

How much structure is best? A comparison of market model, factor model and unstructured equity covariance matrices

Beat G. Briner

MSCI Barra, 75 King William Street, London EC4N 7BE, UK;
email: beat.briner@mscibarra.com

Gregory Connor

London School of Economics, Room A205, Houghton Street, London WC2A 2AE, UK;
email: g.connor@lse.ac.uk

This paper compares three approaches to estimating equity covariance matrices: a factor model, a market model and an unstructured asset-by-asset model. These approaches make different trade-offs between estimation variance and model specification error. We explore this trade-off with a simulation experiment and with an empirical analysis of UK equity portfolios. The factor model is found to perform best for large investment universes and typical sample lengths. The market model underperforms due to excessive specification error while an asset-by-asset model with a short half-life of 22 days underperforms due to high estimation variance. The importance of properly accounting for serial correlation is highlighted.

1 INTRODUCTION

An important issue in covariance matrix estimation is how much structure should be imposed on the estimated matrix. This issue arises because of the need to forecast the risk of equity portfolios in markets with thousands of assets from only a limited amount of historical returns data. Estimating a covariance matrix without imposing any structure allows for any possible relationship between individual equity returns, but this flexibility comes at the cost of high estimation variance. Imposing some structure on the matrix reduces the number of parameters and thereby can decrease the estimation variance. However, structured estimators are likely to exhibit some specification error since it is highly improbable that the imposed structure exactly mirrors the underlying return-generating process for all assets. Assessing how much structure should be imposed on the matrix is therefore equivalent to exploring the trade-off between estimation variance and model specification error.

In the past, several authors have investigated this trade-off by comparing structured and unstructured covariance matrices. Elton and Gruber (1973), Elton *et al* (1978) and Elton *et al* (2006) argue that a tightly structured covariance

The authors would like to thank Nicholas Herbert for technical support. We are also grateful to Lisa Goldberg and Jose Menchero for helpful comments.

matrix with constant correlations between all assets outperforms less structured models. However, Elton *et al* (2006) also find that combining their constant-correlation structure with industry and characteristic-based variables improves the risk forecasting accuracy. Chan *et al* (1999) compare a range of unstructured to tightly structured models. In general, they find that more structured models provide better forecasts, but they also emphasize that the optimal model choice depends upon the particular risk forecasting application.

Ledoit and Wolf (2003, 2004) propose a Bayesian mixing method. They select a structured covariance matrix, which is used as a Bayesian prior and linearly combined with an unstructured sample covariance matrix. The resulting mixture produces a posterior estimate that is “close” to the prior but does not exactly obey the imposed structure. Bayesian mixing is appealing as it provides a theoretically optimal way of combining a misspecified prior with a correctly specified but noisy sample covariance matrix.

Jagannathan and Ma (2003) compare structured and unstructured covariance estimators. They assess how portfolio constraints influence the accuracy of these estimators for portfolio optimization. They show that imposing portfolio constraints such as bounds on individual asset weights can improve the performance of an unstructured estimator. Using a clever change of variables, Jagannathan and Ma show that the portfolio constraints can be reinterpreted as Bayesian priors along the lines of Ledoit and Wolf.

Random matrix theory (RMT) has also been applied to analyze and filter covariance matrices – see, for example, Laloux *et al* (1999), Plerou *et al* (2002) and Papp *et al* (2005). RMT defines bounds for the eigenvalue spectrum of “pure noise” correlation matrices, that is, correlation matrices calculated from samples of uncorrelated time series. By identifying and retaining only eigenvalues beyond the upper limit of the noise spectrum, RMT eliminates components of sample correlation matrices that are indistinguishable from noise. A limitation of this method is that it works only for correlation matrices and cannot provide information about the noise of the estimated variances.

Structured covariance matrices offer an essential advantage from a portfolio construction perspective. They reduce the dimensionality of the covariance estimator to a manageably small number of parameters. If the portfolio construction process uses categories such as regions or sectors to forecast expected returns, it is helpful to budget risk along the same dimensions. The market and factor models both share this practical advantage of low dimensionality. Even though the Bayesian mixing models also use a prior with low dimensionality, they do not offer the benefit of dimension reduction because they combine the prior with the unstructured estimator.

This paper examines three fundamental approaches to covariance matrix modeling that are of particular interest to practitioners. We compare the tightly structured market model, a moderately structured multi-factor model and an unstructured asset-by-asset model. These models are widely used for equity portfolio risk analysis, and they are building blocks for advanced approaches such as Bayesian mixing or RMT. After describing the three approaches from a statistical perspective,

we analyze the trade-off between estimation variance and specification error with a simulation experiment and with empirical results for the UK equity market. To complement the discussion of the three base models, we also provide simulation results and selected empirical data for a Bayesian mixing model that uses the market model as its prior. Bayesian mixing can be applied to a wide variety of priors. For example, a multi-factor prior could be used instead of the market model.

In the empirical part, our performance metric is the ability of an estimated covariance matrix to predict monthly portfolio volatilities over horizons of one to six months. Results are presented for a diversified market portfolio and for concentrated industry, style and random portfolios. We also analyze the performance of the market, multi-factor and asset-by-asset models in portfolio construction by using them to create beta hedged versions of the concentrated portfolios. Beta-hedged portfolios are evaluated by measuring their *ex post* betas and by calculating the portfolio turnover associated with readjusting the beta hedge. As part of the empirical investigation, it is shown how serial correlation between daily asset returns influences the monthly risk forecasts, and how important it is to properly correct for serial correlation.

Section 2 presents the three estimation methods and discusses why structured estimates may sometimes outperform unstructured ones even when the structural restrictions are not exactly correct. In Section 3, we use a simulation experiment to compare the forecasting performance of the three estimation methods for various data frequencies and sample lengths. In Section 4, we calculate monthly covariance matrices for models that represent the three methods, and we apply these matrices to forecast the risk of UK equity portfolios.

2 THREE ESTIMATION METHODS

2.1 Setting the stage: a univariate example

Before confronting the more complex multivariate problem of portfolio risk forecasting it is helpful to illustrate the trade-off between estimation variance and specification error with a simpler univariate example: the forecasting of individual asset betas.

Consider a market with n assets and suppose that the return of each asset i obeys the following process:

$$r_i(t) = c_i + \beta_i r_m(t) + \varepsilon_i(t) \quad (1)$$

where c_i is an arbitrary constant, β_i is the market beta of asset i , which is assumed to be constant over time, $r_m(t)$ is the market return and $\varepsilon_i(t)$ is the non-market return of asset i . Let the set $\{\varepsilon_i(t)\}_{i=1,\dots,n}$ be independent and identically distributed, normally distributed through time, and assume that non-market returns are statistically independent from market returns.

We compare two models for estimating β_i for the time period $[T + 1, T + H]$ using observations of asset and market returns over the interval $[1, T]$. The first

model is the maximum likelihood estimator $\widehat{\beta}_i^{\text{ML}}$. It equals the ordinary least-squares regression coefficient:

$$\widehat{\beta}_i^{\text{ML}} = \frac{\sum_{t=1}^T (r_i(t) - \bar{r}_i)(r_m(t) - \bar{r}_m)}{\sum_{t=1}^T (r_m(t) - \bar{r}_m)^2} \quad (2)$$

where \bar{r}_i denotes the average return over the sample interval $[1, T]$. We define the model specification error as:

$$\Delta_i = E[\widehat{\beta}_i - \beta_i] \quad (3)$$

The maximum likelihood estimator has zero specification error for all assets: $\Delta_i^{\text{ML}} = 0$ for all i . The maximum likelihood estimator is also consistent and efficient. It has minimal estimation variance, $(\Gamma_i)^2$, among all members in the set, $\{C\}$, of unbiased estimators that rely on the same sample data:

$$\begin{aligned} (\Gamma_i)^2 &= E[(\widehat{\beta}_i - E[\widehat{\beta}_i])^2] \\ (\Gamma_i^{\text{ML}})^2 &= \min_{X \in \{C\}} (\Gamma_i^X)^2 \end{aligned} \quad (4)$$

