(9.50)
Fama-French	0.17	0.03	−0.01	−0.02	0.02	0.00	0.11
	(37.62)	(4.94)	(−2.00)	(−5.05)	(4.63)	(2.32)	(10.34)

This table reports estimation results for a pooled OLS regression of the squared deviation of beta forecasts from their cross-sectional mean on a number of firm characteristics:

$$(\beta_{it}^F - \bar{\beta}_t^F)^2 = \gamma_0 + \gamma_1 W_{it} + \nu_{it},$$

where β_{it}^F is the beta forecast for firm i formed using data up to time t, $\bar{\beta}_t^F$ is the cross-sectional average of these beta forecasts, and W_{it} is a vector that contains the firm characteristics size (ME), book-to-market (BE/ME), financial leverage (A/ME), operating leverage (OPL), momentum (MOM), and idiosyncratic volatility (IVOL). Beta forecasts are constructed using the procedure described in the text. We use the log of firm size, book-to-market, financial leverage, and idiosyncratic volatility and standardize all characteristics by subtracting the cross-sectional mean and dividing by the cross-sectional standard deviation in each month. The t-statistics in parentheses are based on standard errors that are clustered by firm and by month following the procedure of Thompson (2011).

estimators are the specification and estimation of the prior. In this section, we study the contribution of both factors to the outperformance of the hybrid estimator. We do so by assessing the forecast accuracy of a number of simplified beta specifications that omit one or more key elements from the full-fledged hybrid model.

We start by dropping conditioning variables from the prior model to assess the importance of incorporating firm fundamentals and economic state variables. Table 7 reports forecasting results for hybrid beta specifications that omit the firm fundamentals, the macroeconomic instrument, or both sets of conditioning variables. Column 2 shows that excluding all instruments increases MSEs by approximately 20% relative to the original hybrid model in Column 1. Dropping firm fundamentals and the macro variable separately (Columns 3 and 4) indicates that both types of instruments matter but that characteristics play the most important role in producing better beta forecasts. These results highlight the importance of incorporating prior knowledge about fundamentals in the estimation of beta and thereby explain why the hybrid estimator beats the Vasicek estimator that uses a common prior that does not exploit information in firm characteristics.

The shrinkage estimator of Karolyi (1992) incorporates more economic information than the Vasicek approach by forming a portfolio-level prior. Nevertheless, average shrinkage weights in the Karolyi estimator are still

Table 7
Decomposition of hybrid beta forecasting performance

Fundamentals	Yes	No	Macro	Firm	Yes	Yes	Yes	Yes
Estimation	Bayesian	Bayesian	Bayesian	Bayesian	TS OLS	Pooled OLS	Bayesian	Pooled OLS
Window	½-Year D	½-Year D	½-Year D	½-Year D	½-Year D	½-Year D	5-Year M	½-Year D
Shrinkage	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Implicit
Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: S&P 100 stocks								
Monthly MSE	0.0604	0.0723	0.0676	0.0652	0.1026	0.1074	0.0919	0.0674
Ratio	1.00	1.20	1.12	1.08	1.70	1.78	1.52	1.12
NW *t*-stat		(6.50)	(4.48)	(4.18)	(9.34)	(18.14)	(15.49)	(4.33)
One-year MSE	0.0513	0.0622	0.0588	0.0554	0.1038	0.0956	0.0756	0.0643
Ratio	1.00	1.21	1.15	1.08	2.02	1.86	1.47	1.25
NW *t*-stat		(2.88)	(1.95)	(1.80)	(7.10)	(4.66)	(4.11)	(2.41)
Five-year MSE	0.0638	0.0731	0.0693	0.0661	0.1402	0.1107	0.0785	0.0960
Ratio	1.00	1.15	1.09	1.04	2.20	1.74	1.23	1.50
NW *t*-stat		(2.14)	(1.69)	(1.31)	(5.02)	(2.95)	(1.84)	(3.92)
Panel B: All stocks								
One-year MSE	0.0970	0.1156	0.1081	0.1058	0.1418	0.1493	0.1424	0.1160
Ratio	1.00	1.19	1.11	1.09	1.46	1.54	1.47	1.20
NW *t*-stat		(3.41)	(2.89)	(2.43)	(6.92)	(5.36)	(7.51)	(5.22)
Five-year MSE	0.0946	0.1103	0.1064	0.0995	0.1730	0.1402	0.1219	0.1226
Ratio	1.00	1.17	1.13	1.05	1.83	1.48	1.29	1.30
NW *t*-stat		(2.63)	(2.04)	(1.60)	(7.98)	(2.55)	(3.61)	(3.43)

This table reports the mean squared error (MSE) of monthly, yearly, and five-year out-of-sample beta forecasts for alternative specifications of the hybrid beta model. Column 1 corresponds to the full-fledged hybrid estimator in Equation (8). Column 2 reports MSEs for a specification that excludes both the firm characteristics and the macroeconomic variable from the prior. The model in Column 3 includes only the macroeconomic instrument as prior information, and the specification in Column 4 contains only firm characteristics. Column 5 corresponds to a specification in which the prior model is estimated by running a time-series regression for each firm. Column 6 is a specification in which the parameters of the prior model are estimated using a pooled OLS regression. In Column 7 the prior is combined with rolling sample betas estimated using a five-year window of monthly returns instead of a semiannual window of daily returns. The estimator in Column 8 employs an implicit form of shrinkage by directly incorporating the semiannual rolling beta as an additional conditioning variable in the panel regression in Equation (5), which is estimated by pooled OLS. For each specification, the table presents the value-weighted MSE (averaged over time), as well as the ratio of the MSE relative to the MSE of the original hybrid model in Column 1. We further report *t*-statistics for a Diebold and Mariano (1995) test that the MSEs generated by the hybrid estimator and an alternative specification are equal. The *t*-statistics are based on Newey-West (1987) standard errors with lag length equal to 1.5 times the number of months in the forecasting period.

lower than in the hybrid approach (0.20 versus 0.65, respectively) because grouping stocks into portfolios leads to a loss of information in firm-level betas.[19] Specifically, in the Karolyi approach, the prior variance is computed as the cross-sectional variance of beta within each of the 48 industries defined by Fama and French (1997). If industry classification were the only relevant determinant of beta, cross-sectional variation in betas within each industry portfolio would be small and the prior would receive a large weight. In practice, however, industry classification is only one of many potential drivers of beta and intra-industry dispersion in betas is large. Consequently, prior precision is low and relatively little shrinkage is applied.

