UBS

UBS Investment Research

Global Quantitative Research Monographs

Global Equity Research

Global

Quantitative

Quantitative

UBS Hybrid Risk Model

17 December 2010

www.ubs.com/investmentresearch

James Sefton
Quantitative Analyst
james.sefton@ubs.com
+44 20 7568 6078

David Jessop
Analyst
david.jessop@ubs.com
+44 20 7567 9882

Giuliano De Rossi
Analyst
giuliano.de-rossi@ubs.com
+44 20 75683072

- **We argue for a new hybrid approach to risk modelling**

The principal advance of this approach is that it can incorporate both economic and characteristic risk factors

- **The risk model is estimated using an Expectation Maximisation Algorithm**

The Expectation Maximisation (EM) algorithm is guaranteed to be locally monotonically convergent.

- **The approach can be generalised to include Bayesian Priors**

By including Bayesian priors we can reduce both sampling errors and speed up convergence of the EM algorithm.

This report has been prepared by UBS Limited
ANALYST CERTIFICATION AND REQUIRED DISCLOSURES BEGIN ON PAGE 27.
UBS does and seeks to do business with companies covered in its research reports. As a result, investors should be aware that the firm may have a conflict of interest that could affect the objectivity of this report. Investors should consider this report as only a single factor in making their investment decision.

A short taxonomy of risk models

We shall start by briefly describing the different approaches to estimating equity risk models. We can then ask the central question of this section; which of these approaches is better? As always, there are advantages and disadvantages to each approach; however we identify three principal objectives.

Firstly, and arguably the most important, is model accuracy. The discussion here focuses on the inherent trade-off between structural and sampling errors – imposing more structure on the model reduces the number of parameters that need estimating, improving accuracy, but possibly at the cost of miss-specifying the model. The other two objectives concern the more general desire for parsimony, and ease of interpretation. We argue at then end of the section that our Hybrid model offers a good compromise over these sometimes conflicting objectives.

The Linear Factor Model

All risk modelling approaches are based on a linear factor model of the return generating process. This model can be written

$$\boldsymbol{r}_t = \boldsymbol{B}_t \boldsymbol{f}_t + \boldsymbol{\varepsilon}_t \tag{1}$$

where $\boldsymbol{r}_t$ denotes the n-vector of returns to the n assets at time t, $\boldsymbol{f}_t$ the k-vector of factor returns, $\boldsymbol{B}_t$ is the n *by* k matrix of factor exposures or betas and $\boldsymbol{\varepsilon}_t$ is the n- vector of residuals. The idea behind the linear factor model is that all the co-movement in returns can be captured by the $k(\ll n)$ factor returns. This reduces scale of the modelling problem. Instead of having to estimate the full n by n covariance matrix, we only need to estimate the n by k factor exposures and the factor covariance matrix. When there are a large number of assets this gain is significant.

To formalise this idea, assume that both $\boldsymbol{f}_t$ and $\boldsymbol{\varepsilon}_t$ are independently and identically distributed normal variates

$$\boldsymbol{f}_t \sim N(0,\boldsymbol{F}) \qquad \varepsilon_t \sim N(0,\boldsymbol{D}) \tag{2}$$

To simplify the notation, we focus on risk only and assume that the random variables have zero mean. This is for convenience only, and in the technical appendix, no such assumption is made. The identifying assumptions underpinning the linear factor model are the

$$Cov(\boldsymbol{f}_t,\boldsymbol{\varepsilon}_t) = 0 \qquad \boldsymbol{D} = diag(d_1, d_2 \ldots d_n) \tag{3}$$

Thus the residuals are independent and uncorrelated with the factors, and so all the covariance in returns is explained by the k factors.

Given these assumptions, asset returns are normally distributed as

$$\boldsymbol{r}_t \sim N(0,\boldsymbol{V}_t) \qquad \text{where} \quad \boldsymbol{V}_t = \boldsymbol{B}_t \boldsymbol{F} \boldsymbol{B}_t^T + \boldsymbol{D} \tag{4}$$

The covariance matrix of returns, $\boldsymbol{V}_t$ is referred to as the risk matrix. Building a Risk model is amounts to simply estimating a covariance matrix of this form.

The Three Basic Modelling Approaches

There are the three different approaches to estimating a risk matrix, which we shall refer to as time-series, cross-sectional and statistical modelling[1]. All approaches use a sample of returns. However, they differ on the assumptions they make concerning the structure of the risk matrix.

Statistical Modelling: This approach makes no additional assumption beyond the fact that there are k factors and the factor exposures and variances are constant over the sample period. Generally principal component analysis is used to identify the k factors from the return sample covariance matrix.

Cross-sectional modelling: This approach assumes that the factor exposures, $\boldsymbol{B}_{t,}$ are observed, though in practise fundamental data is used as a proxy for these exposures. As the factor exposures are 'observed', they can vary over time. The estimation problem is reduced to estimating the factor and residual covariance matrices, $\boldsymbol{F}$ and $\boldsymbol{D}$ respectively.

Time-Series modelling: This approach assumes that the factor returns, $\boldsymbol{f}_{t,}$ are observed and that the factor exposures are constant, i.e. $\boldsymbol{B}_t = \mathbf{B}$. In practise, it uses either time-series of economic data or returns to factor-mimicking portfolios as proxies for the factor returns. The estimation problem is therefore reduced to estimating the constant factor exposures and residual covariance matrix, $\boldsymbol{B}$ and $\boldsymbol{D}$ respectively.

We now assess the merits and weaknesses of these three different approaches to estimating a risk model.

The Design Trade-offs

What is the best approach to building a risk model? As always, there is no straightforward answer. Each approach offers a different balance on a number of objectives or trade-offs:

1) **Structural versus Sampling error.** When estimating a large risk matrix, it is necessary to specify some structure to reduce the number of parameters that need estimating. The more structure that is imposed, the fewer parameters that need estimating; so the more accurately these parameters are likely to estimated. Imposing more structure will reduce sampling or estimation error. However this structure could be mis-specified, for example we omit a significant factor, or incorrectly parameterised, for example we observe the betas only with an error; hence imposing structure will introduce structural errors. So there is a clear trade-off between structural versus sampling errors (this can be understood as a variation on the standard

[1]Time-series models are often called economic models, and cross-sectional are often referred to as random coefficients, fundamental or characteristic factor models. We shall use our descriptive nomenclature, which is related directly to the structural assumptions.

bias versus variance trade off present in most statistical estimation problems). We now briefly discuss the key structural assumptions implicit in the different modelling approaches:

a) **There are *k* factors:** This assumption is necessary whatever the modelling approach. However we clearly want to avoid omitting a significant return factor. A possible advantage of the statistical approach is that by not making any additional assumptions, it will not omit a significant return factor.

   Another possible problem is including insignificant or unnecessary factors. Obviously including too many factors will make the model unnecessarily complex; more importantly, it also runs the risk of over specifying the systematic nature of the risk – that is misattributing stock specific risk as factor risk. This danger can be ameliorated if we can test for the significance of factors in explaining the asset returns.

b) **The factor betas are constant over time:** This assumption is implicit in both the statistical and time-series modelling approaches. In the time-series approach, we know the factor returns, which are either estimated as the returns to factor mimicking portfolios, or are measured directly, and we estimate the betas assuming they are constant over the sample period. In the statistical modelling, both the factors and betas are estimated from the sample covariance matrix.

   Both approaches assume the betas are constant; is this likely to introduce significant errors? Quantitatively this can be investigated by breaking the sample period up and testing whether the estimated betas are equal in the sub-samples. Andersen, Bollerslev, Diebold, Wu (2006), suggests there is a tendency for market, country and sector betas to mean revert over time. Bloomberg specifically incorporate this tendency in their approach, by shrinking the historical market betas back to 1. Further, qualitatively, one would not expect a stock's beta with respect to style risk factors to remain constant, as membership of these style baskets is very time-varying.

c) **The factor exposures are observable**: It is this assumption that underlies the cross-sectional approach to risk modelling. Since the seminal Fama-French paper (1992) investigating the value and size factor, there has a vast volume of papers using this cross-sectional approach to explain the cross-section of stock returns. However they have almost exclusively focused on style risk factors. Papers using this approach to look at market, country and sector risk have been far fewer with Heston and Rowenhorst (1994) being a notable exception. These use an indicator dummy variable – with 1 denoting inclusion in a market, sector or country, 0 otherwise – as a proxy for market, country and sector betas.

   Is this likely to introduce modelling errors? Connor and Linton (2007) investigate the assumption that the value and size betas are a linear function of the book to price ratio and log of market cap; they find evidence of a very mild non-linearity (that can be ignored without very little impairment of performance).

The assumption that all stocks within a sector (or country) have the same sensitivities with respect to sector (or country) risk is more problematic. Certainly this assumption at the stock level is rejected by the data at all standard confidence levels. The problem can be minimised by increasing the sector or regional resolution, but at the cost of complexity and the risk of over-fitting of the model. However, whether this assumption will introduce significant errors at the portfolio level is harder to ascertain.

A number of papers, Tien and Pfleiderer (2005), Scowcroft and Sefton (2006) and Connor and Briner (2008), have examined the trade-off between structural and estimation error. All found that some structure is absolutely necessary to reduce sampling errors. Hence both the time-series and the cross-sectional models out-performed the statistical model in all tests. However, the tests were less powerful in determining precisely how much structure should be imposed; with preference ordering between time-series and cross-sectional modelling varying across tests.

Miller (2006) investigated the claim that statistical models are less likely to 'miss' any risk factors. His results, though, supported the conclusions of the papers cited above. He found that even though statistical models have the potential to pick up transitory sources of risk, in practise they are unable to do so. Statistical models are only successful in identifying pervasive (affecting many stocks) and enduring sources of risk. These are the very sources that will be included in any well built structural model. Statistical models were far less successful at picking out transitory or narrow (affecting only a few stocks) sources of risk – the sources that could possibly be missed by a structural model. Simply speaking, these sources were not strong enough to be identified statistically.

For similar reasons, it is claimed that statistical models are a better short term models of risk; they have the flexibility to capture temporary changes in market structure. However, over the short horizon the data sample is even smaller, demanding one should impose more structure rather than less!

Though this cover the principal statistical trade-offs, there are other objectives which are more pragmatic in nature and concern the simplicity and ease with which the results can be interpreted.

2) **Ease of Risk Attribution:** Portfolios managers not only want to know the total aggregate risk of their portfolio, but they also want to attribute it to various sources. This decomposition adds more value, if the various sources of risk – the factors – have an intuitive economic interpretation. Thus, for example, if one of the factors is associated with the oil price, then the portfolio exposure to this risk factor can be interpreted as exposure to movements in the oil price. But by constraining the directions of risk to lie in certain 'intuitive' directions, one is possibly limiting the ability of the risk model to pick up transitory or new sources of risk.

3) **Parsimony:** Reducing the complexity of the model amounts to keeping the number of factors to a minimum. It also reduces the risk of over-fitting the returns.

4) **Alignment of source of risk with sources of return.** Modern Portfolio Theory (MPT) advocates that in an optimal portfolio the marginal return from a position should equal the marginal risk from a position. The portfolio construction process can be understood as the search for the portfolio that satisfies this condition. This process is both assisted and will be more robust if the risk factors align with the sources of return.

   If the risk factors only align approximately with the sources of return, then an optimiser will focus on a portfolio of stocks that have a positive return, but carry little risk. The optimiser will tend to gear up on such portfolios; possibly leading to an unbalanced portfolio and certainly excess turnover. As a way of avoiding this problem, the return factors should be included as risk factors..

