Option-Implied Correlation and Factor Betas Revisited*

Adrian Buss[†] Grigory Vilkov[‡]

First draft: November 2008
This version: February 2009

Abstract

We propose a forward-looking heterogeneous implied correlation (*HETIC*) model fitted to the prices of options on the S&P100 and its constituents, and use it to construct superior factor beta predictors. *HETIC* market betas are the most efficient and unbiased predictors of realized betas and explain on average 35.47% of the realized betas variance, outperforming other methods by far. Mean *HETIC* market betas explain more than 34% of the cross-sectional variability in mean excess stock returns. We successfully apply *HETIC* in beta' predictions for factors beyond market (where other forward-looking methods cannot be applied), and also show that the predictive power rapidly increases in unstable (volatile) regimes. Our beta does not suffer from using moments of the risk-neutral returns distribution as risk premia on individual/factor variances and on stochastic correlation approximately cancel out. Moreover, we can naturally handle negative risk exposures to a factor. To improve the accuracy of historical and realized beta estimation, we use high-frequency data, and show that this leads to a significant boost (from 22.20% to 33.85% in terms of R^2) in prediction accuracy. The proposed methods are of great importance to fund managers having a factor exposure target (*HETIC* betas outperform other methods by about 30% in risk targeting), for estimating portfolio riskiness or for predicting stock returns.

JEL classification: G11, G12, G14, G17
Keywords: implied correlation, factor beta, return predictability, high-frequency data, CAPM

*We would like to thank Yacine Aït-Sahalia, Raman Uppal, and especially Christian Schlag for comments and stimulating discussions. We received helpful comments from seminar participants at the University of Frankfurt. We are also grateful to Alexandra Hansis and Yuliya Plyakha for their programming expertise. All remaining errors are our own.

[†]Graduate Program 'Finance and Monetary Economics', Goethe University, Grüneburgplatz 1 / Uni-Pf H 25, D-60323 Frankfurt am Main, Germany. Email: buss@finance.uni-frankfurt.de

[‡]Finance Department, Goethe University, Grüneburgplatz 1 / Uni-Pf H 25, D-60323 Frankfurt am Main, Germany. Email: vilkov@vilkov.net.

Option-Implied Correlation and Factor Betas Revisited

First draft: November 2008
This version: February 2009

Abstract

We propose a forward-looking heterogeneous implied correlation (*HETIC*) model fitted to the prices of options on the S&P100 and its constituents, and use it to construct superior factor beta predictors. *HETIC* market betas are the most efficient and unbiased predictors of realized betas and explain on average 35.47% of the realized betas variance, outperforming other methods by far. Mean *HETIC* market betas explain more than 34% of the cross-sectional variability in mean excess stock returns. We successfully apply *HETIC* in beta' predictions for factors beyond market (where other forward-looking methods cannot be applied), and also show that the predictive power rapidly increases in unstable (volatile) regimes. Our beta does not suffer from using moments of the risk-neutral returns distribution as risk premia on individual/factor variances and on stochastic correlation approximately cancel out. Moreover, we can naturally handle negative risk exposures to a factor. To improve the accuracy of historical and realized beta estimation, we use high-frequency data, and show that this leads to a significant boost (from 22.20% to 33.85% in terms of R^2) in prediction accuracy. The proposed methods are of great importance to fund managers having a factor exposure target (*HETIC* betas outperform other methods by about 30% in risk targeting), for estimating portfolio riskiness or for predicting stock returns.

1 Introduction

Return predictability has been the cornerstone of modern empirical asset pricing for a long time, and linear factor models provide us with a simple expected return estimation methodology. Using these models makes sense if we believe in them, have a good estimation of the factor risk premia, and can predict the exposure of a stock with respect to factors, as expressed by the factor betas. In our paper the main emphasis is on this last point: the predictability of betas.

We make two important contributions in this paper. First, we propose a simple stock correlation structure that allows us to construct a heterogeneous implied correlation (*HETIC*) from available stock and index options. A single additive correlation state variable fitted to reconcile the observed variances implied by the index option prices and option prices of its constituents drives the instantaneous correlation matrix process, which is positive definite under minimal assumptions, by determining its deviations from the historical correlation. As options are forward-looking instruments by nature, they contain information about the market's perception of future returns and future return distributions. We show how this information and *HETIC* can be used to make superior beta predictions for the market and other factors constructed from available stock returns. The *HETIC* predicted market betas explains on average 35.47% of the variance in the realized betas for DJ30 stocks over our sample period from January 1996 until June 2007, exhibiting 25% more explanatory power than the best of the rolling window procedures and, 47% more than the closest rival method based on forward-looking option information.

Second, we utilize high-frequency estimators of the second moments of returns to refine historical and realized beta estimations, and show that this procedure alone can boost the predictive power (in terms of R^2) of the estimated betas from 22.20% to 33.85%. Better beta leads to more accurate expectations of returns, and in the cross-sectional regression of mean stock excess returns on mean betas our *HETIC* market beta outperforms the second best method by almost 22%, showing an impressive R^2 of 34.07%. In the individual return time series predictions we also lead the race with an R^2 of 2.41%.

An important by-product of the *HETIC* structure is that we are not limited to one-factor linear models, and can compute forward-looking betas and predict returns using multiple factors. Due to the forward-looking character of our betas, they are also suitable predictors when the return distribution is expected to undergo a structural break, i.e. when returns become non-stationary. During these unstable periods the predictive power of the *HETIC* betas rises to

43.65% as compared with the second best result of 38.08% for the high-frequency rolling window betas.

In recent years several ways to model the correlation between financial assets in continuous time have been proposed to solve portfolio planning, general equilibrium and other problems. The biggest challenge researchers have to solve here beyond preserving the properties of the observed correlations on the market is to guarantee the positive definiteness of the resulting correlation matrix. One solution is to work with the Wishart Autoregressive Process (WAR) that by construction has all necessary and most desirable properties of the correlation matrix, see e.g. Gourieroux and Sufana (2004) and Buraschi, Porchia, and Trojani (2006). However, due to a large number of state variables it is quite complex, and recent work (e.g. Chiriac (2006)) shows it is not straightforward in empirical work. A polar idea to simplify the stochastic correlation dynamics to one multiplicative state variable driving the homogeneous correlation matrix has been implemented by Driessen, Maenhout, and Vilkov (2009). We propose a simple heterogeneous model of pairwise correlations with a single additive state variable. When calibrated to fit the implied variances of options on the index and its constituents, the model becomes *HETIC*.

The notion of beta, as emanating from the seminal work of Sharpe (1964) and Lintner (1965) is one of the corner stones of modern finance and one of the most important concepts in finance theory as well as finance practice. The market beta of a stock represents its sensitivity to movements of the market index and therefore its systematic risk. Typical applications are in modern portfolio management, financial risk management, asset pricing, valuation of cost of capital, or performance measurement. Therefore Wang (2003) and Ghysels and Jacquier (2006) stress the importance of accurate measurement and even more of accurate prediction of individual stock betas, especially for hedge fund or pension fund managers. For instance pension fund managers may want to construct a tracking portfolio with a beta of one, and hedge fund managers may want to have zero or negative beta portfolios. These groups will clearly benefit from better beta forecasts.

One of the simplest, but still frequently used technique is to estimate betas based on historical stock returns and use this estimate as a forecast for the future. Since there is widespread agreement that betas are time-varying (Keim and Stambaugh (1986), Breen, Glosten, and Jagannathan (1989)) the estimation is typically performed on a rolling window of historical returns. However, this technique implicitly assumes that the future is sufficiently similar to the past.

French, Groth, and Kolari (1983) (hereafter FGK) introduce the idea to use option-implied information to improve the performance of beta forecasts. They continue to use historical correlations but replace historical volatilities with implied volatilities from stock and index options. Siegel (1995) is the first one who solely uses option-implied information for beta forecasts. He creates a new derivative, an exchange option, that implicitly reveals the market beta of a stock. Yet these options are not traded, so that the technique is not applicable in practice.

Christoffersen, Jacobs, and Vainberg (2008) (hereafter CJV) pick up and extend the idea of using only forward-looking information for beta forecasts. Their technique is based exclusively on traded stock and index options and does not require a new derivative or historical correlations. Their final expression for the market beta uses only forward-looking estimates of variance[1] and skewness[2]. The ratio of stock skewness to market skewness serves as a proxy for the future realized correlation between the market factor and the stock. However, the proxy ignores the negative correlation risk premium[3] that can make the index risk-neutral skewness more or less pronounced than we would anticipate from a simple individual stock processes aggregation using the anticipated correlations without a premium. Moreover, their derivations are based on the assumption of zero skewness of the market regression residual, and this may contaminate estimations as well, i.e. the ratio of the risk-neutral skew measures may be a biased predictor of future correlations.

As forecasting techniques based on forward-looking information incorporate market expectations, they may be favorable in situations where a company faces a major change, such as a large acquisition or merger. In this paper we use only forward-looking information (i.e. risk-neutral expectations of certain moments) in our estimation, but the correlation risk premium has by construction only a minor effect on the predicted betas, and we do not make any restrictive assumptions on the return distribution.

For the estimation of historical correlation matrices as well as for the estimation of historical and realized betas we rely on high-frequency data and the recent econometric advances for the measurement of integrated volatility and integrated covariance. Especially the widespread

[1]It is estimated in a model-free way with methodology developed by Britten-Jones and Neuberger (2000), Carr and Madan (1998), and Dumas (1995) who build on the seminal work of Breeden and Litzenberger (1978), and extended further by Jiang and Tian (2005) to jump-diffusion settings.

[2]It is computed using the model-free technique proposed by Bakshi and Madan (2000) and Bakshi, Kapadia, and Madan (2003).

[3]The existence, sign, and magnitude of the correlation risk premium in stocks (and hence in stock indices) has been documented by Driessen, Maenhout, and Vilkov (2009).

availability of high-frequency data as well as the work of Andersen and Bollerslev (1998) have triggered a vast amount of research focusing on volatility and covariance estimation using high-frequency data[4].

The *realized volatility (RV)* estimator, as described in Karatzas and Shreve (1991), was the starting point for the high-frequency volatility estimation. The RV estimator is consistent in absence of noise as shown in early work by Jacod and Protter (1998).

Brown (1990), Zhou (1996) and Corsi, Zumbach, Müller, and Dacorogna (2001) emphasize the implications of market microstructure noise in high-frequency data, so current research concentrates on observed stock prices that are contaminated with noise which renders the RV estimator inconsistent and biased. Under an additive *i.i.d.* noise assumption the use of different time scales, such as two time scales (TSRV) or multiple time scales (MSRV) results in consistent and unbiased estimators (Zhang, Mykland, and Aït-Sahalia (2005) and Zhang (2006a)). Kernel-based methods also produce consistent estimates under these conditions (Hansen and Lunde (2006) and Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008)).

The estimation of integrated covariance becomes additionally complicated by the non-synchronicity of the data. Without noise and with synchronous data the simple *realized covariance* (RC) estimator is consistent (Jacod and Protter (1998)). However, non-synchronicity induces a large bias, known as the Epps effect (Epps (1979)) which drives covariances to zero as the sampling frequency increases (see also Zhang (2006b) for a detailed discussion). In absence of microstructure noise Scholes and Williams (1977) introduce an estimator that accounts for this non-synchronicity by using one lagged and one lead return. An unbiased estimator in a setup without noise is the *cumulative covariance* (CC) estimator introduced in Hayashi and Yoshida (2004) and Hayashi and Yoshida (2005). Though in the presence of additive *i.i.d.* noise the CC estimator has a very large variance and becomes biased. Voev and Lunde (2007) show how to correct for the bias and improve the efficiency of the estimator.

Being able to estimate integrated variances as well as integrated covariances it is an easy step to the estimation of betas using high-frequency data as it is done in the early work of Scholes and Williams (1977) and recently by Andersen, Bollerslev, Diebold, and Wu (2006).

For our work we maintain the assumption of additive *i.i.d.* market microstructure noise, but sample at frequencies where the Epps effect can be ignored. For the estimation of variances as

[4]See, e.g. Andersen, Bollerslev, Diebold, and Ebens (2001), Andersen, Bollerslev, Diebold, and Labys (2001) and Barndorff-Nielsen and Shephard (2002).

well as covariances we use a modified version of the second-best estimator presented in Zhang, Mykland, and Aït-Sahalia (2005). We introduce an additional averaging step by computing their proposed estimator with different time scales and then average over these time scales. This step is necessary to arrive at a reliable estimator for integrated covariances even for infrequently traded stocks. From there we can easily compute estimates for correlation matrices as well as betas based on high-frequency data.

The paper is organized as follows: Section 2 presents a theoretical framework that nests the dynamics and the factor structure of stock returns, our proposed heterogeneous correlation model, and the methods of factor beta construction following from the previous structure. The data used in the empirical part are described in section 3. Then section 4 discusses the estimation of the realized and implied second return moments and factor betas. Section 5 provides empirical evidence on the factor beta and return predictability and illustrates further uses of the *HETIC* beta approach. Finally, section 6 concludes.

2 The Model

This section develops a sample framework that we use in testing the stock factor exposure and return predictability. We start from a simple model of a stock market where stock returns are driven by a number of priced systematic factors and idiosyncratic noise[5]. Then we propose a new one-factor model that allows us to describe the heterogeneous correlation matrix in a simple and robust way. This correlation matrix can be easily fitted to match the variances of the index and its components under the actual and risk-neutral measure. In the latter case we call this model Heterogeneous Implied Correlation (*HETIC*). In the last part of this section we show how one can build the factor betas from the estimated correlation matrix and individual stock variances, and discuss different ways of introducing forward-looking (option-implied) information into these estimates.