Next, we shed more light on the importance of the procedure used to estimate the prior model. Prior betas in the Vasicek and Karolyi methods are estimated using an OLS regression for each stock or portfolio, whereas prior betas in the hybrid model are obtained from a Bayesian panel regression. We demonstrate the benefits of the Bayesian procedure by studying the forecasting performance of hybrid beta specifications in which the fundamentals-based prior is estimated using either time-series or pooled OLS regressions. Column 5 in Table 7 shows that estimating the prior using time-series regressions (i.e., using conditional betas as prior) yields significantly larger forecast errors than those generated by the original hybrid model in Column 1 at both long and short horizons. Column 6 shows that estimating the prior using a pooled OLS regression also yields forecast errors that are more than 50% larger than those for the model estimated using Bayesian methods.

The main difference between the pooled model in Column 6 and the hybrid model in Column 1 is the δ_{0i} coefficient on the unscaled market factor in the prior specification in Equation (5), which is assumed to be constant across firms in the pooled OLS estimation. In contrast, in the Bayesian approach this parameter is treated as firm specific and captures cross-sectional heterogeneity in betas unrelated to the fundamentals included in the prior model. The large cross-sectional spread in δ_{0i} in Table 1 highlights the importance of allowing for this flexibility in the prior to reduce misspecification bias.

In addition to shrinking rolling betas toward an economic prior, our hybrid approach differs from the widely used procedure of Fama and MacBeth (1973) by estimating the rolling sample betas using a semiannual window of daily returns instead of a five-year window of monthly returns. We quantify the impact of data frequency and window length on the forecasting performance of the hybrid estimator by combining fundamentals-based prior betas with

---

[19] Apart from a difference in prior precision, the only other possible reason for the lower shrinkage weights in the Vasicek and Karolyi methods is the precision of the rolling beta. The rolling beta in the hybrid model is estimated over a shorter (semiannual) window than the one-year beta used in the Vasicek and Karolyi estimators. Estimating Vasicek and Karolyi betas using a semiannual window leads to slightly higher shrinkage weights (0.24 for Vasicek and 0.27 for Karolyi), but this increase in shrinkage intensities does not translate into better beta forecasts.

five-year rolling betas estimated using monthly returns. The results in Column 7 of Table 7 show that the use of five-year monthly rolling betas leads to a sharp increase in forecast errors. MSEs for the one-month and one-year forecasts are about 50% higher than those for the original hybrid model with daily rolling sample betas, and five-year MSEs increase by 25%. These findings are consistent with the work of Andersen et al. (2006), who show that the use of higher-frequency returns increases the precision of beta estimates. Moreover, because of the shorter estimation window, the rolling betas in the hybrid model are timelier and thus better suited to pick up short-term fluctuations in betas.[20]

Finally, we consider a specification that employs an implicit form of shrinkage as an alternative to the formal shrinkage framework in Equation (8). This simplified model directly incorporates the semiannual rolling beta as an additional conditioning variable in the panel regression in Equation (5) and is estimated by pooled OLS. Results in the last column of Table 7 show that this model underperforms the full-fledged hybrid approach at all horizons. The reasons for this underperformance are twofold. First, the model does not allow for firm-specific shrinkage because the loading on the rolling beta is pooled across stocks. Second, including rolling betas in the pooled regression does not adequately capture the effect of omitted conditioning variables, because it is unlikely that the unobserved heterogeneity in betas is an exact linear function of rolling betas. Ghysels (1998) and Harvey (2001) demonstrate that conditional betas can be biased when the set of conditioning variables is incomplete or when the functional form of conditional expectations is misspecified.

In sum, based on the results in Table 7, we identify three reasons for the outperformance of the hybrid beta estimator over existing estimators. First, by assigning a unique prior to each firm that incorporates a broad set of economic conditioning information, the hybrid model is more effective in reducing estimation noise in betas than conventional shrinkage estimators. Second, the Bayesian approach used to estimate the fundamentals-based prior yields a better bias-variance trade-off than frequentist estimation methods, because it allows for firm-level heterogeneity in some parameters of the prior model to reduce bias, while pooling other parameters to increase precision. Third, by using a semiannual estimation window of daily returns instead of a five-year window of monthly returns, the rolling betas in the hybrid model strike a better balance between timeliness and efficiency than those obtained from the classic rolling sample estimator of Fama and MacBeth (1973).

---

[20] We disentangle the effects of the different data frequency and window length by considering a hybrid model that uses five-year rolling betas estimated with daily returns. Unreported results show that estimating rolling sample betas using a five-year window of monthly returns instead of a five-year window of daily returns increases MSEs by roughly 30% at all forecast horizons. Furthermore, we find that choosing a five-year window of daily returns instead of a semiannual window increases one-month MSEs by 30%, one-year MSEs by 20%, and five-year MSEs by 5%.

5. Market-Neutral Minimum Variance Portfolios

We now explore how the differences in forecast accuracy of the various beta estimators affect the performance of optimal portfolios. Accurate beta forecasts are important for portfolio management because it is common to impose a factor structure on the covariance matrix of stock returns. A factor structure helps to obtain more precise estimates of the covariance matrix when the asset universe is large as it reduces the number of parameters that need to be estimated.

Besides the use of factor models, many other approaches have been proposed to improve the out-of-sample performance of optimized portfolios, such as using shrinkage estimators of the covariance matrix (Ledoit and Wolf 2003) and imposing weight constraints (Jagannathan and Ma 2003). While we acknowledge that some of these techniques may lead to better results, our goal here is to run a horse race between beta estimators rather than to compare various optimization procedures.[21] The use of a factor model allows for a clean comparison of the relative performance of alternative beta estimators because portfolio weights directly depend on predicted betas.

Following Ghysels and Jacquier (2007), our objective is to construct a market-neutral portfolio with minimum return variance. This situation resembles that of a hedge fund manager who follows a typical market-neutral strategy and therefore wants to neutralize the market risk of her portfolio. Because expected returns do not enter the optimization when constructing minimum variance portfolios, differences in portfolio weights only reflect the effect of differences in covariance forecasts.

The setup of the portfolio optimization is as follows. First, we obtain out-of-sample beta forecasts using the procedure explained in Section 4. Following Chan, Karceski, and Lakonishok (1999), we exclude micro-caps and focus on stocks with market capitalization above the 20th percentile of NYSE market cap. We use the beta forecasts of these stocks to predict the covariance matrix implied by the single-factor model

$$S_{t+1|t} = s^2_{Mt+1|t} B_{t+1|t} B'_{t+1|t} + D_{t+1|t}, \qquad (14)$$

where $B_{t+1|t}$ is the $N_t \times 1$ vector of beta forecasts and $s^2_{Mt+1|t}$ is the forecast of the variance of the excess market return. $D_{t+1|t}$ is a diagonal matrix that contains forecasts of the residual variances $d^2_{it+1|t}$. As in Ghysels and Jacquier (2007), we calculate the residuals consistent with the various beta estimates as $R_{it} - \hat{\beta}_{it} R_{Mt}$. We compute the market variance and idiosyncratic variances based on the same window and data frequency that we use to estimate sample betas.