Why we chose our Hybrid modelling approach

We are now in a position to discuss our hybrid approach to estimating the risk matrix. As we have argued, it is necessary to impose some structure on the risk matrix so as to limit likelihood of significant estimation errors. But what structure? We believe this depends on the source of risk.

A list of possible sources of risk would inevitably include

1) Market Risk – the systematic component of risk.

2) Macroeconomic risk factors e.g. commodity prices.

3) Industry Risk – risks that are systematic to a particular industry.

4) Regional or Country risk – risks that are systematic to a particular geographical area.

5) Style risk – risks that are pervasive to stocks that share a common characteristic; such as high debt, high earnings growth potential, heavily capitalised etc.

It is our view that the first four sources differ crucially from final one. Whether a stock is a member of an industry or geography does not change significantly over time. It is therefore reasonable to assume that exposures to these risk factors remain relatively constant over time. Further the industry or country factor returns can be well proxied by the respective index returns. We therefore argue with these factors it is more natural to estimate them in time-series; i.e. assume that we observe the factor returns and estimate the exposures. Are reason are summarised in the following list:

- Estimating these factors in time-series, does allow for a more parsimonious description of the risk – heterogeneity in sensitivities can accommodated by heterogeneity in betas rather than the introduction of more risk factors. For example – global universal banks have a tendency to more sensitive to financial shocks than local retail banks. This can be modelled either by having two factors, one for each sector; with the factor returns being highly correlated but with one having a higher volatility than the other; or a single factor with the stocks have differing sensitivities to this single factor.

- The time series approach can accommodate cross-sensitivities between sector and countries. Thus export orientated companies can have some sensitivity to foreign markets as well as their own domestic market.

- Macroeconomic factors can only easily be incorporated in a risk model using a time-series approach. Given an economic series – e.g. oil prices – it is straightforward to estimate stock sensitivities within a time series regression. It is not immediately obvious how this could be within a cross-sectional model.

In contrast to the first four, sensitivity to style risk is likely to vary over time. There are many possible reasons for this time variation but these would include capital restructuring, change in multiples etc. If we have a reasonable proxy for these time-varying exposures, then the time-varying nature of these sensitivities can be accommodated relatively easily within a cross-sectional framework. We therefore believe style factors are better estimated in a cross-sectional framework.

We therefore suggest the following hybrid approach to estimating a risk model

1) For market, economic, industry and country factors, use a time-series approach. More precisely, assume the risk factor returns, $\boldsymbol{f}_{t,}$ are well proxied by the returns to the respective market, economic, industry and country indices. Then – under the assumption that the betas are constant over the sample – estimate the parameters using a number of time-series regressions. As this means estimating a relatively large number of parameters, we will estimate them within a Bayesian framework. This effectively shrinks the least-squares estimates back to a set of prior values – either '1' or '0' – *but* in proportion to the degree of uncertainty. Thus if the parameter is poorly estimated (as measured by its *t*-stat or sample variance) then it is shrunk back heavily. Conversely if it is well estimated then far more weight is placed on the estimate.

2) For style factors, use a cross-sectional approach. Assume that we observe the time-varying exposures $\boldsymbol{B}_t$ and estimate the factor returns, $\boldsymbol{f}_{t,}$ using a series of cross-sectional regressions.

We refer to this model as UBS hybrid risk model. In the next section we describe briefly how we estimate it. We relegate all technical details to the Appendix.

Later, we shall compare our approach to an alternative, where all but the market factor is estimated in cross-section. We refer to this alternative, rather unimaginatively, as a Cross-Sectional model.

Estimation of the UBS Hybrid Model

The estimation procedure builds on the work of Stroyny (2005). However, we develop his approach in a number of ways. Firstly, and most crucially, we incorporate priors on some of the parameters, in particular the time-series beta exposures. This generalisation is not only likely to reduce sampling error, by shrinking back the parameters that are poorly estimated, it also speeds up significantly the estimation procedure[2].

Our second generalisation is to allow for stochastic nature of the factor returns in the estimation procedure. Some of the early work on factor analysis, Young (1942) and Lawley (1943), treated the factor returns as additional model parameters. The resultant algorithm – often referred to as the Least Squares Method of Factor Analysis (LSMFA) – was straightforward in that it amounted to iterating over a standard set of least squares regressions. However, Whittle (1952) described the approach as “too unstable to be useful”. Rubin and Thayer (1982) suggested using a prior that the factors returns were drawn from a normal distribution, and showed that the resultant model could be estimated using the Expectation Maximisation Algorithm, see McLachlan, Krishnan (1997).

Adapting the Rubin and Thayer (1982) to our problem has three direct payoffs:

1) It stabilises the estimation algorithm.

2) It ensures that our estimated style factors are both industry and regionally neutral.

3) It ensures that our estimate of the factor covariance matrix takes due account of errors in the estimation of the style factor returns.

4) It enables us to test in turn whether each factor significantly adds to the explanatory power of the risk model.

We shall discuss each of these advantages in more detail below.

First though, we describe the estimation procedure. This discussion is relatively involved, but technical details have been deferred to the Appendix. Rewrite equation (1) to separate out the time-series factors form the cross-sectional factors

$$\boldsymbol{r}_t = \begin{bmatrix} \boldsymbol{B}_{TS} & \boldsymbol{B}_{XS,t} \end{bmatrix} \begin{bmatrix} \boldsymbol{f}_{TS,t} \\ \boldsymbol{f}_{XS,t} \end{bmatrix} + \boldsymbol{\varepsilon}_t \tag{5}$$

where the subscript *TS* or XS denote the factors estimated in time-series and cross-section respectively. We assume that we observe the time-series factor returns, $\boldsymbol{f}_{TS,t}$, and the cross-sectional betas, $\boldsymbol{B}_{XS,t}$ and must estimate the time-series betas, $\boldsymbol{B}_{TS}$, and the covariance matrices, $\boldsymbol{D}$ and $\boldsymbol{F}$.

We estimate the risk model in a Bayesian framework in an effort to reduce sampling error. We assume a standard conjugate on the time-series betas,

$$Vec(\boldsymbol{B}_{TS}) \sim N\left(Vec(\boldsymbol{B}_{TS.0}), \tau\boldsymbol{\Omega}\right) \quad \text{where} \quad \boldsymbol{\Omega} = diag(vec(\boldsymbol{\Lambda})) \tag{6}$$

[2] These two advantages are closely related. If a parameter is poorly estimated, its value is likely to ‘jump’ between iterations. By shrinking these parameters back to their priors, one not only reduces estimation error, but stabilise the iterative estimation procedure.

where the $\boldsymbol{B}_{TS,0}$ is the dummy indicator matrix described earlier (an element is '1' if the stock lies in the respective industry or country, and is '0' otherwise), Ω is a diagonal matrix of variances and τ is a scalar. In our default model, we set τ =1. Later we experiment with different values for τ, so as to investigate the sensitivity of the parameter estimates to our prior. The matrix $\boldsymbol{\Lambda}$ has the same dimensions as $\boldsymbol{B}_{TS}$ with elements equal to the variance of the prior estimate. We calibrate these priors using a simple rule of thumb; imagine estimate a single beta by regressing a vector of factor returns on a vector of a stock's returns. In this case, the standard error of the beta estimate will be

$$Var\left(\hat{b}-b\right)=\frac{Var\left(\varepsilon\right)}{TVar\left(f\right)} \tag{7}$$

where the notation should be clear from above. Using typical values of $Var(\varepsilon)/Var(f) = 4$, then with 60 observations the variance of the beta estimate should be 0.25. based on this, in the default model we let $\boldsymbol{\Lambda} =(0.12+0.13\ \boldsymbol{B}_{TS,0i})$. Hence if the mean of prior is a '1', its standard deviation (std.) is 0.5 (or variance of 0.25) and if the mean of the prior is '0' its std. is 0.35. Thus our prior belief is that 95% of stocks have betas with respect to their industry or geography in the range 0 to 2, and betas in the range -0.7 to 0.7 with respect to the other factors.

The estimation procedure is based on an Expectation Maximisation (EM) algorithm. To describe the approach, it will prove useful to write the covariance matrix $\boldsymbol{F}$ in diagonal form where

$$\boldsymbol{F}=\begin{bmatrix}\boldsymbol{F}_{TS} & 0\\ 0 & \boldsymbol{F}_{XS}\end{bmatrix} \tag{8}$$

so that $\boldsymbol{F}_{TS}$ is the covariance matrix of the times-series factor returns

$$\boldsymbol{F}_{TS}=\frac{1}{T}\sum_{t}\boldsymbol{f}_{TS,t}\boldsymbol{f}_{TS,t}^{T} \tag{9}$$

and $\boldsymbol{F}_{XS}$ is the covariance matrix of the cross-sectional factor returns. The EM algorithm is an iterative procedure can now broken down into the following steps:

1) Scale all returns by an estimate of market volatility for that period. This removes a considerable proportion of the observed hetereoskedascity in returns. Initialise the estimation procedure with $\boldsymbol{B}_{TS}$ = $\boldsymbol{B}_{TS,0}$, $\boldsymbol{D}=diag(Cov(\boldsymbol{r}))$ and $\boldsymbol{F}_{XS}^{-1}=\boldsymbol{0}$.

2) Given an estimate for $\boldsymbol{B}_{TS}$, $\boldsymbol{D}$ and $\boldsymbol{F}_{XS}^{-1}$ calculate an estimate for the unobserved cross-sectional factor returns, $\boldsymbol{f}_{XS,t}$ as

$$\hat{\boldsymbol{f}}_{XS,t}=\left(\boldsymbol{B}_{XS,t}^{T}\boldsymbol{D}^{-1}\boldsymbol{B}_{XS,t}+\boldsymbol{F}_{XS}^{-1}\right)^{-1}\boldsymbol{B}_{XS,t}^{T}\boldsymbol{D}^{-1}\left(\boldsymbol{r}_t-\boldsymbol{B}_{TS}\boldsymbol{f}_{TS,t}\right) \tag{10}$$

with variance

$$Var\left(\hat{\boldsymbol{f}}_{XS,t}-\boldsymbol{f}_{XS,t}\right)=\left(\boldsymbol{B}_{XS,t}^{T}\boldsymbol{D}^{-1}\boldsymbol{B}_{XS,t}+\boldsymbol{F}_{XS}^{-1}\right)^{-1} \tag{11}$$

for every period t. Given these, a new estimate of the covariance matrix $\boldsymbol{F}_{XS}$ is

$$\boldsymbol{F}_{XS} = \frac{1}{T}\sum_t \left(\hat{\boldsymbol{f}}_{XS,t} \hat{\boldsymbol{f}}_{XS,t}^T + Var\left(\hat{\boldsymbol{f}}_{XS,t} - \boldsymbol{f}_{XS,t} \right) \right) \tag{12}$$

3) Given the estimate for the cross-sectional factor returns for $\hat{\boldsymbol{f}}_{XS,t}$ calculate a new estimate for the time-series exposures, $\boldsymbol{B}_{TS}$,. This is done asset by asset, effectively performing n time-series regressions. If we use the notation $[X]_{i\bullet}$ to refer to the ith row of the matrix X.

$$\left[\boldsymbol{B}_{TS}\right]_{i\bullet} = \left(\sum_t s_t^{-2} \left(\boldsymbol{r}_{t,i} - \left[\boldsymbol{B}_{XS,t}\right]_{i\bullet} \hat{\boldsymbol{f}}_{XS,t} \right) \boldsymbol{f}_{TS,t}^T + \left[\boldsymbol{B}_{TS,0}\right]_{i\bullet} \boldsymbol{Z}_i \right) \left(\sum_t s_t^{-2} \boldsymbol{f}_{TS,t} \boldsymbol{f}_{TS,t}^T + \boldsymbol{Z}_i \right)^{-1} \tag{13}$$

where $\boldsymbol{Z}_i$ is a diagonal matrix with the jth element on the diagonal equal to $\sigma_i^2 / \Lambda_{ij}$. Finally we update the estimate for the variance of the residual returns as

$$\boldsymbol{D} = \left(\frac{1}{T}\sum_t \left(\boldsymbol{r}_t - \boldsymbol{B}_{TS} \boldsymbol{f}_{TS,t} - \boldsymbol{B}_{XS,t} \boldsymbol{f}_{XS,t} \right) \left(\boldsymbol{r}_t - \boldsymbol{B}_{TS} \boldsymbol{f}_{TS,t} - \boldsymbol{B}_{XS,t} \boldsymbol{f}_{XS,t} \right)^T + \boldsymbol{B}_{XS,t} Var\left(\hat{\boldsymbol{f}}_{XS,t} - \boldsymbol{f}_{XS,t} \right) \boldsymbol{B}_{XS,t}^T \right) \tag{14}$$

4) Given these updated estimates of the parameters go to step (2). Repeat the loop until the parameters converge.

