Factor Representing Portfolios in Large Asset Markets[1]

Enrique Sentana

(CEMFI)

Revised: January 2002

[1]This is an extensively revised version of Sentana (1992). Financial support for that version from the LSE Financial Markets Group as part of the ESRC project "The Efficiency and Regulation of Financial Markets" is gratefully acknowledged. The author would like to thank Manuel Arellano, Gary Chamberlain, Antonis Demos, Steve Durlauf, Gabriele Fiorentini, Andrew Harvey, Jan Magnus, Bahram Pesaran, Paolo Zaffaroni and seminar participants at the XXme Encontre France-Belge de Statisticiens, CEMFI, the Forecasting Financial Markets/Computational Finance 2000 Conference, IES (Oxford), LBS, LSE, and Universidad Carlos III (Madrid) for very useful comments and discussions. The suggestions of two anonymous referees and Bas Werker have also greatly improved the exposition. Special thanks are due to Greg Connor for his invaluable advice. Of course, the usual caveat applies. Address for correspondence: CEMFI, Casado del Alisal 5, 28014 Madrid, Spain (tel.: +34 91 429 0551, fax: +34 91 429 1056, e-mail: sentana@cemfi.es).

Abstract

We study the properties of mimicking portfolios in an intertemporal APT model, in which the conditional mean and covariance matrix of returns vary in an interdependent manner. We use a signal extraction approach, and relate the efficiency of (possibly) dynamic basis portfolios to mean square error minimisation. We prove that many portfolios converge to the factors as the number of assets increases, but show that the conditional Kalman filter portfolios are the ones with both minimum tracking error variability, and maximum correlation with the common factors. We also show that our conclusions are unlikely to change when using parameter estimates.

Keywords: Factor Models, Basis Portfolios, APT, Intertemporal Asset Pricing, Kalman Filter.

JEL: G11, C32

1 Introduction

There is a long tradition of factor or multi-index models in finance, where they were originally introduced to simplify the computation of the covariance matrix of returns in a mean-variance portfolio allocation framework. In this context, the common factors usually correspond to unanticipated innovations in observable economic variables, or unobserved fundamental influences on returns. In addition, the concept of factors plays a crucial role in two major asset pricing theories: the mutual fund separation theory (see e.g. Ross, 1978), of which the standard CAPM is a special case, and the Arbitrage Pricing Theory (see Ross (1976), and Connor (1984) for a unifying approach). In theoretical asset pricing models with a countably infinite collection of primitive assets, it is possible to mimic perfectly the behaviour of the common factors by means of risky, well diversified, limit portfolios (see e.g. Chamberlain and Rothschild (1983), Connor (1984), Huberman (1982), Ingersoll (1984) or Stambaugh (1983)). In these models, agents trade costlessly, so the relative merits of different methods to approximate the common factors are not at stake. However, in empirical applications, such as tests of the model restrictions, asset allocation, hedging decisions, or portfolio performance evaluation, only data on a finite number of assets are available, and, hence, the factors have to be proxied by basis portfolios obtained from the collection of asset at hand. This would also be true in an economy with a finite number of securities, as the available data would never include all existing assets (cf. Roll, 1977).

This problem received some attention in the finance literature a few years ago, and several methods were proposed to construct factor representing portfolios with time-invariant weightings that yield consistent estimates of the common factors as the number of assets increases (see e.g. Chamberlain and Rothschild (1983), Chen (1983), Connor and Korajczyk (1988), Grinblatt and Titman (1987), Ingersoll (1984), Huberman et al. (1987) and Lehmann and Modest (1985)). However, two important issues have not yet been fully investigated:

1. There has been very little discussion on the efficiency of the different procedures; most of the existing results are only concerned with consistency.
2. The methods proposed thus far only consider passive (i.e. static) portfolios, as opposed to active (i.e. dynamic) investment strategies, which would use the available information at the time agents' decisions are taken to form portfolios.

In traditional empirical applications of static asset pricing models, one can ignore these issues at little cost because (i) a very large collection of assets is typically used, and (ii) conditioning information plays no effective role (see e.g. Lehmann and Modest, 1988). In contrast, both these points are particularly important in the rapidly expanding empirical literature modelling the time-variation in the means, variances and covariances of financial assets (see Bollerslev et al. (1992) for a survey). The number of securities considered in this work has typically been small, very much smaller than in the more traditional approach, and often, the applications have been explicitly related to dynamic asset pricing models (see e.g. King et al., 1994).

The purpose of this paper is to fill in this gap in the literature by analysing the statistical properties of alternative ways of creating actively and passively managed mimicking portfolios from a finite number of assets in the context of the dynamic version of the APT developed in King et al. (1994). In this model, the actions of the agents are based on the distribution of returns conditional on their time-varying information set. As a result, both the conditional mean and covariance matrix of asset returns change through time, and furthermore, the former is closely linked to the latter.[1]

[1] In the context of bidimensional stochastic processes (or random fields), their asset pricing

The class of factor scores that we consider is motivated by practical situations in which an agent would like to form portfolios from the finite collection of asset at hand in period $t-1$, whose payoffs in period t closely replicate the factors. Asset allocations that "track" a particular factor, or "basket" hedging decisions constitute obvious examples. But since any conceivable factor representing portfolio will be a linear combination of the assets with (possibly) time-varying stochastic weights, significant improvements may be obtained by considering the conditional distribution of returns when forming mimicking portfolios. For this reason, we explicitly analyse the class of dynamic basis portfolios whose weights depend on the conditioning information. Given that all basis portfolios proposed so far employ constant weightings, such a novel class includes them as special cases, but is far more general. Importantly, our results do not depend on any specific distributional assumptions, although they require correctly specified first and second conditional moments.

From the econometrician's point of view, the factors are effectively unknown. And although they could be regarded as a set of parameters in any given realisation of the asset returns, it is more appropriate to use a signal extraction approach in our context, because the factors can take different values in different realisations. In this respect, note that since we use portfolios to estimate the underlying factors, we implicitly restrict ourselves to the class of conditionally linear predictors with time-varying stochastic weightings, and exclude any non-linear filters and smoothers. A significant advantage of our signal extraction framework is that we can assess the efficiency of the different basis portfolios in terms of their mean square error (MSE), the standard statistical decision theory criterion. As we shall see, the MSE criterion can also be given an intuitive justification in terms of an

approach could also be interpreted as using economic theory to impose restrictions on the time series dependence of asset returns from restrictions on its conditional cross-sectional correlation.

investor with mean-variance preferences, as it corresponds to the so-called "tracking error" variability in the finance literature. In addition, we also investigate the correlation of the different replicating portfolios with the common factors.

The paper is divided as follows. In Section 2 we discuss our dynamic asset pricing restrictions in the theoretical framework employed by King et al. (1994). We also define the factor representing limit portfolios, and introduce a dynamic specification that is compatible with an unconditional factor structure for the innovations in returns, so that the usual passive mimicking portfolio strategies are not meaningless. The statistical properties of the basis portfolios are discussed at length in Section 3. In particular, after introducing them, we study their behaviour both in finite samples and asymptotically, compare their relative efficiency, and analyse the effects of estimation error in the model parameters. This section also contains a small illustrative example that evaluates the different expressions in realistic parameter configurations. Finally, Section 4 concludes. Proofs and formal derivations are gathered in appendices.

