

A Rank Approach to Equity Forecast Construction

S. E. Satchell and S. M. Wright

November 2005

CWPE 0553

Not to be quoted without permission

A Rank Approach to Equity Forecast Construction

S.E. Satchell

Faculty of Economics

University of Cambridge

and Trinity College, Cambridge

S.M. Wright*

Equities Research Department

UBS Investment Bank (UK)

London

October 2005

*The views and opinions expressed in this article are those of the authors and are not necessarily those of UBS. UBS accepts no liability over the content of the article. It is published solely for informational purposes and is not to be construed as a solicitation or an offer to buy or sell any securities or related financial instruments.

ABSTRACT

The purpose of this paper is to present a rank based approach to cross-sectional linear factor modelling. The emphasis is on approximating factor exposures in a consistent manner in order to facilitate the merging of subjective information (from professional investors) with objective information (from accounting data and/or state of the art quantitative models) in a statistically rigorous way without needing to impose the unrealistic simplifying assumptions typical of more standard time series models. We deal with the problems of identifying country and sector returns by an innovative hierarchical factor structure. This is all discussed from the perspective that investment models are not immutable but rather need to be designed with characteristics that are fit for their purpose; for example, returning aggregate county and sector forecasts that are consistent by construction.

Keywords: Linear Factor Models, Ranking, Robustness Exposures, Forecasting.

JEL Classification: G11

1 Introduction

The purpose of this paper is to present a number of arguments favouring a robust cross-sectional approach to forecasting alpha using financial models. We take it as a given that financial markets do not exhibit sufficient regularity to allow us to use the large-sample theory that mainstream statistics is based upon. In mainstream statistics, stochastic processes possess such pleasing properties as stationarity and ergodicity. These properties allow us to get closer to the true, unknown data-generating process, simply by observing long enough time series. However, financial markets are driven by innovation, evolution and exogenous events. Such “one-offs” as leaving the European Monetary System, September 11th, October 1987, the Tequila crisis etcetera etcetera, may be capable of being modelled within a conventional statistical structure but it is probably not helpful to do so in the context of our particular audience, professional investors.

For government purposes, what is of interest is to establish equilibrium states that can guide economic management toward a stable sustainable set of policies, hence for these long term purposes, large sample theory is entirely appropriate. For our audience of professional investors, the transient behaviour of the market and how it moves between states, is all-important as this is what generates investment returns. Hence for this audience, by definition, the one off events are of dominant importance because they are what disturb the equilibrium (if it is ever fully achieved).

Our version of financial markets is that asset returns are highly state dependent non-linear functions of underlying exogenous variables wherein the key parameters, if they can be identified, exhibit profound degrees of time-dependence. Faced with such a bleak description, it may be thought that quantitative finance has very little to offer; in what follows, we describe a methodology that is both easy to use, highly quantitative, and which takes into account the issues discussed above.

Much of what we shall say is motivated by the idea that while professional fund management is practiced as an art and not a science, we can still build rigorous quantitative frameworks within which we can organise our thoughts. The fact that an event is a “one off” in a traditional frequentist sense does not mean that we have nothing to say about its consequences; we bear in mind Fisher Black’s aphorism “In the absence of data I rely on theory”. What is needed is a way to merge theory and data in an effective manner. Whilst we would recognise that

conventional time series analysis has an important role to play, we believe that it was a better tool in 1995 after some years of reasonably regular price behaviour than it is now in 2005 after some years of highly irregular behaviour.

Moreover while we cannot tame the markets, we can hope to design our investment models such that we make the most of all available information, record and discuss our views as concisely and realistically as possible, avoid inconsistencies and inefficiencies in our implementation, manage our risk exposure as realistically as we can, and learn from our experience to maximise our performance in future.

We do not claim to be able to fully meet all these aspirations, they do however provide the context for this paper which is primarily concerned with one stage in this process design, that of forecasting. It is however connected with other stages such as that of robust optimisation which we covered in an earlier paper (see Satchell and Wright, 2003), we shall not delve into these other topics too much; however, we shall refer readers to appropriate papers where necessary.

Our contention is that investment models can be designed to have certain properties and that when you approach the task from this perspective you arrive at some very interesting conclusions that are of crucial importance to the practicing investment professional.

The properties that are of interest for the forecasting task are the ability to merge subjective and objective information, the robustness of the model to noise in the data, the flexibility to accommodate complex non-stationary behaviour and ever changing model form and finally the simplicity and clarity to aid communication of, and build confidence in the conclusions. All of these topics individually are covered by a substantial literature. What we cover here is an innovative approach that brings these ideas together into a practical operational whole.