The EM algorithm has a number of very desirable properties. The principal being that it is guaranteed to converge monotonically to a local maximum. The convergence can be slow but is stable. In practise, we found that with relatively confident priors, $\tau < 100$, convergence was surprisingly fast. The priors shrunk back the poorly estimated parameters – the very parameters that would converge slowly if the priors were diffuse – which dramatically sped up convergence.

A Comparison of Modelling approaches

Ideally, we would try to establish if this modelling approach is better previous standard approaches. This, unfortunately, is not possible as there is no single discriminating test. Instead we must be content with a comparing our Hybrid model to a ‘close’ cross-sectional relative. Our comparison is designed to highlight the principal differences between the two approaches. We take the various components of the model in turn, examining industry and regional exposures, industry and regional factor returns and finally the style returns.

We start off by describing the comparison model. We shall refer to this model as the cross-sectional model though this is not an entirely accurate description for reasons we shall now explain. This cross-sectional model has the same set of factors as our Hybrid model; a market factor, 7 regional factors, 10 industry factors and 7 style factors. These are estimated as follows:

1) **Market Factor**: Firstly the market beta or exposures are estimated. This is done in time-series by regressing each stock return series on the market returns and a constant. The coefficient on the market is then the market beta. The rest of the model is estimated on the residuals of the regression rather than the raw returns. It is for this reason that even our cross-sectional model can be regarded as a sort of Hybrid model.

   An alternative approach adopted by some well-known model providers is to assume that all market exposures are '1' and to estimate the entire model in cross-section. Though we could have adopted this approach, the results are so coloured by the treatment of the market as to reduce the value of the comparison exercise. The market factor is so pervasive that any residual market sensitivity in the stock returns will be captured by other factors; and it is this residual element that dominates the comparison results. Thus when we compare the estimated factor returns of the two models, they can appear to be very different (low correlation). Closer inspection suggests that most of the difference can attributed to the market residual[3].

2) **Regional, Industry and Style Factors**: In a cross-sectional model, all exposures are assumed to be known a priori. The country and industry exposures are the indicator vectors – an element is 1' if the stock lies in the respective industry or country and '0' otherwise – and the style exposures are proxied by a fundamental ratio. Thus in this cross-sectional model the country and industry exposures are the Bayesian priors used in the Hybrid model, and the style exposures are identical to those in the Hybrid model. Given these exposures, the factor returns are estimated by regressing the stock returns net of the market on these exposures in each period.

Both our Hybrid and the Cross-sectional Model are estimated on the MSCI Investible Market Index (IMI) for developed markets of large and medium sized companies. The universe is just over 1600 stocks in each period. The models are estimated on 5 years worth of monthly data finishing in May 2010 – i.e. 60 monthly observations. We also use a larger universe of the FTSE all-world developed market, which is roughly the largest 6000 stocks in developed markets, to see test for any size biases.

Regional and Industry Exposures

In the risk model, there are 7 regional factors – US, UK, Canada, Japan, EMU, Europe ex UK ex EMU, Asia ex Japan – and 10 industry factors – the ICB industry groups.

[3] For example, see the discussion in The BARRA Global Equity Model (GEM2), Research Notes, September 2008, page 13. . In this note, they suggest that sensitivity of high beta stocks to the market is accounted for by high sensitivity to the volatility style factor. However, we stress that our cross-sectional model should *not* be seen as a replicated BARRA model. At a very basic level, the BARRA model has a very different set of factors to the ones we use here. The only similarity is that both are estimated using a cross-sectional approach.

The Hybrid model uses as its prior for the regional and industry exposures the indicator vectors; an element is 1' if the stock lies in the respective region or industry and '0' otherwise. In the following table we assess the impact of these priors. The table is split into two. The left hand side gives results for the case $\tau = \infty$ and the right hand side when $\tau = 1$. As τ tends to infinity, less and less weight is given to the priors, and in the limit the posterior estimates tends towards the same values that would be derived from a least square regression. The 1st column gives the mean of beta for all stocks belonging to the industry or region of that factor. The 3rd column gives the mean of the other beta estimates. The 2nd and 4th columns give the average standard deviation of the corresponding estimates around the means. Columns 5-8 give the same information for the posterior estimates for when $\tau = 1$.

Table 1: A summary of the exposure estimates for each of the industry and region. The first four columns summarise for case $\tau = \infty$, the last four for the case $\tau = 1$. Each line refers to the beta estimates with respect to that factor. We record the mean and standard deviation of the betas for two groups of stocks; a stock is in the first group if it belongs to the respective industry or region and in the 2nd group if it does not.

	Diffuse Prior ($\tau = \infty$)				Conjugate Prior ($\tau = 1$)			
	Stocks within Industry or Region		Stocks not in Industry or Region		Stocks within Industry or Region		Stocks not in Industry or Region	
	Mean Beta	Std. Dev Beta	Mean Beta	Std. Dev Beta	Mean Beta	Std. Dev Beta	Mean Beta	Std. Dev Beta
Oil & Gas	0.83	0.55	0.10	0.49	1.01	0.31	0.04	0.18
Basic Materials	0.77	0.68	0.05	0.45	0.82	0.46	0.03	0.23
Industrials	0.92	0.94	0.20	0.78	0.87	0.35	0.03	0.18
Consumer Goods	0.55	0.90	0.12	0.86	0.84	0.31	0.03	0.18
Health Care	0.67	0.80	0.09	0.65	0.81	0.31	0.02	0.18
Consumer Services	0.85	0.78	0.21	0.76	0.90	0.32	0.04	0.18
Telecommunications	0.65	0.54	0.03	0.63	0.87	0.29	0.01	0.18
Utilities	0.92	0.63	0.14	0.60	0.93	0.32	0.05	0.20
Financials	0.74	1.28	-0.02	1.00	0.74	0.35	-0.03	0.16
Technology	0.76	0.89	-0.03	0.57	0.81	0.44	-0.01	0.19
Canada	0.59	0.67	-0.09	0.63	0.88	0.25	0.01	0.18
UK	0.69	0.85	-0.15	0.82	0.80	0.34	-0.02	0.17
Japan	0.90	0.66	-0.13	0.50	0.94	0.38	-0.02	0.17
United States	1.14	0.94	0.14	0.40	1.02	0.11	0.01	0.04
EMU	0.70	0.93	0.02	0.77	0.94	0.28	0.01	0.14
Europe Ex EMU & UK	-0.05	0.76	-0.08	0.68	0.73	0.27	0.01	0.16
Asia Pacific Ex Japan	0.59	0.70	-0.17	0.63	0.85	0.27	-0.02	0.18

Source: UBS

There are number of clear conclusions from these results:

1) The mean beta of the two groups, those within the respective industry or region (the members) and those that are not (the non-members), are significantly different. The former group's beta is close to but slightly less than 1, with the non-members clearly centred on 0. If we did a market cap

weighted mean of the betas then the mean of the members beta is closer to 1 (0.93) .

2) When the prior is diffuse, the standard deviation within each group of the beta estimates is high. For the first group, it is roughly 0.75. Comparing columns 2 and 5, the impact of strengthening the prior is clear. The prior shrinks the estimates back, reducing the standard deviation of the posterior betas by over 50%.

3) There is some evidence that the industry exposures are more significant than the regional exposures. This comes from comparing the ratio of the standard deviation of the members' beta for the two different models.

Regional and Industry Factor Returns

We shall now compare the factor returns estimated in the cross-sectional model with those used in the Hybrid model (recall that the factor returns used in the Hybrid model are the returns to the respective indices).

The results are given in Table 2. The first column records the correlation coefficient between the respective time series; the second and third columns the annualised volatility of the two series. The 4th and 5th columns record the same number but where the factor returns have been estimated on the larger FTSE All-world index (we omit the volatility numbers of the factors returns of the Hybrid model in this case as they barely differ from the first case).

Again we summarise the main conclusions:

1) Correlations between the industry factor returns are generally much higher than correlations between the estimated regional factors. This is consistent with the notion that the regional factors are less significant.

2) The industry volatilities of the two sets of factors returns are broadly similar, whereas the estimated regional cross-sectional factor returns are far more volatile than the respective index returns. This is particular pronounced in the case of the US returns.

3) For the industry factors the results for the broader universe are very similar. For the regional factors, the correlations are slightly higher on average and the volatilities significantly lower. At first this may appear surprising as broadening the universe means including smaller cap stocks which tend to be more volatile. The results could however be explained if broadening the universe reduces model estimation error.

We also note that these results are all consistent with the earlier observation that the industry factors are more significant in explaining stock returns than the regional factors in this sample.

Table 2: Comparison of the estimated factor returns from the cross-sectional model with the respective index returns used within the Hybrid model.

	MSCI IMI Universe			FTSE All-Word Universe	
	Correlation of factors returns - Hybrid vs. Cross-sectional models	Annualised Volatility of Factors Returns for the Hybrid Model	Annualised Volatility of Factors Returns for the Cross-sectional Model	Correlation of factors returns - Hybrid vs. Cross-sectional models	Annualised Volatility of Factors Returns for the Cross-sectional Model
Oil Gas	0.94	16.10	12.83	0.94	12.41
Basic Materials	0.68	12.85	6.63	0.70	6.47
Industrials	0.60	5.23	5.34	0.45	4.93
Consumer Goods	0.63	4.25	4.29	0.51	3.62
Health Care	0.81	7.85	5.66	0.79	5.06
Consumer Services	0.78	5.68	5.53	0.71	4.50
Telecommunications	0.81	7.44	6.75	0.75	4.89
Utilities	0.78	7.56	6.30	0.82	5.49
Financials	0.79	7.74	6.72	0.77	5.19
Technology	0.64	9.49	7.63	0.64	7.62
Canada	0.36	7.81	15.03	0.51	10.47
UK	0.56	5.02	12.29	0.41	9.70
Japan	0.48	8.25	13.93	0.67	15.01
United States	0.13	1.87	9.63	0.16	9.92
EMU	0.20	4.33	13.16	0.25	8.69
Europe Ex EMU UK	0.24	6.06	13.35	0.24	9.64
Asia Pacific Ex Japan	0.34	5.64	14.25	0.50	9.37

Source: BS

Style Factor Returns

The Hybrid model is built with 7 style risk factors; value, earnings and price momentum, size, volatility, quality and a capital gearing factor. In this section we investigate the explanatory power or significance of each style. We then compare the estimated style returns from both the Hybrid and Cross-sectional models with a set of calculated style index returns.