---

[21] Jagannathan and Ma (2003) show that imposing restrictions on weights can help to lower portfolio variance by mitigating the consequences of forecast errors in covariances. We have a different objective. Instead of correcting the outputs of the optimization by imposing constraints on portfolio weights, we want to measure the effect of improving the inputs by reducing estimation errors in the second moments of returns. Chan, Karceski, and Lakonishok (1999) point out that imposing weight constraints makes it hard to assess the economic impact of forecast errors in second moments because the weights for a constrained portfolio are complicated functions of the forecasts.

For instance, for the Vasicek estimator, we compute market and idiosyncratic variances using a one-year window of daily returns ending in month t and take these estimates as forecasts for month $t+1$.

Subsequently, we use the forecasts of the covariance matrix to construct market-neutral minimum variance portfolios by choosing the portfolio weights that solve the following problem:

$$\min_{w_t} \quad w_t' S_{t+1|t} w_t, \tag{15}$$

$$\text{s.t.} \quad \sum_i w_{it} = 1,$$

$$\sum_i w_{it} \beta_{it+1|t} = 0.$$

The first constraint implies that the portfolio is fully invested. The second constraint states that the ex ante portfolio beta should be zero. We obtain a closed-form solution to this optimization problem along the lines of Scherer (2011),

$$w_{it} = \frac{a}{d_{it+1|t}^2} \left(1 - b\beta_{it+1|t}\right), \tag{16}$$

for constants $a > 0$ and $0 < b < 1$ (see Appendix B). Equation (16) shows that, ceteris paribus, stocks with more extreme beta forecasts receive the largest weight (in absolute terms) in the optimal portfolio.[22] Consequently, the portfolio is particularly sensitive to forecast errors in the highest and lowest betas. To satisfy the market neutrality constraint, stocks with betas lower than $1/b$ receive positive weights, while firms with betas higher than $1/b$ are sold short.[23] We rebalance the portfolio every month because hedge funds typically have a short investment horizon.

Our criterion for evaluating the quality of out-of-sample beta forecasts is the realized beta of the optimized portfolio, which should be close to the target of zero. In contrast, the realized variance of the portfolio is not very informative about the accuracy of beta forecasts because it mainly depends on idiosyncratic risk, not on systematic risk. In fact, if the portfolio is truly market neutral ex post, the portfolio variance is completely idiosyncratic by definition. Consequently, if the beta forecasts are accurate, the systematic risk of the portfolio should be a very small fraction of total portfolio risk.

Table 8 reports the results. We find that the portfolio based on hybrid beta forecasts is the only portfolio that meets the objective of being market neutral

---

[22] Equation (16) shows that weights also depend on idiosyncratic variances. For a given beta estimate, the higher the idiosyncratic variance of a stock, the lower its absolute weight in the minimum variance portfolio.

[23] In Appendix B we demonstrate that the constant b decreases with the cross-sectional variation in betas. Hence, for methods that generate widely dispersed beta estimates, only stocks with very high betas receive negative weights.

Table 8
Ex post risk exposure of market-neutral minimum variance portfolios

	Hybrid	Conditional	OLS monthly	OLS daily	Vasicek	Karolyi	Fama-French
β	0.04	0.75	0.28	0.22	0.19	0.19	0.11
t-stat	(1.14)	(34.83)	(8.06)	(6.31)	(5.31)	(5.34)	(3.06)
σ_ϵ (%)	9.95	9.96	9.88	9.85	10.09	10.12	10.25
σ_{total} (%)	9.97	15.66	10.82	10.43	10.52	10.55	10.41
$\sigma_{syst}^2/\sigma_{total}^2$ (%)	0.33	59.57	16.63	10.92	8.01	8.12	3.08
Min w_i (%)	−0.50	−0.32	−1.55	−0.38	−0.45	−0.44	−0.60
Max w_i (%)	1.90	0.63	1.84	1.27	1.39	1.38	1.35
$w_i < 0$ (%)	48.60	11.58	38.88	39.72	40.99	40.45	42.38

This table reports the ex post risk exposure of minimum variance portfolios that are ex ante market neutral. The portfolios are constructed using NYSE/AMEX/NASDAQ stocks with market capitalization above the 20th percentile of NYSE market cap. The optimization uses forecasts of the covariance matrix of returns produced by the beta models discussed in the text. The first forecast is based on data from August 1964 to July 1984, and the last forecast uses data up to November 2011. The portfolios are formed at the end of each month and their realized return over the next month is recorded. The first set of rows reports the ex post risk exposure of the optimized portfolios. β is the realized market beta, σ_ϵ is the annualized idiosyncratic volatility, and σ_{total} is the annualized total volatility. The table also reports the ratio of systematic variance to total portfolio variance. The t-statistics for the null hypothesis that the ex post portfolio beta is zero are in parentheses and based on Newey-West standard errors. The second set of rows presents descriptive statistics for the weights assigned to the stocks in the optimized portfolios, averaged over time. The table shows the minimum and maximum weights and the percentage of negative weights in the portfolios.

ex post, with a realized beta of 0.04 that is insignificantly different from zero (t-stat = 1.14). Portfolios based on all other estimators have a realized beta that is economically large and statistically significant. For instance, the realized beta of the portfolio corresponding to the daily rolling window estimator equals 0.22 (t-stat = 6.31), which implies a strong upward bias relative to the target of zero. The Vasicek and Karolyi shrinkage methods only yield a marginal improvement over the rolling window estimators. The Fama-French procedure also fails to create a portfolio that is market neutral ex post but does yield a lower ex post beta than these existing shrinkage approaches (0.11 versus 0.19). As in the portfolio-level forecasts in Figure 2, the Fama-French approach fares better than existing estimators in the portfolio-level application in Table 8 because the biases in firm-level betas that arise from assigning the same beta to all stocks in a portfolio tend to cancel out at the portfolio level.[24] However, in the firm-level forecasting tests in Tables 4 and 5, where up- and downward biases in betas cannot offset each other, the Fama-French approach works no better than conventional shrinkage methods.

To get a better understanding of the results in Table 8, we relate the weights of the stocks in the optimized portfolio to their beta forecasts. Specifically, at the

---

[24] More precisely, the Fama-French procedure assigns the same beta to all stocks in each of the 100 size-beta sorted portfolios. As a result, the assigned beta is too high for some stocks, while it is too low for others. However, because all stocks that belong to the same size-beta portfolio receive the same beta, it follows from Equation (16) that all of them are either bought or sold short in the market-neutral minimum variance strategy. Consequently, upward and downward biases in the betas of these stocks can partially offset each other at the portfolio level.

end of each month, we form ten portfolios by sorting stocks on their predicted betas. We then calculate the sum of the minimum variance portfolio weights of the stocks in each beta decile. Figure 3 plots the summed minimum variance weights (black diamonds) and the average forecast error (gray bars) for each decile.[25] The figure shows that the hybrid method outperforms other estimators in the portfolio optimization because it sharply reduces forecast errors for those stocks that receive most weight, that is, the firms with the highest and lowest ex ante estimates of beta. Consistent with Equation (16), we observe that the minimum variance strategy assigns large positive weights to low-beta stocks and takes negative positions in high-beta stocks. Existing beta estimators underestimate the low betas and overestimate the high betas. Consequently, the stocks that are bought have a larger ex post beta than predicted, while the stocks that are sold short have a smaller realized beta than expected. Both effects contribute to the positive market risk exposure of the overall portfolios in Table 8.