[5]For sake of simplicity and intuitive exposition we assume that factors and noise processes are diffusions and that they are driven by standard Wiener processes. However, later estimation procedures does not rely on these assumptions, i.e. factors and noise can also be driven by jump diffusions or more complicated processes. The only critical assumption is that one can create a portfolio from available underlying securities to approximate any factor with a zero error almost surely.

2.1 Factor Structure of Stock Returns

We assume partial equilibrium settings with N stocks driven by K common factors represented by a $K-$dimensional standard Wiener process W^f with individual members $W_k^f,\ k = m, 2, ..., K$[6], and N idiosyncratic shocks represented by a $N-$dimensional standard Wiener process W with individual members $W_d,\ d = 1, ..., N$. We assume that $N \geq K$.

$$\frac{dS_i}{S_i} = \mu_i dt + \sum_{k=m,2..K} \sigma_{i,k}^f dW_k^f + \sigma_i^{id} dW_i,\ \forall i = 1..N. \tag{1}$$

Each stock's i exposure to a given factor k is measured by the respective diffusion coefficient $\sigma_{i,k}^f$. We define the total instanteneous volatility of the stock process S_i as $\sigma_i^S \equiv \sigma_{i,k}^f + \sigma_i^{id}$.

The excess return $\mu_i - r$ represents the risk premium on the priced factors W^f and can be written as

$$\mu_i - r = \sum_{k=m,2..K} \sigma_{i,k}^f \lambda_k, \tag{2}$$

where λ_k denotes the risk premium on the respective factor W_k^f.

We can then rewrite the stock process as

$$\frac{dS_i}{S_i} = rdt + \sigma_i^{id} dW_i + \sum_{k=m,2..K} \sigma_{i,k}^f \left[\lambda_k dt + dW_k^f\right],\ \forall i = 1, ..., N. \tag{3}$$

As factors may not be traded directly, our intention is to replicate them up to a scaling diffusion coefficient using a family of stock portfolio weight vectors $\mathbf{w}_k^F$, $k = m, 2, ..., K$, consisting of the stock weights $\mathbf{w}_k^F = \left(w_{k,1}^F, .., w_{k,N}^F\right)^T$ such that

$$\left(\mathbf{w}_k^F\right)^T \ln \frac{\mathbf{S}_{t+\Delta t}}{\mathbf{S}_t} = \int_t^{t+\Delta t} \left(\sigma_k^f \lambda_k - \frac{1}{2}\left(\sigma_k^f\right)^2\right) dt + \int_t^{t+\Delta t} \sigma_k^f dW_k^f + \varepsilon_k, \tag{4}$$

and assuming that $\varepsilon_k = \int_t^{t+\Delta t} \left(\mathbf{w}_k^F\right)^T diag\left(\boldsymbol{\sigma}^{id}\right) \mathbf{dW}(s) - \frac{1}{2}\left(\mathbf{w}_k^F\right)^T \left(\sigma^{id}\right)^2 ds \to 0\ a.s.$ as $N \to \infty$ and the number of non-zero elements of the weights vector $\mathbf{w}_k^F \to N$.

2.2 Correlations Dynamics

As we discussed in the introduction, modeling correlation in continuous time may be very complex as one has to fulfill several restrictions for the whole correlation matrix.

[6] We assign the letter m to the first factor and assume it is a market factor.

An appealing assumption about only one state variable driving all pairwise correlations has been made by Driessen, Maenhout, and Vilkov (2009) in a recent paper: each pairwise correlation ρ_{ij} is a multiplicative function of a scaling factor $\bar{\rho}_{ij}$ and a state variable $\rho(t)$, i.e. $\rho_{ij} = \bar{\rho}_{ij} \cdot \rho(t)$. This provides us with a simple and robust way of modeling the average level of correlation when we assume that the scaling factor is the same for all pairs of stocks (e.g., 1), and impose some restrictions on the state variable process $\rho(t)$ to guarantee positive definiteness of the correlation matrix. However, even in these simple settings it may not be trivial to find the restrictions for the state variable when the scaling parameter is heterogeneous, i.e. when we want to model the dynamics of each pairwise correlation in the stock universe and not of the average correlation.

We propose a model that extends the idea of one state variable driving all pairwise correlations and allows for a way (surely not as flexible as in WAR) of modeling the heterogeneous correlation matrix. The most important feature of our model is that the resulting correlation matrix stays positive definite by changing the measure, under the assumption of a negative correlation risk premium (i.e. when the expected drift of the process increases by the change from actual to risk neutral measure) if some parametric assumptions are satisfied.

We model the pairwise correlation $\rho_{ij}(t)$ as follows:

$$\rho_{ij}(t) = \begin{cases} 1, & \text{if } i = j; \\ (1-\Delta) \cdot \rho_{ij}^{h} + 2 \cdot \Delta \cdot m \cdot \rho(t), & \text{otherwise}; \end{cases} \tag{5}$$

where ρ_{ij}^{h} denotes the historical pairwise correlation, Δ is a measure of the allowed deviation of the pairwise correlation from its historical mean (e.g. it may be related to the historical standard deviation of the pairwise correlation), $2 \cdot m$ is a scaling parameter, and $\rho(t)$ is the correlation state variable[7]. The following theorem guarantees the positive definiteness of the resulting correlation matrix.

Theorem 1 *The correlation matrix given in (5) is positive definite if the following three conditions are satisfied:* (*a*) $0 \leq m \leq 0.5$, (*b*) $0 \leq \Delta < 1$ *and* (*c*) $0 \leq \rho(t) \leq 1$.

[7] One can think of the pairwise correlation ρ_{ij} as being a linear combination of the historical pairwise correlation ρ_{ij}^{h} with weight $(1-\Delta)$ and the 'aggregated' correlation state variable $\bar{\rho} = 2m \cdot \rho(t)$ with weight Δ.

Proof. We can write the $N \times N$ correlation matrix $\Sigma(t)$ based on model (5) as:

$$\begin{aligned}\Sigma(t) &= I_N + (\Sigma^h - I_N)(1-\Delta) + 2\Delta \cdot m \cdot (\iota \cdot \iota' - I_N) \cdot \rho(t) \\ &= I_N \cdot \Delta \cdot (1 - 2m \cdot \rho(t)) + \Sigma^h \cdot (1-\Delta) + 2\Delta \cdot m \cdot \iota \cdot \iota' \cdot \rho(t), \end{aligned} \quad (6)$$

where Σ^h, I_N, and ι denote the historical correlation matrix, the identity matrix and an $N \times 1$ vector of ones, respectively. Since I_N is positive definite, to guarantee positive semidefiniteness of the first part of expression (6) we need $\Delta \geq 0$ and $(1 - 2 \cdot m \cdot \rho(t)) \geq 0$. If we let $0 \leq m \leq 0.5$ a sufficient condition for this is given by $0 \leq \rho(t) \leq 1$. If we assume that the historical correlation matrix Σ^h is positive definite, the second part is positive definite if $\Delta < 1$. Finally, the matrix $\iota \cdot \iota'$ is positive semidefinite and therefore the last part is positive semidefinite if $\rho(t) \geq 0, \Delta \geq 0$ and $m \geq 0$. As the sum of positive definite and positive semidefinite matrices is positive definite, these conditions together are sufficient for the positive definiteness of $\Sigma(t)$ ■

The correlation state variable ρ controls the level of the pairwise correlations: a high value of ρ implies high (absolute) pairwise correlations whereas a low value of ρ implies low (absolute) pairwise correlations. For instance, if we assume $\rho = 0$, then the pairwise correlation is equal to $\rho_{ij}(t) = (1-\Delta) \cdot \rho_{ij}^h$. For $0 \leq \Delta < 1$ the correlation will be smaller (in absolute value) than the historical correlation ρ_{ij}^h. If, as the other extreme, we assume $\rho = 1$, the correlation will be $\rho_{ij}(t) = (1-\Delta) \cdot \rho_{ij}^h + 2 \cdot \Delta \cdot m$. As we later choose the parameter m to be greater than the average historical correlation to account for extreme correlation values and for the negative correlation risk premium (that would be normally increasing the expected integrated correlation), a correlation state variable $\rho = 1$ implies for most of the pairwise correlations the inequality $\rho_{ij}(t) > (1+\Delta) \cdot \rho_{ij}^h$. Taking into account the two extreme cases, we allow the pairwise correlation $\rho_{ij}(t)$ to be at least in the range $[(1-\Delta) \cdot \rho_{ij}^h, (1+\Delta) \cdot \rho_{ij}^h]$. If we now choose the parameter Δ to be c times the standard deviation of the pairwise historical correlations (e.g. about four standard deviations as we do in our empirical section 4.3) we allow the pairwise correlation to be in the range of c standard deviations of the historical correlation.

The stock market factor (the index) is composed of all N stocks, and given a vector of index weights $\mathbf{w}_m^F$ as well as the index variance $\left(\sigma_m^F\right)^2$ at time t[8] we can write the variance of the market factor as follows:

[8] We omit time as an argument for notational convenience throughout, except when placing particular emphasis on it.

$$\left(\sigma_m^F\right)^2 = \sum_{i=1}^{N} w_{m,i}^2 \left(\sigma_i^S\right)^2 + \sum_{i=1}^{N} \sum_{j \neq i} w_{m,i} w_{m,j} \sigma_i^S \sigma_j^S \rho_{ij}. \tag{7}$$

It is clear from (7) that the index variance is driven by individual volatilities σ_i^S and pairwise correlations ρ_{ij}. If we use our model (5) for ρ_{ij} then we can express the state variable ρ as a function of the other variables in the equation:

$$\rho = \frac{\left(\sigma_m^F\right)^2 - \sum_{i=1}^{N} w_{m,i}^2 \left(\sigma_i^S\right)^2 - \sum_{i=1}^{N} \sum_{j \neq i} w_{m,i} w_{m,j} \sigma_i^S \sigma_j^S \left((1-\Delta)\rho_{ij}^h\right)}{2m \sum_{i=1}^{N} \sum_{j \neq i} w_{m,i} w_{m,j} \sigma_i^S \sigma_j^S \Delta}. \tag{8}$$

All variables on the right-hand side of (8) can be easily calculated from the data. To calibrate the correlation state variable under the actual probability measure, we would use historical realized (co-)variances for index and for individual stocks, and for the risk-neutral probability measure (implied correlation state variable) we would use index and individual variances implied in option prices[9] instead of their realized counterparts. Then the set of all pairwise correlations (5) calculated with the implied correlation state variable is called *HETIC*.

As we will show later in the empirical section, the implied correlation state variable satisfies the conditions stipulated in Theorem 1, and hence we are safe to go further in our investigation.

2.3 Factor Betas Construction

Betas reflect the linear relation between a factor and a stock return. Assuming that the conditional covariance between the stock return and the factor return, and the factor return's conditional variance are known, we can write the *conventional beta* of a stock with respect to the factor k as follows:

$$\beta_{i,k}^{CONV} = \frac{Cov\left(r_i, r_k^F\right)}{Var\left(r_k^F\right)} = \rho_{i,k} \cdot \frac{\sigma_i^S}{\sigma_k^F}, \tag{9}$$

where r^F is the return on the factor portfolio (4). Knowing the weights $\mathbf{w}_k^F$ of the factor k replicating portfolio and the variance-covariance matrix for all stocks, we can rewrite the *conventional beta* as the *expanded beta*:

[9] To calibrate the model in the empirical part of the paper, we will use the Realized (Co-)Variance estimator for the actual probability measure and the Model-Free Implied Variance for the risk-neutral probability measure. Both estimators reflect the total integrated quadratic variation of the stock over a period of time (realized or expected), i.e. empirically we cover a more general class of models than was presented in section 2.1.

$$\beta_{i,k}^{EXP} = \frac{\sum_{j=1}^{N} w_{k,j}^F Cov(r_i, r_j)}{Var\left(\sum_{j=1}^{N} w_{k,j}^F r_j\right)} = \frac{\sum_{j=1}^{N} w_{k,j}^F \sigma_i^S \sigma_j^S \rho_{ij}}{\sum_{j=1}^{N}\sum_{l=1}^{N} w_{k,j}^F w_{k,l}^F \sigma_j^S \sigma_l^S \rho_{jl}}. \tag{10}$$

Thus, to calculate a stock's beta with respect to any factor, we need an estimate of the conditional variance-covariance matrix of either the stock and the factor, or the stock and all components in the factor replicating portfolio (along with the replicating portfolio weights).

There is a myriad of ways and methods to predict future covariances. We can split them roughly into two groups based on the information they are using. The first group uses historical information as a predictor for the future, others mix historical information with forward-looking market instruments such as options to infer investor's expectations of future returns or even solely use forward-looking information. We will use heterogeneous implied correlations (*HETIC*) as in (5) calibrated to match the option-implied variances on individual and index options, as well as option-implied volatilities on individual options for the purpose of estimating betas as in (10), which then become β^{HETIC}.

There have been several other quite successful attempts to use option-implied information in beta estimation. French, Groth, and Kolari (1983) combine historically estimated correlations with option-implied variances to improve the market beta:

$$\beta_{i,m}^{FGK} = \rho_{m,i}\left(\frac{\sigma_i^2}{\sigma_m^2}\right)^{\frac{1}{2}}. \tag{11}$$

In another paper Christoffersen, Jacobs, and Vainberg (2008) suggest not only to use option-implied variances, but also to align the betas with the market's perception of future correlations proxied by one of the forward looking moments (option-implied skewness) of the future return distribution. Their market beta is defined as:

$$\beta_{i,m}^{CJV} = \left(\frac{Skew_i}{Skew_m}\right)^{\frac{1}{3}}\left(\frac{\sigma_i^2}{\sigma_m^2}\right)^{\frac{1}{2}}. \tag{12}$$

These methods have been originally limited to calculating betas for the market factor or a portfolio that has traded options on it, only. Moreover, restrictive distributional assumptions about the market regression residuals have to be made in the CJV case.