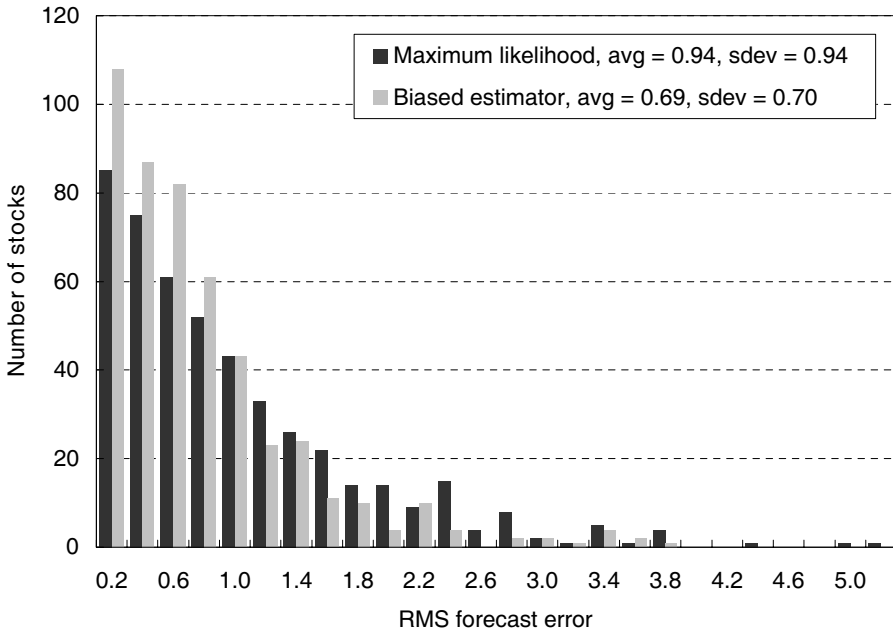
Given these desirable properties, it may seem that $\widehat{\beta}_i^{\text{ML}}$ is the best possible forecast for β_i over the subsequent time period. However, the maximum likelihood estimator has one weakness: it relies only on information contained in the sample of returns. From prior analysis of similar data sets, for example, we might know that the average of β_i across all assets is fairly close to 1. Based on this prior knowledge we could construct a tightly restricted estimator:

$$\widehat{\beta}_i^R = 1 \quad (5)$$

This estimator completely ignores all sample data and relies only on prior knowledge. If we view it as a classical estimator, it has poor statistical properties. It is biased and inconsistent for all $\beta_i \neq 1$, since $E[\widehat{\beta}_i^R - \beta_i] = 1 - \beta_i$ for all sample sizes. Despite its poor statistical properties, this restricted estimator can outperform the maximum likelihood estimator: $\widehat{\beta}_i^R$ can exhibit less forecasting error than the efficient estimate $\widehat{\beta}_i^{\text{ML}}$. This is straightforward to see, remembering that for an arbitrary estimator, X , the squared forecast error is the sum of squared specification error and estimation variance:

$$E[(\widehat{\beta}_i^X - \beta_i)^2] = (\Delta_i^X)^2 + (\Gamma_i^X)^2 \quad (6)$$

If a model introduces some specification error but compensates for this with a sharp reduction of estimation variance, it can outperform a correctly specified model. Figure 1 illustrates this point for the restricted estimator $\widehat{\beta}_i^R$, which obviously has zero estimation variance. Figure 1 shows a histogram of the root-mean-square (RMS) forecast error for 480 UK stocks, based on a 12-month estimation period and a subsequent 12-month forecasting period. It is evident that the restricted estimator $\widehat{\beta}_i^R = 1$ exhibits a lower average forecast error than the maximum likelihood estimator $\widehat{\beta}_i^{\text{ML}}$.

FIGURE 1 RMS error of asset beta forecasts for 480 UK stocks.

12-monthly returns in 2003 were used to calculate forecasts of beta for the subsequent year. The maximum likelihood estimator (dark bars) shows a larger average error and a broader error distribution than the biased estimator (light bars).

The outperformance of the estimator (5) can also be explained from a Bayesian perspective. Suppose that each asset's true beta is known to be the realization of a random variable with expected value of 1 and standard deviation σ^2 . If σ^2 is small relative to the estimation variance of the maximum likelihood estimator (4), then the restricted estimator will usually outperform even though it is misspecified and inconsistent. Although the restriction imposed on beta is false, it is "close enough to true" for improving the estimation accuracy in most cases.

2.2 Three methods of estimating covariance matrices

To forecast covariance matrices of equity returns, we consider three modeling approaches that differ by their amount of structure.

The asset-by-asset model:

$$\hat{C}^U = \frac{1}{T} R' R \quad (7)$$

calculates a covariance matrix from a $(T \times n)$ sample of n de-meaned asset returns, R , observed over a look-back horizon of T periods. This model is unstructured and has zero specification error; the true covariance matrix will always belong to the set of estimates of the form (7). The asset-by-asset model has $\frac{1}{2}n(n+1)$

parameters. For an investment universe of several hundred assets, this model contains a very large number of parameters. Using this approach to cover global equity markets would mean estimating about 10^8 parameters – a truly daunting task. The dependence on so many parameters is the main drawback of the asset-by-asset model – remember that estimation variance generally increases with the number of parameters in a model. For small investment universes and long samples, that is, if $T \gg n$, the asset-by-asset model exhibits good performance because its estimation variance vanishes for $T/n \rightarrow \infty$.

The tightly structured market model:

$$\widehat{C}^{MM} = \beta\beta' \text{var}(r_m) + \text{Diag}(\text{var}(\varepsilon_1), \dots, \text{var}(\varepsilon_n)) \quad (8)$$

assumes that the parameters β and $\text{var}(r_m)$ completely describe all asset covariances. The market model has $(2n + 1)$ parameters: a vector of n market betas, n asset-specific variances, $\text{var}(\varepsilon_i)$, and the market variance, $\text{var}(r_m)$. Betas are estimated via time series regression as indicated in (2) for the univariate case. The market model is parsimonious as it has so few parameters. Its main drawback is the fact that, in practice, it is likely to have considerable specification error. The market model cannot accommodate the fact that non-market returns tend to be more correlated for firms in the same industry than across different industries. The market model also ignores so-called style effects, for example, the observation that non-market returns of firms with similar market capitalizations or similar book-to-price ratios tend to be positively correlated (see, for example, Fama and French (1993)). Even though the exact definition of such sources of systematic non-market covariance may be debatable, the fact that they exist is empirically well established, which means that (8) provides a simplified and misspecified covariance model.

The moderately structured factor model:

$$\widehat{C}^{FM} = BC_F B' + C_\varepsilon \quad (9)$$

describes asset covariances through $(n \times k)$ exposures, B , to k hidden factors. It decomposes the covariance matrix into a common factor part, $BC_F B'$, and an asset-specific part, C_ε , which is assumed to be diagonal. The number of parameters depends on how the factor exposures, B , and the factor returns covariance matrix, C_F , are determined. In Section 4 we use a fundamental factor model that predefines the exposures from external information such as industry classifications or balance sheet data – this is conceptually similar to (5) where a beta of one is predefined. Alternatively, a statistical or macroeconomic factor model could be used to investigate the reduction of estimation noise through application of a factor structure. For a comparison of the different types of factor models we refer the reader to Connor (1995).

With respect to the restrictiveness of its structural assumptions, the factor model is positioned between the market model and the unstructured model. Because the factor model has more parameters, it offers more flexibility than the market model in approximating the true returns-generating process. However, factor models are also likely to exhibit some specification error. To achieve the goal of reducing

the estimation variance, k needs to be much smaller than n , which means that the factor model must focus on the strongest sources of return covariance. For example, a factor model with a banking industry factor might miss a systematically higher covariance between banks with similar regional loan exposures or similar concentrations of mortgage lending versus corporate lending. A factor model could also be misspecified because factors might be of a transitory nature and lose their explanatory power over time. The choice of k is another potential source of misspecification. We already saw that the market model (which could also be described as a factor model with $k = 1$) misses systematic sources of covariance. Including too many factors can be equally detrimental because this could mean overfitting the sample data by keeping factors without any persistent forecasting power.

Bayesian mixing is an alternative approach to control for the optimal amount of structure in a covariance matrix estimator. Ledoit and Wolf (2003, 2004) show that a linear combination of a tightly structured prior and the unstructured covariance matrix can perform better than either of these on its own. Letting λ denote a real number between zero and one, we consider the estimator:

$$\widehat{C}^{LW} = \lambda \widehat{C}^{MM} + (1 - \lambda) \widehat{C}^U \quad (10)$$

The mixing coefficient λ is determined by a Bayesian method following Ledoit and Wolf (2004). An appealing feature of Bayesian mixing is that it holds the promise to improve the accuracy of a simple, tightly structured prior without the need for constructing a more complex factor model. Bayesian mixing is therefore often used in conjunction with simple priors such as the market model or a constant correlation matrix. In line with this practice, we use the market model matrix \widehat{C}^{MM} as a prior.

2.3 Forecasting portfolio risk

The usual purpose of estimating a covariance matrix is to predict the future volatility for a broad range of portfolios. An arbitrary portfolio can be described by an n -vector of asset weights, w . The estimated portfolio volatility is:

$$\widehat{\sigma}_w = (w' \widehat{C} w)^{1/2} \quad (11)$$

In analogy to the univariate case, we write the following expressions for the specification error, estimation variance and squared estimation error of $\widehat{\sigma}_w$:

$$\begin{aligned} \Delta_w &= E[\widehat{\sigma}_w - \sigma_w] \\ (\Gamma_w)^2 &= E[(\widehat{\sigma}_w - E[\widehat{\sigma}_w])^2] \\ E[(\widehat{\sigma}_w - \sigma_w)^2] &= (\Delta_w)^2 + (\Gamma_w)^2 \end{aligned} \quad (12)$$

Here $\sigma_w = (w' C w)^{1/2}$ denotes the true portfolio volatility, which is assumed to be constant over time. Note that both components of the estimation error in (12) now depend not only on the estimation method but also on the particular portfolio being examined. One estimation method may work best for the market portfolio

or other broadly diversified portfolios, but a different estimation method may be advantageous for portfolios that contain only few assets or are highly concentrated in some industries.