We wish to test whether a factor adds to the explanatory power of the risk model. If it does not, we could drop the factor and achieve a more parsimonious description of the data. However, a little care is required. As we have estimated the model in a Bayesian framework, we can not use classical hypothesis tests. Instead we must approach the question, by asking whether one model descriptions is more likely than another. Both the Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC) offer an approach based on these lines. Similar to the classical likelihood ratio test, the criteria are based on twice the likelihood. However the criteria make an adjustment for the number of estimated parameters. These criteria therefore quantify the trade-off between the model fit and parsimony. The preferred model is the one with lowest information criterion. The BIC tends to favour less complex models than the AIC. These statistics are discussed in Lopes and West (2004). They argue that BIC is to be preferred.

Table 3: Testing the parsimony of the Cross-Sectional Factors. Each column refers to a different model. The first column is our default model with all the 7 factors included. The remaining 7 columns drop each of these factors in turn. The first row records the value of likelihood function at the optimum, the second row the BIC, and the 3rd row the difference in the BIC statistic from that of default model (all factors included). The 4th and 5th row record similar statistics for the AIC and the final row gives the estimated volatility in the default model of the dropped factor returns.

	Base Model	Drop Value	Drop Earnings Momentum	Drop Price Momentum	Drop Market Cap	Drop 12 mth Volatility	Drop Quality	Drop Leverage
Log Likelihood	4712.2	4686.7	4687.4	4691.0	4699.3	4702.2	4703.2	4704.6
BIC	50229.1	50267.6	50266.1	50259.0	50242.4	50236.5	50234.5	50231.7
Difference in BIC	0.0	38.5	37.1	29.9	13.3	7.4	5.4	2.7
AIC	57671.7	57708.6	57707.2	57700.0	57683.4	57677.5	57675.5	57672.8
Difference in AIC	0.0	37.0	35.5	28.3	11.8	5.9	3.9	1.1
Volatility		7.21	3.97	5.49	2.45	3.55	1.84	1.85

Source: UBS

Table 3 records the results from the first set of experiments. In the default model there are 7 cross-sectional style exposures. We drop each of these factors in turn, and test whether this reduced factor model is more informative. We shall focus our discussion on the BIC statistics. The number of estimated parameters, p, in each model is nk_{TS} time-series exposures, n idiosyncratic return variances and $k_{XS}(k_{XS}+1)/2$ independent parameters in the cross-sectional covariance matrix; the BIC is equal $-2l+p.log(T)$ where l is the log likelihood and p is the number of parameters. A model with a lower BIC is more probable in a Bayesian sense. Of all the models, the default model has the lowest BIC and so is suggestive that all the factors should be included. However the criterion does rank the factors in some order of explanatory power. The value, momentum and size factors contain the most information, with volatility, quality and leverage the least. The results from the AIC are consistent with these findings.

We have also recorded in the table the volatility of the estimated factor returns in the default model. Perhaps unsurprisingly, the volatility of the factors ranks the factors in almost the same order as the information criteria.

We now compare the estimated cross-sectional style returns from the Hybrid model (HFR) with those from the Cross-Sectional model (XSFR), with a set of calculated style returns (EFR). These calculated style returns are returns to a set of long short factor mimicking portfolios. They are constructed so as to be both sector and regionally neutral. In every period, the stocks are stratified by both industry and macro region. (To ensure that there is a reasonable coverage in each cell, we aggregated our 7 regions into four macro regions; North America, UK, Europe ex UK, and Asia.) Within each cell, the stocks are ranked according the relevant style betas. The equally-weighted industry and regionally neutral style portfolio is defined as long the top third and short the bottom third of the stocks in each cell. The style return is the return to this portfolio.

Table 4 records the correlation coefficients between the various style return series. For the Hybrid model, we have estimated the style returns for four different values of τ (the confidence placed in the prior).

Table 4: Correlation coefficients between respective style return series. The acronyms used are ESR for estimated style returns, HFR for the Hybrid factor returns and XSFR for the Cross-Sectional factor returns. The Hybrid factor returns are estimated for different values of τ, the weight placed on the Bayesian priors; $\tau=\infty$ corresponds to the diffuse prior, and $\tau=0$ corresponds to complete confidence in the priors.

τ		Value	Earnings Momentum	Price Momentum	Market Cap	12 mth Volatility	Quality	Leverage
∞	ESR vs HFR	0.67	0.87	0.02	0.54	0.72	-0.30	0.50
10	ESR vs HFR	0.76	0.86	0.32	0.69	0.80	-0.37	0.60
1	ESR vs HFR	0.83	0.84	0.49	0.72	0.88	-0.29	0.68
0	ESR vs HFR	0.77	0.75	0.85	0.66	0.94	-0.17	0.73
∞	HSF vs XSFR	0.62	0.83	0.47	0.39	0.66	0.48	0.52
10	HSF vs XSFR	0.62	0.83	0.47	0.39	0.66	0.48	0.52
1	HSF vs XSFR	0.87	0.88	0.84	0.71	0.88	0.55	0.72
0	HSF vs XSFR	0.93	0.91	0.94	0.83	0.97	0.64	0.72
	ESR vs XSFR	0.82	0.66	0.77	0.76	0.91	0.34	0.58

Source: UBS

We make a number of observations based on the results in this table:

1) Concerning the results when $\tau=0$: the factor returns estimated in the Hybrid and cross-sectional models are highly correlated, possibly with the sole exception of quality. When $\tau=0$ the factor exposures of the Cross-sectional and Hybrid models are the same. However the two approaches still differ; firstly the cross-sectional model estimates the industry and regional factor returns whereas the hybrid does not, and secondly the Hybrid model is using a maximum likelihood estimator whereas the cross-sectional is a least squares estimator. This result suggests that using industry and regional index returns in the Hybrid model (rather than estimating these returns) does not materially effect the estimates of the style factor returns.

2) Concerning the results when $\tau=1$: the estimated factor returns of the Cross-sectional and Hybrid models are still highly correlated.

3) Concerning the correlation of the estimated price momentum factor returns with the calculated price momentum style returns: As τ increases, the correlation of the hybrid factors returns drops off significantly. To explain this, we note that price momentum is pervasive at the industry and regional level. Relaxing the priors on the stock exposures to these industry and regional factors, allows the Hybrid model to pick up these momentum effects within these factors.

4) Concerning the estimation of the quality style returns: correlation of the estimated factor returns with the style returns is low. Two reasons can explain this observation; firstly this style is relatively uninformative and therefore poorly estimated (see the previous table) and secondly it is negatively correlated with the value style.

Conclusion

In this paper we have argued for a Hybrid approach to building a risk model. We maintain that this imposes sufficient structure so as to reduce sampling error, but retains the flexibility to incorporate all major sources of risk in a parsimonious fashion.

Our Hybrid model assumes the respective index returns are a good proxy for the industry and country risk factors. We prefer to estimate the betas, in the belief that there is substantial heterogeneity in these exposures both within and across sectors and regions. Further this framework can incorporate macroeconomic risk factors naturally.

In contrast, we prefer to estimate style factors in cross-section. Here the major attraction of a cross-sectional approach is that betas can vary over time. Thus we view as critical in that style membership varies from period to period.

Technical Appendix

This appendix contains all technical details on the model estimation procedure and the inference tests used in this research note. The appendix is designed to be self-contained but does focus exclusively on the estimation of the UBS hybrid model. For background on estimation of risk models, the interested reader is referred to the forthcoming book Connor and Korajczyk (2009) – an excellent and detailed exposition.

We start by detailing the model assumptions. Denote the stock returns to the n assets at time t by the n-vector $\boldsymbol{r}_t$. The returns are generated by the linear factor model

$$\begin{aligned} \boldsymbol{r}_t &= \begin{bmatrix} \boldsymbol{B}_{TS} & \boldsymbol{B}_{XS,t} \end{bmatrix} \begin{bmatrix} \boldsymbol{f}_{TS,t} \\ \boldsymbol{f}_{XS,t} \end{bmatrix} + \boldsymbol{\varepsilon}_t \\ &= \boldsymbol{B}_t \boldsymbol{f}_t + \boldsymbol{\varepsilon}_t \end{aligned} \qquad (0.15)$$

where $\boldsymbol{f}_{TS,t}$ denotes the k_{TS} time–series actor returns, $\boldsymbol{B}_{TS}$ refers to the time-series betas, $\boldsymbol{f}_{XS,t}$ to the k_{XS} cross-sectional factor returns, $\boldsymbol{B}_{XS,t,}$to the cross-sectional betas and $\boldsymbol{\varepsilon}_t$ to the idiosyncratic returns. The cross-sectional betas vary over time (and so are indexed by t), whereas the time-series betas are constant. We shall make the standard assumptions concerning the distribution of the returns; factor and idiosyncratic returns are distributed independently and normally as

$$\begin{bmatrix} \boldsymbol{f}_{TS,t} \\ \boldsymbol{f}_{XS,t} \\ \boldsymbol{\varepsilon}_t \end{bmatrix} = N\left(\begin{bmatrix} \boldsymbol{\mu}_{TS} \\ \boldsymbol{\mu}_{XS} \\ 0 \end{bmatrix}, s_t^2 \begin{bmatrix} \boldsymbol{F}_{TS} & \boldsymbol{F}_{TS,XS} & 0 \\ \boldsymbol{F}_{TS,XS}^T & \boldsymbol{F}_{XS} & 0 \\ 0 & 0 & \boldsymbol{D} \end{bmatrix} \right) = N\left(\begin{bmatrix} \boldsymbol{\mu} \\ 0 \end{bmatrix}, s_t^2 \begin{bmatrix} \boldsymbol{F} & 0 \\ 0 & \boldsymbol{D} \end{bmatrix} \right)$$

(0.16)

where $\boldsymbol{D}$ is a diagonal matrix of variances, $\boldsymbol{D}=diag(\sigma_1^2, \sigma_2^2 \ldots\sigma_n^2)=diag(\boldsymbol{\sigma}^2)$ and s_t^2 adjusts for time-varying market volatility. Thus we assume that in periods of high volatility, all variance and covariance increase by the same proportion. We assume that s_t^2 is observed; for example the VIX index would be a reasonable proxy; or backed out from market returns using a GARCH model.

The estimation of this hybrid model is complicated by our assumptions of what is observed and what is to be estimated. In a cross-sectional risk model, it is assumed that the exposures $\boldsymbol{B}$ are observed and the factor returns are estimated. Conversely, in a time-series risk model, it is assumed that the factor returns are observed and the exposures are estimated. Our model is a hybrid of these two polar cases, we assume that some of the exposures and some of factor returns are observed. To be precise, we assume

1) That we observe that economic factor returns, $\boldsymbol{f}_{TS,t}$. These are to be understood as the returns to the market, industry and country indices.