Table 8 further reports the fraction of total portfolio variance that is systematic. As expected, the realized portfolio variance is mostly idiosyncratic. For the portfolios formed using rolling betas, systematic risk makes up about 10.9% (daily OLS) to 16.6% (monthly OLS) of total risk. For the Vasicek and Karolyi estimators, the ratio of systematic to total portfolio variance is approximately 8%, and for the Fama-French approach, it equals 3.1%. The ratio is lowest (0.3%) for the portfolio created using the hybrid estimator, consistent with the better predictive ability of the hybrid model.

The second set of rows in Table 8 presents descriptive statistics for the weights assigned to the stocks in the optimized portfolios. We find that none of the beta models yield extreme positions. For instance, the lowest weight assigned to a stock in the portfolio based on Vasicek betas is -0.45% and the maximum weight is 1.39%. Hence, the significant market risk exposure of the portfolios formed using existing beta estimators is not driven by extreme positions in a few individual stocks.

In summary, the results from the portfolio optimization highlight the practical consequences of inaccurate beta estimation. A hedge fund manager who constructs a portfolio that is ex ante market neutral according to existing beta estimators, may find out ex post that this portfolio has a significant exposure to market risk. We show that traditional estimators fail to produce a market-neutral portfolio due to systematic errors in beta forecasts. Because our hybrid estimator sharply reduces estimation errors in extreme betas, the portfolio constructed based on covariance forecasts from the hybrid model is the only portfolio that is market neutral ex post.

---

25 The plot for the Karolyi approach resembles the plot for the Vasicek estimator and is omitted for brevity.
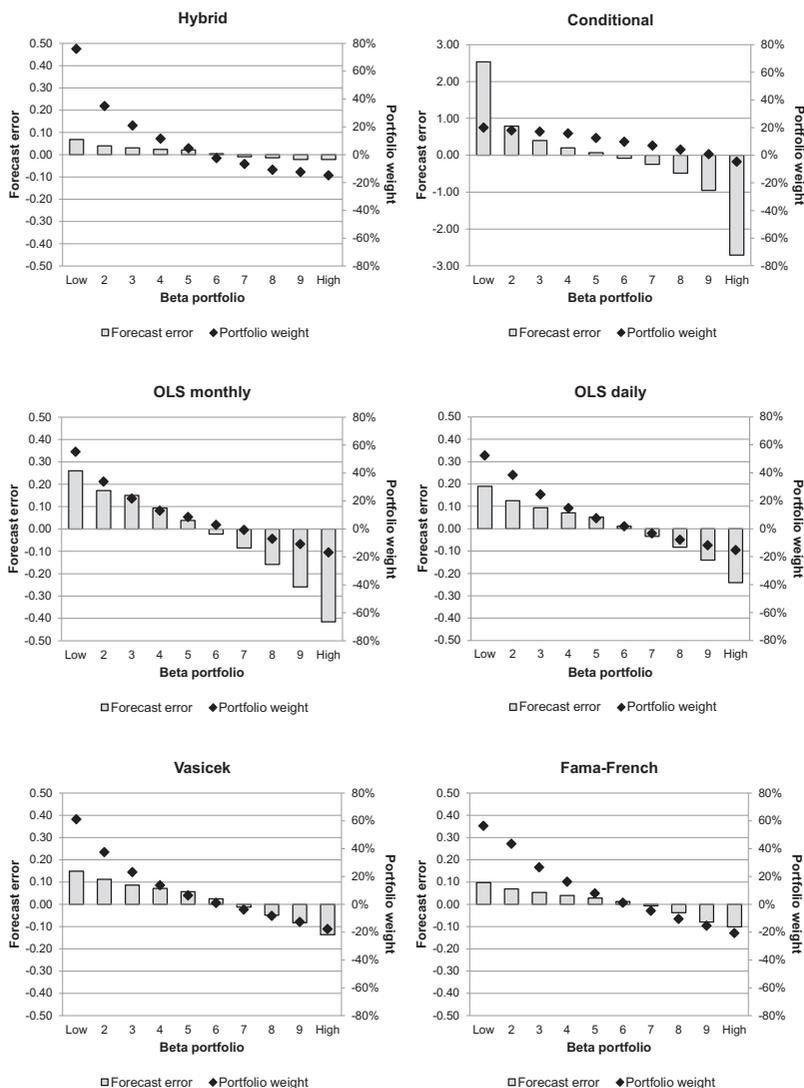
Figure 3
Average forecast errors and minimum variance portfolio weights
This figure shows average forecast errors (gray bars) and minimum variance weights (black diamonds) for decile portfolios formed by sorting stocks on their predicted beta. The portfolios are constructed using NYSE/AMEX/NASDAQ-listed stocks with market capitalization above the 20th percentile of NYSE market cap. Forecast errors are defined as the difference between ex post and ex ante portfolio betas. The minimum variance optimization uses predictions of the covariance matrix of stock returns that are computed based on a one-factor model and the firm-level beta forecasts. The beta portfolios and the market-neutral minimum variance portfolios are held for one month and rebalanced at the end of the month. For each beta decile, the figure shows the average forecast error and the sum of the minimum variance portfolio weights of the stocks in that decile, averaged over time.

6. Cross-Sectional Tests of the Risk-Return Relation

In this section we examine whether the use of hybrid betas changes the outcome of cross-sectional tests of the risk-return relation. This analysis is motivated by the large body of evidence that contradicts the prediction of the CAPM that beta is priced in the cross-section.[26] Following Fama and MacBeth (1973), most of these studies employ rolling window betas and aggregate stocks into portfolios. However, our results in Section 4 show that assigning rolling portfolio betas to individual stocks leads to inaccurate forecasts of firm-level betas. Moreover, Ang, Liu, and Schwarz (2010) show that forming portfolios decreases the precision of risk premium estimates because it results in a loss of information in individual stock betas.[27] We therefore feel that it is important to revisit the risk-return relation using individual stocks and a novel approach to measure betas.