2 A Conditional Factor Model of Asset Returns and Risk Premia

2.1 Conditional moments specification

We base our analysis in an economy with a countably infinite collection of primitive assets, whose payoffs are defined on an underlying probability space. Let R_{it} $(i = 1, 2, \ldots, N, \ldots)$ be the random (gross) holding return from a unit investment in asset i during period t, and let R_{0t} be the (gross) holding return on an asset that is riskless from the agents' point of view at period $t-1$. We work in terms of the conditional distribution of period t asset returns, and assume that the agents' information set, Φ_{t-1}, contains at least the past values of all asset returns. Let L_t^2 denote the collection of all random variables defined on the underlying probability space which are measurable with respect to Φ_t and have finite conditional second moments. Hansen and Richard (1987) show that L_t^2 is the conditional analogue of a Hilbert space under the conditional mean square (MS) inner product and the associated MS norm. We assume that, conditional on Φ_{t-1}, the second moments of the primitive assets are uniformly bounded with probability one (a.s.), so that a fortiori $R_{it} \in L_t^2 \ \forall i$.[2] In addition, we also assume that the minimum eigenvalue of the conditional covariance matrix of any sequence of risky asset returns is uniformly bounded away from zero a.s., so that no primitive risky asset is redundant.

Let u_{it} be the unanticipated component of returns on the i^{th} risky asset, i.e. $u_{it} = R_{it} - \nu_{it}$, where $E(R_{it}|\Phi_{t-1}) = \nu_{it}$, and let $\sigma_{ijt} = E(u_{it}u_{jt}|\Phi_{t-1}) = cov(R_{it}, R_{jt}|\Phi_{t-1})$. We assume that the stochastic structure of *unanticipated* returns is given by:

$$u_{it} = \beta_{i1t}f_{1t} + \ldots + \beta_{ikt}f_{kt} + \varepsilon_{it} \qquad (i = 1, 2, \ldots, N, \ldots) \tag{1}$$

where f_{jt} ($j = 1, \ldots, k$ finite) are common factors that capture economy-wide (i.e. systematic) shocks affecting all assets, $\beta_{ijt} \in \Phi_{t-1}$ ($i = 1, 2, \ldots, N, \ldots; j = 1, \ldots, k$) are the associated factor loadings or betas known in $t-1$, which measure the sensitivity of the assets to the common factors, while ε_{it} are idiosyncratic terms reflecting risks specific to asset i, which by definition are (conditionally) orthogonal to $\mathbf{f}_t$, with $[\mathbf{f}_t]_j = f_{jt}$ ($j = 1, \ldots, k$). To guarantee that common and specific factors are innovations, we assume that they are unpredictable on the basis of Φ_{t-1}. We also assume (without loss of generality at the theoretical level) that the common factors are conditionally orthogonal, and that they have (possibly) time-varying conditional variances, λ_{jt} ($j = 1, \ldots, k$). In this framework, any

[2] The uniformly boundedness condition, i.e. $\sup_{\Phi_{t-1}} E(R_{it}^2|\Phi_{t-1}) < \infty$ a.s., is stronger than required, but simplifies the subsequent discussion. In particular, it implies that $E(R_{it}^2) < \infty$, and hence, that $R_{it} \in L^2$ (cf. Zaffaroni, 2000).

collection of N risky asset returns can be represented as:

$$\mathbf{R}_{Nt} = \boldsymbol{\nu}_{Nt} + \mathbf{u}_{Nt} = \boldsymbol{\nu}_{Nt} + \mathbf{B}_{Nt}\mathbf{f}_t + \boldsymbol{\varepsilon}_{Nt} \tag{2}$$

with $[\mathbf{R}_{Nt}]_i = R_{it}$, $[\boldsymbol{\nu}_{Nt}]_i = \nu_{it}$, $[\mathbf{u}_{Nt}]_i = u_{it}$, $[\boldsymbol{\varepsilon}_{Nt}]_i = \varepsilon_{it}$ and $[\mathbf{B}_{Nt}]_{ij} = \beta_{ijt}$, $(i = 1, 2, \ldots, N, \ldots; j = 1, \ldots, k)$. Let $\boldsymbol{\Sigma}_{Nt}$, $\boldsymbol{\Lambda}_t$ and $\boldsymbol{\Gamma}_{Nt}$ denote the conditional covariance matrices of $\mathbf{R}_{Nt}$ (and $\mathbf{u}_{Nt}$), $\mathbf{f}_t$ and $\boldsymbol{\varepsilon}_{Nt}$ respectively, with $[\boldsymbol{\Sigma}_{Nt}]_{ij} = \sigma_{ijt}$, $\boldsymbol{\Lambda}_t = diag(\lambda_{1t}, \ldots, \lambda_{kt})$ and $[\boldsymbol{\Gamma}_{Nt}]_{ij} = \gamma_{ijt}$. In what follows, we assume for simplicity that rank$(\mathbf{B}_{Nt}) = k$, that each λ_{jt} is uniformly bounded away from zero a.s., and that the same is true of the minimum eigenvalue of $\boldsymbol{\Gamma}_{Nt}$, so that it is not possible to form finite portfolios that contain only factor risk. As a consequence, $\boldsymbol{\Sigma}_{Nt}$ can be written as the sum of two parts: one which is common but of reduced rank k, $\mathbf{B}_{Nt}\boldsymbol{\Lambda}_t\mathbf{B}'_{Nt}$, and one which is specific, $\boldsymbol{\Gamma}_{Nt}$. In order to differentiate one from the other, we follow Chamberlain and Rothschild (1983), and assume that the largest eigenvalue of $\boldsymbol{\Gamma}_{Nt}$ remains uniformly bounded as N increases (as in band-diagonal matrices with uniformly bounded individual elements). We also assume that as $N \to \infty$ this matrix does not become singular, and that all the eigenvalues of the $k \times k$ matrix $\mathbf{B}'_{Nt}\mathbf{B}_{Nt}$ become unbounded, so that the common factors are indeed pervasive, in the sense that they affect most assets in the economy. Notice that (1) is then a statement about the cross-sectional dependence of asset returns, as it implies that k (and only k) eigenvalues of $\boldsymbol{\Sigma}_{Nt}$ grow unboundedly with N (see Chamberlain and Rothschild, 1983).

2.2 Modelling risk premia

Let $\overline{\mathcal{P}}_t$ be the conditional closure of the set of payoffs from all possible portfolios of the primitive assets. It is clear that $\overline{\mathcal{P}}_t$ is a conditionally closed linear subspace of L^2_t, and hence also a conditional analogue to a Hilbert space under the (conditional) MS inner product. In order to model conditional mean returns, we shall use the cost functional $C(.)$, which can be regarded as a linear pricing functional defined on $\overline{\mathcal{P}}_t$ that maps elements of this space onto the information set, Φ_{t-1}. Under a mild no arbitrage condition (see Chamberlain and Rothschild, 1983, or Hansen and Richard, 1987), this functional is conditionally continuous on $\overline{\mathcal{P}}_t$, and extends to limit portfolios the usual definition of cost as the sum of the weights of portfolios of a finite number of primitive assets. Then, a conditional version of the Riesz representation theorem implies that there is a portfolio with payoff $p_{ct} \in \overline{\mathcal{P}}_t$, such that $C(p_t) = E(p_{ct}p_t|\Phi_{t-1})$ for all $p_t \in \overline{\mathcal{P}}_t$, so that p_{ct} represents $C(.)$ in $\overline{\mathcal{P}}_t$ (see Hansen and Richard, 1987). We can interpret p_{ct} as a stochastic discount factor that prices the available portfolios by discounting their uncertain payoffs across different states of the world. We shall also use the conditional mean functional, $E(.|\Phi_{t-1})$, which is always conditionally continuous on $\overline{\mathcal{P}}_t$, and therefore can also be represented by a portfolio with payoff $p_{mt} \in \overline{\mathcal{P}}_t$, so that $E(p_t|\Phi_{t-1}) = E(p_{mt}p_t|\Phi_{t-1})$ for all $p_t \in \overline{\mathcal{P}}_t$. In this respect, we assume that there exists a unique unit-cost, riskless limit portfolio of risky assets, whose (gross) return is equal to ν_{0t} in the conditional MS norm, so that $p_{mt} = R_{0t}/\nu_{0t}$ (see Chamberlain and Rothschild, 1983, for necessary and sufficient conditions).