The specification error of the asset-by-asset model is zero for all portfolios. As we impose a tighter structure by using a factor model with a decreasing number of factors, the estimation variance will decrease but the specification error is likely to increase. Because the optimal compromise between specification error and estimation variance depends on w , it is essential to test any model with a broad range of different portfolios. For example, the market model might produce adequate forecasts for a broadly diversified portfolio but fail to forecast portfolio variance correctly for portfolios that are concentrated in single industries.

For simplicity of notation, the analysis so far has been based on the idealized assumption of a constant true covariance matrix. Knowing that real-world variances and correlations exhibit temporary fluctuations, practitioners often prefer to use exponentially weighted averages for calculating sample covariances. This method places more weight on recent observations and therefore responds more quickly to any unknown dynamic variation in the true covariance matrix.¹ However, the advantage of better responsiveness comes at the price of reducing the effective sample size because distant observations receive only little weight. We can write the effective sample size of an exponentially weighted sum with weights $w_t = \lambda^{t-1}$, $t = 1, \dots, T$, $\lambda < 1$ as:

$$T_{eff} = \frac{(1 - \lambda^T)^2(1 - \lambda^2)}{(1 - \lambda)^2(1 - \lambda^{2T})} \quad (13)$$

Exponential weighting is used for the empirical analysis in Section 4 to account for the dynamic variations of variances and correlations exhibited by real-world financial returns. It could also be used in the theoretical and simulation parts but this would not change any of the basic findings.

3 A SIMULATION-BASED ANALYSIS

In this section we use a simulation experiment to examine the trade-off between estimation variance and specification error. We assume in this experiment that the true return-generating process follows an 11-factor model with 10 industry factors and one market factor:

$$r_i(t) = \beta_{im}f_m(t) + \beta_{ij}f_j(t) + \varepsilon_i(t) \quad (14)$$

Each asset, $i = 1, \dots, 200$, is associated with one industry, $j = 1, \dots, 10$, that is, each asset has a single non-zero industry beta in addition to its market beta. We repeatedly draw a set of T random returns for a universe of $n = 200$ assets and then apply different models to estimate the true covariance matrix from the time series of simulated returns.

¹This method has been popularized by RiskMetrics, see Mina and Xiao (2001).

To fully specify the model (14), we first define a constant reference covariance matrix of market and industry factor returns. Factor returns are assumed to be independent and follow a normal distribution with zero mean. Market and industry factor annual volatilities are set to 20%. We note that factor models that are calibrated to empirical data typically exhibit inter-industry correlations. Inclusion of such correlations is, however, not necessary to demonstrate the relative importance of estimation variance and specification error. Therefore, we select the simplest possible factor model covariance matrix in these simulations.

Market and industry betas are drawn from a normal distribution with mean of one and standard deviation of 10%. This parameter choice approximately corresponds to the dispersion of betas that is found for empirical equity data.² Asset-specific returns $\varepsilon_i(t)$ are assumed to be independent and have a constant annualized volatility of 35%.

The simulated $(T \times n)$ matrix of sample returns is then used to build the following five models that represent the three estimation methods discussed in the last section:

- (a) an unstructured asset-by-asset model with equal weighting;
- (b) the true 11-factor model (14) with market and industry betas estimated via time series regression;
- (c) a market model that estimates the market betas, β_{im} , via time series regression and ignores the industry betas, ie, this model assumes that all non-market returns are asset-specific and independent;
- (d) a Bayesian mixing model that uses (c) as a prior and combines it with the asset-by-asset model (a); and
- (e) an 11-factor model that replaces the market and industry betas by dummy values of zero or one. This model is misspecified in a similar way as (5). It ignores the fact that the true betas are not uniformly equal to one. It is, however, quite close to the true process in the sense that it uses the correct number of industries and “knows” the correct industry of each asset.

Models (a) and (b) are correctly specified while models (c)–(e) exhibit some degree of misspecification. The dispersion of the randomly drawn true betas in (14) is intentionally set to a relatively high level to include a sizable amount of misspecification in model (e).

Simulations are run for different sample lengths to explore how the estimation variance depends on the sample size. We approximate a year with 250 trading days and draw returns for five sample sizes corresponding to 24 months, 60 months, 250 days, 500 days and 2500 days. A big advantage of the simulations is the fact that the true covariance matrix, C , is available (conditioning on the true random draw of betas). This allows us to directly compare the different models (a)–(e) with the true covariance matrix. Following Ledoit and Wolf (2003), we use the Frobenius

²Note that the dispersion of estimated betas is substantially higher than the dispersion of true betas due to estimation variance in the time series regression.

norm $|M|$ as a metric of estimation error between true and estimated covariance matrices:

$$M = \widehat{C} - C$$

$$|M| = \left[\left(\frac{1}{n} \right)^2 \sum_{j=1}^n \sum_{i=1}^n M_{ij}^2 \right]^{\frac{1}{2}} \quad (15)$$

$$\mu = E[|M|]$$

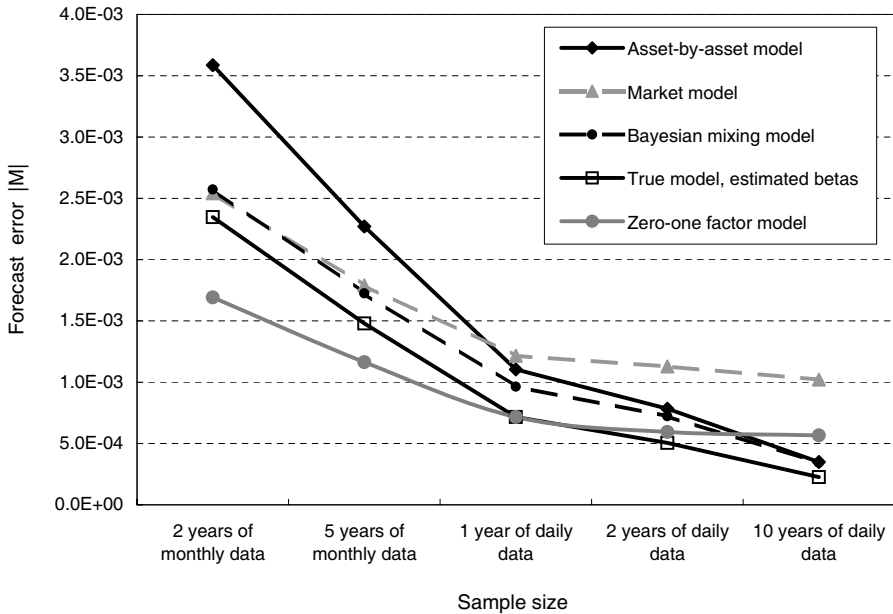
The expected value of the Frobenius norm, μ , can be interpreted as the average standard deviation of estimation error in a typical component of the estimated covariance matrix. This interpretation is valid in the following sense: if all components of the error matrix M have the same standard deviation, then μ will equal this standard deviation. For each model, $|M|$ is averaged over 100 simulation runs.

Figure 2 compares the five models for five different sample sizes. For sample lengths of up to one year of daily data, the zero-one factor model (e) outperforms all other models. This indicates that the reduction of estimation variance by predefining zero-one betas more than compensates for the specification error of this simplified factor model. For two years of daily data, the zero-one factor model (e) is marginally outperformed by the true factor model (b) with estimated betas. Only for the longest sample, which uses 10 years of daily data, do the asset-by-asset model and the true factor model clearly outperform the zero-one factor model. With such a long sample, the estimation variance of the correctly specified models (a) and (b) becomes small.

The market model (c) also outperforms the asset-by-asset model for up to five years of monthly data. This again emphasizes the advantage of reducing estimation variance through selection of a suitably structured estimator. However, for one year of daily data, the specification error of the market model becomes too large, making (c) the least accurate estimate. It is clearly advantageous to include the industry factors in the model as can be seen by comparing models (c) and (e).

The Bayesian mixing model (d) outperforms the asset-by-asset model (a) for all but the longest sample. This illustrates the ability of Bayesian mixing to improve estimates by reducing estimation variance. The Bayesian mixing model also outperforms its prior except for the shortest sample, where it is almost identical to the prior. These results confirm the merits of Bayesian mixing in optimally combining the unstructured estimator (a) with the structured prior (c). We note, however, that the zero-one factor model (e) outperforms the Bayesian mixing model for sample lengths of up to two years of daily data. The reason for this outperformance lies in the fact that the market model (c) prior exhibits a large specification error. This highlights the need for selecting a sufficiently accurate prior for Bayesian mixing.

In summary, these simulations illustrate the trade-off between estimation variance and model specification error. They indicate an advantage to selecting a moderately structured factor model to estimate risk for large investment universes and moderate sample lengths, that is, for $T/n < 1$. For large samples and small

FIGURE 2 Simulation results for $n = 200$ assets.

Forecast errors of asset-by-asset, factor, Bayesian mixture and market models. Results indicate how specification error and estimation variance contribute to the overall forecast errors as a function of the sample size.

investment universes, the unstructured asset-by-asset estimator performs best. We note that these simulation results depend on the exact investment universe and structure as well as on the sample length.