We study the risk-return relation by estimating monthly cross-sectional regressions of stock returns on beta estimates and a set of controls,

$$R_{it} = \lambda_{0t} + \lambda_{1t}\beta_{it} + \lambda_{2t}W_{it} + \upsilon_{it}, \tag{17}$$

where W_{it} is a vector of firm characteristics. We calculate the Fama-MacBeth estimator of the risk premium and its standard error from the time series of monthly coefficient estimates and correct the standard error for measurement error in betas following Shanken (1992). We also adjust the risk premium estimate itself for a bias that can occur when the first-stage beta estimation uses the month in which the second-stage cross-sectional regression is run. In particular, Chordia, Goyal, and Shanken (2012) show that the risk premium estimate is biased if measurement errors in betas are contemporaneously cross-sectionally correlated with the error terms in the first-stage regression. We control for this bias by reestimating all beta models T times, each time excluding the return in the month of the cross-sectional regression from the estimation (T is the number of months in our sample). Thus, the beta used in the cross-sectional regression for month t is estimated using the full sample of returns, except for the return in month t. The aim of this procedure is to ensure that the contemporaneous correlation between the measurement errors in the betas at time t and the time t regression residuals is zero, thereby avoiding the potential bias in risk premium estimates.[28]

Table 9 reports estimates of the market risk premium for the different beta estimators.[29] The left-hand side of the table corresponds to univariate

---

[26] See Fama and French (2004) for a comprehensive overview of this literature.

[27] In addition, they demonstrate that running cross-sectional regressions using the Fama-French estimates of individual stock betas effectively boils down to using the size-beta sorted portfolios themselves as test assets.

[28] We are grateful to an anonymous referee for suggesting this procedure.

[29] For completeness, we also report the average of the monthly cross-sectional R^2s. However, the mean R^2 should be interpreted with caution because Kan, Robotti, and Shanken (2013) point out that it can be high even if the ex ante risk premium is zero. This can happen when ex post risk premia are positive for some months and negative for others. Therefore, the average R^2 should not be used to infer whether beta is priced in the cross-section.

Table 9
Cross-sectional asset pricing tests

	No controls			With controls		
	λ_0	λ_1	Adj. R^2 (%)	λ_0	λ_1	Adj. R^2 (%)
Hybrid - Bayesian	0.28	0.47	3.61	0.26	0.49	6.76
	(1.70)	(2.08)		(1.55)	(2.23)	
Hybrid - Pooled OLS	0.45	0.28	1.93	0.56	0.23	5.57
	(2.33)	(1.09)		(2.90)	(0.82)	
Hybrid - TS OLS	0.59	0.16	2.99	0.67	0.15	6.16
	(3.99)	(1.01)		(4.48)	(0.97)	
Conditional	0.57	0.15	1.58	0.65	0.13	5.05
	(4.94)	(1.32)		(4.88)	(1.13)	
OLS monthly	0.61	0.11	2.65	0.70	0.08	5.63
	(4.74)	(0.79)		(5.00)	(0.59)	
OLS daily	0.73	0.00	2.98	0.79	0.01	6.04
	(4.70)	(0.04)		(5.18)	(0.10)	
OLS daily - Dimson	0.64	0.10	3.03	0.73	0.06	6.20
	(4.46)	(0.70)		(4.83)	(0.47)	
Vasicek	0.72	0.01	3.00	0.77	0.03	6.45
	(4.78)	(0.11)		(5.26)	(0.14)	
Karolyi	0.72	0.01	3.04	0.76	0.03	6.48
	(4.76)	(0.08)		(5.24)	(0.13)	
Fama-French	0.73	0.00	2.75	0.78	0.03	6.16
	(4.70)	(0.02)		(5.32)	(0.15)	

This table reports Fama-MacBeth estimates for the cross-section of stocks listed on the NYSE, AMEX, and NASDAQ. In the first step, betas are estimated using the procedures described in the text. In the second stage, monthly cross-sectional regressions are run of excess stock returns on the beta estimates and a set of control variables to obtain an estimate of the market risk premium λ_1:

$$R_{it} = \lambda_{0t} + \lambda_{1t}\beta_{it} + \lambda_{2t}W_{it} + \upsilon_{it},$$

where W_{it} is a vector of control variables that includes the firm characteristics size, book-to-market, momentum, financial leverage, operating leverage, and idiosyncratic volatility. The left-hand side of the table shows estimation results for cross-sectional regressions without control variables, and the right-hand panel corresponds to cross-sectional regressions with controls. Coefficient estimates are corrected for a contemporaneous EIV bias following the procedure described in the text. Shanken-adjusted t-statistics are in parentheses. The reported adjusted R^2 is the average of the monthly cross-sectional adjusted R^2s. The sample period is August 1964 to December 2011.

regressions of stock returns on betas. For hybrid betas, we find a positive risk premium estimate of 0.47% that is statistically significant (t-stat = 2.08) and close in magnitude to the average excess market return over the sample period (0.42%). In contrast, betas obtained from all other estimators do not carry a significant price of risk. For instance, the risk premium on rolling sample betas varies from 0.00% per month for one-year daily OLS betas to 0.11% for five-year monthly OLS betas.

The right-hand panel in Table 9 presents results for Fama-MacBeth regressions that include firm characteristics as control variables, motivated by the conclusion of Fama and French (1992) that beta is no longer priced after controlling for characteristics. We add firm size, book-to-market, momentum,

financial leverage, operating leverage, and idiosyncratic volatility to the regressions. Controlling for these characteristics has little impact on the estimates of the market risk premium. In particular, hybrid betas continue to carry a positive price of risk that is significant at the 5% level. Consistent with prior literature, the risk premium on rolling betas remains insignificant. As pointed out by Buss and Vilkov (2012), it is hard to detect a significant risk-return relation with rolling betas due to the upward bias in high beta estimates and downward bias in low betas (see Figure 2). Because of the systematic underestimation of low betas, returns on low-beta stocks are too high relative to the predictions of the CAPM, while the returns on high-beta stocks are too low due to the overestimation of high betas. As a result, the risk-return relation estimated with rolling betas is too flat. In line with Fama and French (1992), we find that assigning portfolio betas to individual stocks also yields an insignificant risk premium estimate.

Table 9 further shows that existing shrinkage methods yield little improvement over rolling estimators in the cross-sectional tests, even though they improve forecast accuracy. This discrepancy in performance reflects the different nature of the forecasting exercise and the asset pricing test. Specifically, the predictability tests in Section 4 are time series tests in which it is crucial that beta estimates are not too extreme. By squaring forecast errors, the MSE criterion penalizes forecasts that are far off target. As a result, methods that shrink extreme beta estimates toward a central value (as in Vasicek 1973), and thereby limit the cross-sectional dispersion in betas, tend to do better in such tests than standard rolling sample estimators. In contrast, in the asset pricing tests cross-sectional variation in betas is needed to capture the observed spread in returns. Hence, assigning a common prior to all firms does not help in explaining cross-sectional variation in returns.