The fundamental building blocks for our approach will be listed next. The principal tool will be the linear factor model; this is central to all quantitative portfolio analysis. Another important notion is what we could call local normality; this is not to claim that returns are normal, but that returns over short periods of time and/or conditional upon particular states of the world are normal. Such an approach allows for unconditional returns being non-normal; it also encompasses log normality and the style of analysis used in the option pricing literature. The third key idea we shall utilise is that of ranks. In previous work, we used ranked means in mean variance analysis, see Satchell and Wright (2003). There, we showed that using ranked forecasts was a robust procedure that dealt with many of the

substantial difficulties encountered when constructing portfolios using knowledge about forecast alphas. Here, we extend that principle to construction of linear factor models, focusing on the use of rank information on the factor structure. Building models based around these three ideas leads to a robust procedure that is easy to implement and straightforward to understand whilst taking on board many of the theoretical problems raised earlier. In section 2 we discuss our approach to linear factor models, in section 3 we consider issues of modelling the returns distribution, in section 4 we investigate the ranking of factor information and the interrelationship between accuracy of the approximation and quality of the information availability. Section 5 discusses how these ideas can be put together in a model and discusses certain examples. Section 6 discusses problems in model building such as lack of independence of our leading indicators and what to do about them; our conclusions follow in section 7.

2 Linear Factor Models

In this section, we first review linear factor models.

Let y_{it} be the return/rate of return to asset i at time t . Let β_{ijt} be the exposure of the return of asset i to factor j at time t . Let f_{jt} be the value of factor j at time t . We assume n assets, $i = 1, \dots, n$. We assume k factors, $j = 1, \dots, k$. The linear factor model (LFM) takes the following form

$$y_{it} = \sum_{j=1}^k \beta_{ijt} f_{jt} + V_{it} \quad (1)$$

The error terms V_{it} represent the idiosyncratic component of the LFM. This is usually modelled as a random variable, V_{it} , which has a mean α_{it} and a variance σ_i^2 . The key assumption of (1), other than linearity is that it is the factors f_{jt} that pick up the common covariation in y_{it} , so that the V_{it} are uncorrelated. By use of vectors and matrix notion, (1) can be re-written as

$$\underset{\sim}{y}_t = \underset{\sim}{\beta} \underset{\sim}{f}_t + \underset{\sim}{V}_t \quad (2)$$

where $\underset{\sim}{y}_t$ and $\underset{\sim}{V}_t$ are $(n \times 1)$ vectors, $\underset{\sim}{\beta}$ is an $(n \times k)$ vector.

It is clearly unlikely that such an attractive structure is an accurate description of reality, however much we may wish it to be. Our perspective on this is that of the mathematician interested in linearising highly non-linear dynamic equation

systems. We are motivated by the fact that however complex and non linear the real world may be, the operational decision facing the investment committee is always the same i.e. “what do we do next given our understanding of current local linearities”. A detailed discussion of such procedures can be found in Dutton et al. (1997); here we shall content ourselves with an overview.

Generally, suppose that, following the notation of (2)

$$\underline{y}_t = H_t \left(\underline{f}_t, \underline{\varepsilon}_t \right) \quad (3)$$

Here H_t is a $(n \times 1)$ vector function of \underline{f}_t and $\underline{\varepsilon}_t$, where ε_t is some $(m \times 1)$ source of noise in the system.

By taking the usual Taylors-series expansions around zero vectors, we see that

$$\underline{y}_t = H_t(\underline{0}, \underline{0}) + H_{t1}(\underline{0}, \underline{0})\underline{f}_t + H_{t2}(\underline{0}, \underline{0})\underline{\varepsilon}_t + \text{Rem}_t \quad (4)$$

where $H_{t1}(\underline{0}, \underline{0})$ is the $(n \times k)$ matrix of partial-derivatives with respect to \underline{f}_t , $H_{t2}(\underline{0}, \underline{0})$, is the $(n \times m)$ matrix of partial derivatives with respect to $\underline{\varepsilon}_t$; both are evaluated at zero; and Rem_t corresponds to higher-order terms in the expansion. To see the links between (4) and (2) we equate β_t to $H_{t1}(\underline{0}, \underline{0})$ and similarly,

$$\underline{V}_t = H_t(\underline{0}, \underline{0}) + H_{t2}(\underline{0}, \underline{0})\varepsilon_t + \text{Rem}_t \quad (5)$$

It is immediate from the above that we can incorporate more complex schemes than (3) i.e. where y_t is *implicitly* connected to f_t .

Furthermore, we can, and probably would want to expand, the Taylor’s series around some point rather than $(\underline{0}, \underline{0})$.