4 EMPIRICAL ANALYSIS USING UK EQUITY DATA

Simulations provide clear insight into the subtle balance between specification error and estimation variance, but they rely on an artificial return-generating process and cannot fully reflect the behavior of real-world financial markets. Therefore, it is important to explore the forecasting accuracy of the different modeling methods as they are applied to real data. In this section we demonstrate that the trade-off between specification error and estimation variance strongly influences indeed the accuracy of empirical risk forecasts. Empirical results also reveal how differences between the “imperfect” real-world data and the idealized theoretical assumptions impact the forecasts. In particular, we discuss the influence of serial correlation of daily stock returns on the risk forecasting accuracy.

4.1 Experimental framework

The empirical analysis uses daily UK equity returns over the period from 1997 to 2006. The universe includes all stocks of the FTSE All-Share index for which

at least 900 successive daily returns are available. Stocks with fewer returns are excluded from the universe to avoid problems with missing data when calculating asset-by-asset covariance matrices.³ The resulting universe varies in size between 485 and 684 assets over the backtesting period.

We calculate and analyze four different sets of covariance matrices that represent the tightly structured, moderately structured and unstructured modeling approaches. All matrices are calculated monthly, and their forecasting power is explored over a look-ahead horizon of one to six months. Two data points with large negative outlier returns (August 1998 and September 2001) are excluded from the bias tests because they would unduly distort the results.

To explore the multi-factor modeling approach, we use a commercial factor model for the UK market (MSCIBarra (2006)). This model contains 43 industry factors and the following 11 style factors: size, momentum, volatility, trading activity, leverage, value, yield, foreign sensitivity, growth, midcap indicator and non-estimation universe indicator. The model uses a 90-day half-life for the diagonal elements of the factor covariance matrix. Off-diagonal elements are modeled with a longer half-life of 180 days, based on the notion that correlations between factors tend to vary less rapidly than single factor variances. Specific risks are forecast with a structural model of cross-sectional asset-specific return heteroskedasticity. We henceforth refer to this model as the factor model.

Unstructured asset-by-asset matrices are calculated with exponential weighting. This is done to allow for the fact that real-world variances and correlations change over time, in contrast to the theoretical framework where true covariances are assumed constant. For this analysis we calculate asset-by-asset matrices using 900 trading days of returns data. Returns are truncated to $\pm 20\%$ to limit the influence of outliers on the covariance matrices. Daily risk forecasts are then scaled to a monthly forecasting horizon using a serial correlation correction, which will be further discussed below.

We present results for two sets of asset-by-asset covariance matrices. The first set uses the same 90- and 180-day half-lives as the factor model. This facilitates comparison between the models and ensures that both models exhibit the same responsiveness to sudden volatility changes. The second set of matrices uses a single half-life of 22 days. This half-life corresponds to a decay constant of $\lambda = 0.97$, a value that has been suggested as best practice for one-month forecasts by Mina and Xiao (2001) and is widely used by practitioners. We henceforth refer to these two models as the slow asset model (90/180-day split half-lives) and the fast asset model (22-day half-life), respectively.

Changing the half-life also has the effect of altering the effective sample size. According to (13), the effective sample size is only 63 days for a half-life of 22 days and $T = 900$ data points. For a half-life of 180 days we get $T_{eff} = 488$. Differences

³The inclusion of time series with missing data would introduce specification error into the covariance matrices. This is a common problem in the practical application of asset-by-asset matrices, which we want to avoid here. The reader should, however, be aware of this shortcoming when using asset-by-asset matrices to model real-world data.

TABLE 1 Characteristics and half-lives of models investigated in the empirical section.

Model	Characteristics	Half-life
Slow asset model	Unstructured	90 and 180 days
Fast asset model	Unstructured	22 days
Factor model	Moderately structured	90 and 180 days
Market model	Tightly structured	90 and 180 days

between the performance of the fast and slow asset models reflect the impact of sample size on estimation variance.

Finally, we define a market model to investigate the real-world performance of the tightly structured modeling approach. We extract predicted betas, $\hat{\beta}_i$, for each stock, i , using the covariance forecasts of the slow asset model:

$$\hat{\beta}_i = \frac{\text{cov}(r_i, r_m)}{\text{var}(r_m)} \quad (16)$$

The market returns, r_m , are defined as the returns of the capitalization-weighted universe portfolio. To complete the market model as described in (8), we calibrate its specific variances such that the market model and the slow asset model exhibit the same total asset variances. This is done to create a close relationship between both models. The market model can be interpreted as the result of imposing the maximum amount of structure on the matrices of the slow asset model. Table 1 gives an overview of the four models.

4.2 Risk forecasts of long-only portfolios

Before we apply the different models to forecast the risk of concentrated portfolios, we show that all four models generate reasonably accurate risk forecasts for a diversified market portfolio. The market is defined here as the cap-weighted universe portfolio. Figure 3 compares market risk forecasts of all four models with the realized volatility, which is calculated using six months of equally weighted daily returns. A serial correlation correction is applied to scale the realized volatility from a daily to a monthly horizon. Realized volatility data is then lagged by six months. This means that the realized volatility is a truly out-of-sample measure that has no overlap with the data used for deriving the forecasts. The slow asset model, market model and factor model exhibit very similar characteristics. On average these models predict market volatility well. Their forecasts do not overshoot in the high-volatility phase from 1999 to 2003 but they fail to quickly track the rapid volatility drop in early 2003. The factor model has a slight edge as it tracks the volatility peak in 2002 better than the slow asset model, but the differences are rather small. Owing to its short half-life, the fast asset model is able to quickly follow the volatility drop in early 2003, but this responsiveness comes at the price of additional noise. This is particularly evident in 1999 and 2002 where the forecast

FIGURE 3 Comparison of forecasts (continuous lines) with realized volatility (dashed lines) of the UK market portfolio. Realized volatility is lagged by six months and uses six months of equally weighted daily returns.

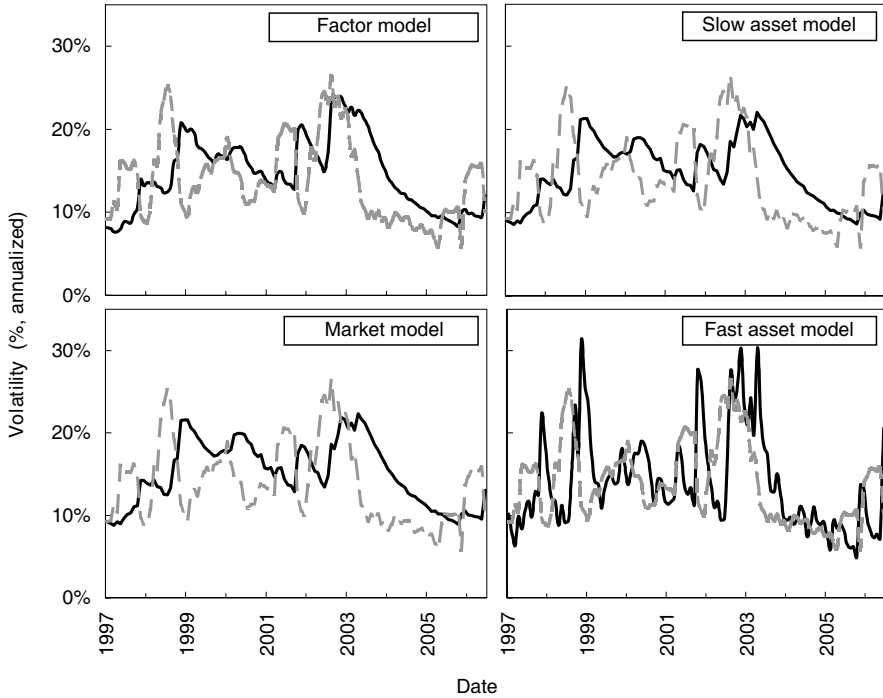


TABLE 2 Overview of portfolio categories investigated in the empirical part. Industry and style portfolios use the classification scheme of the factor model.

Portfolio group	Number of assets	Number of portfolios
Random portfolios	30	100
Industry portfolios	7–98	19
Style portfolios	48–69	18

overshoots considerably. However, if we set aside its higher noise level, the fast asset model also produces usable market volatility forecasts.

In the next step we investigate the forecasting accuracy of all models for the three groups of concentrated long-only portfolios listed in Table 2.

All portfolios are equally weighted to avoid dominance of a few large-cap stocks. Random portfolios are built by selecting 30 stocks from the 1997 universe and keeping these stocks as long as they remain in the universe. When a stock drops out of the universe it is replaced by another random draw. Industry portfolios are

based on the industry scheme of the factor model. Only industries with at least six assets over the entire 10-year backtesting period are included. This explains why we use only 19 industry portfolios, although the factor model contains 43 industry factors. Some industry portfolios (eg, banks or pharmaceutical companies) are very concentrated with 7–13 assets, whereas other industry portfolios such as retail or trusts are more diversified with up to 98 assets. Style portfolios use nine of the 11 style factors of the factor model. They contain either the top or the bottom decile of all assets ranked according to their exposure to a style factor. The style portfolios arguably could put the factor model at an advantage because they use its native factors. While industry portfolios also use the industry scheme of the factor model, we regard this subdivision as a more generic and model-independent categorization than in the case of the style portfolios. Random portfolios are of course entirely model independent.