In the empirical section apply our method and the two other mentioned methods to calculate forward-looking factor betas and predict both realized betas and realized returns based on those estimates.

3 Data Description

3.1 Data on Stocks

The stock data used in this paper consist of transaction prices of the S&P100 constituents from the NYSE's trades and quotes (TAQ) database, tick data for the S&P100 index from tickdata.com, daily stock prices from CRSP, and S&P100 index weights from Bloomberg L.P. The sample period extends from January 4, 1996 to June 30, 2007, for a total of 2894 trading days. TAQ prices were filtered from the official opening 9:30 EST until 16:00 EST and only include valid entries. In addition, we remove obvious outliers, such as transaction prices reported at zero and errors of shifting a decimal place.

We use a calendar time sampling scheme, that is we artificially construct a regularly spaced one minute price grid for every trading day (i.e. typically 390 data points per day) by using the size-weighted average[10] of all transactions within one minute[11]. Finally, we fill empty data points, as a result of no trading activity for a stock within a minute, with the price from the last available data point before it. On average we have, before filling the empty data points, 257 data points for each stock and day for the period 1996-2000, and 344 data points for the period 2001-2007.

The S&P 100 is a value-weighted index with rebalancing taking place on every third Friday of the last month in each quarter where the index shares are fixed for the following quarter. Nevertheless the value-based weights of the constituents can change due to stock price movements. Moreover the list of constituents may also change whenever a company is removed from the index and a new one is added, and hence new weights are introduced. For the period January 1996 to mid-December 2000 we reconstruct the S&P 100 weights using the market capitalizations of all companies currently in the index based on daily market capitalizations from CRSP. For the period from mid-December 2000 to June 2007 we use the official index shares as published by Bloomberg L.P. to calculate the real index weights on each day.

For ease of presentation we showcase our obtained results in the following sections only for

[10] For the S&P100 index we take simple averages.
[11] For instance the data point for 10:30 contains all transactions from 10:29:01 up to 10:30:00.

a subsample of the S&P 100 stocks, namely the stocks included in the Dow Jones Industrial Average Index (DJ30) as of February 18, 2008. A list of the corresponding stocks is presented in Table 1.

3.2 Data on Stock Options

For stock options we use the Ivy DB that contains data on all US exchange-listed and NASDAQ equities and market indices, as well as all US listed index and equity options, for our sample period from January 1996 until June 2007.

We select all options on the S&P100 index and all its components with maturities from 30 to 90 days, underlying stock prices for each day, discrete dividends history, and certificate of deposit rates as the riskfree rate proxy. Then we apply several filters to the options data: we remove all in-the-money options to diminish the influence of an early exercise premium on our estimations, all options with zero open interest on a given day or with zero bid prices. To eliminate outliers and options with non-standard features we also discard options with implied volatilities (IV) higher than 100%, and options with missing implied volatilities[12]. Then we select on each day the maturities for which, after filtering, there are at least 2 call and 2 put options available, so that we have at least 4 options.

To get the riskfree rate proxy of the exact required maturity, we take known certificate of deposit yields for maturities between 1 day and 1 year and interpolate them linearly to get the appropriate yield. As we have options on dividend-paying stocks in our sample, to make life easier, we adjust the stock price on each day by the discounted value of future dividends. Discrete dividends are a bit complicated, we do not know them in advance and hence need to forecast them from the previous dividend payments, i.e. from the common payment schedule. We adjust the stock prices on each day by the predicted or declared regular dividends (using the actual yield as a proxy for the anticipated yield), and by the declared special dividend (using its actual dividend yield). In general the effect of dividends on implied volatilities is minor, so that the assumptions we make for predicting the dividends should not introduce any substantial bias.

[12]The implied volatility in OptionMetrics is calculated from fitting a binomial tree with discrete dividends. The IV will be missing in the database in case of non-standard settlement of an option, or in case of various arbitrage conditions violations.

4 Second Moments and Beta Estimations

As we have seen in the previous sections, to estimate the factor betas we need information on the second moments of stock returns. Depending on the beta construction method, we will use historical moments, forward-looking moments, or we will mix both types into one estimation. In the current section we describe in detail the methods and procedures we use to construct the realized variances/covariances using daily and high-frequency data, the forward-looking variances from options data, the *HETIC* from both realized and implied moments, and how we use all the calculated variables to construct a variety of predicted betas.

4.1 Realized Measures

Our estimation procedure using high-frequency data is based on the structure of stock returns as in (1) which implies the following structure for the logarithmic price process $X_i = \log(S_i)$

$$dX_i = \mu_i^* dt + \sigma_i^f dW^f + \sigma_i^{id} dW_i, \forall i = 1..N;$$

where μ_i^* captures all drift dynamics, σ_i^f denotes the 1xK vector $(\sigma_{i,m}^f, \ldots, \sigma_{i,K}^f)$ of sensitivities and dW^f denotes the Kx1 vector $(dW_1^f, \ldots, dW_K^f)$ of standard Brownian motions.

We assume that the observed logarithmic price process $Y_i(t)$ is given by $Y_i(t) = X_i(t) + \varepsilon_i(t)$, where $\varepsilon_i(t)$ captures market microstructural noise effects, including, but not limited to, bid-ask bounces, asynchronous and discrete trading.

Our assumptions regarding the noise process $\varepsilon_i(t)$ are given by:

1. $\varepsilon_i(t)$ is *i.i.d.* with $E[\varepsilon_i(t)] = 0$; and

2. ε is independent of X.

The observed stock return $r_i(t)$ for a time interval of length δ is then given by $r_i(t) = Y_i(t) - Y_i(t - \delta)$. For estimation purposes the quantity of interest is the integrated variance/covariance of the interval 0 to T, given by:

$$\langle X_i, X_j \rangle = \int_0^T \sigma_i^f \left(\sigma_j^f\right)' dt = \int_0^T \sum_{k=m..K} \sigma_{i,k}^f \sigma_{j,k}^f dt, \tag{13}$$

Defining $n \equiv T/\delta$ a common estimator for this quantity is given by the *realized variance* (RV) / *realized covariance* (RC)

$$\widehat{\langle Y_i, Y_j \rangle}^{(\delta)} = \sum_{k=1}^{n} r_i(k\delta) \cdot r_j(k\delta) = \sum_{k=1}^{n} \Big(Y_i\big(k\delta\big) - Y_i\big((k-1)\cdot\delta\big) \Big) \cdot \Big(Y_j\big(k\delta\big) - Y_j\big((k-1)\cdot\delta\big) \Big). \quad (14)$$

Without market microstructure noise, i.e. $Y_i(t) = X_i(t)$, it is well known that this estimator is consistent because of the theoretical result that

$$p\lim \sum_{k=1}^{n} r_i(k\delta) r_j(k\delta) = \int_0^T \sigma_i^f \; \left(\sigma_j^f\right)' dt$$

as the sampling frequency n increases (see Jacod (1994) and Jacod and Protter (1998)) whereas the estimator becomes inconsistent and biased in the presence of noise (see Zhang, Mykland, and Aït-Sahalia (2005), and Bandi and Russell (2005)).

Zhang, Mykland, and Aït-Sahalia (2005) (ZMAS henceforth) discuss several methods to estimate the quantity in (13) in the presence of noise. Their first-best estimator, the two time scale realized volatility (TSRV) estimator, uses information from two different time scales: In a first step the authors compute $\widehat{\langle Y, Y \rangle}^{(all)}$ which is estimator (14) using information at a very high frequency (typically at the highest frequency available). They show that this quantity can be used to correct for the bias since it is actually a consistent estimator of the variance of the noise. Using estimator (14) computed at a second (slower) time scale and constructing a suitable linear combination of these two quantities the authors derive an consistent and unbiased estimator of the integrated variance/covariance in (13).

Unfortunately we cannot apply the TSRV estimator here since our one minute grid of stock prices is not fast enough to qualify as a fast time scale. This can be seen in Figure 1 (a), where we estimate the variance of CSCO for the year 1998 using the TSRV estimator with a fast time scale of one minute. When choosing the second (slower) time scale faster than 20 minutes, we get a sharp decline in variance which is due to an overcorrection using $\widehat{\langle Y, Y \rangle}^{(1\,\mathrm{min})}$. This indicates that the estimator $\widehat{\langle Y, Y \rangle}^{(1\,\mathrm{min})}$ does not represent pure noise and therefore cannot be used for bias correction.

So, instead of using the first-best estimator of ZMAS we rely on their second-best estimator which is basically computing the RV/RC estimator (14) at a slower frequency and improving efficiency by subsampling and averaging. The problem with using a slower time scale is that one throws away information and therefore loses efficiency. To overcome this problem the authors propose to use subsamples and average over the subsamples. To illustrate the idea assume we have stock returns at a one minute grid, i.e. for times $1, 2, 3, \ldots$. If we decided to sample at a ten minute frequency we can do this for different subsamples: for the first subsample we can use the information from minutes $1, 11, 21, \ldots$; for the second subsample we can use the information from minutes $2, 12, 22, \ldots$ and so on. For each of the $K = 10$ resulting subsamples we can compute estimator (14). Finally, ZMAS propose to average over the subsamples to improve the efficiency of the estimator. The resulting estimator is then given by:

$$\widehat{\langle Y_i, Y_j \rangle}^{(avg,K)} = \frac{1}{K} \sum_{k=1}^{K} \widehat{\langle Y, Y \rangle}^{(\delta,k)}. \tag{15}$$

There remains the problem of choosing the optimal sampling frequency δ. A common way to identify δ is to use signature plots, as introduced by Andersen, Bollerslev, Diebold, and Labys (2000) and extended to covariance analysis by Griffin and Oomen (2006). In a signature plot the estimator for the desired quantity is plotted against different sampling frequencies and one chooses as the optimal frequency the frequency where the estimator starts to level out. As one can see from Figures 1 (b) and (c) we could choose relatively fast frequencies for the estimation of the variance, i.e. the estimator starts to level out at frequencies of about 30 minutes. But it is well known and can be seen in Figure 1 (d) that the covariance estimator does not level out at such fast frequencies and may exhibit an unstable behavior over a range of frequencies.

To overcome this problem with the covariance estimation we introduce an additional averaging step. That is we compute estimator (15) at frequencies $\delta = 75\text{min.}, 100\text{min.}, \ldots, 300\text{min.}$ and finally average over the different sampling frequencies to obtain our final estimator of quantity (13):

$$\widehat{\langle Y_i, Y_j \rangle} = \frac{1}{10} \sum_{l=3}^{12} \widehat{\langle Y_i, Y_j \rangle}^{(avg,25l)}. \tag{16}$$

Hansen and Lunde (2005) show that overnight returns contain substantial information for variance estimation. Since we are typically estimating the integrated volatility and covariance over a period of about 60 days, we implicitly include overnight returns here.

As outlined in the introduction non-synchronicity of the data results in an Epps effect for integrated covariances, i.e. covariances approach zero as the sampling frequency increases. Since the fastest time scale we are using is 75 minutes, this problem does not arise here, so that we abstain from corrections for the Epps effect.

Based on the estimator in (16) for the integrated variance/covariance $\int_0^T \sigma_i^f \left(\sigma_j^f\right)' dt$ for stocks i and j we can easily compute estimates of the correlation matrix $\Sigma = \left(\rho_{i,j}\right)$ using high-frequency data:

$$\rho_{i,j} = \frac{\widehat{\langle Y_i, Y_j \rangle}}{\sqrt{\widehat{\langle Y_i, Y_i \rangle}}\sqrt{\widehat{\langle Y_j, Y_j \rangle}}},$$

where the estimators are computed over the desired time period. Moreover we can use the estimator in (16) to compute estimators for historical as well as realized high-frequency betas with respect to a given factor k. This reduces to the estimator

$$\beta_{i,k}^{\widehat{CONV}} = \frac{\widehat{\langle Y_i, Y_k \rangle}}{\widehat{\langle Y_k, Y_k \rangle}}. \tag{17}$$

In the same fashion an estimator for the expanded beta (10) can be constructed.

Since we are comparing the beta and return prediction capabilities of our *HETIC* model in section 5.1 with the capabilities of a model using historical return information on a daily basis, we also have to estimate the historical variance-covariance matrix using daily information. Therefore we apply the simple RV/RC estimator (14) using daily return data (close-to-close), i.e. $\delta = 390$ min. From there we easily get correlation matrices and betas in the same fashion as outlined above.

4.2 Forward-Looking (Implied) Measures

To compute the implied moments needed for the different beta methodologies we first prepare the options database to infer the value of the variance and cubic contracts, or the proxies for the forward-looking risk neutral variance ($MFIV$) and skewness ($SKEW$) respectively, as described in detail by Bakshi, Kapadia, and Madan (2003). Obviously, to calculate the integrals in the formulas precisely, we need a continuum of option prices. In reality we approximate them from available option data: using cubic splines we interpolate the implied volatilities of the options inside the available moneyness range, and extrapolate using the last known (boundary for each side) value to fill in 1001 grid points in the moneyness range from 1/3 until 3[13]. Then we calculate the option prices from the interpolated volatilities using the known interest rate for a

[13] The reason for choosing such a wide grid is that our simulation studies have shown that with a narrower grid we may not be estimating the skew and kurtosis of the risk-neutral distribution well enough.

given maturity, and use these resulting prices to compute $MFIV$ and $SKEW$.

The formulas for these implied measures and the approximation procedure description are provided in Appendix A. Summary statistics for the quantities of interest for the S&P 100, the S&P 100 constituents as well as the DJ30 components are presented in Table 2. As expected from previous research, the S&P 100 index displays a clearly higher, in absolute terms, model-free skewness while exhibiting a lower model-free variance compared to the S&P 100 index constituents.