Let $R_{ct} = p_{ct}/C(p_{ct})$ be the return on a unit investment on the pricing representing portfolio, and define p_{at} as an arbitrage (i.e. zero-cost) portfolio conditionally orthogonal to R_{ct} with payoffs $p_{mt} - C(p_{mt})R_{ct}$. Then, if we rule out risk neutral pricing, so that not all conditional expected returns are equal, it is possible to prove that all the portfolios on the *zero-cost* conditional mean variance (MV) frontier generated from *all* primitive assets will be spanned by p_{at}, and that all those on the *unit-cost* conditional MV frontier generated from *all* primitive assets are spanned by R_{ct} and p_{at} as

$$R_{et}(\bar{\nu}_t) = R_{ct} + \frac{\bar{\nu}_t - E(R_{ct}|\Phi_{t-1})}{E(p_{at}|\Phi_{t-1})}p_{at},$$

where $\bar{\nu}_t = E(R_{et}(\bar{\nu}_t)|\Phi_{t-1})$ (see Hansen and Richard, 1987). The simplest arbitrage portfolios are those that hold a unit long position in each primitive asset and an equivalent short position in the riskless asset, so that their payoff vector $\mathbf{r}_{Nt} = \mathbf{R}_{Nt} - R_{0t}\boldsymbol{\iota}_N$, where $\boldsymbol{\iota}_N$ is a vector of N ones and $[\mathbf{r}_{Nt}]_i = r_{it}$, coincides with the vector of returns on the N primitive risky assets in excess of the riskless asset. The main advantage of working with excess returns is that their conditional means, $\boldsymbol{\mu}_{Nt} = \boldsymbol{\nu}_{Nt} - \nu_{0t}\boldsymbol{\iota}_N$, with $[\boldsymbol{\mu}_{Nt}]_i = \mu_{it}$, can be identified with the assets' risk premia, while preserving the structure of the conditional second moments.

In this framework, it is possible to prove that there is a conditionally affine relationship between expected returns and betas with respect to any (conditional) frontier "asset" other than the minimum (conditional) variance one (see Hansen and Richard, 1987). In particular, we have

$$\nu_{it} - \nu_{0t} = \beta_{iet}(\bar{\nu}_t - \nu_{0t}) \tag{3}$$

where $\beta_{iet} = cov(R_{it}, R_{et}(\bar{\nu}_t)|\Phi_{t-1})/V(R_{et}(\bar{\nu}_t)|\Phi_{t-1})$.

Our crucial asset pricing assumption is that there exists a conditional MV frontier risky "asset" in $\overline{\mathcal{P}}_t$ which contains only factor risk. More formally, we assume that:

$$R_{et}(\bar{\nu}_t) = \bar{\nu}_t + \beta_{e1t}f_{1t} + \beta_{e2t}f_{2t} + \ldots + \beta_{ekt}f_{kt} \tag{4}$$

Given that $R_{et}(\bar{\nu}_t)$ is not generally observable, though, expression (3) has to be expanded further. Since we know that:

$$\beta_{iet} = \frac{\beta_{i1t}\lambda_{1t}\beta_{e1t} + \ldots + \beta_{ikt}\lambda_{kt}\beta_{ekt}}{\beta_{e1t}^2\lambda_{1t} + \ldots + \beta_{ekt}^2\lambda_{kt}}$$

we can re-write (3) as

$$\mu_{it} = \beta_{i1t}\pi_{1t} + \ldots + \beta_{ikt}\pi_{kt} \tag{5}$$

where $\pi_{jt} = \lambda_{jt}\tau_{jt}$, and

$$\tau_{jt} = \frac{\beta_{ejt}}{\beta_{e1t}^2\lambda_{1t} + \ldots + \beta_{ekt}^2\lambda_{kt}}(\bar{\nu}_t - \nu_{0t})$$

2.3 Factor representing limit portfolios

If we combine expressions (1) and (5), we finally obtain that in terms of excess returns:

$$r_{it} = \beta_{i1t} r_{f_1 t} + \ldots + \beta_{ikt} r_{f_k t} + \varepsilon_{it}$$

where $r_{f_j t}$ is short-hand for $\pi_{jt} + f_{jt}$. In order to interpret this expression further, we follow Chamberlain (1983) in defining the set of well diversified portfolios, $\mathcal{D}_t$, as those elements of $\overline{\mathcal{P}}_t$ whose returns are the limits of sequences of finite portfolios $p^{+}_{Nt} = w_{0t} R_{0t} + \sum_{i=1}^{N} w_{it} R_{it}$ whose weights $w_{it} \in \Phi_{t-1}$ are such that $\sum_{i=1}^{N} w_{it}^2$ goes to 0 as N increases. For our purposes, it is also convenient to introduce a different type of portfolio, which holds one unit of the i^{th} primitive risky asset R_{it}, and $-\nu_{it}/\nu_{0t}$ units of R_{0t}. This portfolio mimics (in the conditional MS norm) the behaviour of the innovations in the returns of asset i, u_{it}, at a cost $-\mu_{it}/\nu_{0t}$. Since all such portfolios are conditionally orthogonal to the safe asset, then the conditional closure of the set of payoffs spanned by them with weights determined in $t-1$, $\overline{\mathcal{U}}_t \subset \overline{\mathcal{P}}_t$, is the (conditionally) orthogonal complement of $\{R_{0t}\}$ in $\overline{\mathcal{P}}_t$. But our assumptions about $\mathbf{\Sigma}_{Nt}$ imply that the limit riskless asset must be well diversified. Hence, since we are mainly interested in portfolios with uncertain payoffs, we can define the set of well diversified risky portfolios $\mathcal{F}_t$ as the (conditionally) orthogonal complement of $\{R_{0t}\}$ in $\mathcal{D}_t$, which by construction, must be a subset of $\overline{\mathcal{U}}_t$. Then, it is easy to see that although the conditional variance of any portfolio in $\overline{\mathcal{U}}_t$ will generally contain both a common component and a specific one, the bounded eigenvalue condition on $\mathbf{\Gamma}_{Nt}$ implies that well-diversified risky portfolios contain only factor variance. Therefore, it is not surprising that under our assumptions about $\mathbf{\Sigma}_{Nt}$, the dimension of $\mathcal{F}_t$ is precisely k (see Chamberlain, 1983), which is the dimension of systematic risk in our model.