2) The parameters $\boldsymbol{B}_{TS}$ are to be estimated. However, we allow for an informative prior on these parameters; let the mean of this prior be the

indicator matrix $\boldsymbol{B}_{TS,0}$ which has an element '1' if the stock lies in the industry or country, and is '0' otherwise. Our prior on $\boldsymbol{B}_{TS}$ is that

$$vec\left(\boldsymbol{B}_{TS}\right) \sim N\left(vec\left(\boldsymbol{B}_{TS,0}\right), \tau\boldsymbol{\Omega}\right) \tag{0.17}$$

It is straightforward to impose constraints on these exposures by using a degenerate prior in the direction of the constraint. In this way we shall constrain the beta of the market portfolio with respect to the market factor to be equal to 1. Denote the market portfolio weights as $\boldsymbol{w_m}$. Assume the market factor is the first of the k_{TS} time-series factors, and select this factor with the k_{TS} -vector $\boldsymbol{e}$ which has 1 as the first element and zeros elsewhere, then let

$$\boldsymbol{\Omega} = \boldsymbol{\Omega}_0 - \tfrac{1}{(1+\delta)} \frac{\boldsymbol{\Omega}_0 ww^T \boldsymbol{\Omega}_0}{w^T \boldsymbol{\Omega}_0 w} \tag{0.18}$$

where

$$\boldsymbol{\Omega}_0 = diag(vec(\boldsymbol{\Lambda})) \;\text{ and }\; \boldsymbol{w} = \boldsymbol{e} \otimes \boldsymbol{w}_m \tag{0.19}$$

and $\boldsymbol{\Lambda}$ is n by k_{TS} matrix of variances and δ is scalar. Thus the prior is degenerate as δ tends to zero in the direction $\boldsymbol{w}$ (the market portfolio beta). We write it this way, as then the matrix Ω is invertible;

$$\boldsymbol{\Omega}^{-1} = \boldsymbol{\Omega}_0^{-1} + \tfrac{1}{\delta} \frac{ww^T}{w^T \boldsymbol{\Omega}_0 w} \tag{0.20}$$

and look at the solution as δ tends to zero . The matrix $\boldsymbol{\Lambda}$ can be thought as a matrix of the variances of each element of the priors, $\boldsymbol{B}_{TS}$. For the results in the text we let $\boldsymbol{\Lambda}$ = (0.12 +0.13*$\boldsymbol{B}_{TS,0}$). Later we shall also need the reciprocal of each of these variances. We shall denote this element-by-element inversion as (1/$\boldsymbol{\Lambda}$), *i.e*, (1/$\boldsymbol{\Lambda}$) = (25/3 – 13/3*$\boldsymbol{B}_{TS,0}$).

3) We observe the cross-sectional $\boldsymbol{B}_{XS,t,}$. These are understood to be the fundamental exposures to style factors such value, growth, size etc.

4) The stochastic factors returns $\boldsymbol{f}_{XS,t}$ are latent and are estimated as an intermediary step in the algorithm.

5) The parameter to be estimated will be denoted $\boldsymbol{\theta}$ where $\boldsymbol{\theta}=\{ \boldsymbol{B}_{TS}\ \boldsymbol{D}, \boldsymbol{F}\}$. On all parameters except $\boldsymbol{B}_{TS}$ we shall hold the standard diffuse prior.

To estimate the risk model, we use the EM algorithm of Dempster, Laird and Rubin (1977). In this presentation, we follow the treatment in McLachlan and Krishnan (2008). The approach is to estimate the mode of the posterior distribution.

$$\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}} \log p\left(\boldsymbol{\theta} \middle| \boldsymbol{r}_t, \boldsymbol{f}_{TS,t}\right) \tag{0.21}$$

where p denotes the posterior distribution function. We have written, this distribution conditional on $\boldsymbol{f}_{TS,}$ but not $\boldsymbol{B}_{XS,t}$. This is because $\boldsymbol{f}_{TS\ ,}$is a stochastic variate (hence we must consider its distribution), whereas $\boldsymbol{B}_{XS,t}$ are observed parameters and direct dependence is assumed to simplify the notation.

Rather than estimate the maximum of this conditional distribution, we follow standard procedures and maximise the joint distribution[4],

$$\begin{aligned}\boldsymbol{\theta}^* &= \arg\max_{\theta} \log p\left(\boldsymbol{\theta}, \boldsymbol{r}_t, \boldsymbol{f}_{TS,t}\right) \\ &=: \arg\max_{\theta} L\left(\boldsymbol{\theta}\right)\end{aligned} \tag{0.22}$$

and define the Likelihood function, *L(θ)*, as a equal to this joint density distribution. The first 'trick' of the EM algorithm is to rewrite the joint distribution as a function of the unobserved or latent variates, $f_{XS,t}$.

$$L\left(\boldsymbol{\theta}\right) = \log \frac{p\left(\boldsymbol{r}_t, \boldsymbol{f}_{TS,t}, \boldsymbol{f}_{XS,t}, \boldsymbol{\theta}\right)}{p\left(\boldsymbol{f}_{XS,t} \middle| \boldsymbol{r}_t, \boldsymbol{f}_{TS,t}, \boldsymbol{\theta}\right)} = \log p\left(\boldsymbol{r}_t, \boldsymbol{f}_{TS,t}, \boldsymbol{f}_{XS,t}, \boldsymbol{\theta}\right) - \log p\left(\boldsymbol{f}_{XS,t} \middle| \boldsymbol{r}_t, \boldsymbol{f}_{TS,t}, \boldsymbol{\theta}\right) \tag{0.23}$$

Note that even though the variables $\boldsymbol{f}_{XS,t}$ cancel out across the two terms on the right hand side, each individual term is a function of $\boldsymbol{f}_{XS,t}$. We now integrate both sides of this equation with respect to a probability distribution $p(\boldsymbol{f}_{XS,t})$ to derive the following identity

$$L\left(\boldsymbol{\theta}\right) = \int p\left(\boldsymbol{f}_{XS,t}\right) \log p\left(\boldsymbol{r}_t, \boldsymbol{f}_{TS,t}, \boldsymbol{f}_{XS,t}, \boldsymbol{\theta}\right) d\boldsymbol{f}_{XS,t} - \int p\left(\boldsymbol{f}_{XS,t}\right) \log p\left(\boldsymbol{f}_{XS,t} \middle| \boldsymbol{r}_t, \boldsymbol{f}_{TS,t}, \boldsymbol{\theta}\right) df_{XS,t} \tag{0.24}$$

This identity is satisfied for any probability distribution $p(\boldsymbol{f}_{XS,t})$. However, we shall pick a particular distribution shortly. The EM algorithm approximates the likelihood function by evaluating the first term only. Clearly, this approximation will be better the smaller the second term. The second 'trick' or observation of the EM algorithm is to note that the second term is similar to the entropy of a distribution and will be minimised when

$$p\left(\boldsymbol{f}_{XS,t}\right) = p\left(\boldsymbol{f}_{XS,t} \middle| \boldsymbol{r}_t, \boldsymbol{f}_{TS,t}, \boldsymbol{\theta}\right) \tag{0.25}$$

Unfortunately the parameters $\boldsymbol{\theta}$, are unknown. However if we have an estimate for the parameters, say $\boldsymbol{\theta}^{(j)}$, we can use these as a proxy for $\boldsymbol{\theta}$ and set $p\left(\boldsymbol{f}_{XS,t}\right) = p\left(\boldsymbol{f}_{XS,t} \middle| \boldsymbol{r}_t, \boldsymbol{f}_{TS,t}, \boldsymbol{\theta} = \boldsymbol{\theta}^{(j)}\right)$. To this end, define the two functions,

$$\begin{aligned}Q\left(\boldsymbol{\theta}; \boldsymbol{\theta}^{(j)}\right) &= \int p\left(\boldsymbol{f}_{XS,t} \middle| \boldsymbol{r}_t, \boldsymbol{f}_{TS,t}, \boldsymbol{\theta} = \boldsymbol{\theta}^{(j)}\right) \log p\left(\boldsymbol{r}_t, \boldsymbol{f}_{TS,t}, \boldsymbol{f}_{XS,t}, \boldsymbol{\theta}\right) d\boldsymbol{f}_{XS,t} \\ H\left(\boldsymbol{\theta}; \boldsymbol{\theta}^{(j)}\right) &= \int p\left(\boldsymbol{f}_{XS,t} \middle| \boldsymbol{r}_t, \boldsymbol{f}_{TS,t}, \boldsymbol{\theta} = \boldsymbol{\theta}^{(j)}\right) p\left(\boldsymbol{f}_{XS,t} \middle| \boldsymbol{r}_t, \boldsymbol{f}_{TS,t}, \boldsymbol{\theta}\right) d\boldsymbol{f}_{XS,t}\end{aligned} \tag{0.26}$$

then from (0.27), we have

$$L\left(\boldsymbol{\theta}\right) = Q\left(\boldsymbol{\theta}; \boldsymbol{\theta}^{(j)}\right) - H\left(\boldsymbol{\theta}; \boldsymbol{\theta}^{(j)}\right) \tag{0.28}$$

Then EM algorithm is an iterative procedure for maximizing the Log-likelihood. It consists of the following two steps which are repeated until convergence:

[4] Equivalence follows from Bayes Theorem as the condition is equal to the joint distribution divided by the probability distribution of the given variables. However this latter distribution is independent of the parameters and can be ignored.

1) The E-step: Given an estimate of the parameters $\boldsymbol{\theta}^{(j)}$ calculate the integral $\boldsymbol{Q}(\boldsymbol{\theta},\boldsymbol{\theta}^{(j)})$.

2) The M-step: Maximise the function $\boldsymbol{Q}(\boldsymbol{\theta},\boldsymbol{\theta}^{(j)})$ with respect to θ to derive a new estimate for the parameter vector, $\boldsymbol{\theta}^{(j+1)}$. Return to step 1 and iterate.

The EM algorithm is guaranteed to converge monotonically to a local maximum of the Likelihood function. This can be demonstrated rather neatly and quickly. Note that

$$L\left(\boldsymbol{\theta}^{(j+1)}\right)-L\left(\boldsymbol{\theta}^{(j)}\right)=\left(Q\left(\boldsymbol{\theta}^{(j+1)};\boldsymbol{\theta}^{(j)}\right)-Q\left(\boldsymbol{\theta}^{(j)};\boldsymbol{\theta}^{(j)}\right)\right)-\left(H\left(\boldsymbol{\theta}^{(j+1)};\boldsymbol{\theta}^{(j)}\right)-H\left(\boldsymbol{\theta}^{(j)};\boldsymbol{\theta}^{(j)}\right)\right) \tag{0.29}$$

The first term on the right hand side is greater than 0 by virtue of the M-step. The second term is often called the relative entropy of two distributions and is always non-positive. Therefore $L\left(\boldsymbol{\theta}^{(j+1)}\right)\geq L\left(\boldsymbol{\theta}^{(j)}\right)$ and this plus some technical regularity conditions are sufficient to ensure monotonic convergence of the algorithm.