4.3 HETIC

To estimate the *HETIC* model (5), we first calculate the correlation state variable ρ as in (8). For this we need several values like implied variances of the index and its components', the index weights, and the historical pairwise correlations. For the computation of the implied measures we use the methodology outlined in the previous section, the index weights are obtained as described in section 3.1, and the historical pairwise correlations are computed using high-frequency stock returns over the last 60 trading days. In addition, we need to specify the parameters of the model that define the variability of the resulting state variable: first, Δ, the measure of the allowed relative deviation of the implied from the historical pairwise correlation, and second, the scaling parameter m of the correlation state variable in *HETIC*. In our estimations we set both values to 0.5, i.e. each implied pairwise correlation can at maximum deviate from its historical value by 50% and the resulting correlation will consist of pairwise historical correlation and correlation state variable with equal weights (50/50):

$$\rho_{ij}^{HETIC} = 0.5\rho_{ij}^{h} + 2 \cdot 0.5 \cdot 0.5\rho = 0.5\rho_{ij}^{h} + 0.5\rho, \quad \text{if } i \neq j.$$

For the admissible range of the state variable $\rho \in [0,1]$ (to guarantee the positive definiteness of the resulting *HETIC* matrix), the ρ_{ij}^{HETIC} value is restricted to the bounds $\left[0.5\rho_{ij}^{h}, 0.5\rho_{ij}^{h} + 0.5\right]$.

We do not offer a formal procedure for choosing the parameters, and rely more on intuition. The choice of Δ should be linked to the stability of an average pairwise correlation over time, and as we estimate the implied correlation matrix for a period of less than two months on average, we do not expect it to deviate much from the historical counterpart. The mean standard deviation of the time-series of the historical pairwise correlations is relatively low (0.1250), and so with our choice of Δ we should capture most of the variation in correlations. The choice of

m is related to the average anticipated correlation level. In our sample the average pairwise historical correlation is 0.3308, but for the implied correlation after adding the correlation risk premium we expect a higher value, and hence select $m = 0.5$.

After plugging in all known and assumed variables, we first get the time series of ρ. The dynamics of the state variable are shown in Figure 2. It fits well with its permissible range of values, with a time series mean of 0.4727 and 0.2049 standard deviation. Moreover, it has negative correlation with market returns (-0.1321) and market level (-0.4359) which confirms earlier findings that the correlation increases in bad times, i.e. when the market goes down. For each date we check if our ρ estimate violates the permissible bounds of $[0, 1]$ - it does marginally violate the upper bound in about 2% of all cases, and we truncate it at the upper bound to satisfy the technical conditions[14].

Having obtained an estimate of ρ we can use equation (5) to compute the whole *HETIC* matrix. As expected, the mean *HETIC* (0.4204) is higher than the mean historical pairwise correlation (0.3303). What matters most for us is not the mean level of *HETIC*, but rather its performance in predicting realized correlations as this will drive the performance of the beta predictions. We carry out two tests using as predicted values the *HETIC* and the historical high-frequency correlation matrices. First, we regress the time series of the cross-sectional mean realized correlations on the time series of cross-sectional mean predicted correlations; second, we do the same for individual stock pairs, i.e. we run a number of time-series regressions of realized pairwise correlations on predicted values. The R^2 for the *HETIC* based regression is 45.05%, while the mean historical high-frequency correlation gives us only an $R^2 = 41.68\%$. The mean explanatory power in individual regressions is 21.29% for *HETIC* vs. 14.63% for historical high-frequency correlations.

4.4 Estimating Betas

As described in the introduction our main emphasis is to predict the future linear exposure of a stock return to a factor, expressed by its beta. Therefore we estimate several types of betas as described in section 2.3 to see how these betas can predict the realized betas. We consider historical rolling window betas using high-frequency as well as daily stock return data, the FGK betas, the CJV betas and the betas emanating from our *HETIC* model. All betas are calculated

[14] In the cases of violations the estimated *HETIC* matrix remains positive definite even without truncation of the correlation state variable ρ.

for the market factor; moreover, the historical (realized) and *HETIC* betas are also calculated with respect to other factors that are given by certain portfolios of stocks.

We compute the Historical HF betas as well as the Historical Daily betas using the methodology outlined in section 4.1. Our final expressions are then given by (17) and the corresponding expanded version of it where the integrated volatilities and covariances are computed over the last 60 trading days, i.e. using about 23,400 one minute returns, as well as the modified second-best estimator from ZMAS for high-frequency data and using 60 close-to-close returns and formula (14) for daily data, respectively.

The computations of the FGK betas, the CJV betas as well as our *HETIC* betas depend on implied volatilities, while the CJV betas moreover depend on implied skewness. For the computations of these implied moments we use the methodology outlined in section 4.2. Therefore we draw on options with a maturity between 30 and 90 days and if several are available within this range, we choose the options with maturity closest to 60 days.

For the computation of the FGK betas we follow equation (11) where the historical correlations are computed over the last 60 trading days using daily stock return data. The CJV betas are computed with formula (12). As Christoffersen, Jacobs, and Vainberg (2008) we also do not compute CJV betas for stocks on a specific date if they display a positive skewness (due to mathematical reasons). This highlights a problem with the CJV betas as they can only take on positive values. However, if a stock's return is negatively correlated with the market return the beta should be negative[15].

The computation of the *HETIC* betas relies on the computation of the heterogeneous implied correlation matrix (5) for the specific day as described in section 4.3. Together with the implied volatilities we can then compute the implied covariance matrix from which we get the implied betas using the conventional (9) or the expanded (10) formula.

Since we want to forecast realized betas using option-implied information, a natural choice for the time period over which the realized betas are computed is the option maturity. Here we use the average option maturity for the S&P 100 constituents and the methodology outlined in section 4.1 to get high-frequency realized as well as daily realized betas.

Almost all predictive market betas are highly correlated with both the realized beta and with other predictive betas as shown in Table 3. *HETIC* has a correlation to realized beta of

[15]In our full sample 1.58% of the realized HF betas are negative.

0.57 and to Historical HF beta a correlation of 0.54. Note that the CJV beta has the highest correlation (0.39) not with the realized beta, but with *HETIC* (probably due to the use of forward-looking variances by both methods), and FGK beta has the second highest correlation (0.73) with *HETIC* (probably due to the mixing in some manner the forward-looking variances and historical correlations in both estimation methods).

As we want to use our *HETIC* betas to predict realized betas, we have to be careful with the change of measure and the influence of risk premia. Since we are explicitly modeling the implied correlations, we need to take the risk premia on single stock volatility, on index volatility, and finally on the correlation between a stock and the market into account. If we simply associate the risk premia with a scaling factor, defined by the ratio of the average implied quantity and the average realized quantity, we can analyze the effect of these risk premia on our *HETIC* betas. Table 2 includes the scaling factors for single stock volatility (1.17) as well as for index volatility (1.32). The scaling factor for the correlation between a stock and the market can be easily computed from Table 5, and is equal to 1.16. If we use this simple risk premia presentation we get, on average, the following relationship between implied (superscript Q) and realized (superscript P) betas:

$$\beta_i^Q = \frac{\sigma_i^Q \cdot \rho_{iM}^Q}{\sigma_M^Q} = \frac{1.17\ \sigma_i^P \cdot 1.16\ \rho_{iM}^P}{1.32\ \sigma_M^P} = 1.03 \cdot \frac{\sigma_i^P \cdot \rho_{iM}^P}{\sigma_M^P} = 1.03 \cdot \beta_i^P;$$

so that we expect the relation between implied and realized betas to be very close to one. This is quite important as only for values very close to one implied betas can be used as meaningfull predictors for the realized betas.

In Table 4 we present the mean betas for the different forecasting methodologies as well as the mean realized betas for the DJ30 stocks. The mean values confirm our expectation that the *HETIC* betas only include a marginal risk premium, in this case a scaling factor of 1.02 (and an absolute difference between the mean *HETIC* and mean realized beta of only 0.0184). Moreover for 12 out of 30 stocks the *HETIC* and the realized betas are not significantly different. The mean beta for the Historical Daily, Historical HF, Realized Daily and Realized HF methodologies are almost the same. The small difference is induced by the fact that the historical betas are always computed over a 60 trading days period whereas the realized betas are computed over the average option maturity.

The absolute difference between the mean realized betas and the mean FGK betas (0.0742) as well as the mean CJV betas (0.2360), respectively, is clearly more pronounced. As FGK, CJV and *HETIC* all use the same implied volatilities, the differences in betas can only be due to the different correlation models. Table 5 presents the average correlations between the DJ30 stocks and the market index as being used in the computations of the market betas. As expected from the results in Driessen, Maenhout, and Vilkov (2009) the historical correlations used for FGK betas are lower than the implied correlations from the *HETIC* model because of the correlation risk premium. In contrast to this, the artificial correlations, i.e. the ratio of stock and index skewness, as used by CJV, are clearly higher than the *HETIC* implied correlations which explains the higher betas. The distinct difference between the average beta for the CJV model in our setup (1.16) and the average beta in Christoffersen, Jacobs, and Vainberg (2008) (0.92) stems from the fact that we are using options with an average maturity of 60 days whereas CJV use much longer options with an average maturity of 180 days. Stock options with longer maturity have usually a lower implied volatility whereas this effect is not that prominent for index options. The same effect can be observed for implied skewness and this is maybe one of the reasons why the CJV betas in our calculation are higher compared to their setup.

5 Betas and Return Predictability

After using the available methods to estimate the betas, we now turn to studying their comparative power in predicting future betas and expected returns. We do this for both the market factor, and statistical factors derived from a principal component analysis. We also study how the prediction capabilities of the beta methodologies differ between stable and unstable regimes.

5.1 Market Factor: Predictability Regressions

At first we run a battery of time-series and cross-sectional tests to isolate the best methodology in predicting the market betas and excess returns.

First, for each of the DJ30 stocks we run the time-series regression of realized market beta on its predicted value:

$$\beta_{i,t}^{Realized} = \alpha_i + \lambda_i \beta_{i,t}^{Predicted} + \varepsilon_{i,t}, \forall t. \tag{18}$$

Subscripts i, t refer here and henceforth to stock i as well as time t, respectively. The individual results and summary statistics for the regression in (18) are given in Table 6. Although predicted betas are calculated with differently sampled data from the past (high-frequency returns for *HETIC* and Historical HF, daily returns for Historical Daily and FGK), we show in Table 6(b) that all betas predict the realized betas calculated with high-frequency returns better than the realized betas from daily return data. This is most probably due to more precise estimates from high-frequency data, and so in the following we show the results only for high-frequency realized betas.

As expected from the correlation prediction results in section 4.3, *HETIC* outperforms other betas with an average R^2 of 35.47%. Furthermore, it exhibits the best explanatory power in 14 out of 30 in individual regressions. The second best beta (Historical HF) explains 33.85% of the realized beta variance, and delivers the highest R^2 in 12 cases. The second best method using the forward-looking information is FGK, and though it loses the race to Historical Daily beta on average (with an R^2 of 24.12% vs. 28.38%), it still delivers the remaining 4 best individual results. The slope value test ($H_0 : \lambda_i = 1$) for individual regressions shows how close the realized beta is to its predicted value, and here *HETIC* is the absolute champion with 18 cases not rejected. Thus, there is clear evidence that for our sample the *HETIC* market beta is the most efficient (in terms of explained variability) and the least unbiased (in terms of the predicted value) predictor of the realized market beta. In Figures 3 (a), (b), and (c) we show the time-series of the *HETIC* market beta and of the realized beta for the time period January 2, 1996 to June 30, 2007 for AA (a), XOM (b) and MRK (c), respectively. The predicted *HETIC* market betas fit the realized betas in general quite well.

Second, to see how the prediction of the betas is affecting return predictability, we estimate for each of the DJ30 stocks the regression of the realized returns on the predicted betas:

$$r_{i,t}^{Realized} = \alpha_i + \gamma_i \beta_{i,t}^{Predicted} + \varepsilon_{i,t}, \forall t. \tag{19}$$

The results for the individual estimations and an overview are provided in Table 7. The leadership is still in *HETIC*'s camp with an average R^2 of 2.41%, though Historical HF is better in a larger number of individual cases (10 vs. 8). In general, the prediction results confirm that stock returns are hardly predictable on an individual level, but better beta still makes a difference.

Third, to see how well our estimated betas explain the cross-sectional differences in expected stock returns, we run a regression of mean stock returns on the mean predicted betas:

$$\bar{r}_i = \alpha_i + \lambda \bar{\beta}_{m,i} + \varepsilon_i. \tag{20}$$

Table 8 shows that the mean *HETIC* beta stands out with a predictive power of 34.07%, while all other methods deliver comparable results in the range of about 24% to 28%. In Figure 4 we plot the mean excess returns $\bar{r}_i$ against the mean predicted market betas $\bar{\beta}_{m,i}$ for the *HETIC* methodology and show the fitted regression. In a recent study Buss, Schlag, and Vilkov (2009) use *HETIC* betas to test the conditional CAPM and find evidence that the truly conditional *HETIC* market betas (as they are based on forward-looking information) can account for some of the cross-sectional differences in expected returns typically attributed to firm characteristics like size, book-to-market or stock momentum. The traditional historical market betas are not that flexible and the use of additional factors becomes a necessity.

5.2 Market Factor: Asset Allocation

As market betas are very important for asset allocation and risk management we now analyze the performance of the different beta methodologies in an empirical asset allocation setup. We create portfolios with a target beta based on the predicted betas of the individual stocks and then compare this target beta with the realized beta of the portfolios.