Given that $\mathbf{B}_{Nt}$ has full column rank, it is possible to find elements of $\overline{\mathcal{U}}_t$ whose payoffs have a unit loading on factor j ($j = 1, \ldots, k$), and zero loadings on the others, at a cost equal to $-\pi_{jt}/\nu_{0t}$ (e.g. $(\mathbf{B}'_{Nt}\mathbf{B}_{Nt})^{-1}\mathbf{B}'_{Nt}(\mathbf{R}_{Nt}-\boldsymbol{\nu}_{Nt}R_{0t}/\nu_{0t})$). Then, if we compute the conditional least squares projections of such portfolios on $\mathcal{F}_t$, it is clear that we obtain portfolios that mimic (in the conditional MS norm) the common risk components f_{jt} at exactly the same cost. We shall refer to these portfolios as *factor representing limit portfolios*, and shall often denote them in terms of the associated "assets" as $R_{f_jt}-[(\nu_{0t}+\pi_{jt})/\nu_{0t}]R_{0t}$. Not surprisingly, their excess returns exactly replicate r_{f_jt}.

Since $p_{mt}=R_{0t}/\nu_{0t}$ and $p_{ct}\in\mathcal{F}_t$, an interesting property of $R_{f_1t},\ldots,R_{f_kt}$ is that, together with the safe asset, they span the conditional MV frontier obtained from *all* the primitive assets. Specifically, given that $\boldsymbol{\pi}_t=\boldsymbol{\Lambda}_t\boldsymbol{\tau}_t$ and $V(\mathbf{r}_{ft}|\Phi_{t-1})=\boldsymbol{\Lambda}_t$, we can prove that

$$p_{at}=\frac{1}{1+\boldsymbol{\tau}'_t\boldsymbol{\Lambda}_t\boldsymbol{\tau}_t}\boldsymbol{\tau}'_t\mathbf{r}_{ft}$$

and

$$R_{ct}=R_{0t}-\frac{\nu_{0t}}{1+\boldsymbol{\tau}'_t\boldsymbol{\Lambda}_t\boldsymbol{\tau}_t}\boldsymbol{\tau}'_t\mathbf{r}_{ft},$$

Hence, an investor with conditional MV preferences will effectively invest all her wealth in a convex combination of $R_{0t},R_{f_1t},\ldots,R_{f_kt}$.

2.4 Unconditional moments specification

Let $\boldsymbol{\Gamma}_N=V(\boldsymbol{\varepsilon}_{Nt})$ be the unconditional covariance matrix of the idiosyncratic terms, which can be computed as $E(\boldsymbol{\Gamma}_{Nt})$ by the law of iterated expectations. Given our assumptions, this matrix remains positive definite (p.d.) with uniformly bounded diagonal elements for all N. In principle, though, it may not preserve the bounded eigenvalue condition. Our first simplifying assumption is that it does. The obvious example is when the specific risk terms are conditionally orthogonal to each other, so that the conditional factor structure is exact.

But even when $\boldsymbol{\Gamma}_{Nt}$ is diagonal, the unconditional covariance matrix of a multivariate stochastic process characterised by a zero conditional mean and a conditional factor representation may very well lack an unconditional factor structure for any $k < N$ (see e.g. Hansen and Richard (1987) or Lehmann (1992)). The intuition is as follows: the contribution of the common factors to the unconditional variance is $E(\mathbf{B}_{Nt}\mathbf{\Lambda}_t\mathbf{B}'_{Nt})$, which is the scalar weighted average, or more precisely, the Riemann-Stieltjes integral, of many (possibly infinite) rank k positive semi-definite (p.s.d.) matrices. Unless all those matrices share the same nullspace (except perhaps for a subset of measure 0), the rank of their average will exceed k, and may well be N. In practical terms, what this means is that for $E(\mathbf{B}_{Nt}\mathbf{\Lambda}_t\mathbf{B}'_{Nt})$ to be of rank k, (almost) all $\mathbf{B}_{Nt}$ matrices must be of the form $\mathbf{B}_N\mathbf{\Psi}_t^{1/2}$, where $\mathbf{B}_N$ is a rank k matrix of fixed coefficients, and $\mathbf{\Psi}_t$ a p.d. matrix of order k, not necessarily diagonal. In view of this discussion, we follow the standard solution in the empirical literature, and assume that, for any given unconditional normalisation of the factors, such as $\mathbf{\Lambda} = V(\mathbf{f}_t) = E(\mathbf{\Lambda}_t) = \mathbf{I}_k$, the factor loadings are time-invariant. As Engle et al. (1990) pointed out, such an assumption is observationally equivalent to a model in which the conditional variance of the factors is constant, but the betas of different assets on a given factor change proportionately over time.[3] Then, it is not difficult to see that the unconditional covariance matrix of the *innovations in returns*, $\mathbf{\Sigma}_N = V(\mathbf{u}_{Nt}) = E(\mathbf{\Sigma}_{Nt})$, will have an unconditional k factor structure, namely $\mathbf{B}_N\mathbf{B}'_N + \mathbf{\Gamma}_N$. Similarly, if we call $\mu_i = E(\mu_{it})$ and $\pi_j = E(\pi_{jt})$ the (temporal) average "risk premium" for asset i and factor j respectively, the assumption of constant betas implies that our linear factor pricing model also holds on average, i.e.

$$\mu_i = \beta_{i1}\pi_1 + \ldots + \beta_{ik}\pi_k \tag{6}$$

Note, though, that the unconditional covariance matrix of the vector of *excess returns*, $\mathbf{r}_{Nt} = \mathbf{R}_{Nt} - \boldsymbol{\iota}R_{0t} = \mathbf{B}_N\mathbf{r}_{ft} + \boldsymbol{\varepsilon}_{Nt}$, will be of the form $\mathbf{B}_N\mathbf{\Xi}\mathbf{B}_N + \mathbf{\Gamma}_N$,

[3] See Sentana and Fiorentini (2001) for the implications of this assumption on the identification of the common factors and their loadings.

where $\Xi = \mathbf{I}_k + V(\boldsymbol{\pi}_t)$. In addition, the unconditional covariance matrix of *gross returns* will be different from the two previous ones unless the returns on the conditionally riskless asset are constant over time.

Finally, we would need to specify the temporal variation in the volatility of common and idiosyncratic factors to complete the model. In what follows, we simply note that by definition, $\boldsymbol{\Lambda}_t$, $\boldsymbol{\Gamma}_{Nt}$ and $\boldsymbol{\pi}_t$ are measurable functions of the agents' information set, Φ_{t-1}, but shall not use any particular parametrisation in our discussion, except for the example in Section 3.7.[4]

3 Factor Representing Portfolios with a Finite Number of Assets

In Section 2.3 above, we have seen that the factors, $\mathbf{f}_t$, can be exactly replicated (in the conditional MS norm) by means of well diversified limit portfolios in $\overline{\mathcal{U}}_t$. In empirical applications, though, only data on a finite number of assets are available, and hence *from the econometrician's point of view*, the factors are effectively *unknown*. And although in any given realisation of the asset returns, the factors might be regarded as a set of k unknown parameters, in fact, because they could take different values in different realisations, we should more appropriately regard them as *unobservable random variables*, and use a signal extraction approach (see Section 3.5 of Bartholomew, 1987).

Such an approach is particularly appropriate in the following practical situation. Suppose that in period $t-1$, we would like to form portfolios from the finite collection of asset at hand, R_{0t} and $\mathbf{R}_{Nt}$, whose payoffs in period t closely replicate $\mathbf{f}_t$. Or in financial markets parlance, we would like to construct k portfolios that

[4]Some popular parametric time-series models for financial returns specify a data generating process in which the volatility of the factors is a function of unobservables. Nevertheless, what is important for our purposes is not the latent volatility itself, but rather the variance of the factors conditional on the agents' information set, Φ_{t-1}.