At first sight, it may seem surprising that the betas estimated using monthly returns (conditional betas and five-year rolling betas) yield larger risk premium estimates than betas computed from daily returns (one-year rolling betas and Vasicek and Karolyi betas), even though daily betas tend to be more precise. However, our finding that betas estimated from lower-frequency data carry a higher price of risk than betas estimated from higher-frequency returns is consistent with prior evidence reported by Handa, Kothari, and Wasley (1989). As pointed out by Kothari, Shanken, and Sloan (1995), one potential explanation for the lower risk premium obtained with higher-frequency betas is nonsynchronous trading, which can lead to biased beta estimates. To account for this bias, we have estimated Dimson (1979)-corrected daily OLS betas, obtained by adding four lags of the market return to the regression. We find that this adjustment leads to higher MSEs in the forecasting tests because the decrease in the efficiency of beta estimates resulting from the additional lags is larger than the reduction in bias. However, the correction could be important for the cross-sectional tests because the bias in daily betas likely varies across stocks (a uniform bias does not distort the risk premium estimate).

Table 9 confirms that correcting daily betas for nonsynchronous trading effects increases the risk premium on daily OLS betas, which becomes similar in magnitude to the premium obtained with monthly OLS betas. Nevertheless, even after the adjustment for nonsynchronous trading, the risk premium on rolling betas remains insignificant.

The beta models that we compare not only differ in terms of data frequency but also in terms of estimation method. We study the effect of the estimation approach by comparing the risk premium estimate obtained with full-fledged hybrid betas to the premium obtained with hybrid specifications in which the prior model is estimated using either time-series or pooled OLS regressions instead of Bayesian methods. The results in Table 9 confirm that the estimation method plays a crucial role in the asset pricing tests. The model estimated using pooled OLS yields an insignificant risk premium of 0.23% per month. Estimating prior betas using a separate time-series regression for each firm also yields a risk premium that is too small (0.15%) and statistically insignificant.

These findings highlight a fundamental trade-off in cross-sectional asset pricing tests. On the one hand, we must reduce measurement noise in betas to limit errors-in-variables problems. On the other hand, we need sufficient cross-sectional spread in betas to explain variation in returns. Conventional beta estimators fall short on at least one of these dimensions. Conditional betas estimated using time-series regressions exhibit large cross-sectional dispersion but suffer from large measurement errors. Grouping stocks into portfolios or running pooled regressions mitigates noise but leads to a loss of cross-sectional information. Our findings suggest that the hybrid model strikes a better balance between reducing noise and preserving information than existing methods. By pooling some coefficients in the prior and leaving other parameters firm specific, the Bayesian approach not only increases precision but also allows for unobserved heterogeneity in betas. As a result, with hybrid betas we find a positive risk-return relation even after controlling for characteristics.

Our goal is not to resurrect the CAPM. Most characteristics in the Fama-MacBeth regressions are significant, in line with the conclusion of Lewellen and Nagel (2006) that a conditional CAPM cannot fully explain anomalies like the value premium. However, our finding that the market risk premium is positive and significant for hybrid betas contradicts the common belief in the literature that beta is dead and provides a strong rationale for the widespread usage of beta in practice.

7. Conclusion

Firm-specific betas are key inputs for financial decision making but are often strongly time varying and hard to estimate with precision. In this article, we propose a new approach for estimating individual security betas that improves precision by incorporating prior information based on firm fundamentals. Our method is a hybrid of a data-driven rolling sample estimator and the

parametric method for modeling beta dynamics proposed by Shanken (1990). The rolling window approach is robust to misspecification of beta dynamics, while the parametric specification is theoretically appealing because it links beta to economic state variables and firm fundamentals. Our hybrid procedure exploits the strengths of both methods by combining them in a formal shrinkage framework.

We provide direct evidence on the merits of the hybrid model by predicting betas out-of-sample. Our results show that the gains of the hybrid estimator relative to existing methods are economically large and statistically significant. For instance, MSEs produced by conventional shrinkage estimators are 20% larger at the monthly horizon and more than 50% higher for a five-year period than MSEs generated by the hybrid model. The strong outperformance at long horizons supports our conjecture that incorporating prior knowledge about fundamentals can improve beta forecasts because in the long run beta tends toward a level determined by the fundamentals of the firm.

We quantify the economic benefits of the hybrid beta estimator by examining its performance in two applications. First, we consider a portfolio manager who aims to run a market-neutral minimum variance strategy and uses a factor model to forecast return covariances. We find that the minimum variance portfolio based on hybrid betas is the only portfolio that is market neutral ex post. Portfolios based on covariance forecasts derived from other estimators have a significantly positive exposure to market risk due to systematic errors in beta forecasts. Second, we study the impact of hybrid betas on the outcome of cross-sectional tests of the risk-return relation. We find that hybrid betas yield a market risk premium estimate that is significantly positive and in line with theoretical predictions, even after controlling for characteristics. In contrast, existing rolling betas, conditional betas, and shrinkage estimates of beta are not priced in the cross-section.

We identify three key aspects of the hybrid approach that drive its consistent outperformance. First, shrinkage is more powerful in the hybrid model because the prior is unique to each firm and incorporates multiple firm characteristics and economic conditioning variables. Other shrinkage methods are less effective in correcting noisy sample estimates of beta because they employ a common or portfolio-level prior that contains too little economic information. Second, the Bayesian approach we use to estimate the fundamentals-based prior increases precision by pooling the loadings on the conditioning variables, while also allowing for heterogeneity in betas unrelated to fundamentals to mitigate misspecification bias. Alternative models with firm-level conditioning variables estimated using standard frequentist methods produce beta forecasts that are either noisy or biased. Third, the use of mixed frequency data enables us to estimate rolling betas more precisely using daily returns, while at the same time exploiting prior information embedded in monthly fundamentals.

Our results demonstrate that the method used for measuring beta has important implications, both for academic tests and for practical applications.

Given these consequences, it is striking that the five-year rolling window estimator that is widely used in the literature and in practice turns out to be one of the worst predictors of beta. We hope that these findings encourage researchers and practitioners to question the routine use of rolling beta estimators and to consider using a better approach for applications that rely on accurate beta estimates. The hybrid estimator that we propose is useful in many such applications. In the paper we demonstrate its benefits for asset pricing tests and portfolio construction, but the approach should also be well suited for cost-of-capital calculations and for computing abnormal returns in event studies.

The hybrid model can be extended in various directions. For instance, the MIDAS approach of Ghysels, Santa-Clara, and Valkanov (2005) can be implemented to attach optimal weights to the daily returns in the window used for estimating rolling sample betas. In addition, methods other than the Bayesian weighting scheme of Vasicek (1973), such as weighting schemes driven by past forecasting performance, can be used to combine rolling betas and prior betas.[30] Furthermore, while we focus on the estimation of market betas because of the widespread usage of the CAPM, our work can be readily extended to estimate firm-level betas in multifactor models.