It is worth enquiring as to the origins of (3), where should such a complex structure come from? The short answer is that if we start with data-generating processes that are, in the short run, exogenous to the market such as productivity, climate etc. and we know the utility/decisions functions of all individuals and organisations that participate in the model we can compute an equilibrium which will link return distribution parameters to utility characteristics. These will take the form of highly non-linear restrictions which can be substituted back into the data generating processes to arrive at an equilibrium pricing process, which is (3). Equation (3) will also encapsulate all the heterogeneous assessments of probability and behavioural quirks of the participants, not to mention the current state of institutions. An example of the above which is still highly simplified is Cox, Ingersoll and Ross (1985).

Inspecting the previous argument, we see that local linearity seems both reasonable and plausible except that:

1. The point about which the linearisation occurs will change with time.
2. The partial derivatives that become the coefficients in the linear system will change their values as the points in 1. change.

Using linear models successfully requires one to be aware of 1. and 2.. In particular, procedures that assume that coefficients in the linear factor model are fixed are likely to suffer from severe inaccuracies through time, unless the underlying structure is actually linear. Practitioners are aware of these difficulties and build their models using rolling windows; this means that models will change their linear coefficients through time. However, this will not always work. For example, virtually all linear factor models failed in the late nineties due to their inability to pick up the internet factor. With a rolling window of say seven years of monthly data, it will take several years for the internet factor to manifest itself in the model. Even when it does, the factor may have then ceased to be of any importance. The industry response to this has been to move to higher frequency data to make the models more responsive to evolutionary changes. This brings its own problems of increased sensitivity to noise and fails to accommodate the fact that these parameters may not vary smoothly over time, but exhibit jumps from one value to another.

Furthermore measuring at high frequency measures something different i.e. correlation of hourly data is *not* the same as correlation of monthly data because for example the hourly data will pick up the end of day effects where traders need to leave their books balanced over night. These dynamics are entirely absent from monthly data, hence we can see the technical issue which is that of temporal aggregation; it is not always possible to recover the density of daily prices by convoluting the density of hourly prices.

The approach that we shall advocate essentially allows the factor exposures to be recalculated every period while being agnostic as to how the recalculation is done. This accommodates all three types of assumption at the user's discretion, i.e.

- 1) Stationary parameters.
- 2) Smoothly varying parameters over time.

3) Discontinuous parameter behaviour.

In this way our model structure can be thought of as a way of summarising the conclusions about “the state of the market” at each moment in time, derived from a much wider set of possible model forms either explicitly quantitative in form or implicitly so having being constructed by a mixture of theoretical inference and quantitative support. In fact it can be shown how a range of modelling techniques that are often thought of as independent of each other actually lie on a spectrum of increasing complexity as your simplifying assumptions are relaxed from 1) to 3) above.

Some readers may at this point conclude that we are in some sense cheating in that our proposed model form throws no additional light on how the markets work. However our contention is that the only certainty in forecasting is change, and the only decision that is necessary to take at each rebalancing point is what to do next. From our perspective we wish to design an investment process that accommodates the former and illuminates the latter in a simple understandable manner. Hence we are agnostic as to which detailed model structure may best model market conditions over time, or even which model is currently performing best. What we aim to provide is a standard form that summarises (and allows the user to compose and/or merge) the conclusions from different models in a systematic and statistically rigorous manner while still being widely intelligible to all the participants in the investment decision i.e. our approach is a framework for comparing contrasting and communicating the results of other models rather than a model in its own right.

In its basic form, our approach results in a vector of relative preferences for each asset at each rebalancing period to feed into that period’s diversification calculation. In Section 4 and 5 we show how these relative preferences can be calculated using a rank score card. In the next section we will see how it can be extended to cover the risk input to these calculations as well as the return inputs.

3 A Mixture of Normals Approximates Risk

In this section we discuss return distributions. The usual assumption of multivariate normality for conditional or unconditional returns does not strike us as reasonable although it may hold at the level of well-diversified portfolios with a long investment horizon due to the operation of the central limit theorem. We

prefer to assume that conditional expected returns are multivariate normal, conditionally for each single period. We could weaken this assumption to allow returns to follow mixtures of multivariate normals. Thus we can structure the problem by assuming that there are S states of the world, and conditional upon each state we get a multivariate normal distribution but with different means and variances. Practitioners are probably aware of this framework without the necessary formal mathematics. A simple but important example might be where $S = 4$.

The four states being:

1. High alpha, high risk
2. High alpha, low risk
3. Low alpha, high risk
4. Low alpha, low risk.

Each such state would have a probability attached to it. One may be able to identify the states, as we could in the above example, or the states may be hidden, as in the examples of hidden Markov models, or as they are more frequently known, regime-switching models, see Hamilton (1989). Arguably in the above example, we are currently in a low alpha, low risk environment (regime 4). However, circa 2000-2002 we were in a low alpha, high risk environment (regime 3).