"track" each element of $\mathbf{f}_t$ as their "target", either because we want to "tilt" our investment decision towards a particular factor, or because we want to hedge its risk by means of a "basket" of assets (see e.g. Sheikh (1996) and Sorensen et al. (1993) respectively). In any case, since we would be effectively using portfolios in $\mathcal{U}_{Nt}$ to estimate the underlying factors, where $\mathcal{U}_{Nt}$ is the analogue of $\overline{\mathcal{U}}_t$ constructed from R_{0t} and $\mathbf{R}_{Nt}$ alone (i.e. the set of payoffs spanned by $\mathbf{u}_{Nt}$ with weights determined in $t-1$), we would be implicitly restricting ourselves to the class of conditionally linear predictors with time-varying stochastic weightings, and excluding any non-linear filters and all smoothers. In this respect, it is crucial for our purposes to distinguish between $\mathcal{U}_{Nt}$, and its subset, $\mathcal{V}_{Nt}$, which is the set of payoffs spanned by the same portfolios but with *fixed weights*. As we shall see below, $\mathcal{V}_{Nt}$ includes all the basis portfolios suggested so far in the literature, but excludes the payoffs on dynamic portfolio strategies based on information in Φ_{t-1}.

In this framework, our efficiency analysis shall be based on comparing the conditional MSEs of the different basis portfolios, which is the appropriate metric in the conditional Hilbert space setting of Section 2. In financial markets parlance, this simply means that, *ceteris paribus*, we would always prefer factor representing portfolios with a smaller degree of "tracking error". Given its widespread use by practitioners, it is not surprising that such a criterion can be given an intuitive justification. In particular, it is easy to show that an investor with unit wealth and conditional MV preferences of the form $U(R_t;\Phi_{t-1}) = E(R_t|\Phi_{t-1}) - (\alpha/2)E(R_t^2|\Phi_{t-1})$ will hold one unit of the conditionally riskless asset, and $1/\alpha$ units of the arbitrage portfolio (i) p_{at} if she can invest in all the primitive assets, or (ii) $p_{aNt} = \mathbf{w}'_{aNt}\mathbf{r}_{Nt}$, where:

$$\mathbf{w}_{aNt} = \frac{\mathbf{\Sigma}_{Nt}^{-1}\boldsymbol{\mu}_{Nt}}{1+\boldsymbol{\mu}'_{Nt}\mathbf{\Sigma}_{Nt}^{-1}\boldsymbol{\mu}_{Nt}}$$

if she can only hold a finite subset of N risky assets. It is then straightforward to prove that p_{aNt} is precisely the arbitrage portfolio that is closest to p_{at} in the conditional MS norm. But since as we saw before, the whole conditional MV frontier generated by all the available assets can be spanned by R_{0t} and $R_{f1t}, \ldots, R_{fkt}$, it is not surprising that the payoffs to p_{aNt} can be exactly replicated by means of a combination of the riskless asset and the factor representing portfolios that are closest to $f_{1t}, \ldots f_{kt}$ in the conditional MS norm. Finally, note that since the average of the conditional MSE is the unconditional MSE by virtue of the law of iterated expectations, we can keep in line with the more standard practice in the statistical time series literature without additional effort.[5]

3.1 Basis portfolios with time-varying weights

3.1.1 Conditional Kalman Filter Basis Portfolios

If the joint conditional distribution of $\mathbf{u}_{Nt}$ and $\mathbf{f}_t$ given Φ_{t-1} were normal, the model derived in Section 2 would have a natural conditionally Gaussian linear state-space representation (see Harvey, 1989). Taking the common factors as state variables, equation (2) could be understood as the measurement equation which relates them to the observable variables $\mathbf{u}_{Nt}$, with the idiosyncratic terms, $\boldsymbol{\varepsilon}_{Nt}$, corresponding to the measurement error.[6] In this well known, ideal, conditionally Gaussian framework, the Kalman filter would be perfectly suited to "extract" estimates of the unknown factors because the *updated* estimates of $\mathbf{f}_t$,

$$\mathbf{f}_{Nt}^K = \boldsymbol{\Lambda}_t \mathbf{B}_N' \boldsymbol{\Sigma}_{Nt}^{-1} \mathbf{u}_{Nt} = (\mathbf{B}_N' \boldsymbol{\Gamma}_{Nt}^{-1} \mathbf{B}_N + \boldsymbol{\Lambda}_t^{-1})^{-1} \mathbf{B}_N' \boldsymbol{\Gamma}_{Nt}^{-1} \mathbf{u}_{Nt}, \tag{7}$$

would coincide with the conditional expectation of $\mathbf{f}_t$ given $\mathbf{u}_{Nt}$ and Φ_{t-1}, which is *best* in the (conditional) MS sense over the class of all predictors that use the same

[5] The unconditional MSE criterion would break down if the unconditional second moments of the assets were unbounded. And although our assumption of uniformly bounded conditional second moments implies that this is not so (see footnote 2), our main results (i.e. Theorems 1, 2 and 3) would remain valid.

[6] The transition equation would be somewhat unusual, though, because $\mathbf{f}_t$ contains no mean dynamics by construction.

information, whether linear or not (see e.g. Harvey, 1989).[7] In fact, the optimality of $\mathbf{f}_{Nt}^{K}$ holds under the more general assumption that, conditioned on Φ_{t-1}, $\mathbf{u}_{Nt}$ and $\mathbf{f}_t$ follow a joint multivariate elliptically symmetric distribution (see Sentana, 1991), of which the multivariate normal and the multivariate t are rather special cases. But although the elliptical class is rather broad (see e.g. Fang et al., 1990), ideally we would like to have a similar result that did not depend on distributional assumptions. In this respect, the following theorem characterises the optimality of the Kalman filter updated estimates for any conditional distribution:

Theorem 1 *For any $k \times M$ matrix $\mathbf{W}(\Phi_{t-1})$ of measurable functions of the information set, $\mathbf{W}'(\Phi_{t-1})\mathbf{f}_{Nt}^{K}$ is best in the conditional and unconditional MSE sense within the class of "conditionally affine" predictors of $\mathbf{W}'(\Phi_{t-1})\mathbf{f}_t$ of the form $\mathbf{c}(\Phi_{t-1}) + \mathbf{D}'(\Phi_{t-1})\mathbf{u}_{Nt}$, where $\mathbf{c}()$ is a $M \times 1$ vector and $\mathbf{D}()$ a $N \times M$ matrix of measurable functions of the information set.*