Bias tests are conducted to determine the one-month forecasting accuracy of the different models. The bias statistic, $b_p(T, h)$, of portfolio p , at time T over a look-back horizon h , is defined as the sample standard deviation of the standardized outcomes $z_p(t)$. A standardized outcome is calculated by dividing the realized monthly portfolio return, $r_p(t)$, by the forecast portfolio volatility, $\hat{\sigma}_p(t-1)$, at previous month-end:

$$z_p(t) = \frac{r_p(t)}{\hat{\sigma}_p(t-1)} \quad (17)$$

$$b_p(T, h) = \sqrt{\frac{1}{h-1} \sum_{t=T-h+1}^T (z_p(t) - \bar{z}_p)^2} \quad (18)$$

If risk forecasts are exact, and if the true volatility is not time dependent, the bias statistic has an expected value of one. Assuming normality of returns and no estimation error in $\hat{\sigma}_p$, the bias statistic is approximately normal for large sample sizes, with standard error $1/\sqrt{2h}$, where h denotes the number of sample points (see Connor (2000) for a derivation). Using this large sample approximation and determining b_p for a set of different portfolios, 95% of all observations should fall within the two-sided confidence interval⁴ of $1 \pm \sqrt{2/h}$. If observations of b_p are clustered above (below) the confidence interval, this is evidence of model specification error that creates a downward (upward) bias in the risk forecasts. If observations are spread widely across both tail regions, this is evidence of estimation variance in the risk forecasts.

Figures 4(a) (industry and style portfolios) and 4(b) (random portfolios) present bias test results over the entire backtesting period of $h = 10$ years. Dark bars indicate observations outside the 95% confidence interval. Neither the factor model nor the slow asset model exhibit any statistically significant bias; no more than 5% of portfolios fall outside the confidence limits. In contrast, the market model and

⁴For notational simplicity we write the confidence interval as 2.0 standard deviations above and below the mean rather than using the exact value of 1.96 standard deviations.

the fast asset model both underforecast risk for a sizable fraction of all portfolios. The forecast error is larger for the industry and style portfolios than for the random portfolios. The fact that the tightly structured market model exhibits substantially more bias than the factor model is fully in line with our expectations: it has been explained in Section 2 that the market model is likely to have a fair amount of specification error, which can lead to biased risk forecasts.

Note that the fast asset model also exhibits bias. This may seem surprising: the fast asset model is unstructured and has no specification error; it should be noisy, but unbiased on average. Recall our remark earlier that the bias test relies on a constant true risk level though time. If true variances and correlations fluctuate over the estimation period and/or forecast period, the bias test results can exhibit one-sided tails outside the confidence bands even for a correctly specified model. To illustrate this point, we consider the univariate case of forecasting a single asset variance. Suppose for simplicity that, over a finite sample and forecast period, the true variance increases linearly through time:

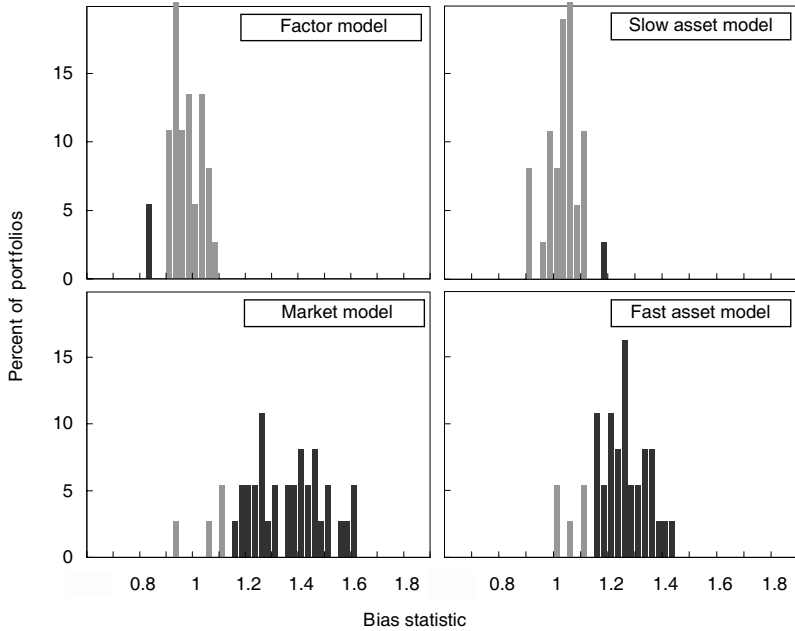
$$\sigma^2(t) = \sigma_0^2 + \gamma t \quad (19)$$

If we estimate $\sigma^2(t)$ using an equally weighted sample on the interval $[t - h, t - 1]$, we find that this estimator has an expected value of $\sigma_0^2 + \frac{1}{2}\gamma h$, that is, it underestimates the true variance at time t by $\frac{1}{2}\gamma h$. The dynamic behavior of real-world market returns is of course more complex than this simplistic example. Nevertheless, this example illustrates how temporary volatility fluctuations can induce temporary biases and hamper the interpretation of bias test results. In principle, a given model can underforecast over one observation period and overforecast over the following period. If a bias test aggregates over both periods, it could cancel these forecast errors with opposite sign, and the results could deceptively suggest the absence of any bias. To avoid such misinterpretations, it is essential that bias tests are performed using different look-back horizons. Figure 5 shows bias tests for the random portfolios that were generated with a sliding look-back window of $h = 2$ years. The upper panel of Figure 5 shows the average bias of all 100 random portfolios, the lower panel displays the percentage of portfolios that are either underforecast or overforecast. Values below the 5% threshold are insignificant in the lower panel, because confidence bounds were set to 95%.

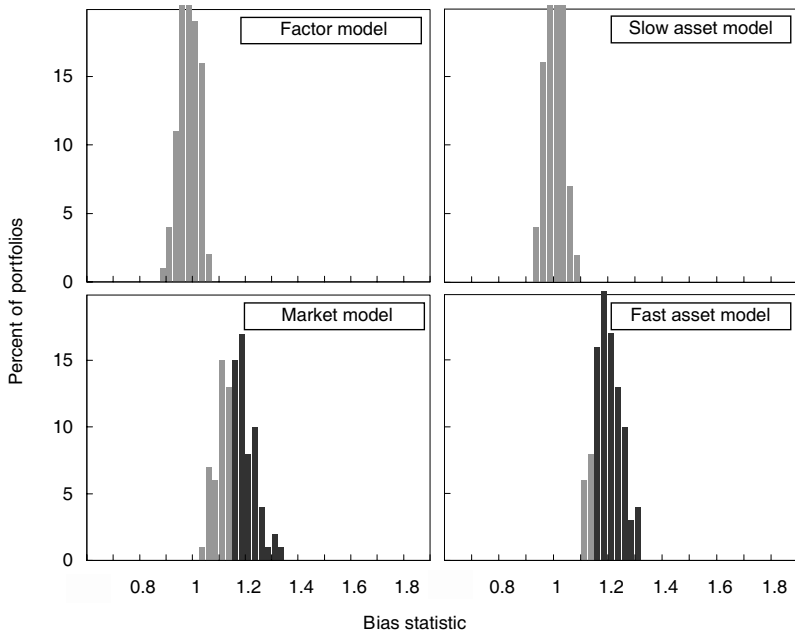
We find that the factor model and the slow asset model provide the best forecasts most of the time. In 1999 and in 2005 the factor model slightly outperforms the slow asset model. These two models both exhibit a tendency towards overforecasting in late 2000 and in early 2005. In contrast, the fast asset model and the market model heavily underforecast in 1999 and in 2003, but perform well in 2005. Figure 5 illustrates the dynamic nature of the bias tests, but it also confirms our previous findings that the factor model and the slow asset model provide more accurate forecasts. For the sake of brevity we do not show sliding-window bias tests of the industry and style portfolios because they exhibit very similar characteristics.

To illustrate how well the different models perform if the forecasting horizon is extended beyond one month, we also present bias statistics that were calculated

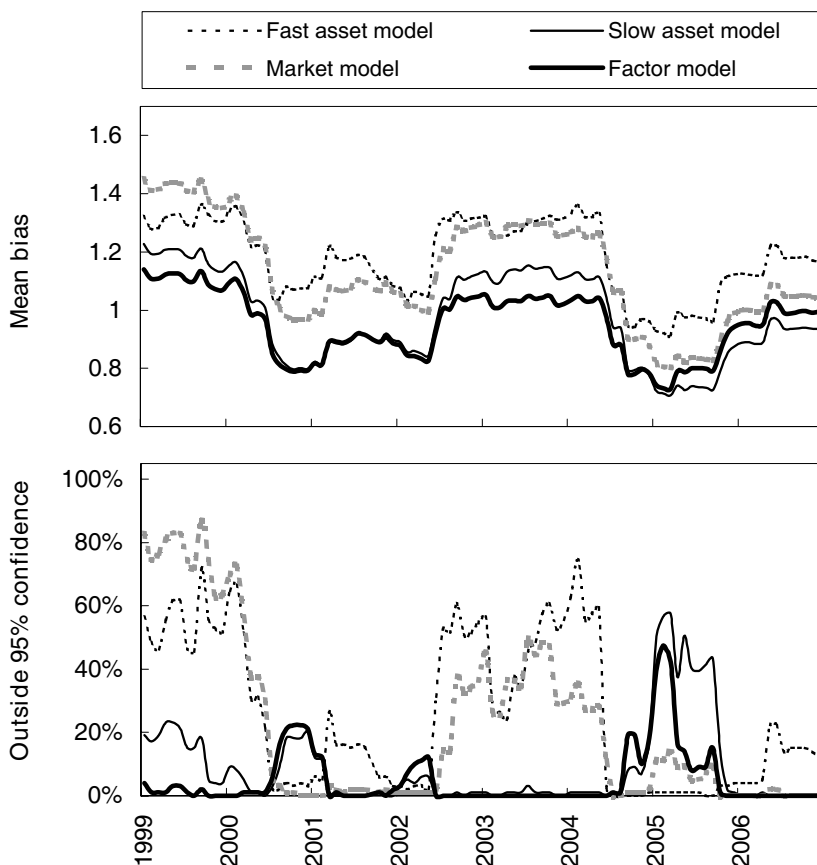
FIGURE 4 Histogram of bias statistics for (a) 37 long-only industry and style portfolios and (b) 100 random portfolios with 30 stocks, determined from 1997 to 2006. Dark bars indicate portfolios outside the 95% confidence interval.