Formally, our returns model could be described by the following:

Suppose there are m regimes, $l = 1 \dots m$. For each regime, returns are described by the following probability density function (pdf).

$$pdf_l(x) \sim N(\mu_l, \Omega_l) \tag{6}$$

where μ_l and Ω_l are the mean vector and covariance matrix of returns in regime l and $N(\cdot)$ denotes multivariate normality.

For each regime l , there is a regime probability π_l such that $\sum_{l=1}^m \pi_l = 1$. The unconditional returns *pdf* follows a mixture of normals; symbolically,

$$pdf(x) = \sum_{l=1}^m \pi_l pdf_l(x). \tag{7}$$

Such a specification follows naturally in a Bayesian framework with multiple priors, in which case the prior and the posterior follow (7), assuming that the data are multivariate normal.

Another important property of mixture of normal and log-normal distributions is their ability to approximate, with very close accuracy, arbitrary return distributions.

This approximation property is widely used in the options pricing literature to construct estimates of risk-neutral distributions and also in the stochastic volatility literature where many of the joint densities whose formulation is not known in closed form are replaced by low-order mixtures of normals or log-normals.

From the perspective of a cross-sectional linear factor model regression, all of the above adds no complexity to the econometrics as long as we are implicitly or explicitly conditioning on the regime information.

Of course, if there are dynamics in the system, this regime information will itself be serially correlated as with GARCH or Kalman filter models, see Diderrich (1985). Conceptually this is little different to the regime switching models above except that the parameters are assumed to move smoothly through a series of states (or adjacent regimes), hence needing a larger number of such states to adequately represent a mixture of normals approximation to the risk. (In practice of course this does complicate the parameter estimation as the quantity of data for each state is reduced but that is a practical consideration outside the scope of this paper.)

Last but by no means least, we started the discussion in this paper talking about the importance of one off events and the difficulties that these presented to frequentist modelers. In the discussion so far we have in the main discussed a model evolving through time hence generating a mixture of normals. An alternative view is where we may have a number of candidate models but not be confident which applies at any moment in time. If we give each alternative a degree of belief, then we are again in a mixture of normals environment. This also provides an entry into the world of Bayesian statistics where we update our model not just over time, but from a parsimonious prior to our posterior probability in a cumulative manner as we successively take into account each piece of information (subjective or objective) we hold together with our confidence in that item of information. In this case, we would arrive at equation (7).

From this we can see that the basic mixture of normals formulation provides a standard mathematical form that is compatible with a range of underlying risk models. In particular it allows us to systematically include subjective information in our analysis and merge it with all our information from other sources. In practical terms, we might have a precise scorecard based on our subjective information which is updated by detailed model information where it is present, and in proportion to the confidence in that model.

Hence we have achieved for our risk input a similar generalised form which can capture the conclusions on risk from a range of underlying models while being agnostic about each particular source. Unfortunately the standard form in its most general incarnation is not designed for manual interpretation in that multivariate normal distributions are intimidating when presented in isolation, when presented as a mixture, simply interpreting the numbers to identify the detailed behaviour implied therein becomes unrealistic. Fortunately because the underlying structure is conceptually simple, flexible, and easy to describe, computer aided tools can come to our aid by letting the user work with the key investment interpretations while hiding the necessary but confusing detail. This is after all routine algebra that can be safely taken as read.

The main advantage of this unifying form is that it allows us to see the strengths and weaknesses of different approaches not as unconnected alternatives but as part of a spectrum with properties that vary smoothly as we move along it and where we can change our position on this spectrum as our circumstances change over time without needing to radically change our underlying philosophy or support tools.

In practice we will often adopt some simplifying assumptions to radically simplify this presentation. E.g we might choose to model the four state model above by having four rank scorecards one for each scenario, give each a probability, and assume that the covariance structure is the same for each. We can then add increasing degrees of sophistication when and only when they are needed to adequately reflect important detail in the underlying problem.

4 Information, Quality and Model Resolution

In this section we focus on the components of the linear factor model described in equation (1) and ask ourselves which elements we are likely to have useful information about so that we can construct a working model. The answer depends upon who we are. Considering the three entities in turn we have the exposures, we have the parameters of the factor distribution, and we have the volatilities of the error terms.

From the perspective of institutional investors it seems to us that it is the exposures that are of the greatest importance, but even this answer could be conditioned on the choice of factors.