Theorem 1 generalises to a conditional setting the well-known result that the Kalman filter updated estimates of an unconditionally linear model are best within the class of linear predictors that use information up to, and including, period t (see Harvey, 1981, and Theorem 4 below). Intuitively, the optimality of the Kalman filter estimates derives from the fact that

$$E\left[\left(\mathbf{f}_t - \mathbf{f}_{Nt}^{K}\right)\left[\mathbf{c}(\Phi_{t-1}) + \mathbf{D}'(\Phi_{t-1})\mathbf{u}_{Nt}\right]' \middle| \Phi_{t-1}\right] = \mathbf{0}, \tag{8}$$

which means that they can be interpreted as the conditionally linear least squares projection of $\mathbf{f}_t$ on $\mathcal{U}_{Nt}$ (see Appendix 1). In this respect, it is important to mention that the conditional MSE of $\mathbf{f}_{Nt}^{K}$ provided by the usual Kalman filter recursions

$$\mathbf{\Omega}_{Nt}^{K} = \mathbf{\Lambda}_t - \mathbf{\Lambda}_t\mathbf{B}_N'\mathbf{\Sigma}_{Nt}^{-1}\mathbf{B}_N\mathbf{\Lambda}_t = (\mathbf{\Lambda}_t^{-1} + \mathbf{B}_N'\mathbf{\Gamma}_{Nt}^{-1}\mathbf{B}_N)^{-1} \tag{9}$$

only yields $V(\mathbf{f}_t|\,\mathbf{u}_{Nt}, \Phi_{t-1})$ under conditional normality (e.g. for any other elliptical distribution $V(\mathbf{f}_t|\,\mathbf{u}_{Nt}, \Phi_{t-1}) \propto g(\mathbf{u}_{Nt}'\mathbf{\Sigma}_{Nt}^{-1}\mathbf{u}_{Nt})\mathbf{\Omega}_{Nt}^{K}$, where $g()$ is a scalar func-

[7] This property provides the rationale for using $\mathbf{f}_{Nt}^{K}$ as factor estimates in King et al. (1994).

tion whose form depends on the particular member of the elliptical class used; see Harvey et al. (1992) for the multivariate t, Sentana (1991) for the general case).

Like all conditionally linear predictors, $\mathbf{f}_{Nt}^K$ is an (un)conditionally unbiased predictor of $\mathbf{f}_t$. More importantly, (8) also implies that forecast, $\mathbf{f}_{Nt}^K$, and forecast error, $\mathbf{f}_{Nt}^K - \mathbf{f}_t$, are (un)conditionally uncorrelated. This confirms that $\mathbf{\Omega}_N^K = V\left[\mathbf{f}_{Nt}^K - \mathbf{f}_t\right] = E\left[\mathbf{\Omega}_{Nt}^K\right]$ constitutes the lower bound for the MSE of all (conditionally) linear predictors forecasting errors. In fact, this lower bound is not achieved in finite N samples by any other (conditionally) linear predictor. The orthogonality of predictor and prediction error also implies that the Kalman filter estimates are (un)conditionally smoother than the factors, but since $V\left[\mathbf{f}_{Nt}^K \middle| \Phi_{t-1}\right] = \mathbf{\Lambda}_t[\mathbf{\Lambda}_t + (\mathbf{B}_N'\mathbf{\Gamma}_{Nt}^{-1}\mathbf{B}_N)^{-1})^{-1}\mathbf{\Lambda}_t$, the estimators for different factors would not be conditionally uncorrelated unless $\mathbf{B}_N'\mathbf{\Gamma}_{Nt}^{-1}\mathbf{B}_N$, and thus $\mathbf{\Omega}_{Nt}^K$, were diagonal, and it is unlikely that they would be unconditionally uncorrelated otherwise.

One interesting property of $\mathbf{f}_{Nt}^K$ that also derives from Theorem 1 is that as we include more assets in our sample, the conditional variance of the prediction error cannot increase, and generally decreases (i.e. $V\left[\mathbf{f}_{Nt}^K - \mathbf{f}_t \middle| \Phi_{t-1}\right] - V\left[\mathbf{f}_{N+1t}^K - \mathbf{f}_t \middle| \Phi_{t-1}\right] = \mathbf{\Omega}_{Nt}^K - \mathbf{\Omega}_{N+1t}^K$ is p.s.d.), which is also true unconditionally. Since predictor and prediction error are uncorrelated, the monotonicity extends to the conditional and unconditional correlations between f_{jNt}^K and f_{jt} for $j = 1, \ldots, k$. In fact, we can prove a much stronger result:

Theorem 2 *For any $k \times 1$ vector $\mathbf{w}(\Phi_{t-1})$ of measurable functions of the information set, $\mathbf{w}'(\Phi_{t-1})\mathbf{f}_{Nt}^K$ has maximum conditional correlation with $\mathbf{w}'(\Phi_{t-1})\mathbf{f}_t$ within the class of "conditionally affine" predictors of the form $c(\Phi_{t-1}) + \mathbf{d}'(\Phi_{t-1})\mathbf{u}_{Nt}$, where $c()$ is a scalar and $\mathbf{d}()$ a $N \times 1$ vector of measurable functions of the information set.*

Therefore, we can obtain a dynamic portfolio in $\mathcal{U}_{Nt}$ that maximises conditional correlation with any given element of the set of well diversified, risky portfolios $\mathcal{F}_t$, by choosing the time-varying portfolio weights that minimise our conditional

MSE criterion. Nevertheless, it is important to bear in mind that the opposite result is not true, since any conditionally affine transformation of $\mathbf{w}'(\Phi_{t-1})\mathbf{f}^K_{Nt}$ will maintain the conditional correlation with $\mathbf{w}'(\Phi_{t-1})\mathbf{f}_t \in \mathcal{F}_t$, but will increase the conditional MSE. The reason is that unlike what happens with the conditional MSE, a metric cannot be based on conditional correlations.

Finally, it is worth mentioning that these mimicking portfolios are often obtained as a by-product of the ML estimation method, so that their calculation does not increase the computational burden (see Harvey et al., 1992).

3.1.2 Conditional GLS basis portfolios

Since any conceivable factor representing portfolio will effectively be a combination of R_{0t} and a portfolio in $\mathcal{U}_{Nt}$, in view of Theorems 1 and 2, a natural question at this stage is what justification can be given for using any other factor extraction procedure. The following result characterises the optimality of the conditional Generalised Least Squares (GLS) estimator:

$$\mathbf{f}^G_{Nt} = (\mathbf{B}'_N\mathbf{\Gamma}^{-1}_{Nt}\mathbf{B}_N)^{-1}\mathbf{B}'_N\mathbf{\Gamma}^{-1}_{Nt}\mathbf{u}_{Nt} \tag{10}$$

which, importantly, does not make use of the information given by $E(\mathbf{f}_t|\,\Phi_{t-1}) = \mathbf{0}$ and $V(\mathbf{f}_t|\,\Phi_{t-1}) = \mathbf{\Lambda}_t$.

Theorem 3 $\mathbf{W}'(\Phi_{t-1})\mathbf{f}^G_{Nt}$ *is best in the conditional and unconditional MSE sense within the subclass of "conditionally linear" predictors of* $\mathbf{W}'(\Phi_{t-1})\mathbf{f}_t$ *of the form* $\mathbf{D}'(\Phi_{t-1})\mathbf{u}_{Nt}$ *that satisfy the restriction* $\mathbf{D}'(\Phi_{t-1})\mathbf{B}_N = \mathbf{W}'(\Phi_{t-1})$.

As we mentioned before, in any given realisation of the asset returns, $\mathbf{f}_t$ could be regarded as a set of k parameters. From this point of view, the difference between $\mathbf{f}^K_{Nt}$ and $\mathbf{f}^G_{Nt}$ would be the difference between a Bayesian cross-sectional GLS estimator of $\mathbf{f}_t$ which uses the proper prior $\mathbf{f}_t|\Phi_{t-1} \sim D(\mathbf{0}, \mathbf{\Lambda}_t)$, and another one which uses a diffuse prior instead. In this context, the constraint

$\mathbf{D}'(\Phi_{t-1})\mathbf{B}_N = \mathbf{W}'(\Phi_{t-1})$ that appears in Theorem 3 could be regarded as imposing the unbiasedness restriction $E\left[\mathbf{f}_{Nt}^G \middle| \mathbf{f}_t, \Phi_{t-1}\right] = \mathbf{f}_t$.