Appendix

A. Prior Distributions for Parameters in the Panel Model

In this Appendix we provide details about the prior distributions that we assume for the parameters in the panel model in Equation (5), which we repeat here for convenience:

$$R_{it} = \alpha_i^* + (\delta_{0i} + \delta_{1i} X_{t-1} + \delta_2' Z_{it-1} + \delta_3' Z_{it-1} X_{t-1}) R_{Mt} + \eta_{it}. \tag{A.1}$$

We stack the firm-specific parameters α_i^*, δ_{0i}, and δ_{1i} in the vector θ_i and the pooled parameters δ_2 and δ_3 in the vector ξ. We specify a common prior for the parameters in θ_i:

$$\theta_i \sim \mathcal{N}(\bar{\theta}, \Omega_\theta). \tag{A.2}$$

Imposing very sharp priors on the θ_i parameters by setting the diagonal elements in Ω_θ equal to small values is equivalent to pooling the parameters in θ_i across stocks. Alternatively, specifying flat priors by setting the diagonal elements in Ω_θ equal to very large values corresponds to running a separate time-series regression for each firm. Instead, we treat the vector of prior means $\bar{\theta}$ and the precision matrix Ω_θ^{-1} as unknown parameters and assume the following uninformative prior densities for these hyperparameters

$$\bar{\theta} \sim \mathcal{N}(\kappa, \Omega_\kappa), \tag{A.3}$$

$$\Omega_\theta^{-1} \sim W\left([\psi_\theta S_\theta]^{-1}, \psi_\theta\right), \tag{A.4}$$

where ψ_θ and $[\psi_\theta S_\theta]^{-1}$ are the degrees of freedom parameter and the scale matrix in the Wishart distribution W, respectively. We specify an uninformative prior for $\bar{\theta}$ by setting κ equal to the

---

[30] See Timmermann (2006) for a comprehensive survey and discussion of forecast combination schemes and Andreou, Ghysels, and Kourtellos (2013) for a recent application of these methods. Preliminary analyses show that combination schemes in which weights are inversely proportional to past MSEs do not outperform the weighting scheme of Vasicek (1973). However, more advanced methods to estimate the combination weights may do better.

zero vector and Ω_κ to $100I$, with I denoting the identity matrix. For this flat prior, the conditional posterior mean of $\bar{\theta}$ reduces to the sample average of all of the θ_i. We set ψ_θ equal to the dimension of Ω_θ, because this value gives the lowest possible weight to the prior information (Gelman et al. 2004). The matrix S_θ is the prior mean of Ω_θ^{-1} and is set equal to the identity matrix. This uninformative prior implies that the conditional posterior mean of the covariance matrix Ω_θ is approximately equal to the sample variability of the θ_i around their common mean $\bar{\theta}$.

For the pooled parameters in ξ, we specify an uninformative prior

$$\xi \sim \mathcal{N}\left(\bar{\xi}, \Omega_\xi\right), \tag{A.5}$$

with the vector of prior means $\bar{\xi}$ set equal to zero and the prior covariance matrix Ω_ξ to $100I$. We complete the specification of the model by adding a diffuse prior for the idiosyncratic volatility σ_{η_i}:

$$\sigma_{\eta_i} \propto \frac{1}{\sigma_{\eta_i}}. \tag{A.6}$$

B. Market-Neutral Minimum Variance Portfolios

In this Appendix we solve the optimization problem in Section 5, which can be characterized as

$$\min_w \quad w'Sw, \tag{B.1}$$

$$\text{s.t.} \quad w'\iota = 1,$$

$$w'\beta = 0,$$

where w is the N-vector of portfolio weights, S the covariance matrix of stock returns, and β the N-vector of beta forecasts. The constraints can be written in matrix notation as

$$C'w = c, \tag{B.2}$$

with $C = (\iota \ \beta)$ and $c = (1 \ 0)'$. The general solution to this problem is:

$$w = S^{-1}C\left(C'S^{-1}C\right)^{-1}c. \tag{B.3}$$

Assuming a single-factor structure for S and using the matrix inversion lemma, we obtain

$$S^{-1} = D^{-1} - \frac{s^2}{1+s^2 S_{\beta\beta}}D^{-1}\beta\beta'D^{-1}, \tag{B.4}$$

where we define $S_{gh} = g'D^{-1}h$ for $g, h = \iota, \beta$. D is a diagonal matrix with idiosyncratic variances d_i^2 on the main diagonal and s^2 is the factor variance. This implies that

$$S^{-1}C = D^{-1}C\left(I - \frac{s^2}{1+s^2 S_{\beta\beta}}e_2\left(\beta'D^{-1}C\right)\right), \tag{B.5}$$

where e_2 is a (2×1) vector that selects the second column (i.e., β) from C. The matrix post-multiplication $D^{-1}C$ is of order (2×2). Hence, expression (B.5) shows that the portfolio weights are in a two-dimensional space spanned by the vectors $D^{-1}C$ that have elements $\frac{1}{d_i^2}$ and $\frac{\beta_i}{d_i^2}$. That means the weights can be represented as

$$w_i = \frac{a}{d_i^2}(1 - b\beta_i), \tag{B.6}$$

with

$$a = \frac{S_{\beta\beta}}{S_{\iota\iota} S_{\beta\beta} - S_{\iota\beta}^2},$$

(B.7)

$$b = \frac{S_{\iota\beta}}{S_{\beta\beta}}.$$

(B.8)

Both the numerator and denominator in (B.7) are positive, and therefore $a > 0$. Under the assumptions that the average beta is one and that β_i and $\frac{1}{d_i^2}$ are cross-sectionally independent, for large N the constant b converges in probability to

$$\frac{E\left[\frac{1}{d^2}\right] E[\beta]}{E\left[\frac{1}{d^2}\right] E[\beta^2]} = \frac{1}{1 + \omega_\beta^2},$$

(B.9)

where ω_β^2 represents the cross-sectional variance of betas. We therefore find that $0 < b < 1$.

References

Andersen, T., T. Bollerslev, and N. Meddahi. 2005. Correcting the errors: Volatility forecast evaluation using high-frequency data and realized volatilities. *Econometrica* 73:279–96.

Andersen, T., T. Bollerslev, F. Diebold, and J. Wu. 2000. A framework for exploring the macroeconomic determinants of systematic risk. *American Economic Review* 95:398–404.

———. 2006. Realized beta: Persistence and predictability. In *Advances in econometrics: Econometric analysis of economic and financial time series*. Eds. T. Fomby and D. Terrell. Amsterdam: Elsevier.

Andreou, E., E. Ghysels, and A. Kourtellos. 2013. Should macroeconomic forecasters use daily financial data and how? *Journal of Business and Economic Statistics* 31:240–51.

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

Ang, A., J. Liu, and K. Schwarz. 2010. Using stocks or portfolios in tests of factor models. Working Paper, Columbia University.