All of these factors are likely to vary by sectors, regions, styles. For institutional stock selection models, regions and sectors are probably of the greatest importance, followed by styles. The macroeconomic factors tends to be of greater interest at the asset class level and also to regulators and academics. The poor quality of macroeconomic data, and its low frequency, tend to make it less useful for monthly or weekly stock returns. However, most news that impacts upon analysts and strategists is in fact macroeconomic which may not be easily quantified but sensitivity of asset returns to those inputs might be and so, in an ideal world, we would like to be able to merge this qualitative information with a quantitative model. We also need to note that many factors are only relevant at certain times and so we need the ability to “switch” factors on and off and to consider them in a range of combinations and proportions. For example, it is conventional wisdom that momentum factors are more closely watched by the market during Bull markets, and quality factors are more important in bear phases of the market cycle. This leads us to the cross sectional form of our linear factor model as described in Section 5 which allows the user to enter projected macroeconomic scenarios or one-off events as well as allowing the user to explore the effect of changing individual factor relevance over time,

We now consider sectors and regions as our factor choice and ask how we might be able to present information about exposures. One possible solution might be to assign a value of one for a stock in a particular region or sector, or zero if it is not in that region or sector. Variations of this methodology are quite widespread, see for example Heston, Rowenhorst and Wessels (1995). The difficulty with this is that it may not necessarily capture all country or regional effects, since for example, a particular company in the UK may buy its inputs in Australia and sell its products in France. Its exposure to both these countries would be non-zero. Again, region and sector returns data can be contaminated by the fact that many companies have multi-national and multi-sectoral activities. Notwithstanding the comments above, most institutional investment companies organize their research around countries and sectors, sometimes sectors within countries. In the next section we show how some of these problems can be mitigated by introducing more structure into our factor assumptions. For example, by having separate terms for costs and sales we can assign different sensitivities to Australian wage inflation and French GDP growth for the same stock. There have been attempts to purge regions of their global content but these methods seem computationally

demanding; region and sector definitions remain a substantial problem. In, the next section we propose a model structure where the user can vary the region and sector aggregation level and detailed definition to reflect the quality of information available and the underlying dynamics being modelled.

Similarly, styles, such as value and growth, are clearly factors of great importance to investors and computing exposures to these styles can be done in a number of ways, the detailed discussion of which is beyond the scope of this paper. All we need to observe here is that style and sector can be used interchangeably in linear factor models. So all our observations on sectors apply equally to style.

Summarizing the above discussion, we see that the structure of research and the flow of information in institutional investment companies is focused on regional sector and style research and responds to a mixture of macroeconomic information and company accounting data. The challenge we shall address in Section 5 is how to construct methodologies that merge this diagnostic information in a consistent and easily understood manner.

There are deeper issues to do with information that may be relevant to this discussion. These focus on where information is produced/processed within financial organisations. It is not possible to deal in detail with this vast topic but we will look at a related topic, which is the quality of information.

In the world of the full time series model, we have histories to all our inputs which are either asset and factor returns or asset returns and factor exposures. Typically this will involve five or more years of monthly data or a shorter amount of weekly data. What is required for the above is a certain constancy of data definition. However, such models will suffer greatly, whenever there is a reform in data definition. Unfortunately these revisions are all too frequent, for example; we are about to experience a huge change in accounting information, known as International Financial Reporting Standards (IFRS). In 2005, 7000 EU companies are expected to comply with IFRS; this will change in profound ways such hoary old chestnuts as earnings - revisions and book values. Time series models will have major problems with this; cross-sectional models should be able to deal with these problems relatively easily. Our format allows one to combine the best of both worlds by using time series models to calculate exposures where this can be done reliably and fill in the rest from the cross sectional results where needed.

Even where the definitions are stable the accuracy of exposures may be questionable. It is an interesting feature of various proprietary stock-selection models

that exposures are computed at an integer level, typically values between 3 and -3 . The actual factors are not explicitly known but their returns are computed by cross-sectional regression for each period. Likewise, hedonic models use firm specific characteristics, typically based on accounting information, to implicitly specify factors, the factor returns being again estimated by similar methods. The characteristics are often ranked, or normalized to have within group mean of zero and standard deviation of 1 at each point in time, where the group is often a region or sector. The key point that comes out of these practises is that cross-sectional models seem to work without requiring a precision or accurate representation of the exposures.

To take this last point forward more formally consider at time t the vector y_t and the columns of β_t as defined in equation (2). We can number them as $\beta_{1t}, \beta_{2t}, \dots, \beta_{kt}$ and i is a vector of ones. We can then as a purely theoretical exercise consider the best linear prediction (*blp*) of y_t on the columns of β_t and i . This is in effect a calculation determined by the joint cross-sectional distributions of the exposures and (the asset) returns. Changes to β_t via approximations, smoothings and rankings will change the coefficients of the *blp* and the residuals. Non-singular transformations of the columns of β_t and i will leave the residuals unchanged and the forecast of y_t unchanged.