First, we create portfolios consisting of two individual stocks from the universe of the S&P100 stocks. The portfolio weights are chosen such that the target portfolio beta is equal to zero. For instance, such a setup is typical for *pairs trading strategies* of hedge funds who go short one stock and long the other one while being market neutral[16].

To implement this portfolio strategy we randomly choose $25,000$ pairs. For each pair we compute the portfolio weights at the end of every month in our sample period as well as the realized portfolio betas over the following two months. After pooling the observations for each pair and each month we compute the *Mean Absolute Error (MAE)* for the different beta prediction methodologies as a measure of overall fit.

HETIC Betas yield by far the best results, showing an impressive improvement of about 30% compared to the second best method. They exhibit the lowest MAE (0.2921) followed by the CJV betas (0.4103) and Historical HF betas (0.4172). The Historical Daily Betas and FGK

[16] See Vidyamurthy (2004) and Whistler (2004).

betas yield MAEs of 0.5491 and 0.7805, respectively. A better performance of the CJV betas compared to section 5.1 is probably due to the fact that the biases across the two stocks cancels out.

Second, we create portfolios consisting of six individual stocks from the universe of the S&P100 stocks. Again, the portfolio weights are chosen such that the target portfolio beta is equal to zero. This setup is compabtible with long/short market neutral strategies. For implementation we randomly choose $50,000$ combinations of six stocks: three stocks which we go long and three stocks which we go short. Then follow the procedure outlined above.

We would expect the forecasting errors to be higher in this second setup as we now have six instead of just two stocks. However, we would expect the differences in the performance for the different beta methodologies to be not as pronounced compared to the two stocks case as possible biases will be 'diversified away' better. Again, the *HETIC* betas yield the best results with an MAE of 0.5261 followed by the Historical HF betas with MAE of 0.5744 and the other three methologies with MAEs around 0.68.

In all, the empirical asset allocation application confirms the results from section 5.1 and shows that *HETIC* betas are the best predictor for future market betas. As asset allocation is important not only for hedge fund managers but also for mutual fund manager who typically want to achieve a portfolio beta of one, we have performed a similar analysis with s target beta of one. The results are similar to the ones presented above.

5.3 Multifactor Linear Models

The linear pricing models are widely used in predicting expected returns, and it has been shown that adding factors beyond the market portfolio in general increases their power. We want to see if the use of the forward-looking information also boosts beta and return predictability for the linear multifactor models. In short, we want to add more factors to our model, and to perform an analysis similar to the one shown above. In constructing the stock factors we are limited to the S&P100 constituents, and hence our use of economic factors, such as size or book-to-market, is limited. Instead, we rely on the idea of statistical factors obtained from a principal component analysis (PCA).

We compute the variance-covariance matrix of the returns for all S&P100 constituents for the period January 2, 1996 to December 31, 2000 using high-frequency data and the estimation methodology outlined in section 4.1. Based on this covariance matrix we then perform a principal component analysis and finally transform the PCA coefficients into portfolio weights. As the return of the first statistical factor portfolio has a correlation of 95.35% with the S&P100 index, we associate the first factor with the market. In addition to this market factor we use two additional statistical factors from the PCA, so that we end up with a three factor model, in total explaining about 28.23% of the total variation in realized stock returns.

The FGK and CJV betas cannot be computed for factors that do not have traded options on them. In contrast, the *HETIC* model as well as the Historical HF and Historical Daily methodologies are appropriate to predict betas with respect to these statistical factor portfolios. These methodologies estimate a whole variance-covariance matrix for the stocks, and using this information we can compute the predicted betas with the expanded beta expression (10). Realized betas for the statistical factors are computed using the methodology from section 4.1 and formula (10).

To see how good these models are in predicting realized betas for the statistical factors, we run over the period January 2, 2001 to June 30, 2007 the following time-series regression for all DJ30 stocks:

$$\beta_{i,t}^{(k),Realized} = \alpha_i + \lambda_i \beta_{i,t}^{(k),Predicted} + \varepsilon_{i,t}, \forall t; \tag{21}$$

where $k = 2, 3$ denotes the statistical factors, excluding the market factor.

Table 9 shows that the *HETIC* model outperforms the Historical HF as well as the Historical Daily methods for the second factor, with an average R^2 of 65.48% vs. 60.10% and 56.37%, respectively and delivering the highest R^2 for 29 out of 30 stocks. For the third factor the Historical HF methodology delivers the best results (with an average R^2 of 17.93% and the highest R^2 in 13 cases) followed by *HETIC* (with an average R^2 of 16.04% and the highest R^2 in 10 cases) and Historical Daily (with an average R^2 of 15.49% and the highest R^2 in 7 cases).

Second, to analyse the return predictability we run the time-series regression in (19), representing a one-factor model, as well as the time-series regression based on a three-factor model given by:

$$r_{i,t}^{Realized} = \alpha_i + \gamma_i^{(M)} \beta_{i,t}^{(M),Predicted} + \gamma_i^{(2)} \beta_{i,t}^{(2),Predicted} + \gamma_i^{(3)} \beta_{i,t}^{(3),Predicted} + \varepsilon_{i,t}, \forall t;$$

Both regressions are performed for the DJ30 stocks over the period from January 2, 2001 to June 30, 2007. The results are presented in Table 10. The *HETIC* model delivers the best results for both types of time-series regressions, with average R^2's of 4.47% of 8.95% for the one factor and the three factor model, respectively, followed by the Historical HF methodology (3.81% and 8.85%) and the Historical Daily methodology (3.55% and 7.90%).

5.4 Enhancing Predictability

In contrast to historical methodologies (high-frequency and daily) the FGK, the CJV and the *HETIC* methods use forward-looking information from option data. An investor would not rely too much on historical information, and rather use information reflecting updated market beliefs, if he anticipates some sudden changes (a structural break) in the return distributions. Such a structural break may take place for the whole economy (and hence for a return driving factor), e.g. when the political situation is unstable, or it may affect one company only, e.g. when a company undergoes a restructuring. In the first case we would expect that both components of the linear pricing model change, i.e. the exposure of the stock to the factor and the risk premium; in the second case we would only expect the beta to change (for a good review see Damodaran (2008)). Thus, we would expect forward-looking betas to have an advantage in highly volatile and unstable periods.

For each stock we split our sample period into two regimes based on the magnitude of the implied volatility using a two-state Markov regime switching model driven by an asset-specific state variable s_{ti}[17]:

$$\sigma_{it}^{MF} = \mu_{s_{ti}} + \kappa_{s_{ti}} u_{it},$$

where σ_{it}^{MF} denotes the model-free implied volatility of company i at time t and $u_{it} \sim N(0,1)$. Such model allows for a state-dependent mean $\mu_{s_{ti}}$ and a state-depedent variance $\kappa_{s_{ti}}$.

The asset-dependent state variable s_{ti} is unobservable for the investor and treated as latent. The realizations of the state S_{ti} are governed by a first order Markov chain with constant 2×2 transition probability matrix P with generic element:

$$P(S_{it} = s_{it} | S_{it-1} = s_{it-1}) = p_{s_{it} s_{it-1}}, \qquad s_{it}, s_{it-1} = 1, 2.$$

[17]For detailed descriptions on model specification and estimation for Markov-Switching models see Hamilton (1989) and Hamilton (1994).

As our main goal is to isolate asset-specific periods of high model-free implied volatility, we impose the additional restriction that the mean of the second regime, the volatile one, is higher than two times the mean of the overall time-series of the model-free implied volitlities σ_{it}^{MF}. That way, on average, only the highest 20% of the implied volatilities fall into the volatile regime.

For each of the two regimes, we then run regression (18) for the market beta separately. The results are presented in Table 11. For the more stable regime the *HETIC* model yields the best results (average R^2 of 30.04% and highest R^2 in 13 cases) closely followed by the Historical HF methodology (average R^2 of 28.92% and highest R^2 in 9 cases). In the case of the more volatile regime the *HETIC* model clearly performs best, with an R^2 of 43.65% and being best in 19 cases, followed by the Historical HF methodology (average R^2 of 38.08% and highest R^2 in 4 cases). As shown in Table 11 (b) the difference between the *HETIC* model and the Historical HF as well as Historical Daily methodology is significant at the 5% level for the volatile regime when performing a one-sided test for *HETIC* Betas being better than Historical HF as well as Historical Daily betas.

6 Conclusion

Linear factor models are by far the most popular means of predicting stock returns. It is clear that for a good performance of a model accurate predictions of the stock factor exposures, i.e. the betas, are crucial.

We propose a new method of factor betas construction which is based on the estimated heterogeneous implied correlation (*HETIC*) model. The *HETIC* allows us to construct a positive definite implied correlation matrix that is fitted to the model-free implied variances of options on the index and its constituents. The estimated correlations are forward-looking by nature, and hence are based on the most current anticipation of future return dynamics. We show that *HETIC* has a better performance in predicting realized correlations than have historical correlations, and hence it is not surprising that the *HETIC* betas (that in addition to predicted correlations use option-implied variances) also shows the best performance in realized betas prediction. Moreover, we show that the forward-looking correlation structure matters a lot by itself, as the betas constructed using implied variances and historic correlations still lag behind. Various tests demonstrate that *HETIC* betas are the most efficient and unbiased predictor of realized betas from five models under consideration. Moreover, *HETIC* betas explain the most time-series variability in stock returns among competing methods, and account for a vast part of the cross-sectional variance in mean stock returns. In addition, they deliver the best results in an empirical asset allocation application (risk level targeting), outperforming the second-best methodology by about 30%.

From *HETIC* estimation and implied variances we have the whole forward-looking variance-covariance matrix at our disposal, and hence we can construct betas for any factor portfolio consisting of the available stocks. It turns out that the *HETIC* beta is the only one from the methods which uses solely forward-looking information that can be used to construct betas for factors other than the market. We use a PCA to get statistical factors from the return space, and test beta and return predictability using this artifical factor structure. Again, *HETIC* outperforms the other available methods by far.

For the estimation of historical (and realized) variances and covariances we use daily and high-frequency returns. From comparing these two methods we conclude that using high-frequency data leads to more accurate results.

There is one more important use of *HETIC* that requires further investigation. We show that the performance of *HETIC Betas* increases a lot in economically turbulent times, and we associate this predictability boost with the more precise correlation structure that becomes particularly important in such periods. Moreover, to be able to predict returns in different regimes one should have a way to estimate a factor risk premium associated with a given regime. A recent paper by Buss, Schlag, and Vilkov (2009) takes the first step in this direction, and shows promising results.

In all, *HETIC* delivers very promising results for correlation as well as betas and return predictability within the linear factor models framework.

Appendix A Construction of Risk-Neutral Moments[18]

Let the τ-period return be given by the log price relative:

$$R(t,\tau) \equiv ln[S(t+\tau) - S(t)].$$

Define a variance, a cubic and a quartic contract with the following payoffs:

$$H[S] = \begin{cases} R(t,\tau)^2, & \text{volatility contract;} \\ R(t,\tau)^3, & \text{cubic contract.} \\ R(t,\tau)^4, & \text{quartic contract.} \end{cases}$$

Let $V(t,\tau) \equiv \mathcal{E}_t^*\{e^{-r\tau}R(t,\tau)^2\}$, $W(t,\tau) \equiv \mathcal{E}_t^*\{e^{-r\tau}R(t,\tau)^3\}$, $X(t,\tau) \equiv \mathcal{E}_t^*\{e^{-r\tau}R(t,\tau)^4\}$ represent the fair value of the respective payoff.

The price of the variance contract is given by

$$V(t,\tau) = \int_{S(t)}^{\infty} \frac{2\left(1-\log\left(\frac{K}{S(t)}\right)\right)}{K^2} \cdot C(t,\tau;K)dK + \int_0^{S(t)} \frac{2\left(1-\log\left(\frac{K}{S(t)}\right)\right)}{K^2} \cdot P(t,\tau;K)dK, \quad (22)$$

the price of the cubic contract is

$$\begin{aligned} W(t,\tau) = \int_{S(t)}^{\infty} & \frac{6\log\left(\frac{K}{S(t)}\right) - 3\left(\log\left(\frac{K}{S(t)}\right)\right)^2}{K^2} \cdot C(t,\tau;K)dK \\ & - \int_0^{S(t)} \frac{6\log\left(\frac{K}{S(t)}\right) + 3\left(\log\left(\frac{K}{S(t)}\right)\right)^2}{K^2} \cdot P(t,\tau;K)dK, \end{aligned} \quad (23)$$

and the price of the quartic contract is

$$\begin{aligned} X(t,\tau) &= \int_{S(t)}^{\infty} \frac{12\left(ln[\frac{K}{S(t)}]\right)^2 - 4\left(ln[\frac{K}{S(t)}]\right)^3}{K^2} \cdot C(t,\tau;K)dK \\ &+ \int_0^{S(t)} \frac{12\left(ln[\frac{S(t)}{K}]\right)^2 + 4\left(ln[\frac{S(t)}{K}]\right)^3}{K^2} \cdot P(t,\tau;K)dK. \end{aligned} \quad (24)$$

[18] The formulas in this appendix closely follow the exposition in Bakshi, Kapadia, and Madan (2003) and are given for completeness. Only approximation procedure is our own.