Avramov, D., and T. Chordia. 2006. Asset pricing models and financial market anomalies. *Review of Financial Studies* 19:1001–40.

Berk, J., and J. van Binsbergen. 2016. Assessing asset pricing models using revealed preference. *Journal of Financial Economics* 119:1–23.

Black, F., M. Jensen, and M. Scholes. 1972. The capital asset pricing model: Some empirical tests. In *Studies in the theory of capital markets*. Ed. M. Jensen. New York: Praeger.

Bollerslev, T., R. Engle, and J. Wooldridge. 1988. A capital asset pricing model with time-varying covariances. *Journal of Political Economy* 96:116–31.

Buss, A., and G. Vilkov. 2012. Measuring equity risk with option-implied correlations. *Review of Financial Studies* 25:3113–40.

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.

Carlson, M., A. Fisher, and R. Giammarino. 2004. Corporate investment and asset price dynamics: Implications for the cross-section of returns. *Journal of Finance* 59:2577–603.

Chan, L., J. Karceski, and J. Lakonishok. 1999. On portfolio optimization: Forecasting covariances and choosing the risk model. *Review of Financial Studies* 12:937–74.

Chordia, T., A. Goyal, and J. Shanken. 2012. Cross-sectional asset pricing with individual stocks: Betas versus characteristics. Working Paper, Emory University.

Clark, T., and M. McCracken. 2013. Advances in forecast evaluation. In *Handbook of economic forecasting*. Eds. G. Elliott and A. Timmermann. Amsterdam: Elsevier.

Diebold, F., and R. Mariano. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13:253–63.

Dimson, E. 1979. Risk measurement when shares are subject to infrequent trading. *Journal of Financial Economics* 7:197–226.

Engle, R., E. Ghysels, and B. Sohn. 2013. Stock market volatility and macroeconomic fundamentals. *Review of Economics and Statistics* 95:776–97.

Fama, E., and K. French. 1992. The cross-section of expected stock returns. *Journal of Finance* 47:427–65.

———. 1997. Industry costs of equity. *Journal of Financial Economics* 43:153–93.

———. 2004. The capital asset pricing model: Theory and evidence. *Journal of Economic Perspectives* 18:25–46.

———. 2008. Dissecting anomalies. *Journal of Finance* 63:1653–78.

Fama, E., and J. MacBeth. 1973. Risk, return and equilibrium: Empirical tests. *Journal of Political Economy* 71:607–36.

Ferson, W., and C. Harvey. 1998. Fundamental determinants of national equity market returns: A perspective on conditional asset pricing. *Journal of Banking and Finance* 21:1625–65.

———. 1999. Conditioning variables and the cross section of stock returns. *Journal of Finance* 54:1325–60.

Foster, D., and D. Nelson. 1996. Continuous record asymptotics for rolling sample variance estimators. *Econometrica* 64:139–74.

Gelman, A., J. Carlin, H. Stern, and D. Rubin. 2004. *Bayesian data analysis*, 2nd ed. Boca Raton, FL: Chapman and Hall/CRC.

Ghysels, E., and E. Jacquier. 2007. Market beta dynamics and portfolio efficiency. Working Paper, University of North Carolina and HEC Montreal.

Ghysels, E., P. Santa-Clara, and R. Valkanov. 2005. There is a risk-return trade-off after all. *Journal of Financial Economics* 76:509–48.

Ghysels, E. 1998. On stable factor structures in the pricing of risk: Do time-varying betas help or hurt? *Journal of Finance* 53:549–73.

Giacomini, R., and H. White. 2006. Tests of conditional predictive ability. *Econometrica* 74:1545–78.

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

Goyal, A., and I. Welch. 2008. A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies* 21:1455–508.

Graham, J., and C. Harvey. 2001. Theory and practice of corporate finance: Evidence from the field. *Journal of Financial Economics* 60:187–243.

Grundy, B., and J. Martin. 2001. Understanding the nature of the risks and the source of the rewards to momentum investing. *Review of Financial Studies* 14:29–78.

Gulen, H., Y. Xing, and L. Zhang. 2009. Value versus growth: Time-varying expected stock returns. Working Paper, University of Michigan.

Hamada, R. 1972. The effect of the firm's capital structure on the systematic risk of common stocks. *Journal of Finance* 27:435–52.

Handa, P., S. P. Kothari, and C. Wasley. 1989. The relation between the return interval and betas: Implications for the size effect. *Journal of Financial Economics* 23:79–100.

Harvey, C. 2001. The specification of conditional expectations. *Journal of Empirical Finance* 8:573–637.

Harvey, D., S. Leybourne, and P. Newbold. 1997. Testing the equality of prediction mean squared errors. *International Journal of Forecasting* 13:281–91.

Jagannathan, R., and T. Ma. 2003. Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance* 58:1651–83.

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

Kan, R., C. Robotti, and J. Shanken. 2013. Pricing model performance and the two-pass cross-sectional regression methodology. *Journal of Finance* 68:2617–49.

Karolyi, A. 1992. Predicting risk: Some new generalizations. *Management Science* 38:57–74.

Korteweg, A., and M. Sorensen. 2010. Risk and return characteristics of venture capital-backed entrepreneurial companies. *Review of Financial Studies* 23:3738–72.

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

Ledoit, O., and M. Wolf. 2003. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance* 10:603–21.

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. 2010. A skeptical appraisal of asset pricing tests. *Journal of Financial Economics* 96:175–94.

Livdan, D., H. Sapriza, and L. Zhang. 2009. Financially constrained stock returns. *Journal of Finance* 64:1827–62.

Newey, W., and K. West. 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55:703–8.

Pastor, L., and R. Stambaugh. 1999. Costs of equity capital and model mispricing. *Journal of Finance* 54:67–121.

Petkova, R., and L. Zhang. 2005. Is value riskier than growth? *Journal of Financial Economics* 78:187–202.

Scherer, B. 2011. A note on the returns from minimum variance investing. *Journal of Empirical Finance* 18:652–60.

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

———. 1992. On the estimation of beta-pricing models. *Review of Financial Studies* 5:1–34.

Thompson, S. 2011. Simple formulas for standard errors that cluster by both firm and time. *Journal of Financial Economics* 99:1–10.

Timmermann, A. 2006. Forecast combinations. In *Handbook of economic forecasting*. Eds. G. Elliott, C. Granger, and A. Timmermann. Amsterdam: Elsevier.

Vasicek, O. 1973. A note on using cross-sectional information in Bayesian estimation of security betas. *Journal of Finance* 28:1233–9.

West, K. 1997. Another heteroskedasticity- and autocorrelation-consistent covariance matrix estimator. *Journal of Econometrics* 76:171–91.

White, H. 2000. A reality check for data snooping. *Econometrica* 68:1097–126.

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