(a)



(b)

FIGURE 5 Bias statistics for random portfolios, using a sliding two-year look-back window.

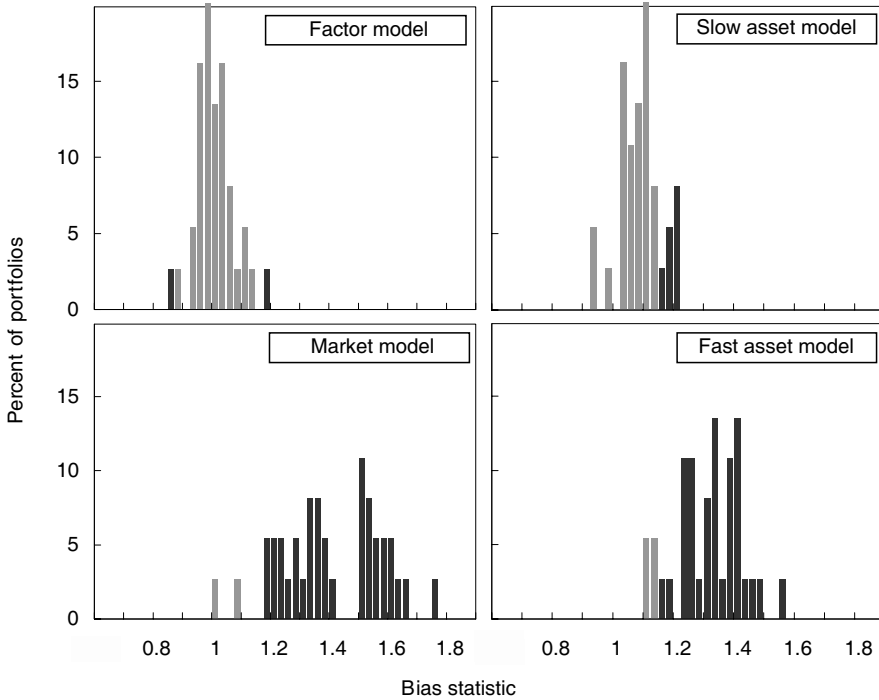
The upper panel displays the average bias statistic of all 100 portfolios. The lower panel indicates the percentage of all portfolios that exhibit bias statistic values outside the 95% confidence band.

from standardized outcomes with six months lag between forecast volatility and realized return:

$$z_p^{(6)}(t) = \frac{r_p(t+5)}{\widehat{\sigma}_p(t-1)} \quad (20)$$

Results for the industry and style portfolios are shown in Figure 6. They are directly comparable to the one-month bias statistics of Figure 4(a). We find that, for the six-month forecasting horizon, the factor model continues to produce unbiased forecasts while the quality of the slow asset model deteriorates somewhat: it now underforecasts risk for six out of 37 portfolios. We note that, for the random portfolios, the slow asset model and the factor model both generate accurate six-month forecasts. These findings indicate that the forecasting advantage of the factor

FIGURE 6 Forecast accuracy over a six-month horizon: a histogram of six-month bias statistics for 37 long-only industry and style portfolios, 1997 to 2006. Dark bars indicate portfolios outside the 95% confidence interval.



model over longer time horizons is not universal but related to the fact that the industry and style portfolios were built using its native factors. Industry and style factors are routinely used to budget active risk in portfolio management by over or underweighting exposures to these factors. The results shown in Figure 6 suggest that it is advantageous to use the same risk model for portfolio construction and risk control.

Bias test results for the three base models can be summarized as follows. Overall, the factor model and the slow asset model produce the most accurate forecasts. The market model is limited by its specification error, whilst the fast asset model suffers from high estimation variance. In Section 3 we have shown with a simulation experiment that Bayesian mixing can improve the forecasting accuracy of the base models. To explore how Bayesian mixing performs with empirical data, we build a set of covariance matrices that combine the market model prior with the slow asset model. Calculation of the mixing coefficient λ follows Ledoit and Wolf (2004) and adjusts for the fact that exponential weighting is used for the prior and for the slow asset model. We find an average λ of 0.45 over the 10-year backtesting period.

Risk forecasts from the Bayesian mixing matrices are then generated for the same equal-weighted industry and style portfolios shown in Figure 4(a).

Table 3 presents one-month bias test results for the entire 10-year backtesting period and for three shorter subperiods. The table indicates the percentage of all portfolios that exhibit bias statistic values outside the 95% confidence band. Values below 5% in Table 3 are not statistically significant. Table 3 compares the performance of Bayesian mixing with the factor, market and slow asset models. Over the entire backtesting period the slow asset model and the factor model perform best. Bayesian mixing performs worse as it is limited by the modest accuracy of the market model, which serves as its prior. Results for the low-volatility period from 2004 to 2006 shed further light on the subtle interdependence between the accuracy of the prior and the performance of Bayesian mixing. In this period Bayesian mixing outperforms all other approaches. From 2004 to 2006, the market model underforecasts the risk of some portfolios while the slow asset model tends to overforecast. This creates an ideal situation for Bayesian mixing, which, in this case, can counterbalance the errors of both base models. The opposite situation is observed in the high-volatility period from 2000 to 2003. Here the slow asset model provides accurate forecasts whereas the market model exhibits a strong tendency towards underforecasting. Even though Bayesian mixing yields much better results than the market model, it still underperforms the factor model and the slow asset model in the high-volatility period. These results are in line with the simulation results where we also found that the performance of Bayesian mixing strongly depends upon the choice of an accurate prior. With empirical data, Bayesian mixing can underperform one of its constituents. This somewhat contrasts with the simulations where Bayesian mixing is always at least as accurate as the prior and the unstructured asset model. The reason for this different behavior lies in the fact that empirical covariances change over time while the simulations work with constant true covariances. We note that the results shown here are not generally representative of Bayesian mixing, they only indicate how Bayesian mixing performs with a market model prior. Selecting a prior with a more complex structure and less specification error is likely to improve the forecasting performance of Bayesian mixing. Indeed, the combination of a factor model prior with Bayesian mixing may hold promising potential in terms of forecasting accuracy.

4.3 Influence of serial correlation between daily returns

It has become common practice to calculate asset-by-asset covariance matrices for risk forecasting from daily returns because daily returns are now readily available for many equity markets. A stylized fact of daily stock returns is, however, that they exhibit a fair amount of serial correlation, whereas monthly returns tend to be serially uncorrelated (see, for example, Lo and MacKinlay (1988) and Campbell *et al* (1993)). Serial correlation needs to be taken into account when creating monthly covariance forecasts from daily data. The covariance matrices that are

TABLE 3 Performance of the Bayesian mixing model compared to the asset, market and factor models by one-month bias tests for 37 long-only industry and style portfolios. The table indicates the percentage of all portfolios with bias statistic values outside the 95% confidence band.