Define

$$\mu(t,\tau) = e^{r\tau} - 1 - \frac{e^{r\tau}}{2}V(t,\tau) - \frac{e^{r\tau}}{6}W(t,\tau) - \frac{e^{r\tau}}{24}X(t,\tau). \tag{25}$$

Then we can calculate τ-period risk-neutral return skew as:

$$SKEW(t,\tau) = \frac{e^{r\tau}W(t,\tau) - 3\mu(t,\tau)e^{r\tau}V(t,\tau) + 2(\mu(t,\tau))^3}{(e^{r\tau}V(t,\tau) - (\mu(t,\tau))^2)^{\frac{3}{2}}}. \tag{26}$$

To calculate the integrals in (22), (23) and (24) precisely, we need a continuum of option prices. We discretize the respective integrals and approximate them from the available options. As we mentioned earlier, we have at least 4 options at our disposal for each maturity. Using cubic splines we interpolate the implied volatilities of these options inside the available moneyness range, and extrapolate using the last known (boundary for each side) value to fill in 1001 grid points in the moneyness range from 1/3 until 3 . Then we calculate the option prices from the interpolated volatilities using known interest rate for a given maturity, and use these prices to compute the model-free variance ($MFIV$) and risk-neutral skew ($SKEW$) as in (22) and (26).

References

Andersen, Torben G., and Tim Bollerslev, 1998, Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts, *International Economic Review* 39, 885–905.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Heiko Ebens, 2001, The Distribution of Stock Return Volatility, *Journal of Financial Economics* 61, 43–76.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys, 2000, Great Realisations, *Risk Magazine* 18, 105–108.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys, 2001, The Distribution of Exchange Rate Volatility, *Journal of the American Statistical Association* 96, 42–55.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Jin Wu, 2006, *Advances in Econometrics: Econometric Analysis of Economic and Financial Time Series* vol. B chap. Realized Beta: Persistence and Predictability, pp. 1–40.

Bakshi, Gurdip S., Nikunj Kapadia, and Dilip B. Madan, 2003, Stock Return Characteristics, Skew Laws, and the Differential Pricing of Individual Equity Options, *The Review of Financial Studies* 16, 101–143.

Bakshi, Gurdip S., and Dilip B. Madan, 2000, Spanning and Derivative Security Valuation, *Journal of Financial Economics* 55, 205–238.

Bandi, Federico M., and Jeffrey R. Russell, 2005, Realized covariation, realized beta, and microstructure noise, *Working paper series, The University of Chicago.*

Barndorff-Nielsen, Ole E., Peter R. Hansen, Asger Lunde, and Neil Shephard, 2008, Designing Realised Kernels to Measure the Ex-Post Variation of Equity Prices in the Presence of Noise, *Econometrica* 76, 1481–1536.

Barndorff-Nielsen, Ole E., and Neil Shephard, 2002, Estimating Quadratic Variation Using Realised Variance, *Journal of Applied Econometrics* 17, 457–477.

Breeden, Douglas T., and Robert H. Litzenberger, 1978, Prices of State-Contingent Claims Implicit in Option Prices, *Journal of Business* 51, 621–652.

Breen, William, Lawrence R. Glosten, and Ravi Jagannathan, 1989, Economic Significance of Predictable Variations in Stock Index Returns, *Journal of Finance* 44, 1177–89.

Britten-Jones, Mark, and Anthony Neuberger, 2000, Option Prices, Implied Price Processes, and Stochastic Volatility, *Journal of Finance* 55, 839–866.

Brown, Stephen J., 1990, *Financial Options: From Theory to Practice* chap. Estimating Volatility, pp. 516–537.

Buraschi, Andrea, Paolo Porchia, and Fabio Trojani, 2006, Correlation Risk and Optimal Portfolio Choice, *SSRN eLibrary.*

Buss, Adrian, Christian Schlag, and Grigory Vilkov, 2009, CAPM with Option-Implied Betas: Another Rescue Attempt, *Working Paper.*

Carr, Peter P., and Dilip B. Madan, 1998, Towards a Theory of Volatility Trading, *in Robert A. Jarrow, ed.: Volatility: New Estimation Techniques for Pricing Derivatives (RISK Publications, London).*

Chiriac, Roxana, 2006, Estimating Realized Volatility Wishart Autoregressive Model, *Working Paper.*

Christoffersen, Peter, Kris Jacobs, and Gregory Vainberg, 2008, Forward-Looking Betas, *McGill University Working Paper.*

Corsi, Fulvio, Gilles Zumbach, Ulrich Müller, and Michel Dacorogna, 2001, Consistent high-precision volatility from high-frequency data, *Economic Notes* 30, 183–204.

Damodaran, Aswath, 2008, Equity Risk Premiums (ERP): Determinants, Estimation and Implications, *SSRN eLibrary.*

Driessen, Joost, Pascal Maenhout, and Grigory Vilkov, 2009, The Price of Correlation Risk: Evidence from Equity Options, *Journal of Finance.*

Dumas, Bernard, 1995, The Meaning of the Implicit Volatility Function in Case of Stochastic Volatility, *unpublished note, HEC.*

Epps, Thomas W., 1979, Comovements in Stock Prices in the Very Short Run, *Journal of the American Statistical Association* 74, 291–298.

French, Dan W., John C. Groth, and James W. Kolari, 1983, Current Investor Expectations and Better Betas, *Journal of Portfolio Management* 10, 12–18.

Ghysels, Eric, and Eric Jacquier, 2006, Market Beta Dynamics and Portfolio Efficiency, *Working Paper, University of North Carolina.*

Gourieroux, Christian, and Razvan Sufana, 2004, Derivative Pricing with Multivariate Stochastic Volatility: Application to Credit Risk, *Working paper CREF 04-09.*

Griffin, Jim E., and Roel C. Oomen, 2006, Covariance Measurement in the Presence of Non-Synchronous Trading and Market Microstructure Noise, *Working Paper, University of Warwick.*

Hamilton, James D, 1989, A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle, *Econometrica* 57, 357–84.

Hamilton, James D, 1994, *Time Series Analysis.* (Princeton University Press).

Hansen, Peter R., and Asger Lunde, 2005, A Realized Variance for the Whole Day Based on Intermittent High-Frequency Data, *Journal of Financial Econometrics* 3, 525–554.

Hansen, Peter R., and Asger Lunde, 2006, Realized Variance and Market Microstructure Noise, *Journal of Business & Economic Statistics* 24, 127–161.

Hayashi, Takaki, and Nakahiro Yoshida, 2004, On Covariance Estimation for High-Frequency Financial Data, *Proceeding of IASTED/Financial Engineering and Applications 2004* 437-801, 282–286.

Hayashi, Takaki, and Nakahiro Yoshida, 2005, On Covariance Estimation of Non-Synchronously Observed Diffusion Processes, *Bernoulli* 11, 359–379.

Jacod, Jean, 1994, Limit of Random Measures Associated with the Increments of a Brownian Semimartingale, *Preprint University Paris VI.*

Jacod, Jean, and Philip Protter, 1998, Asymptotic Error Distributions for the Euler Method for Stochastic Differential Equations, *Annals of Probability* 26, 267–307.

Jiang, George J., and Yisong S. Tian, 2005, The Model-Free Implied Volatility and Its Information Content, *Review of Financial Studies* 18, 1305–1342.

Karatzas, Ioannis, and Steven E. Shreve, 1991, *Brownian Motion and Stochastic Calculus.* (Springer).

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

Lintner, John, 1965, The Valuation of Risky Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets, *Review of Economics and Statistics* 47, 13–37.

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

Sharpe, William F., 1964, Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk, *Journal of Finance* 19, 425–442.

Siegel, Andrew F., 1995, Measuring Systematic Risk Using Implicit Beta, *Management Science* 41, 124–128.

Vidyamurthy, Ganpathy, 2004, *Pairs Trading - Quantitative Methods and Analysis.* (John Wiley & Sons, Inc.).

Voev, Valeri, and Asger Lunde, 2007, Integrated Covariance Estimation using High-Frequency Data in the Presence of Noise, *Journal of Financial Econometrics* 5, 68–104.

Wang, Kevin Q., 2003, Asset Pricing with Conditioning Information: A New Test, *Journal of Finance* 58, 161–196.

Whistler, Mark, 2004, *Trading Pairs - Capturing Profits and Hedging Risk with Statistical Arbitrage Strategies.* (John Wiley & Sons, Inc.).

Zhang, Lan, 2006a, Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach, *Bernoulli* 12, 1019–1043.

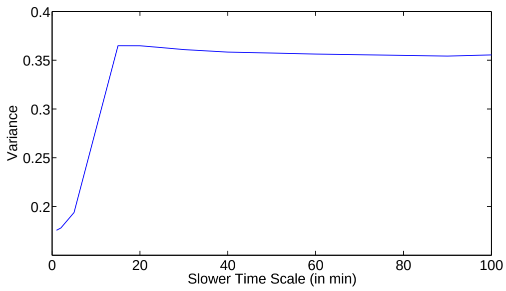
Zhang, Lan, 2006b, Estimating Covariation: Epps Effect, Microstructure Noise, *Working Paper, University of Illinois at Chicago.*

Zhang, Lan, Per A. Mykland, and Yacine Aït-Sahalia, 2005, A Tale of Two Time Scales: Determining Integrated Volatility With Noisy High-Frequency Data, *Journal of the American Statistical Association* 100, 1394–1411.

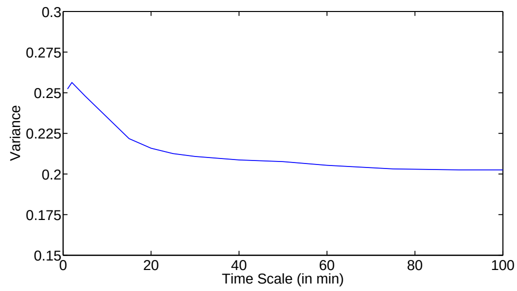
Zhou, Bin, 1996, High-Frequency Data and Volatility in Foreign-Exchange Rates, *Journal of Business and Economic Statistics* 14, 45–52.

Figure 1: Variance / Covariance Signature Plots

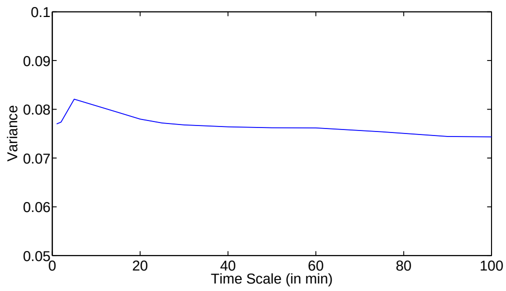
In Figure 1 we present variance/covariance signature plots. In (a) we plot the TSRV estimator for the variance of CSCO for the year 1998 against different choices for the slower time scale. The fast time scale is one minute. In (b) and (c) we plot the second-best estimator from Zhang, Mykland, and Aït-Sahalia (2005) (ZMAS) for the variance of DIS for the year 2002 and the variance of JNJ for the year 1998, respectively, against different choices for the time scale. In (d) we plot the second-best estimator from ZMAS for the covariance between CHV and WMT for the year 2000 against different choices for the time scale.



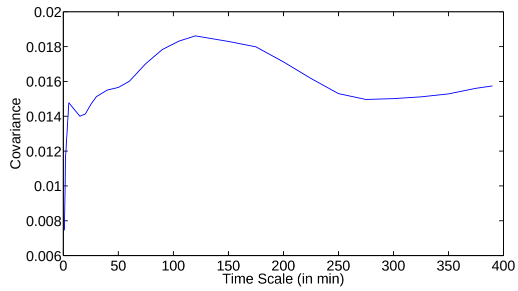
(a) Variance of CSCO for 1998.



(b) Variance of DIS for 2002.



(c) Variance of JNJ for 1998.



(d) Covariance for CHV and WMT for 2000.

Figure 2: Correlation State Variable versus Mean HETIC Correlation

Figure 2 presents the time-series of the implied correlation state variable and of the mean HETIC correlation for the S&P 100 constituents. The implied correlation state variable is computed using the methodology oulined in section 4.3. Then the pairwise *HETIC* correlation follows directly from (5). Finally we compute for each day the mean over all available pairwise correlations for the S&P 100 constituents to arrive at the mean *HETIC* correlation.

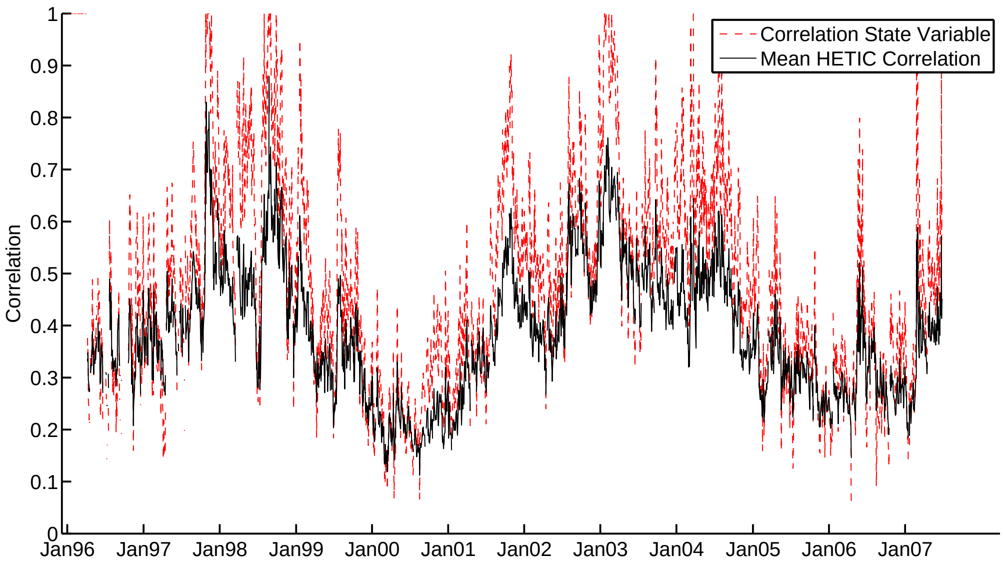
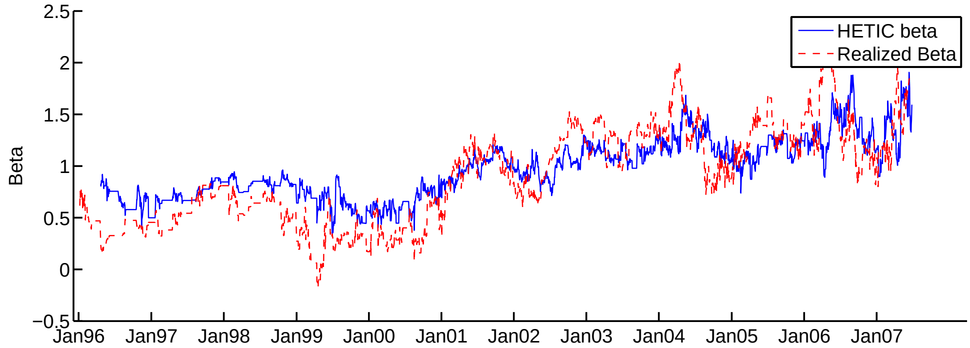
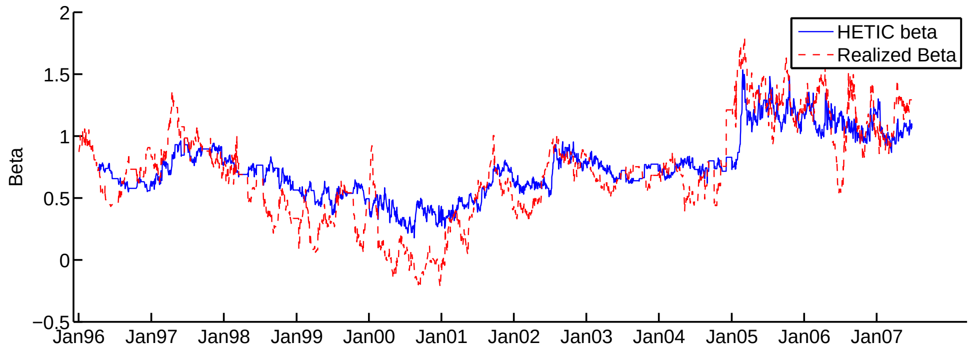