In any case, given that

$$\mathbf{f}_{Nt}^G = \mathbf{f}_t + (\mathbf{B}_N'\mathbf{\Gamma}_{Nt}^{-1}\mathbf{B}_N)^{-1}\mathbf{B}_N'\mathbf{\Gamma}_{Nt}^{-1}\boldsymbol{\varepsilon}_{Nt}$$

a noticeable property of $\mathbf{f}_{Nt}^G$ is that they constitute "sufficient statistics" of the information in $\mathbf{u}_{Nt}$ regarding $\mathbf{f}_t$, in the sense that the conditional Kalman filter estimates of $\mathbf{f}_t$ based on $\mathbf{f}_{Nt}^G$ coincide with $\mathbf{f}_{Nt}^K$. Namely:

$$\begin{aligned}\mathbf{f}_{Nt}^K &= \mathbf{\Lambda}_t[\mathbf{\Lambda}_t + (\mathbf{B}_N'\mathbf{\Gamma}_{Nt}^{-1}\mathbf{B}_N)^{-1}]^{-1}\mathbf{f}_{Nt}^G \\ &= [\mathbf{\Lambda}_t^{-1} + (\mathbf{B}_N'\mathbf{\Gamma}_{Nt}^{-1}\mathbf{B}_N)]^{-1}(\mathbf{B}_N'\mathbf{\Gamma}_{Nt}^{-1}\mathbf{B}_N)\mathbf{f}_{Nt}^G\end{aligned}$$

(see Gourieroux et al. (1991), and Fiorentini et al. (2001) for a stronger result under conditional Gaussianity).[8] A useful way of looking at this relationship is in terms of the joint distribution of $\mathbf{f}_t$ and $\mathbf{f}_{Nt}^G$ conditioned on Φ_{t-1}:

$$\begin{pmatrix} \mathbf{f}_t \\ \mathbf{f}_{Nt}^G \end{pmatrix} |\Phi_{t-1} \sim D\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \mathbf{\Lambda}_t & \mathbf{\Lambda}_t \\ \mathbf{\Lambda}_t & \mathbf{\Lambda}_t + (\mathbf{B}_N'\mathbf{\Gamma}_{Nt}^{-1}\mathbf{B}_N)^{-1} \end{pmatrix}\right]$$

In this context, $\mathbf{f}_{Nt}^K$ and $\mathbf{\Omega}_{Nt}^K$ are the fitted value and residual variance from the conditional least squares regression of $\mathbf{f}_t$ on $\mathbf{f}_{Nt}^G$, while $\mathbf{f}_t$ and $\mathbf{\Omega}_{Nt}^G = (\mathbf{B}_N'\mathbf{\Gamma}_{Nt}^{-1}\mathbf{B}_N)^{-1}$ are the corresponding quantities in the conditional least squares regression of $\mathbf{f}_{Nt}^G$ on $\mathbf{f}_t$.

Also, it is worth mentioning that on the basis of $\mathbf{f}_{Nt}^G$ one can write what Gourieroux et al. (1991) call an endogenous factorial representation

$$\mathbf{u}_{Nt} = \mathbf{B}_N\mathbf{f}_{Nt}^G + \boldsymbol{\varepsilon}_{Nt}^G$$

[8] As a consequence, $\mathbf{f}_{Nt}^K$, $\mathbf{f}_{Nt}^G$ or indeed any conditionally linear non-singular transformation of them, span the same linear subspace, which is important if the object of interest is the whole of $\mathcal{F}_t$ itself, not one of its elements (cf. Huberman et al., 1987). Nevertheless, Theorems 1 and 2 imply that the Kalman filter yields the best predictor of every element of that subspace, and the most highly correlated. In particular, as we mentioned before, $\boldsymbol{\tau}'\mathbf{f}_{Nt}^K/(1+\boldsymbol{\tau}'\mathbf{\Lambda}_t\boldsymbol{\tau})$ provides the best predictor of (the unanticipated component of) p_{at}, which is also the one with the highest conditional correlation.

where $\boldsymbol{\varepsilon}_{Nt}^G = \left[\mathbf{I}_N - \mathbf{B}_N(\mathbf{B}_N'\boldsymbol{\Gamma}_{Nt}^{-1}\mathbf{B}_N)^{-1}\mathbf{B}_N'\boldsymbol{\Gamma}_{Nt}^{-1}\right]\mathbf{u}_{Nt}$ and $\mathbf{f}_{Nt}^G$ are conditionally uncorrelated. This property is related to the fact that $\mathbf{f}_{Nt}^G$ can be written as a conditionally linear combination of the conditional principal components of $\mathbf{u}_{Nt}^\dagger = \boldsymbol{\Gamma}_{Nt}^{-1/2}\mathbf{u}_{Nt}$, as shown in Appendix 2.

Since $V\left[\mathbf{f}_{Nt}^G \middle| \Phi_{t-1}\right] = \boldsymbol{\Lambda}_t + (\mathbf{B}_N'\boldsymbol{\Gamma}_{Nt}^{-1}\mathbf{B}_N)^{-1}$, the factor estimates are again (conditionally) correlated with each other in general, but now they are less smooth than the factor themselves. Furthermore, if $\mathbf{B}_N'\boldsymbol{\Gamma}_{Nt}^{-1}\mathbf{B}_N$ were a diagonal matrix (e.g. if $k=1$), then each f_{jNt}^G would be proportional to f_{jNt}^K for $j = 1, \ldots, k$, and the conditional correlations with f_{jt} would then be identical, although the factors of proportionality will change over time.

Not surprisingly, though, predictor and prediction error are conditionally correlated in this case, which reflects that $\mathbf{f}_{Nt}^G$ does not use the available information efficiently. Nevertheless, since $\mathbf{f}_{Nt}^G = \left[\begin{array}{cc} (\mathbf{B}_N'\boldsymbol{\Gamma}_{Nt}^{-1}\mathbf{B}_N)^{-1}\mathbf{B}_N'\boldsymbol{\Gamma}_{Nt}^{-1} & \mathbf{0} \end{array}\right]\mathbf{u}_{N+1t}$ and $\left[\begin{array}{cc} (\mathbf{B}_N'\boldsymbol{\Gamma}_{Nt}^{-1}\mathbf{B}_N)^{-1}\mathbf{B}_N'\boldsymbol{\Gamma}_{Nt}^{-1} & \mathbf{0} \end{array}\right]\mathbf{B}_{N+1} = \mathbf{I}_k$, $\boldsymbol{\Omega}_{Nt}^G$ is also monotonically nonincreasing with the cross-sectional sample size.

3.2 Basis Portfolios with Constant Weightings

3.2.1 Unconditional Kalman filter basis portfolios

Given that our assumptions are compatible with an unconditional k factor structure for $\mathbf{u}_{Nt}$, we can also consider unconditional counterparts to the previous predictors of $\mathbf{f}_t$. These portfolios will be elements of $\mathcal{V}_{Nt}$, the set of payoffs spanned by u_{it} $(i = 1, \ldots, N)$ whose weights are time-invariant. To the best of our knowledge, all the factor representing portfolios discussed so far in the existing literature fall within this class.