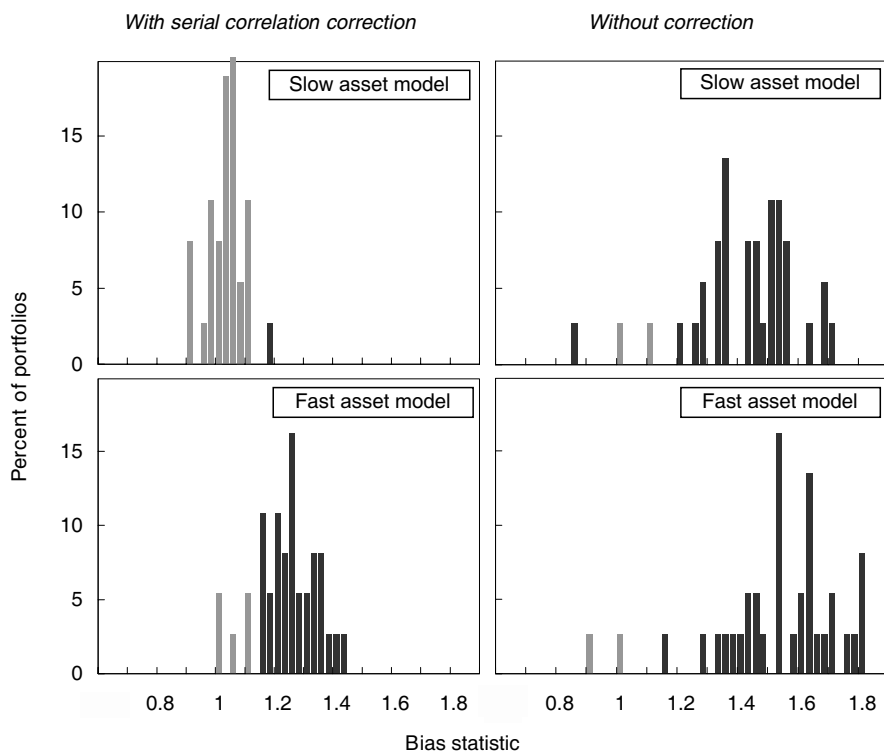
Period	Model	% Outside confidence	% Over forecast	Under forecast
1997–2006	Slow asset model	2.7	0.0	2.7
	Market model	89.2	0.0	89.2
	Bayesian mixing	51.4	0.0	51.4
	Factor model	5.4	5.4	0.0
1997–1999	Slow asset model	22.9	0.0	22.9
	Market model	91.7	0.0	91.7
	Bayesian mixing	64.6	0.0	64.6
	Factor model	2.1	2.1	0.0
2000–2003	Slow asset model	0.0	0.0	0.0
	Market model	66.7	0.0	66.7
	Bayesian mixing	7.7	0.0	7.7
	Factor model	5.1	5.1	0.0
2004–2006	Slow asset model	14.6	14.6	0.0
	Market model	12.2	2.4	9.8
	Bayesian mixing	4.9	4.9	0.0
	Factor model	9.8	9.8	0.0

generated from daily returns correspond to one-day forecasts. To scale these one-day matrices to a larger forecasting horizon we follow Newey and West (1987):

$$C^{(m)} = n \left[C_0^{(d)} + \sum_{l=1}^{n-1} \left(1 - \frac{l}{n} \right) (C_l^{(d)} + C_{-l}^{(d)}) \right] \quad (21)$$

Here $C^{(m)}$ denotes the monthly covariance matrix, $C_0^{(d)}$ the contemporaneous daily covariance matrix and $C_l^{(d)}$ the lagged daily matrix for lag l . The aggregation period, n , is set to 21, assuming an average of 21 trading days per month. All matrices in this empirical analysis have been corrected for serial correlation. Terms for $l > 5$ were ignored because it was found that serial correlations between asset returns tend to be negligible for lags beyond five days.

The existence of serial correlation is well known in the academic community. Nevertheless, from a practitioner's perspective it may seem tempting to reduce model complexity by omitting the calculation of lagged covariance matrices. This approximation corresponds to using the popular "square root of time" rule for scaling volatilities. Figure 7 demonstrates that this simplification introduces strong forecasting biases, which can dominate the more subtle biases shown in Figures 4–6. Figure 7 compares 10-year bias test results for industry and style portfolios that were generated with serial correlation correction (left column) and with square-root-of-time scaling (right column). It is evident that the good forecasting accuracy

FIGURE 7 Influence of serial correlation correction on the bias statistics.

The left column depicts results for industry and style portfolios that were obtained after applying serial correlation correction. The right column depicts results that were obtained without correcting for serial correlation.

of the slow asset model is completely lost when serial correlations are ignored. With serial correlation correction, the slow asset model produces far more accurate forecasts than the fast asset model. This is expected because the effective sample size is so different between these models. However, when square-root-of-time scaling is applied, the slow and fast asset models both produce inadequate forecasts. The error introduced by assuming daily returns to be uncorrelated is much larger than the error due to a small effective sample size.

Without providing further results we note that the same behavior was found for the market model and for the random portfolios. The factor model already includes a serial correlation correction in the calculation of the factor returns covariance matrix. We did not reverse this correction to explore how much the factor model would deteriorate, but we expect results that are similar to the findings for the other models.

4.4 Beta-hedged long-short portfolios

Beta-hedging is a method to neutralize the directional market exposure of a portfolio. It involves estimating the beta of a portfolio and then taking an offsetting position in the market index, or equivalently in an index future contract, to set the beta of the combined portfolio equal to zero. With the increasing popularity of long-short equity portfolios, full or partial beta hedging is now routinely used as a means to reduce the market exposure of an equity portfolio.⁵

Ex post empirical analysis of beta hedging is a powerful method to evaluate a risk model as a tool for portfolio construction, since beta hedging relies on a risk model to provide a beta forecast. The model performance can be determined along two dimensions:

- 1) by measuring the *ex post* beta of a beta hedged portfolio, that is, by testing how much market exposure the portfolio had in hindsight, given that the model predicted zero exposure; and
- 2) by investigating the turnover needed to periodically readjust the hedge, that is, by measuring how costly it is to maintain a hedged portfolio.

Beta-hedged versions of the industry, style and random portfolios that were discussed earlier are created by subtracting $\widehat{\beta}_p$ times the market portfolio. Each of the four risk models is used to construct a different beta hedged version of these portfolios. All hedges are readjusted monthly. As before, we use the cap-weighted universe portfolio as a proxy for the market.

Bias test results for the beta hedged portfolios are not shown here as the findings simply confirm the trends already discussed for the long-only portfolios. The factor model and the slow asset model provide unbiased forecasts, whereas the fast asset model and the market model tend to underforecast risk for a significant fraction of all portfolios.

Results of the investigation of *ex post* betas are shown in Figure 8 and Table 4. *Ex post* betas were determined by regressing the monthly portfolio returns against the market returns over the entire 10-year backtesting period. If the risk models were perfect, the *ex post* betas would be close to zero. In reality, the *ex post* betas will deviate from zero due to estimation variance and specification error in the risk model. Figure 8 illustrates that *ex post* betas for the beta hedged industry and style portfolios vary considerably. We note, for example, that all models over-hedge the pharmaceutical industry portfolio and the top10 momentum portfolio. Table 4 shows summary statistics of the *ex post* betas separately for each portfolio group. It indicates average, minimum and maximum *ex post* beta, and also lists the RMS forecast errors.

Table 4 corroborates our finding that the factor model and the slow asset model provide more accurate forecasts than the fast asset model and the market model, but the differences between the results of the four models are relatively

⁵Equity long-short products have gained much popularity after the decision of several European regulatory bodies to permit the application of so-called 130/30 strategies in mutual funds.

FIGURE 8 *Ex post* betas of (a) beta hedged industry and (b) style portfolios, 1997 to 2006.

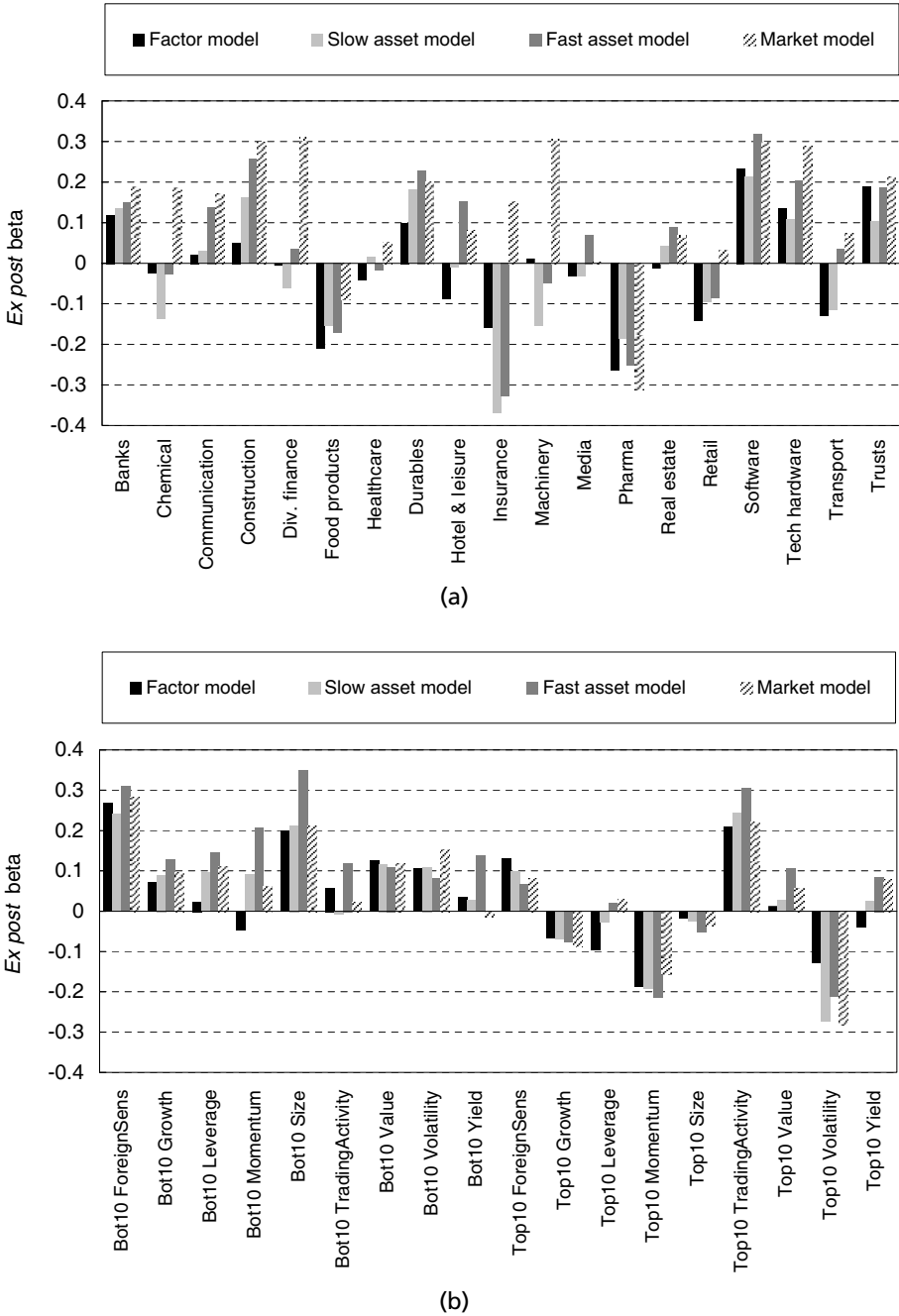


TABLE 4 Summary statistics of *ex post* betas for beta hedged industry, style and random portfolios.