Formally, we can compute “risk-premia” as implied factor returns $\lambda_{0t}, \lambda_{1t}, \dots, \lambda_{kt}$ by the *blp*.

$$y_{it} = \lambda_{0t} + \sum_{j=1}^k \beta_{ijt} \lambda_{jt} + \varepsilon_{it} \quad i = 1, N \quad (8)$$

$$\text{where } \sum_{i=1}^N \varepsilon_{it} \beta_{ijt} = 0 \text{ for } j = 1, \dots, k$$

$$\text{and } \sum_{i=1}^N \varepsilon_{it} = 0$$

Our forecast of the mean of y_{it} is $y_{it} - \varepsilon_{it}$.

Adequate forecasts for y_{it} seem to follow from robust choices of β_{ijt} . Overspecification of β_{ijt} seems to lead to poor forecast measurement, perhaps due to the measurement of noise rather than signal when calculating the exposures. Hence we see that ranking, n-tile, or simple scores can be seen as an effective way of reflecting the quality of information in a measure and avoiding over fitting or spurious accuracy.

However practical they may prove in day to day application, this diversity of scoring conventions complicates model comparison and the choice of theoretical justification can be somewhat arbitrary. Simple ranking of our assets is not perfect either. As well as the possibility of spurious accuracy discussed above, it can distort the information that we do have by forcing us to distinguish between assets that may be essentially identical in the context under consideration. Equally, it gives us no way of capturing the information that there may be much bigger differences between some assets than others.

Both of these problems can be addressed by choosing to rank groups of assets rather than single assets. If these groups are allowed to have an arbitrary number of members (including zero members), then we can smoothly vary the precision of our description from simple binary scores through to near linearity. The choice of resolution is chosen with a consistent justification linked to the quality of the information contained in each model term. We refer to this scoring convention as stylised ranking.

5 Our Scorecard Format

So far we have argued that a cross sectional linear factor model with coefficients represented in a stylised ranking form provides the simplest and most versatile way of representing the user's views on short term investment return and its drivers. In order to complete the derivation of the details of our rank scorecard format we need to now discuss how our format is affected by our assumptions about the chosen asset set.

A typical use of our scorecard format is to calculate return expectations across a global country sector portfolio (i.e. where our asset set consists of all possible permutations of countries and sectors) so we will couch our discussion in these terms even though it could equally apply to a sector versus style or a region versus asset class analysis.

If we had to calculate a beta for every country sector and factor permutation, then compared to a simple regional (or sector) split, this structure would vastly expanded the number of coefficients that we are trying to calculate in our model with all of the dangers of over fitting and lack of clarity that that brings with it. To restore some parsimony to our model we restrict ourselves to considering only betas that are independent of each other (i.e. the same sector has the same beta

in all regions) or where the ratio of the betas is independent of each other (i.e. the overall beta is the product or interaction of a country term and a sector term which are independent of each other).

The logic of allowing this type of interactive structure is perhaps most clearly seen in macroeconomic analysis. Here a macro-factor like consumer spending varies with geography, while the sensitivity to that variable varies by sector. E.g. the consumer discretionary sector is likely to be much more sensitive to variation in consumer spending than is the utility sector. Hence the product of these two terms allows an approximation to be made for a factor beta for each country sector permutation while still requiring a small number of user inputs.

Of course some factors vary independently by sector or region rather than interacting as discussed above. However from a purely presentational point of view these are a special case of interaction where one of the dimensions, either region or sector, has zero variation so can still be captured in the same format scorecard.

The assumption of independence of beta ratios across region and sector may be seen as a limitation, however. In the same way that we argued that the relative importance of factors varies with time, and that quality of informational content varies between those factors, we now argue that choice of appropriate aggregation structure is also factor- dependant. Hence we argue that the aggregation structure of the model should be varied until the independence assumption is an acceptable approximation.

An example of this issue could be in the industrial sectors where aggregation is traditionally based on the type of production skills needed. For example, aircraft and washing machines both use engineering and assembly skills and so are often put together in the same broad sector. For a cost shock such as engineering labour wage inflation, this might be quite appropriate, but the markets for their products are totally dissimilar. Here the supply side is common but the demand side is different. For a shock to consumer spending or defence procurement aggregating them together is entirely inappropriate. Similarly, factor models will often have an interest rate sensitivity term, but what do you do when US interest rates are moving in a different direction to European rates? In our model, the answer is simple, you disaggregate the factors and each asset then has a beta to US rates and another to European rates.

In this way we have arrived at all the main features of our scorecard format.

It is a cross-sectional linear factor model with betas represented in a stylised ranking form where the ratio of the factor betas across sectors are assumed to be independent of region and vice versa.