Figure 3: HETIC Beta versus Realized Beta

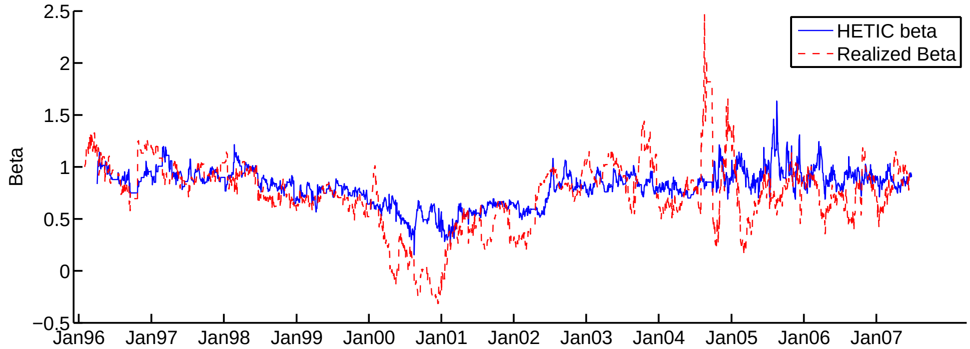
Figure 3 presents the time-series of the *HETIC* betas and of the realized betas for AA (a), XOM (b) and MRK (c) for the period January 2, 1996 to June 30, 2007. The *HETIC* betas are computed using options within the range of 30 to 90 days and choosing the options with maturity closest to 60 days if several are available. The parameters for the *HETIC* computations are $m = 0.5$ and $\Delta = 0.5$. The realized betas are calculated on each date for the period until the mean maturity of the available individual options on that day used for *HETIC* calculation and using high-frequency data.



(a) AA.



(b) XOM.



(c) MRK.

Figure 4: Mean Excess Return versus Mean HETIC Beta

In Figure 4 we plot the mean excess return for each of the S&P 100 constituents against the mean *HETIC* beta. The excess returns are calculated on each date for the period until the mean maturity of the available individual options on that day used for *HETIC* calculation. Averages are computed over the period January 2, 1996 to June 30, 2007. The solid line represents the regression of the mean excess return on a constant and the mean *HETIC* beta. The adjusted R^2 of the regression is 0.3407, the constant is -0.0100 with a t-statistic for significance of -2.0597 and the coefficient for the mean *HETIC* beta is 0.0332.

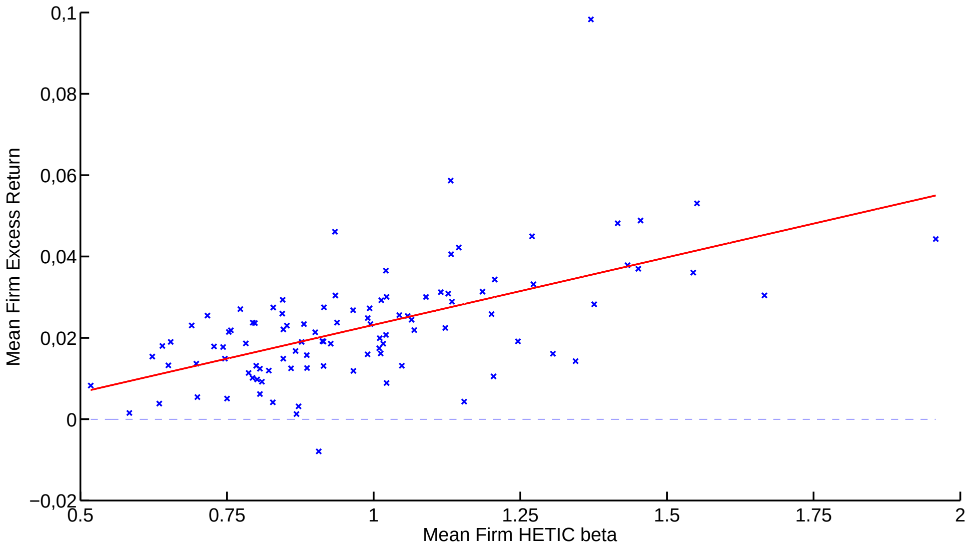


Table 1: DJ30 Components

The Table provides the list of the stocks that we use in our predictability analysis in the paper. The selected stocks were all components of DJ30 on February 19, 2008. For an easier link with the Ivy DB we provide in addition to the ticker the matching identifier of each stock in the database (Secid).

Secid	Ticker	Company Name
107616	MMM	3M CO
101204	AA	ALCOA INC
101375	AXP	AMERICAN EXPRESS CO
101397	AIG	AMERICAN INTERNATIONAL GROUP INC
109775	T	AT&T INC
101966	BAC	BANK OF AMERICA CO
102265	BA	BOEING CO
102796	CAT	CATERPILLAR INC
102968	CHV	CHEVRON CORP
103049	C	CITIGROUP INC
103125	KO	COCA-COLA CO
103969	DD	DU PONT E I DE NEMOURS & CO
104533	XOM	EXXON MOBIL CORP
105169	GE	GENERAL ELECTRIC CO
105175	GM	GENERAL MOTORS CORP
105700	HPQ	HEWLETT-PACKARD CO
105759	HD	HOME DEPOT INC
106203	INTC	INTEL CO
106276	IBM	INTERNATIONAL BUSINESS MACHINES
106566	JNJ	JOHNSON & JOHNSON
102936	JPM	J.P. MORGAN CHASE & CO
107318	MCD	MCDONALDS CORP
107430	MRK	MERCK & CO INC
107525	MSFT	MICROSOFT CORP
108948	PFE	PFIZER INC
109224	PG	PROCTER & GAMBLE CO
111459	UTX	UNITED TECHNOLOGIES CORP
111668	VZ	VERIZON COMMUNICATIONS INC
111860	WMT	WAL MART STORES INC
103879	DIS	WALT DISNEY CO

Table 2: Summary Statistics

The Table provides summary statistics for the S&P 100 index, the S&P 100 constituents as well as the DJ30 components. For the computation of the model-free variance and the model-free skewness we draw on options with a maturity within the range of 30 to 90 days and if several are available we choose the options with a maturity closest to 60 days. The average maturity of these options is given in the Table. The realized excess returns and the realized variances are calculated on each day for the period until the mean maturity of the available individual options using daily stock return data and are annualized. Moreover we give the ratio of the mean model-free variance to the mean realized variance and the average number of available, before filling empty data points, one minute transaction prices per trading day.

	S&P 100 constituents	DJ 30 constituents	S&P 100 Index
Average number of 1-minute transaction prices per day	277	353	387
Mean excess return, p.a.	0.1181	0.1099	0.0482
Average maturity of options, days	55.09	54.27	60.32
Mean Model-free variance (annualized)	0.3765^2	0.3438^2	0.2232^2
Mean Model-free skewness	-0.5628	-0.6310	-1.4009
Mean Realized variance	0.3359^2	0.2934^2	0.1689^2
Ratio of Model-free to Realized variance	1.1208^2	1.1715^2	1.3218^2

Table 3: Estimated Beta Correlations: Market Factor

The Table provides the summary of the correlation between the daily time series of the mean realized and of the mean predicted market factor betas constructed with different methods. The averages are computed over the DJ30 stocks. The *Realized* beta as well as the predicted *HETIC* and *Historical HF* betas are calculated using the high-frequency data. The *Historical Daily* and *GFK* betas are constructed with daily returns, while *CJV* beta does not require any historical variance-covariance matrix for estimations.

Market β	Realized	HETIC	Hist. HF	Hist. Daily	CJV	GFK
Realized	1.0000	0.5703	0.5399	0.4989	0.1329	0.4507
HETIC		1.0000	0.7587	0.7010	0.3924	0.7264
Hist. HF			1.0000	0.8596	0.2190	0.6989
Hist. Daily				1.0000	0.2056	0.7871
CJV					1.0000	0.3325
GFK						1.0000

Table 4: Betas Calculation Methods: A Comparison

The Table provides the summary of the market beta computations for different methodologies over the period January 2, 1996 to June 30, 2007. The predicted *HETIC*, *Historical HF* and *Realized HF* betas are using high-frequency data, *Historical Daily*, *Realized Daily* and *FGK* betas are using daily sampled returns, while *CJV* beta does not rely on any stock data for estimation. *Realized* beta is calculated on each date for the period until the mean maturity of the available individual options on that day used for *HETIC* calculation. For each day within the sample period we compute the mean market beta over the DJ30 stocks. The Table then reports the mean, median as well as standard deviation of the resulting time-series of average DJ30 betas. We report the difference between the mean of the predicted time-series and the mean of the realized time-series of average DJ30 betas. In addition we test for each DJ30 stock separately if the predicted and the realized betas are significantly different. As we have overlapping observations, we use Newey-West standard errors with 30 lags to account for autocorrelation in t-statistic calculation. We report for how many cases this hypothesis is not rejected.

	Forward-Looking (predicted)			Historical (predicted)		Realized	
	HETIC	CJV	FGK	HF	Daily	HF	Daily
Mean	0.9433	1.1609	0.8497	0.9288	0.9276	0.9249	0.9239
Median	0.9580	1.1418	0.8559	0.9545	0.9455	0.9533	0.9508
Standard Deviation	0.0825	0.1836	0.1398	0.1073	0.1107	0.1201	0.1241
Mean of Predicted-Realized	0.0184	0.2360	-0.0742	0.0039	0.0037	-	-
H0: Realized=Predicted not rej.	12	3	12	30	29	-	-
Number of observations	59170	59149	59229	59259	58434	59806	59106

Table 5: Correlation Calculation Methods: A Comparison

The Table provides the summary of the correlations between the DJ30 stocks and the market index as used in the computations of the market betas over the period January 2, 1996 to June 30, 2007. *Realized* and *historical* are using high-frequency data whereas *FGK* use daily sampled returns for the historical correlation calculations. *HETIC* uses high-frequency data and formula (5). For *CJV* we report the artificial correlations, as being the ratio of stock to index skewness: $\left(\frac{Skew_i}{Skew_m}\right)^{\frac{1}{3}}$. For each day within the sample period we compute the mean correlation between the DJ30 stocks and the market index. The Table then reports the mean, median as well as standard deviation of the resulting time-series.

	Correlation calculation method				
	HETIC	Realized	Historical	Historical Daily (FGK)	Artificial (CJV)
Mean	0.6207	0.5365	0.5379	0.5571	0.7497
Median	0.6251	0.5393	0.5443	0.5650	0.7344
Standard Deviation	0.1120	0.1101	0.1062	0.1125	0.0978
Number of Observations	59170	58434	59259	59229	59149

Table 6: Stock Beta Predictability: Market Factor

The Table provides the summary of the market factor beta predictability for the DJ30 stocks in terms of the coefficient of determination (adjusted R^2), i.e. in terms of the explained variability of the *Realized* beta. Over the whole sample period from January 2, 1996 until June 30, 2007 we run the regression of the daily time series of the *Realized* market beta on the predicted market beta and a constant for several methods of beta construction ($\beta_{i,t}^{Realized} = \alpha_i + \lambda_i \beta_{i,t}^{Predicted} + \varepsilon_{i,t}, \forall t$). As we have overlapping observations, we use Newey-West standard errors with 30 lags to account for autocorrelation in t-statistic calculation. *Realized* beta is calculated on each date for the period until the mean maturity of the available individual options on that day used for *HETIC* calculation. The average R^2 values are simple averages over all stocks for a given method. For each method we note the data sampling frequency used for *Realized* and predicted beta calculation. In Panel (a) we then provide detailed statistics for each stock, always using high-frequency data (HF) for beta calculation. In Panel (b) we provide summary statistics, including also the regressions results for the realized betas from daily returns. The best R^2 column shows in how many cases out of 30 regressions the best predictability was achieved.