Portfolio group	Model	Average <i>ex post</i> beta	RMS error of betas	Minimum <i>ex post</i> beta	Maximum <i>ex post</i> beta
Industry	Factor model	-0.013	0.121	-0.264	0.233
	Slow asset model	-0.017	0.136	-0.367	0.214
	Fast asset model	0.049	0.163	-0.327	0.319
	Market model	0.133	0.189	-0.311	0.310
Style	Factor model	0.036	0.112	-0.188	0.268
	Slow asset model	0.043	0.125	-0.274	0.244
	Fast asset model	0.089	0.160	-0.214	0.348
	Market model	0.053	0.129	-0.282	0.283
Random	Factor model	-0.053	0.267	-0.305	0.241
	Slow asset model	-0.054	0.287	-0.356	0.215
	Fast asset model	-0.029	0.293	-0.439	0.288
	Market model	-0.032	0.296	-0.469	0.256

Ex post betas are determined by equally weighted time series regression from 1997 to 2006.

small. For the industry portfolios, the factor model produces the smallest average *ex post* betas and the smallest RMS errors. Average *ex post* beta is worst for the market model, which tends to underestimate the betas. This is in line with its tendency of underforecasting risk seen in Figure 4(a). As expected, the factor model also performs best for the style portfolios that are tilted along its native factors. Surprisingly, here the market model outperforms the fast asset model by producing a smaller average *ex post* beta and RMS error. For the random portfolios, the results are somewhat inconclusive. On one hand, the market model and the fast asset model exhibit the lowest average *ex post* betas; on the other hand, these models show substantially larger extreme values of *ex post* beta than the factor model, which again has the smallest RMS error of all models.

Finally, we look at the amount of turnover needed to readjust the hedges. Table 5 presents average turnover data for all four models. Turnover is defined here as the average weight of each portfolio that needs to be traded per month, with the average taken over the 10-year backtesting period and over all portfolios. Table 5 indicates that the turnover of the fast asset model is always much higher than the turnover of the other three models; for the industry portfolios the fast asset model generates more than three times the turnover of the market model. High turnover is indicative of the use of a noisy covariance matrix for hedging. As explained in Section 3, the fast asset model is expected to exhibit a much higher estimation variance than the other models because it uses only a short effective sample size. By imposing a tight structure, the market model suppresses noise most radically and generates by far the lowest turnover for the industry portfolios. However, this desirable property of the market model comes at the price of less accurate beta forecasts as shown in Table 4. The factor model and the slow asset model generate very similar turnover for these portfolios, which gives the factor model a slight edge because it produces more

TABLE 5 Average monthly turnover for readjusting the hedges of the beta hedged industry, style and random portfolios.

Model	Industry portfolios	Style portfolios	Random portfolios
Factor model	3.27%	9.97%	2.57%
Slow asset model	3.28%	10.09%	2.55%
Fast asset model	8.13%	12.83%	5.78%
Market model	2.20%	10.12%	2.50%

accurate beta hedges. Note that the level of turnover is much lower for the industry and random portfolios than for the style portfolios. This is a consequence of the quasi-static character of the industry and random portfolios. The style rankings can change every month and consequently more turnover is required to rebalance the more dynamic style portfolios.

5 SUMMARY AND CONCLUSIONS

The trade-off between misspecification and estimation variance is a fundamental and easily underappreciated issue in risk forecasting. Risk models that impose no structure on returns are superficially appealing since they avoid any misspecification associated with structural restrictions. However, this advantage usually comes at the cost of higher estimation variance because unstructured models need to estimate a large number of parameters. Structured models always have some specification error, since no structural assumption can hold exactly across the entire set of security returns, but this disadvantage can be compensated by a sharp reduction in estimation variance.

In this paper we compare unstructured asset-by-asset covariance matrices, moderately structured multi-factor models and a tightly structured market model. In addition, we present results for a Bayesian mixing model combining the market model and the asset-by-asset model. Simulation results show that imposing some structure can be beneficial even if the imposed structure is not exactly correct. Empirical results for UK equity portfolios reinforce these findings and emphasize their practical relevance. They illustrate that the high estimation variance of an unstructured model with short half-life and the excessive specification error of a tightly structured market model both can produce inaccurate risk forecasts. We find that a moderately structured factor model tends to outperform the other two approaches. The performance of an unstructured model with long half-life trails the performance of the factor model only slightly. This is a consequence of the moderate-sized investment universe ($n \approx 600$) used in the empirical experiments; for large investment universes, the performance of the unstructured model would deteriorate further. Of course, all empirical results depend upon the selected market, the backtesting period and the model parameters. It therefore is not possible to state with complete generality that any one covariance matrix model will outperform in every possible circumstance.

In the case of Bayesian mixing, empirical results and simulations both point to the importance of selecting an accurate prior. With the market model prior, Bayesian mixing is outperformed by the factor model.

Besides comparing the accuracy of different model approaches, the empirical analysis also highlights the importance of correcting for serial correlation between daily returns when building monthly forecasts from daily returns data. It is shown that the omission of serial correlation corrections results in severe estimation errors that can mask the more subtle relationship between structural restrictions and forecasting accuracy.

Finally, the performance of the different models for beta hedging is investigated. It is shown that the use of a noisy, unstructured model results in excessive trading due to the need to periodically rebalancing the hedge.

So, how much structure is best? If we had the luxury of very long time series in markets with only few assets, an unstructured model would be the optimal choice. However, forecasting risk in larger investment universes and with samples of moderate lengths benefits from the application of a structured model as long as its misspecification does not outweigh the reduction of estimation variance.

REFERENCES

- Campbell, J. Y., Grossman, S. J., and Wang, J. (1993). Trading volume and serial correlation in stock returns. *The Quarterly Journal of Economics* **108**, 905–939.
- Chan, L. K. C., Karceski, J., and Lakonishok, J. (1999). On portfolio optimization: forecasting covariances and choosing the risk model. *The Review of Financial Studies* **12**, 937–974.
- Connor, G. (1995). The three types of factor models: a comparison of their explanatory power. *Financial Analysts Journal* **51**, 42–46.
- Connor, G. (2000). Robust confidence intervals for the bias test of risk forecasts. Technical Report, MSCI Barra.
- Elton, E. J., and Gruber, M. J. (1973). Estimating the dependence structure for share prices: implications for portfolio selection. *Journal of Finance* **28**, 1203–1232.
- Elton, E. J., Gruber, M. J., and Spitzer, J. (2006). Improved estimates of correlation coefficients and their impact on the optimum portfolios. *European Financial Management* **12**, 303–318.
- Elton, E. J., Gruber, M. J., and Ulrich, T. J. (1978). Are betas best? *Journal of Finance* **33**, 1375–1384.
- Fama, E. F., and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* **33**, 3–56.
- Jagannathan, R., and Ma, T. (2003). Risk reduction in large portfolios: why imposing the wrong constraints helps. *Journal of Finance* **58**, 1651–1684.
- Laloux, L., Cizeau, P., Bouchaud, J.-P., and Potters, M. (1999). Noise dressing of financial correlation matrices. *Physical Review Letters* **83**, 1467–1470.
- Ledoit, O., and Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance* **10**, 603–621.

- Ledoit, O., and Wolf, M. (2004). Honey, I shrunk the sample covariance matrix. *Journal of Portfolio Management* **30**, 110–119.
- Lo, A. W., and MacKinlay, A. C. (1988). Stock market prices do not follow random walks: evidence from a simple specification test. *Review of Financial Studies* **1**, 41–66.
- Mina, J., and Xiao, J. Y. (2001). *Return to RiskMetrics: the Evolution of a Standard*. RiskMetrics Group, New York.
- MSCIBarra (2006). *UKE7S – United Kingdom Short-horizon Equity Risk Model*. MSCI Barra, New York.
- Newey, W. K., and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55**, 703–708.
- Papp, G., Pafka, S., Nowak, M. A., and Kondor, I. (2005). Random matrix filtering in portfolio optimization. *Acta Physica Polonica B* **36**, 2757–2765.
- Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L. A. N., Guhr, T., and Stanley, H. E. (2002). Random matrix approach to cross correlations in financial data. *Physical Review E*, **65** 066126-1–066126-18.