More formally, the details of our Scoring System are described as follows.

We slightly alter equation (1); we now write,

$$y_{it} = \alpha_i + \sum_{j=1}^k \beta_{ijt} f_{jt} + V_{it}, \quad (9)$$

where “ y_{it} ” is the current return, “ α_i ” is the long term average return for each asset “ i ”. “ β_{ijt} ” is the sensitivity of that asset return to the factor “ f_j ” and “ V_{it} ” is a noise or disequilibrium term. Our objective is to construct asset (or stock) forecasts based on $\alpha_i + \sum \beta_{ijt} \lambda_{jt}$ as in (8).

In our suggested model all the individual assets are classified by sector and country, and ranking them over these dimensions captures our description of how those betas coefficients vary across both dimensions. Hence we can expand equation (9) above by indexing over “ p ” and “ q ” to re-create the conventional index “ i ”. This then allows us to write

$$\beta_{ijt} = s_{pjt} g_{qjt} \quad (10)$$

and hence substituting into (9) and dropping time subscripts for Betas.

$$y_{it} = \alpha_i + \sum_j^n \sum_{p,q} s_{pj} g_{qj} f_j + V_{it} \quad (11)$$

Where “ p ” is the sector index, “ q ” is the region index, and “ j ” is the factor index. “ s ” and “ g ” are the sector and geographic terms for each factor. “ n ” is the number of factors.

As we are interested in relative performance over our investment horizon, we can drop the constant term “ α_i ” and the unknown noise term, “ V_{it} ” from our return calculation, although of course the magnitude and correlation of the latter will figure prominently in our later risk calculations.

The terms $(s_{pj} g_{qj} f_j)$ above can now be conveniently collected together in the matrix [S:G:f]. This matrix describes the relative importance of each factor, and a first order approximation to how the sensitivity to this factor varies by geography and sector.

At this point in our derivation [S:G:f] is comprised of linear coefficients, however as explained earlier, these can now be approximated to any appropriate level of

precision using a stylised ranking convention combined with a mapping from score space back into coefficient space.

While this mapping can be described in a number of ways, we have chosen a simple convention where the coefficient values of the maximum and minimum ranks are provided so that all the intermediate coefficients can be recovered by interpolation. As we are applying our stylised ranking to both sectors and regions for each factor we can construct a (4 by “n”) matrix “M”.

In practice, the factor sector and region terms are multiplied together. Hence the combined scaling implicit in this can, as a matter of convenience, be left to the final mapping between scores and return expectations described in equation (12). In this case, the coefficients chosen for “M” will usually be a zero to one range, or a simple permutation on it such as minus one to plus one. By normalising the scores in this way, the investment interpretation is clear without requiring the user to keep track of numerous intermediate-scaling factors. However, by retaining the full “M” matrix in the rank scorecard format, users can choose to use a more detailed convention if for instance they want to compare intermediate variables with a standard linear model.

We can now collect “M” together with the ranked version of [S:G:f] to give our full rank scorecard format $[\widehat{S} : \widehat{G} : M : \widehat{f}]$

While the intermediate calculation of return scores for each permutation of region and sector was driven by the need to cater for interaction between region and sector effects, it has a very useful additional consequence. Viz when we re-aggregate our country and sector views, because we know the weights of each country sector combination, we can ensure that our aggregate country and sector views are consistent by construction.

6 Independence of Factors

Ex post we can fill in our scorecard from achieved relative performance using simple linear regression. When we do this, we have the problem that our factors are not always independent of each other. In these circumstances if we do a stepwise regression the fitted value of the coefficients will depend on the order in which the different terms are introduced.

Suppose two factors are highly correlated. In this case whichever explanatory factor is introduced first will pick up most of the variation that is associated with

both of them. Hence the matrix of partial derivatives is not a unique set of values even for a known fixed set of relative returns. Ex ante the same problem manifests itself in a similar way in that if you enter high scores for two correlated factors you are in effect double counting that score.

There are a number of palliative actions which can be adopted to minimize these problems. These range from careful selection of your factor set to be as independent as possible through to the formulation of Grinold and Kahn (“Active Portfolio Management” second edition page 263) where you assume that you have a knowledge of the full covariance between forecast and returns. As follows

$$E(y|g) = E(y) + Cov(y, g). Var^{-1}(g)(g - E(g)) \quad (12)$$

Where g is the vector of raw forecast scores and y is the return vector. In words this is saying that our expected return given our scores is consensus expected return plus a scaled term proportional to the deviation of our score from its consensus value. The scaling term being derived from the covariance of observed return with the forecast scores.

A detailed discussion of how best to minimize problems of correlated factors in linear factor models is outside the scope of this paper. However having demonstrated the equivalence of our rank scorecard formulation with a standard linear factor model, all of the recognized techniques can be applied as appropriate.