(a) Individual Regressions Explanatory Power

	Adjusted R^2 for beta prediction				
Ticker	HETIC	Hist. HF	Hist. Daily	CJV	FGK
MMM	0.5276	0.4771	0.3590	0.0047	0.2643
AA	0.6206	0.7139	0.5551	0.0738	0.4026
AXP	0.3050	0.3896	0.2576	-0.0003	0.0427
AIG	0.3438	0.2733	0.2814	0.0504	0.2430
T	0.4118	0.4448	0.3580	0.2024	0.3616
BAC	0.3622	0.2417	0.1293	0.0408	0.0839
BA	0.3811	0.3519	0.2548	0.0002	0.2086
CAT	0.2632	0.1022	0.2746	0.0305	0.3123
CHV	-0.0025	0.0472	0.0090	0.0182	0.0165
C	0.2281	0.1574	0.1515	0.0118	0.0813
KO	0.5998	0.6263	0.5174	0.0069	0.5777
DD	0.4881	0.5409	0.4694	0.0225	0.3641
XOM	0.7429	0.7117	0.6672	0.1860	0.5703
GE	0.4102	0.4518	0.3624	0.0964	0.1236
GM	0.2323	0.1836	0.0748	0.1651	0.0815
HPQ	0.1451	0.1743	0.1076	0.0436	0.0659
HD	0.1218	0.0595	0.0920	0.0020	0.0886
INTC	0.1531	0.1038	0.1166	0.1717	0.2177
IBM	0.1966	0.1156	0.1643	0.0992	0.0713
JNJ	0.5224	0.5369	0.5422	0.0123	0.5458
JPM	0.3149	0.2852	0.1163	0.0861	0.1057
MCD	0.4876	0.4142	0.2376	0.0729	0.2363
MRK	0.3353	0.3298	0.2465	0.0005	0.3126
MSFT	0.1849	0.1244	0.0847	0.0809	0.0606
PFE	0.4359	0.3699	0.4025	0.0010	0.4320
PG	0.6058	0.6150	0.5255	0.0899	0.4830
UTX	0.3074	0.3182	0.2495	-0.0001	0.1361
VZ	0.3942	0.3977	0.2980	0.0370	0.2566
WMT	0.0851	0.1167	0.1606	0.0372	0.1683
DIS	0.4363	0.4817	0.4485	0.0406	0.3206
Average	0.3547	0.3385	0.2838	0.0561	0.2412
Predicted/Realized β	HF/HF	HF/HF	Daily/HF	../HF	Daily/HF

Table 6: Stock Beta Predictability: Market Factor (cont.)

(b) Individual Regressions Explanatory Power: Summary

Market β calculation Method	Predicted/Realized	Best R^2	$\alpha = 0$ not rejected	$\lambda = 1$ not rejected	Average R^2	Average correlation	Average MAE
HETIC	HF/HF	14	16	18	0.3547	0.5703	0.2018
Hist. HF	HF/HF	12	2	0	0.3385	0.5399	0.2038
Hist. Daily	Daily/HF	0	0	0	0.2838	0.4989	0.2370
CJV	../HF	0	2	1	0.0561	0.1329	0.3618
FGK	Daily/HF	4	0	1	0.2412	0.4507	0.2647
Hist. Daily	Daily/Daily	-	0	1	0.2220	0.4381	0.2717
CJV	../Daily	-	2	0	0.0386	0.1163	0.3938
FGK	Daily/Daily	-	0	4	0.1861	0.3903	0.3024

Table 7: Stock Returns Predictability: Market Factor

The Table provides the summary of the stock return predictability in terms of the adjusted R^2 from regressing daily the series of realized stock excess returns on the predicted market beta $r_{i,t}^{Realized} = \alpha_i + \gamma_i \beta_{i,t}^{Predicted} + \varepsilon_{i,t}, \forall t$. The realized returns are calculated each day over the period to the average maturity of options used in the *HETIC* estimation. The predicted *HETIC*, *Historical HF* betas are using high-frequency data, *Historical Daily* and *FGK* betas are using daily sampled returns, while *CJV* beta does not rely on any historical data for estimation. The average R^2 values are simple averages over all DJ30 stocks for a given method. Panel (a) provides detailed information stock by stock. Panel (b) gives a summary and different methods predictive power comparison in terms of the number of individual regressions where a given method achieves the best R^2 across all methods.

(a) Individual Regressions Explanatory Power

			Adjusted R^2		
Ticker	HETIC	Hist. HF	Hist. Daily	CJV	FGK
MMM	0.0067	0.0109	0.0054	0.0068	0.0104
AA	-0.0003	0.0048	0.0043	0.0029	0.0107
AXP	0.0376	0.0059	0.0049	0.0080	0.0068
AIG	0.0007	0.0065	0.0167	0.0131	0.0182
T	0.0247	0.0004	-0.0002	-0.0022	0.0024
BAC	0.0142	0.0033	0.0008	0.0009	0.0005
BA	-0.0002	-0.0003	0.0060	0.0065	0.0058
CAT	0.2145	0.0838	0.1420	0.0016	0.0998
CHV	0.0191	-0.0027	0.0088	0.1538	0.0056
C	0.0250	0.0195	0.0400	0.0083	0.0187
KO	0.0354	0.0285	0.0283	0.0024	0.0336
DD	0.0088	0.0000	0.0012	0.0135	-0.0003
XOM	0.0066	0.0136	0.0098	0.0060	0.0085
GE	0.0076	0.0117	0.0040	0.0387	0.0005
GM	0.0147	0.0275	0.0105	0.0080	0.0028
HPQ	0.0160	0.0375	0.0255	0.0013	0.0071
HD	0.0500	0.0766	0.0676	0.0222	0.0014
INTC	0.0047	0.0318	0.0303	0.0135	0.0219
IBM	0.0103	0.0009	0.0001	-0.0003	-0.0004
JNJ	-0.0005	0.0131	0.0094	0.0183	-0.0006
JPM	0.0157	0.0103	0.0328	0.0076	0.0060
MCD	0.0616	0.0696	0.0376	-0.0006	0.0340
MRK	0.0526	0.0506	0.0600	0.0064	0.0391
MSFT	0.0097	-0.0004	0.0072	0.0089	0.0008
PFE	0.0052	0.0295	0.0230	0.0098	0.0087
PG	0.0197	0.0229	0.0072	0.0185	0.0072
UTX	0.0196	0.0302	0.0303	0.0019	0.0080
VZ	0.0013	0.0087	0.0157	0.0185	0.0359
WMT	0.0429	0.0164	0.0067	0.0252	0.0051
DIS	-0.0003	0.0011	0.0005	-0.0001	-0.0005
Average	0.0241	0.0204	0.0212	0.0140	0.0133

Table 7: Stock Returns Predictability: Market Factor (cont.)

(b) Individual Regressions: Summary

Market Factor β	Best R^2	Average R^2	Average Corr.	Average MAE
HETIC	8	0.0241	0.0588	0.0840
Hist. HF	10	0.0204	0.0488	0.0839
Hist. Daily	4	0.0212	0.0468	0.0843
CJV	5	0.0140	-0.0646	0.0838
FGK	3	0.0133	0.0206	0.0841

Table 8: Mean Market Factor Beta Explanatory Power

The Table provides the summary of the cross-sectional stock return predictability with the market factor beta using different methods of beta calculation, i.e. we regress cross-sectionally the excess return on the mean predicted beta for each method ($\bar{r}_i = \alpha + \lambda\bar{\beta}_{m,i} + \varepsilon_i, \forall i$). The realized excess returns are calculated each day over the period to the average maturity of options used in the *HETIC* estimation. The predicted *HETIC*, *Historical HF* betas are using high-frequency data, *Historical Daily* and *FGK* betas are using daily sampled returns, while *CJV* beta does not rely on any historical data for estimation.

Market Factor β	R^2	α (per month)	Risk Premium (λ) (per month)
HETIC	0.3407	-0.0050	0.0166
Hist HF	0.2580	0.0016	0.0107
Hist Daily	0.2797	0.0007	0.0117
CJV	0.2367	-0.0076	0.0156
FGK	0.2755	-0.0006	0.0142

Table 9: Stock Beta Predictability Overview: Statistical Factors

The Table provides the summary of the statistical factor betas predictability for the DJ30 stocks in terms of the coefficient of determination (adjusted R^2), i.e. in terms of the explained variability of the *Realized* beta. We derive the statistical factors by running a Principal Component Analysis on the variance-covariance matrix of high-frequency returns of the S &P 100 stocks in the period from January 2, 1996 until December 30, 2000. The first principal component is highly correlated with the market factor, hence we present in this table the analysis for the second and the third principal components. Over the sample period from January 2, 2001 until June 30, 2007 we run the regression of the daily time series of the *Realized* factor beta on the predicted factor beta and a constant for several methods of beta construction ($\beta_{i,t}^{(k),Realized} = \alpha_i + \lambda_i \beta_{i,t}^{(k),Predicted} + \varepsilon_{i,t}, \forall t, k = 2, 3$). As we have overlapping observations, we use Newey-West standard errors with 30 lags to account for autocorrelation in t-statistic calculation. *Realized* beta is calculated using the high-frequency data. We calculate it on each date for the period until the mean maturity of the available individual options on that day used for *HETIC* calculation. The average R^2 values are simple averages over the DJ30 stocks for a given method. The best R^2 column shows in how many cases out of 30 regressions the best predictability was achieved.

Factor β	Best R^2	$\alpha = 0$ not rejected	$\lambda = 1$ not rejected	Average R^2	Average Correlation	Average MAE
Second PCA factor β results						
HETIC	29	14	3	0.6548	0.8030	0.1911
Hist. HF	1	15	10	0.6010	0.7595	0.1613
Hist. Daily	0	16	5	0.5637	0.7323	0.1758
Third PCA factor β results						
HETIC	10	23	1	0.1604	0.3651	0.1487
Hist. HF	13	11	1	0.1793	0.3844	0.0849
Hist. Daily	7	10	0	0.1549	0.3352	0.1052

Table 10: Stock Returns Predictability: Market vs. Multifactor Model

The Table provides the summary of the stock return predictability in terms of the adjusted R^2 from regressing daily the series of realized stock excess returns on the predicted market beta: $r_{i,t}^{Realized} = \alpha_i + \gamma_i \beta_{i,t}^{Predicted} + \varepsilon_{i,t}, \forall t$ as well as regressing daily the series of realized stock excess returns on the predicted market beta and the additional two statistical factor betas: $r_{i,t}^{Realized} = \alpha_i + \gamma_i^{(M)} \beta_{i,t}^{(M),Predicted} + \gamma_i^{(2)} \beta_{i,t}^{(2),Predicted} + \gamma_i^{(3)} \beta_{i,t}^{(3),Predicted} + \varepsilon_{i,t}, \forall t$ in the period from January 2, 2001 until June 30, 2007. The predicted *HETIC*, *Historical HF* betas are using high-frequency data, *Historical Daily* betas are using daily sampled returns. The realized excess returns are calculated each day over the period to the average maturity of options used in the *HETIC* estimation. The second and the third principal components are calculated using the variance-covariance matrix of high-frequency S&P 100 stocks' returns from January 2, 1996 until December 30, 2000. The average R^2 values are simple averages over all DJ 30 stocks for a given method.

Factor	Average R^2 for Factor beta regressions	
β	Market Factor	Three Factors
HETIC	0.0447	0.0895
Hist HF	0.0381	0.0885
Hist Daily	0.0355	0.0790

Table 11: Stock Beta Predictability with Regimes: Market Factor

The Table provides the summary of the market factor beta predictability for the DJ30 stocks in terms of the coefficient of determination (adjusted R^2), i.e. in terms of the explained variability of the *Realized* beta. Over the whole sample period from January 2, 1996 until June 30, 2007 we identify two different regimes (stable and volatile) using a two-state Markov-Switching model based on the model-free implied volatilities. For each regime we run the regression of the daily time series of the *Realized* market beta on the predicted market beta and a constant for several methods of beta construction ($\beta_{i,t}^{Realized} = \alpha_i + \lambda_i \beta_{i,t}^{Predicted} + \varepsilon_{i,t}, \forall t$). *Realized* beta is calculated on each day for the period until the mean maturity of the available individual options on that day used for *HETIC* calculation. The predicted *HETIC*, *Historical HF* betas are using high-frequency data, *Historical Daily* and *FGK* betas are using daily sampled returns, while *CJV* beta does not rely on any historical data for estimation. The *Realized* betas are computed using the high-frequency data. Panel (a) gives an overview for all methods. The average R^2 values are simple averages over all DJ30 stocks for a given method. Best R^2 column shows the number of cases in which the given method has the best predictability across all methods. For two stocks out the DJ30 the Markov-Switching model esimates such a low number of time perids being in the volatile regime that the regression results were not representative. We therefore do not report R^2 and the best method in these cases. In Panel (b) the t-statistics for the difference in mean R^2 across the DJ30 individual stocks for selected methods are provided. To account for heteroskedasticity we use White standard errors in t-statistic calculation.

(a) Individual Regressions: Summary

	Regime 1 (stable)		Regime 2 (volatile)	
Market β method	Average R^2	Best R^2 cases	Average R^2	Best R^2 cases
HETIC	0.3004	13	0.4365	19
Hist. HF	0.2892	9	0.3808	4
Hist. Daily	0.2468	3	0.3107	2
CJV	0.0470	0	0.0996	1
FGK	0.2128	5	0.2644	2

(b) Explanatory Power Difference

	Hist. HF	Hist. Daily
Regime 1 (stable)		
HETIC	0.8438	3.5143
Hist HF		3.0615
Regime 2 (volatile)		
HETIC	1.8973	3.3724
Hist HF		4.0583