The Solution procedure

We shall now detail the E and M steps for our particular problem. We start by writing down the joint probability density function in (0.30)

$$\begin{aligned}
p\left(\boldsymbol{r}_t,\boldsymbol{f}_{TS,t},\boldsymbol{f}_{XS,t},\boldsymbol{\theta}\right) &= p\left(\boldsymbol{r}_t\middle|\boldsymbol{\theta},\boldsymbol{f}_{XS,t},\boldsymbol{f}_{TS,t}\right)p\left(\boldsymbol{f}_{XS,t}\middle|\boldsymbol{f}_{TS,t},\boldsymbol{\theta}\right)p\left(\boldsymbol{f}_{TS,t}\middle|\boldsymbol{\theta}\right)p(\boldsymbol{\theta})\\
&= K-\frac{3}{2}\sum_t\log\left(s_t^2\right)-\frac{T}{2}\log\left|\boldsymbol{D}\right|-\frac{1}{2}\sum_t s_t^{-2}\left(\boldsymbol{r}_t-\boldsymbol{B}_t\boldsymbol{f}_t\right)^T\boldsymbol{D}^{-1}\left(\boldsymbol{r}_t-\boldsymbol{B}_t\boldsymbol{f}_t\right)\\
&\quad-\frac{T}{2}\log\left|\boldsymbol{F}_{XS|TS}\right|-\frac{1}{2}\sum_t s_t^{-2}\left(\boldsymbol{f}_{XS,t}-\boldsymbol{\mu}_{XS|TS,t}\right)^T\boldsymbol{F}_{XS|TS}^{-1}\left(\boldsymbol{f}_{XS,t}-\boldsymbol{\mu}_{XS|TS,t}\right)\\
&\quad-\frac{T}{2}\log\left|\boldsymbol{F}_{TS}\right|-\frac{1}{2}\sum_t s_t^{-2}\left(\boldsymbol{f}_{TS,t}-\boldsymbol{\mu}_{TS}\right)^T\boldsymbol{F}^{-1}\left(\boldsymbol{f}_{TS,t}-\boldsymbol{\mu}_{TS}\right)\\
&\quad-\frac{1}{2}\log\left|\boldsymbol{\Omega}\right|-\frac{1}{2}\left(vec\left(\boldsymbol{B}_{TS}\right)-vec\left(\boldsymbol{B}_{TS,0}\right)\right)^T\boldsymbol{\Omega}^{-1}\left(vec\left(\boldsymbol{B}_{TS}\right)-vec\left(\boldsymbol{B}_{TS,0}\right)\right)
\end{aligned} \tag{0.31}$$

where K is just a constant of integration and the conditional distribution of $\boldsymbol{f}_{XS,t}$ given $\boldsymbol{f}_{TS,t}$ and parameters $\boldsymbol{\theta}$ is given in terms of the expressions

$$\begin{aligned}
\boldsymbol{\mu}_{XS|TS,t} &= \boldsymbol{\mu}_{XS}-\boldsymbol{F}_{XS,TS}\boldsymbol{F}_{TS}^{-1}\left(\boldsymbol{f}_{TS,t}-\boldsymbol{\mu}_{TS}\right)\\
\boldsymbol{F}_{XS|TS} &= \boldsymbol{F}_{XS}-\boldsymbol{F}_{XS,TS}\boldsymbol{F}_{TS}^{-1}\boldsymbol{F}_{XS,TS}^{T}
\end{aligned} \tag{0.32}$$

This form of the joint pdf, though standard, is difficult to use. We can rearrange it using Lemma 1 (at the end of the appendix) into a much more useful form. To this end write

$$
\begin{aligned}
&p\left(\boldsymbol{r}_t, \boldsymbol{f}_{TS,t}, \boldsymbol{f}_{XS,t}, \boldsymbol{\theta}\right) = \\
&\quad = K - \frac{3}{2}\sum_t \log\left(s_t^2\right) - \frac{T}{2}\log\left|\boldsymbol{D}\right| - \frac{1}{2}\sum_t s_t^{-2}\left(\boldsymbol{r}_t - \boldsymbol{B}_t \boldsymbol{f}_t - \boldsymbol{B}_{XS,t}\widehat{\boldsymbol{f}}_{XS,t}\right)^T \boldsymbol{D}^{-1}\left(\boldsymbol{r}_t - \boldsymbol{B}_t \boldsymbol{f}_t - \boldsymbol{B}_{XS,t}\widehat{\boldsymbol{f}}_{XS,t}\right) \\
&\qquad - \frac{T}{2}\log\left|\boldsymbol{F}_{XS|TS}\right| - \frac{1}{2}\sum_t s_t^{-2}\left(\widehat{\boldsymbol{f}}_{XS,t} - \boldsymbol{\mu}_{XS|TS,t}\right)^T \boldsymbol{F}_{XS|TS}^{-1}\left(\widehat{\boldsymbol{f}}_{XS,t} - \boldsymbol{\mu}_{XS|TS,t}\right) \\
&\qquad - \frac{1}{2}\sum_t s_t^{-2}\left(\boldsymbol{f}_{XS,t} - \widehat{\boldsymbol{f}}_{XS,t}\right)^T\left(\boldsymbol{B}_{XS,t}^T \boldsymbol{D}^{-1}\boldsymbol{B}_{XS,t} + \boldsymbol{F}_{XS|TS}^{-1}\right)\left(\boldsymbol{f}_{XS,t} - \widehat{\boldsymbol{f}}_{XS,t}\right) \\
&\qquad - \frac{T}{2}\log\left|\boldsymbol{F}_{TS}\right| - \frac{1}{2}\sum_t s_t^{-2}\left(\boldsymbol{f}_{TS,t} - \boldsymbol{\mu}_{TS}\right)^T \boldsymbol{F}^{-1}\left(\boldsymbol{f}_{TS,t} - \boldsymbol{\mu}_{TS}\right) \\
&\qquad - \frac{1}{2}\log\left|\boldsymbol{\Omega}\right| - \frac{1}{2}\left(vec\left(\boldsymbol{B}_{TS}\right) - vec\left(\boldsymbol{B}_{TS,0}\right)\right)^T \boldsymbol{\Omega}^{-1}\left(vec\left(\boldsymbol{B}_{TS}\right) - vec\left(\boldsymbol{B}_{TS,0}\right)\right)
\end{aligned}
\tag{0.33}
$$

where

$$
\begin{aligned}
&\widehat{\boldsymbol{f}}_{XS,t} = \left(\boldsymbol{B}_{XS,t}^T \boldsymbol{D}^{-1}\boldsymbol{B}_{XS,t} + \boldsymbol{F}_{XS|TS}^{-1}\right)^{-1}\left(\boldsymbol{B}_{XS,t}^T \boldsymbol{D}^{-1}\left(\boldsymbol{r}_t - \boldsymbol{B}_{TS}\boldsymbol{f}_{TS,t}\right) + \boldsymbol{F}_{XS|TS}^{-1}\boldsymbol{\mu}_{XS|TS,t}\right) \\
&Var\left(\boldsymbol{f}_{XS,t} - \widehat{\boldsymbol{f}}_{XS,t}\right) = s_t^2\left(\boldsymbol{B}_{XS,t}^T \boldsymbol{D}^{-1}\boldsymbol{B}_{XS,t} + \boldsymbol{F}_{XS|TS}^{-1}\right)^{-1}
\end{aligned}
\tag{0.34}
$$

Given these slightly unwieldy expressions, we are finally in a position to detail the algorithm.

E-step: The E-step of the algorithm amounts to assuming a distribution for the latent factors $f_{XS,t}$ and then taking the expectation of the likelihood in (0.31) with respect to these factors. However by rewriting the probability distributing in form of (0.33), only the 3rd term is a function of these latent factors, $f_{XS,t}$. Therefore taking expectation (or integrating out with respect to the pdf of $f_{XS,t}$.) gives

$$
\begin{aligned}
&Q\left(\theta;\theta^{(j)}\right) = \\
&\quad = K - \frac{3}{2}\sum_t \log\left(s_t^2\right) - \frac{T}{2}\log\left|\boldsymbol{D}\right| - \frac{1}{2}\sum_t s_t^{-2}\left(\boldsymbol{r}_t - \boldsymbol{B}_t \boldsymbol{f}_t - \boldsymbol{B}_{XS,t}\widehat{\boldsymbol{f}}_{XS,t}\right)^T \boldsymbol{D}^{-1}\left(\boldsymbol{r}_t - \boldsymbol{B}_t \boldsymbol{f}_t - \boldsymbol{B}_{XS,t}\widehat{\boldsymbol{f}}_{XS,t}\right) \\
&\qquad - \frac{T}{2}\log\left|\boldsymbol{F}_{XS|TS}\right| - \frac{1}{2}\sum_t s_t^{-2}\left(\widehat{\boldsymbol{f}}_{XS,t} - \boldsymbol{\mu}_{XS|TS,t}\right)^T \boldsymbol{F}_{XS|TS}^{-1}\left(\widehat{\boldsymbol{f}}_{XS,t} - \boldsymbol{\mu}_{XS|TS,t}\right) \\
&\qquad - \frac{1}{2}\sum_t trace\left(s_t^{-2}\left(\boldsymbol{B}_{XS,t}^T \boldsymbol{D}^{-1}\boldsymbol{B}_{XS,t} + \boldsymbol{F}_{XS|TS}^{-1}\right)Var\left(\boldsymbol{f}_{XS,t} - \widehat{\boldsymbol{f}}_{XS,t}\right)\right) \\
&\qquad - \frac{T}{2}\log\left|\boldsymbol{F}_{TS}\right| - \frac{1}{2}\sum_t s_t^{-2}\left(\boldsymbol{f}_{TS,t} - \boldsymbol{\mu}_{TS}\right)^T \boldsymbol{F}^{-1}\left(\boldsymbol{f}_{TS,t} - \boldsymbol{\mu}_{TS}\right) \\
&\qquad - \frac{1}{2}\log\left|\boldsymbol{\Omega}\right| - \frac{1}{2}\left(vec\left(\boldsymbol{B}_{TS}\right) - vec\left(\boldsymbol{B}_{TS,0}\right)\right)^T \boldsymbol{\Omega}^{-1}\left(vec\left(\boldsymbol{B}_{TS}\right) - vec\left(\boldsymbol{B}_{TS,0}\right)\right)
\end{aligned}
\tag{0.35}
$$

where

$$\widehat{f}_{XS,t} = \left(B_{XS,t}^T D^{(j)-1} B_{XS,t} + F_{XS|TS}^{(j)-1} \right)^{-1} \left(B_{XS,t}^T D^{(j)-1} \left(r_t - B_{TS}^{(j)} f_{TS,t} \right) + F_{XS|TS}^{(j)-1} \mu_{XS|TS,t}^{(j)} \right)$$
$$Var\left(f_{XS,t} - \widehat{f}_{XS,t} \right) = s_t^2 \left(B_{XS,t}^T D^{(j)-1} B_{XS,t} + F_{XS|TS}^{(j)-1} \right)^{-1}$$
$$(0.36)$$

where we have made clear the dependence on the current estimate of the parameters by the suffix *(j)*.

M-step: This step is to maximise the expectation in (0.35) with respect to the parameters. We first derive a set of first order condition for the time-series betas is

$$\frac{\delta Q}{\delta B_{TS}} = \sum_t s_t^{-2} D^{-1} \left(\left(r_t - B_{XS,t} \widehat{f}_{XS,t} \right) f_{TS,t}^T - B_{TS} f_{TS,t} f_{TS,t}^T \right) + \left(B_{TS,0} - B_{TS} \right) \bullet (1/\Lambda)$$
$$+ \frac{1}{\delta w^T \Omega_o w} w_m w_m^T \left(B_{TS,0} - B_{TS} \right) ee^T$$
$$(0.37)$$

with the conditions for the other parameters being

$$\frac{\delta Q}{\delta D} = -\frac{T}{2} D^{-1} + \frac{D^{-2}}{2} \sum_t s_t^{-2} \left(r_t - B_t f_t - B_{XS,t} \widehat{f}_{XS,t} \right) \left(r_t - B_t f_t - B_{XS,t} \widehat{f}_{XS,t} \right)^T$$
$$+ \frac{D^{-2}}{2} \sum_t s_t^{-2} B_{XS,t} Var\left(f_{XS,t} - \widehat{f}_{XS,t} \right) B_{XS,t}^T$$
$$(0.38)$$

$$\frac{\delta Q}{\delta F_{TS}} = -\frac{T}{2} F_{TS}^{-1} + \frac{F_{TS}^{-2}}{2} \sum_t s_t^{-2} \left(f_{TS,t} - \mu_{TS} \right) \left(f_{TS,t} - \mu_{TS} \right)^T \qquad (0.39)$$

$$\frac{\delta Q}{\delta F_{XS|TS}} = -\frac{T}{2} F_{XS|TS}^{-1} + \frac{F_{XS|TS}^{-2}}{2} \sum_t s_t^{-2} \left(\left(\widehat{f}_{XS,t} - \mu_{XS|TS,t} \right) \left(\widehat{f}_{XS,t} - \mu_{XS|TS,t} \right)^T + Var\left(f_{XS,t} - \widehat{f}_{XS,t} \right) \right)$$
$$(0.40)$$