7 Conclusions

In this paper we have assumed that an ever-changing kaleidoscope of factors and one off events drives equity returns. Hence the forecasting of these returns must involve a mixture of objective analysis for those aspects that are amenable to quantitative modelling combined with theory, experience and judgement for those aspects that are not.

This perspective has led us to focus not on an individual fixed model of return, but on how to design an investment process or modelling framework in order to track this evolving problem most effectively.

From this perspective, the key characteristics of this framework are flexibility to accommodate different model forms, robustness to noise in the data, simplicity and parsimony and an intuitive accessible reporting format to aid comprehension and communication of the results, combined with theoretical rigour to maximise

confidence in the output from the process. In this context we have discussed four main ideas in this paper.

1. The role of cross sectional linear factor models as a standard intermediate form (or scorecard) that allows presentation discussion and merging of information from a variety of different sources, without needing to introduce stationarity as a simplifying assumption.
2. The mixture of normals approach to approximating return distributions of arbitrary form, and how this allows cross sectional linear factor models to be constructed that are consistent with a range of more complex Bayesian (e.g. Black Litterman, Scenario or subjective models) or other non stationary time series models (Kalman filters, GARCH or Hidden Markov models etc) in order to present a reduced form relevant to the operational decision of “what do we do next”.
3. Ranking or binning of data to reflect the quality of the underlying information
4. Construction of hierarchies in region, sector, style and factor aggregation structure to allow the user to select the most relevant level of resolution in both value and scope for each factor.

These ideas have then been brought together in the form of an innovative rank score card approach that exploits the fact that the sensitivities to each factor may themselves be approximated as a linear combination of sector style and regional terms which specifically avoids the need for explicit return forecast. This combination of hierarchical presentation, and implicit factor construction combine to allow a highly parsimonious practical representation of the information available at a point in time without needing to make unrealistic assumptions about wider market behaviour, or limiting the user’s ability to adopt more detailed structure where this is called for.

This cross sectional rank scorecard form then allows us to build sector style and region forecasts that are inherently consistent with each other by construction as they are derived by aggregation of our disaggregate scores. This seems to us, at least, to be a highly desirable feature particularly for large organisations where inconsistencies in investment policy often reflect communication problems rather

than intellectual differences and where the lack of consistency can result in reduced investment performance at higher cost.

This work complements the forecast structure work of Sefton, Bulsing and Scowcroft (2002) by including the aspects of non parametric representation of data and the need for simple intuitive reduced form models as a key component of an operational forecasting and diversification process. It also broadens the range of model forms that can be represented within the framework.

What it does not do is to provide any “miracles” in the continual search for predictability and understanding of market behaviour. The efficient market theory tells us that this will always be a confusing non-stationary and non-linear world. However by proposing a simple understandable and statistically rigorous framework in which we can organise our thoughts, while avoiding spurious accuracy and implausible simplifying assumptions, it does increase the chance that we will be able to apply our experience and judgement to add value to the investment process.

References

- [1] Cox, J.C., J.E. Ingersoll and S.A. Ross (1985) A Theory of the Term Structure of Interest Rates, *Econometrica* 53, 385-407.
- [2] Diderrich, G.T. (1985) The Kalman Filter from the Perspective of Goldberger-Theil Estimators. *The American Statistician* 39, No. 3
- [3] Dutton, K., S. Thompson and B. Barraclough (1997) *The Art of Control Engineering*, Prentice Hall ISBN 0-201-17545-2.
- [4] Hamilton, J.D. (1989) Analysis of Time Series Subject to Changes in Regimes. *Journal of Econometrics* 45, pp. 39-70.
- [5] Heston, S.L., K.G. Rowenhorst and R.E. Wessels (1995) The Structure of International Stock Returns and the Integration of Capital Markets. *Journal of Empirical Finance* Vol. 2, pp. 173-197.
- [6] Satchell, S., S. Wright , S. Hwang (2003) Assessing the Merits of Rank-based Optimisation for Portfolio Construction , in S. Satchell and A. Scowcroft eds., *New Advances in Portfolio Construction and Implementation*, Butterworth-Heinemann, London.
- [7] Sengupta, J.K. and P. Fanchon (1997) *Control Theory Methods in Economics*, Kluwer Academic Publishers ISBN 0-7923-9845-9.
- [8] Sefton, J, A. Scowcroft and M. Bulsing (2002) “On the Nature and application of equity return forecasts” UBS Warburg Equity research note October 2002
- [9] Volteenalno, T. (2002). “What Drives Firm-Level Stock Returns? *Journal of Finance* 57(1), February, 2002.