We derive an updated estimate of the parameters by solving these first order conditions. Of these, only solving for the parameters B_{TS} needs some thought as the prior is degenerate. However, the analysis is greatly simplified if the time-series factors have previously been orthognalised with respect to the market or first factor – implying $f_{TS,t} f_{TS,t}^T$ is block diagonal. This implies that

$$e^T \left(\sum_t s_t^{-2} f_{TS,t} f_{TS,t}^T \right)^{-1} = \frac{1}{\sum_t s_t^{-2} \left(e^T f_{TS,t} \right)^2} e^T \qquad (0.41)$$

Under these conditions we can solve for the next estimate of the parameters B_{TS} using the following two steps:

1) Derive a first preliminary estimate, denoted $\tilde{B}_{TS}^{(j+1)}$, by solving the *n* time-series regressions ignoring the final term in (0.37), that is

$$\left[\tilde{\boldsymbol{B}}_{TS}^{(j+1)}\right]_{i\bullet} = \left(\sum_t s_t^{-2}\left(\boldsymbol{r}_{t,i} - \left[\boldsymbol{B}_{XS,t}\right]_{i\bullet}\widehat{\boldsymbol{f}}_{XS,t}\right)\boldsymbol{f}_{TS,t}^T + \left[\boldsymbol{B}_{TS,0}\right]_{i\bullet} diag\left(\left[\boldsymbol{D}(1/\boldsymbol{\Lambda})\right]_{i\bullet}\right)\right)$$
$$\left(\sum_t s_t^{-2}\boldsymbol{f}_{TS,t}\boldsymbol{f}_{TS,t}^T + diag\left(\left[\boldsymbol{D}(1/\boldsymbol{\Lambda})\right]_{i\bullet}\right)\right)^{-1} \tag{0.42}$$

where we use the notation $[X]_{i\bullet}$ to refer to the *i*th row of the matrix *X*.

1) Enforce the constraint on the market betas by letting, then

$$\left[\boldsymbol{B}_{TS}^{(j+1)}\right]_{\bullet 1} = \tilde{\boldsymbol{B}}_{TS}^{(j+1)}\boldsymbol{e} - \frac{\boldsymbol{w}_m^T\left(\tilde{\boldsymbol{B}}_{TS}^{(j+1)} - \boldsymbol{B}_{TS,0}\right)\boldsymbol{e}}{\boldsymbol{w}_m^T\boldsymbol{D}\boldsymbol{w}_m}\boldsymbol{D}\boldsymbol{w}_m \tag{0.43}$$

where notation $[X]_{\bullet j}$ refers to the jth column. The other columns of $\boldsymbol{B}_{TS}^{(j+1)}$ are equal to the respective columns of $\tilde{\boldsymbol{B}}_{TS}$, i.e. $\left[\boldsymbol{B}_{TS}^{(j+1)}\right]_{\bullet j} = \left[\tilde{\boldsymbol{B}}_{TS}^{(j+1)}\right]_{\bullet j}$ for $j \neq 1$.

The solutions to the other first order conditions follow immediately.

$$\boldsymbol{D}^{(j+1)} = \frac{1}{T}\sum_t\left(s_t^{-2}\left(\boldsymbol{r}_t - \boldsymbol{B}_t\boldsymbol{f}_t - \boldsymbol{B}_{XS,t}\widehat{\boldsymbol{f}}_{XS,t}\right)\left(\boldsymbol{r}_t - \boldsymbol{B}_t\boldsymbol{f}_t - \boldsymbol{B}_{XS,t}\widehat{\boldsymbol{f}}_{XS,t}\right)^T + \boldsymbol{B}_{XS,t}Var\left(\boldsymbol{f}_{XS,t} - \widehat{\boldsymbol{f}}_{XS,t}\right)\boldsymbol{B}_{XS,t}^T\right) \tag{0.44}$$

$$\boldsymbol{F}_{TS}^{(j+1)} = \frac{1}{T}\sum_t s_t^{-2}\left(\boldsymbol{f}_{TS,t} - \boldsymbol{\mu}_{TS}\right)\left(\boldsymbol{f}_{TS,t} - \boldsymbol{\mu}_{TS}\right)^T \tag{0.45}$$

$$\boldsymbol{F}_{XS|TS}^{(j+1)} = \frac{1}{T}\sum_t s_t^{-2}\left(\left(\widehat{\boldsymbol{f}}_{XS,t} - \boldsymbol{\mu}_{XS|TS,t}\right)\left(\widehat{\boldsymbol{f}}_{XS,t} - \boldsymbol{\mu}_{XS|TS,t}\right)^T + Var\left(\boldsymbol{f}_{XS,t} - \widehat{\boldsymbol{f}}_{XS,t}\right)\right) \tag{0.46}$$

With this next set of parameter estimates, we can now return to the **M**-step. This procedure is repeated until convergence.

An Intermediate Result: Lemma 1

***Lemma 1**: Let y=Bx+ε where $x \sim N(\mu, F)$, $\varepsilon \sim N(0, D)$ and Cov(x,ε)=0. Define the following quadratic function of these stochastic variables*

$$L = (y - Bx)^T D^{-1}(y - Bx) + (x - \mu)^T F^{-1}(x - \mu) \tag{0.47}$$

The stochastic variate x can be decomposed as

$$\begin{aligned} x &= \left(B^T D^{-1} B + F^{-1}\right)^{-1}\left(B^T D^{-1} y + F^{-1}\mu\right) + \tilde{x} \\ &:= \hat{x} + \tilde{x} \end{aligned} \tag{0.48}$$

where $Cov(\tilde{x}, y) = 0$ and $\tilde{x} \sim N\left(0, \left(B^T D^{-1} B + F^{-1}\right)^{-1}\right)$. Using this decomposition the quadratic function (0.47) can be written as

$$L = \left(y - B\hat{x}\right)^T D^{-1}\left(y - B\hat{x}\right) + \left(\hat{x} - \mu\right)^T F^{-1}\left(\hat{x} - \mu\right) + \tilde{x}^T\left(B^T D^{-1} B + F^{-1}\right)\tilde{x}$$

(0.49)

Proof: By construction

$$\tilde{x} = \left(B^T D^{-1} B + F^{-1}\right)^{-1}\left(F^{-1}\left(x - \mu\right) - B^T D^{-1}\varepsilon\right) \qquad (0.50)$$

It therefore follows that $E\left(\tilde{x}\right) = 0$ and that

$$\begin{aligned} Var\left(\tilde{x}\right) &= \left(B^T D^{-1} B + F^{-1}\right)^{-1}\left(B^T D^{-1} B + F^{-1}\right)\left(B^T D^{-1} B + F^{-1}\right)^{-1} \\ &= \left(B^T D^{-1} B + F^{-1}\right)^{-1} \end{aligned} \qquad (0.51)$$

Similarly

$$\begin{aligned} Cov\left(\tilde{x}, y\right) &= \left(B^T D^{-1} B + F^{-1}\right)^{-1}\left(F^{-1} Var\left(x\right) B^T - B^T D^{-1} Var\left(\varepsilon\right)\right) \\ &= \left(B^T D^{-1} B + F^{-1}\right)^{-1}\left(B^T - B^T\right) = 0 \end{aligned} \qquad (0.52)$$

Now we can rearrange *(0.53) so*

$$\begin{aligned} &\left(B^T D^{-1} B + F^{-1}\right)\tilde{x} = \left(F^{-1}\left(\hat{x} + \tilde{x} - \mu\right) - B^T D^{-1}\varepsilon\right) \Rightarrow \\ &B^T D^{-1}\left(B\tilde{x} - \varepsilon\right) = F^{-1}\left(\hat{x} - \mu\right) \Rightarrow \\ &B^T D^{-1}\left(y - B\hat{x}\right) = F^{-1}\left(\hat{x} - \mu\right) \end{aligned} \qquad (0.54)$$

The identity now follows substituting in $x = \hat{x} + \tilde{x}$ into *(0.55)* and multiplying out.

References

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Wu, J. (2006), "Realized Beta: Persistence and Predictability." in T. Fomby, ed., Advances in Econometrics: Econometric Analysis of Economic and Financial Time Series.

Fama, E.F. and French, K.R. (1992). "The cross-section of expected stock returns", Journal of Finance, 47(2), 427-465.

Heston, S.L. and Rouwenhorst, K.G. (1995). "Industry and country effects in international stock returns, The Journal of Portfolio Management, 21(3), pp. 53-58.

Tien, D. and Pfleiderer, P. and Maxim, R. and Marsh, T. (2005). "Decomposing Factor Exposure for Equity Portfolios" in 'Linear Factor Models in Finance'..

Scowcroft A., Sefton J. (2006). "Understanding Factor Models", UBS Global Quantitative Research.

Briner, B.G. and Connor, G. (2008). "How much structure is best? A comparison of market model, factor model and unstructured equity covariance matrices", Journal of Risk, 10(4), pp. 3-30.

Miller, G. (2006). "Needles, Haystacks, and Hidden Factors", The Journal of Portfolio Management, 32(2), pp. 25-32.

Stroyny, A.L. (2005). "Estimating factor models" in Linear Factor Models in Finance, pp. 210-215.

Young, G. (1940) "Maximum likelihood estimation and factor analysis", Psychometrika, 6, pp. 49:53.

Whittle, P. (1952) "On principal components and least squares methods of factor analysis" Skandinavisk Aktuarietidskrift, 36, pp. 223-239.

Rubin, Donald B. and Dorothy T. Thayer. (1982), "*EM* Algorithms for ML Factor Analysis." Psychometrika, 47, pp. 69-76.

Lawley, D. N. (1941). "Further investigations in factor estimation". Proc. Roy. Soc. Edinburgh Sec. A, 61, pp. 176-185.

McLachlan, G.J. and Krishnan, T. (1997). "The EM algorithm and extensions", Wiley, New York.

Connor, G. Korajczyk R. (2009). "Factor Models in Portfolio and Asset Pricing Theory", http://ssrn.com/abstract=1139062.

■ Analyst Certification

Each research analyst primarily responsible for the content of this research report, in whole or in part, certifies that with respect to each security or issuer that the analyst covered in this report: (1) all of the views expressed accurately reflect his or her personal views about those securities or issuers; and (2) no part of his or her compensation was, is, or will be, directly or indirectly, related to the specific recommendations or views expressed by that research analyst in the research report.

Required Disclosures

This report has been prepared by UBS Limited, an affiliate of UBS AG. UBS AG, its subsidiaries, branches and affiliates are referred to herein as UBS.

For information on the ways in which UBS manages conflicts and maintains independence of its research product; historical performance information; and certain additional disclosures concerning UBS research recommendations, please visit www.ubs.com/disclosures. The figures contained in performance charts refer to the past; past performance is not a reliable indicator of future results. Additional information will be made available upon request.

UBS Securities Co. Limited is licensed to conduct securities investment consultancy businesses by the China Securities Regulatory Commission.

UBS Investment Research: Global Equity Rating Allocations

UBS 12-Month Rating	Rating Category	Coverage[1]	IB Services[2]
Buy	**Buy**	51%	37%
Neutral	**Hold/Neutral**	40%	33%
Sell	**Sell**	9%	22%
UBS Short-Term Rating	**Rating Category**	**Coverage[3]**	**IB Services[4]**
Buy	**Buy**	less than 1%	20%
Sell	**Sell**	less than 1%	0%

1:Percentage of companies under coverage globally within the 12-month rating category.
2:Percentage of companies within the 12-month rating category for which investment banking (IB) services were provided within the past 12 months.
3:Percentage of companies under coverage globally within the Short-Term rating category.
4:Percentage of companies within the Short-Term rating category for which investment banking (IB) services were provided within the past 12 months.

Source: UBS. Rating allocations are as of 30 September 2010.

UBS Investment Research: Global Equity Rating Definitions

UBS 12-Month Rating	Definition
Buy	FSR is > 6% above the MRA.
Neutral	FSR is between -6% and 6% of the MRA.
Sell	FSR is > 6% below the MRA.
UBS Short-Term Rating	**Definition**
Buy	Buy: Stock price expected to rise within three months from the time the rating was assigned because of a specific catalyst or event.
Sell	Sell: Stock price expected to fall within three months from the time the rating was assigned because of a specific catalyst or event.

KEY DEFINITIONS

Forecast Stock Return (FSR) is defined as expected percentage price appreciation plus gross dividend yield over the next 12 months.
Market Return Assumption (MRA) is defined as the one-year local market interest rate plus 5% (a proxy for, and not a forecast of, the equity risk premium).
Under Review (UR) Stocks may be flagged as UR by the analyst, indicating that the stock's price target and/or rating are subject to possible change in the near term, usually in response to an event that may affect the investment case or valuation.
Short-Term Ratings reflect the expected near-term (up to three months) performance of the stock and do not reflect any change in the fundamental view or investment case.
Equity Price Targets have an investment horizon of 12 months.

EXCEPTIONS AND SPECIAL CASES

UK and European Investment Fund ratings and definitions are: Buy: Positive on factors such as structure, management, performance record, discount; Neutral: Neutral on factors such as structure, management, performance record, discount; Sell: Negative on factors such as structure, management, performance record, discount.
Core Banding Exceptions (CBE): Exceptions to the standard +/-6% bands may be granted by the Investment Review Committee (IRC). Factors considered by the IRC include the stock's volatility and the credit spread of the respective company's debt. As a result, stocks deemed to be very high or low risk may be subject to higher or lower bands as they relate to the rating. When such exceptions apply, they will be identified in the Company Disclosures table in the relevant research piece.

Research analysts contributing to this report who are employed by any non-US affiliate of UBS Securities LLC are not registered/qualified as research analysts with the NASD and NYSE and therefore are not subject to the restrictions contained in the NASD and NYSE rules on communications with a subject company, public appearances, and trading securities held by a research analyst account. The name of each affiliate and analyst employed by that affiliate contributing to this report, if any, follows.

UBS Limited: James Sefton; David Jessop; Giuliano De Rossi.

Unless otherwise indicated, please refer to the Valuation and Risk sections within the body of this report.

Global Disclaimer

This report has been prepared by UBS Limited, an affiliate of UBS AG. UBS AG, its subsidiaries, branches and affiliates are referred to herein as UBS. In certain countries, UBS AG is referred to as UBS SA.

This report is for distribution only under such circumstances as may be permitted by applicable law. Nothing in this report constitutes a representation that any investment strategy or recommendation contained herein is suitable or appropriate to a recipient's individual circumstances or otherwise constitutes a personal recommendation. It is published solely for information purposes, it does not constitute an advertisement and is not to be construed as a solicitation or an offer to buy or sell any securities or related financial instruments in any jurisdiction. No representation or warranty, either express or implied, is provided in relation to the accuracy, completeness or reliability of the information contained herein, except with respect to information concerning UBS AG, its subsidiaries and affiliates, nor is it intended to be a complete statement or summary of the securities, markets or developments referred to in the report. UBS does not undertake that investors will obtain profits, nor will it share with investors any investment profits nor accept any liability for any investment losses. Investments involve risks and investors should exercise prudence in making their investment decisions. The report should not be regarded by recipients as a substitute for the exercise of their own judgement. Past performance is not necessarily a guide to future performance. The value of any investment or income may go down as well as up and you may not get back the full amount invested. Any opinions expressed in this report are subject to change without notice and may differ or be contrary to opinions expressed by other business areas or groups of UBS as a result of using different assumptions and criteria. Research will initiate, update and cease coverage solely at the discretion of UBS Investment Bank Research Management. The analysis contained herein is based on numerous assumptions. Different assumptions could result in materially different results. The analyst(s) responsible for the preparation of this report may interact with trading desk personnel, sales personnel and other constituencies for the purpose of gathering, synthesizing and interpreting market information. UBS is under no obligation to update or keep current the information contained herein. UBS relies on information barriers to control the flow of information contained in one or more areas within UBS, into other areas, units, groups or affiliates of UBS. The compensation of the analyst who prepared this report is determined exclusively by research management and senior management (not including investment banking). Analyst compensation is not based on investment banking revenues, however, compensation may relate to the revenues of UBS Investment Bank as a whole, of which investment banking, sales and trading are a part.

The securities described herein may not be eligible for sale in all jurisdictions or to certain categories of investors. Options, derivative products and futures are not suitable for all investors, and trading in these instruments is considered risky. Mortgage and asset-backed securities may involve a high degree of risk and may be highly volatile in response to fluctuations in interest rates and other market conditions. Past performance is not necessarily indicative of future results. Foreign currency rates of exchange may adversely affect the value, price or income of any security or related instrument mentioned in this report. For investment advice, trade execution or other enquiries, clients should contact their local sales representative. Neither UBS nor any of its affiliates, nor any of UBS' or any of its affiliates, directors, employees or agents accepts any liability for any loss or damage arising out of the use of all or any part of this report. For financial instruments admitted to trading on an EU regulated market: UBS AG, its affiliates or subsidiaries (excluding UBS Securities LLC and/or UBS Capital Markets LP) acts as a market maker or liquidity provider (in accordance with the interpretation of these terms in the UK) in the financial instruments of the issuer save that where the activity of liquidity provider is carried out in accordance with the definition given to it by the laws and regulations of any other EU jurisdictions, such information is separately disclosed in this research report. UBS and its affiliates and employees may have long or short positions, trade as principal and buy and sell in instruments or derivatives identified herein.

Any prices stated in this report are for information purposes only and do not represent valuations for individual securities or other instruments. There is no representation that any transaction can or could have been effected at those prices and any prices do not necessarily reflect UBS's internal books and records or theoretical model-based valuations and may be based on certain assumptions. Different assumptions, by UBS or any other source, may yield substantially different results.

United Kingdom and the rest of Europe: Except as otherwise specified herein, this material is communicated by UBS Limited, a subsidiary of UBS AG, to persons who are eligible counterparties or professional clients and is only available to such persons. The information contained herein does not apply to, and should not be relied upon by, retail clients. UBS Limited is authorised and regulated by the Financial Services Authority (FSA). UBS research complies with all the FSA requirements and laws concerning disclosures and these are indicated on the research where applicable. **France:** Prepared by UBS Limited and distributed by UBS Limited and UBS Securities France SA. UBS Securities France S.A. is regulated by the Autorité des Marchés Financiers (AMF). Where an analyst of UBS Securities France S.A. has contributed to this report, the report is also deemed to have been prepared by UBS Securities France S.A. **Germany:** Prepared by UBS Limited and distributed by UBS Limited and UBS Deutschland AG. UBS Deutschland AG is regulated by the Bundesanstalt fur Finanzdienstleistungsaufsicht (BaFin). **Spain:** Prepared by UBS Limited and distributed by UBS Limited and UBS Securities España SV, SA. UBS Securities España SV, SA is regulated by the Comisión Nacional del Mercado de Valores (CNMV). **Turkey:** Prepared by UBS Menkul Degerler AS on behalf of and distributed by UBS Limited. **Russia:** Prepared and distributed by UBS Securities CJSC. **Switzerland:** Distributed by UBS AG to persons who are institutional investors only. **Italy:** Prepared by UBS Limited and distributed by UBS Limited and UBS Italia Sim S.p.A.. UBS Italia Sim S.p.A. is regulated by the Bank of Italy and by the Commissione Nazionale per le Società e la Borsa (CONSOB). Where an analyst of UBS Italia Sim S.p.A. has contributed to this report, the report is also deemed to have been prepared by UBS Italia Sim S.p.A.. **South Africa:** UBS South Africa (Pty) Limited (Registration No. 1995/011140/07) is a member of the JSE Limited, the South African Futures Exchange and the Bond Exchange of South Africa. UBS South Africa (Pty) Limited is an authorised Financial Services Provider. Details of its postal and physical address and a list of its directors are available on request or may be accessed at http:www.ubs.co.za. **United States:** Distributed to US persons by either UBS Securities LLC or by UBS Financial Services Inc., subsidiaries of UBS AG; or by a group, subsidiary or affiliate of UBS AG that is not registered as a US broker-dealer (a **'non-US affiliate'**), to major US institutional investors only. UBS Securities LLC or UBS Financial Services Inc. accepts responsibility for the content of a report prepared by another non-US affiliate when distributed to US persons by UBS Securities LLC or UBS Financial Services Inc. All transactions by a US person in the securities mentioned in this report must be effected through UBS Securities LLC or UBS Financial Services Inc., and not through a non-US affiliate. **Canada:** Distributed by UBS Securities Canada Inc., a subsidiary of UBS AG and a member of the principal Canadian stock exchanges & CIPF. A statement of its financial condition and a list of its directors and senior officers will be provided upon request. **Hong Kong:** Distributed by UBS Securities Asia Limited. **Singapore:** Distributed by UBS Securities Pte. Ltd [mica (p) 039/11/2009 and Co. Reg. No.: 198500648C] or UBS AG, Singapore Branch. Please contact UBS Securities Pte Ltd, an exempt financial advisor under the Singapore Financial Advisers Act (Cap. 110); or UBS AG Singapore branch, an exempt financial adviser under the Singapore Financial Advisers Act (Cap. 110) and a wholesale bank licensed under the Singapore Banking Act (Cap. 19) regulated by the Monetary Authority of Singapore, in respect of any matters arising from, or in connection with, the analysis or report. The recipient of this report represent and warrant that they are accredited and institutional investors as defined in the Securities and Futures Act (Cap. 289). **Japan:** Distributed by UBS Securities Japan Ltd to institutional investors only. Where this report has been prepared by UBS Securities Japan Ltd, UBS Securities Japan Ltd is the author, publisher and distributor of the report. **Australia:** Distributed by UBS AG (Holder of Australian Financial Services License No. 231087) and UBS Securities Australia Ltd (Holder of Australian Financial Services License No. 231098) only to 'Wholesale' clients as defined by s761G of the Corporations Act 2001. **New Zealand:** Distributed by UBS New Zealand Ltd. An investment adviser and investment broker disclosure statement is available on request and free of charge by writing to PO Box 45, Auckland, NZ. **Dubai:** The research prepared and distributed by UBS AG Dubai Branch, is intended for Professional Clients only and is not for further distribution within the United Arab Emirates. **Korea:** Distributed in Korea by UBS Securities Pte. Ltd., Seoul Branch. This report may have been edited or contributed to from time to time by affiliates of UBS Securities Pte. Ltd., Seoul Branch. **Malaysia:** This material is authorized to be distributed in Malaysia by UBS Securities Malaysia Sdn. Bhd (253825-x).

The disclosures contained in research reports produced by UBS Limited shall be governed by and construed in accordance with English law.

UBS specifically prohibits the redistribution of this material in whole or in part without the written permission of UBS and UBS accepts no liability whatsoever for the actions of third parties in this respect. Images may depict objects or elements which are protected by third party copyright, trademarks and other intellectual property rights. © UBS 2010. The key symbol and UBS are among the registered and unregistered trademarks of UBS. All rights reserved.

UBS