In this framework, we can repeat the analysis of Section 3.1.1 for the class of static portfolios to show that the best (in the *unconditional* MSE sense) predictor of $\mathbf{f}_t$ with time-invariant weightings is given by the standard Kalman filter updated estimate

$$\mathbf{f}_{Nt}^{UK} = \mathbf{B}_N' \mathbf{\Sigma}_N^{-1} \mathbf{u}_{Nt} = (\mathbf{B}_N' \mathbf{\Gamma}_N^{-1} \mathbf{B}_N + \mathbf{I}_k)^{-1} \mathbf{B}_N' \mathbf{\Gamma}_N^{-1} \mathbf{u}_{Nt}$$

More formally:[9]

Theorem 4 *For any $k \times M$ constant matrix $\mathbf{W}$, $\mathbf{W}'\mathbf{f}_{Nt}^{UK}$ is best in the unconditional MSE sense within the class of "affine" predictors of $\mathbf{W}'\mathbf{f}_t$ of the form $\mathbf{c} + \mathbf{D}'\mathbf{u}_{Nt}$, where $\mathbf{c}$ is a $M \times 1$ vector and $\mathbf{D}$ a $N \times M$ matrix of constants.*

In this case, the prediction error unconditional MSE is

$$\mathbf{\Omega}_N^{UK} = (\mathbf{B}_N' \mathbf{\Gamma}_N^{-1} \mathbf{B}_N + \mathbf{I}_k)^{-1}$$

(see Harvey, 1981). Not surprisingly, we can understand $\mathbf{f}_{Nt}^{UK}$ as the *unconditional* least squares projection of $\mathbf{f}_t$ on $\mathcal{V}_{Nt}$, and for that reason, these estimators are usually known as the regression scores in the factor analysis literature (see e.g. Lawley and Maxwell, 1971).

Similarly, every element of $\mathbf{f}_{Nt}^{UK}$ is also a static portfolio with maximum unconditional correlation with the corresponding element of $\mathbf{f}_t$. Specifically:[10]

Theorem 5 *For any $k \times 1$ constant vector $\mathbf{w}$, $\mathbf{w}'\mathbf{f}_{Nt}^{UK}$ has maximum unconditional correlation with $\mathbf{w}'\mathbf{f}_t$ within the class of "affine" predictors of the form $c + \mathbf{d}'\mathbf{u}_{Nt}$, where c is a scalar and $\mathbf{d}$ a $N \times 1$ vector of constants.*

Therefore, by minimising the variability in tracking error, we obtain as a by-product portfolios that maximise correlation with the common factors. But again, the converse is not true, as any affine transformation of $\mathbf{w}'\mathbf{f}_{Nt}^{UK}$ will increase MSE, despite the fact that it leaves the correlation unchanged.

Given that the results in Theorems 4 and 5 are well known, it is surprising that these mimicking portfolios have never been used in empirical applications of the static APT (to the best of the author's knowledge), especially if we take into account the fundamental role they play in maximum likelihood estimation of traditional factor models via the EM algorithm (see Rubin and Thayer, 1982).

[9] See e.g. Lawley and Maxwell (1972) for a proof, who attribute these factor scores to Thomson (1951).

[10] See Ingersoll (1984) or Huberman et al. (1987) for a proof.

3.2.2 Unconditional GLS basis portfolios

A less efficient estimator is the unconditional generalised least squares estimator

$$\mathbf{f}_{Nt}^{UG} = (\mathbf{B}_N'\mathbf{\Gamma}_N^{-1}\mathbf{B}_N)^{-1}\mathbf{B}_N'\mathbf{\Gamma}_N^{-1}\mathbf{u}_{Nt},$$

which does not make use of the information given by $E(\mathbf{f}_t) = \mathbf{0}$ and $V(\mathbf{f}_t) = \mathbf{I}$. Nevertheless, this estimator still has some optimality properties with respect to a particular subclass, as specified by the following result:[11]

Theorem 6 $\mathbf{W}'\mathbf{f}_{Nt}^{UG}$ *is best in the unconditional MSE sense within the subclass of "linear" predictors of* $\mathbf{W}'\mathbf{f}_t$ *of the form* $\mathbf{D}'\mathbf{u}_{Nt}$ *that satisfy the restriction* $\mathbf{D}'\mathbf{B}_N = \mathbf{W}'$.

Again, there is a one-to-one relationship between $\mathbf{f}_{Nt}^{UK}$ and $\mathbf{f}_{Nt}^{UG}$ given by

$$\mathbf{f}_{Nt}^{UK} = [\mathbf{I} + (\mathbf{B}_N'\mathbf{\Gamma}_N^{-1}\mathbf{B}_N)^{-1}]^{-1}\mathbf{f}_{Nt}^{UG}$$

which means that the unconditional Kalman filter estimates based on $\mathbf{f}_{Nt}^{UG}$ coincide with those based on $\mathbf{u}_{Nt}$. Similarly, these mimicking portfolios can also be used to interpret the first order conditions of the maximum likelihood estimation of traditional factor models (see Grinblatt and Titman, 1987). Finally, the unconditional MSE of $\mathbf{f}_{Nt}^{UG}$ is given by:

$$\mathbf{\Omega}_N^{UG} = (\mathbf{B}_N'\mathbf{\Gamma}_N^{-1}\mathbf{B}_N)^{-1}$$

3.3 Other Basis Portfolios

Many other conditionally linear factor representing portfolios are conceptually possible. For instance, there are at least two other Kalman-based portfolios that can be considered, for which the analysis in the previous sections can be easily

[11] See e.g. Lawley and Maxwell (1971) for a proof, who refer to these estimators as Barlett's scores after Barlett (1938).

adapted. One results from applying the unconditional Kalman filter to $\mathbf{f}_{Nt}^{G}$, yielding $\mathbf{f}_{Nt}^{UKG}$; the other one, which we shall term $\mathbf{f}_{Nt}^{KUG}$, is obtained if we apply the conditional Kalman filter to $\mathbf{f}_{Nt}^{UG}$.

An alternative unconditionally linear predictor of $\mathbf{f}_t$ is the Ordinary Least Squares estimate:

$$\mathbf{f}_{Nt}^{O} = (\mathbf{B}_N'\mathbf{B}_N)^{-1}\mathbf{B}_N'\mathbf{u}_{Nt} \tag{11}$$

whose prediction errors unconditional variance

$$\mathbf{\Omega}_N^{O} = (\mathbf{B}_N'\mathbf{B}_N)^{-1}\mathbf{B}_N\mathbf{\Gamma}_N\mathbf{B}_N'(\mathbf{B}_N'\mathbf{B}_N)^{-1}$$

cannot be smaller than $\mathbf{\Omega}_N^{UG}$. The intuition is that while $\mathbf{f}_{Nt}^{UG}$ ignores the information in the mean and variance of $\mathbf{f}_t$, $\mathbf{f}_{Nt}^{O}$ ignores information about $\mathbf{\Gamma}_N$ as well. Also, the correlations between $\mathbf{f}_t$ and $\mathbf{f}_{Nt}^{O}$ will be generally smaller than the correlations between $\mathbf{f}_t$ and $\mathbf{f}_t^{UG}$. Unfortunately, it is not clear that the predictor error variance of $\mathbf{f}_{Nt}^{O}$ is monotonically decreasing with the sample size except in special cases.

When $k = 1$, we can also consider cross-sectional weighted averages of the assets at hand. In particular, we could consider what Lehmann and Modest (1988) call minimum idiosyncratic risk portfolios, which in this case simplify to:

$$f_{Nt}^{UI} = (\boldsymbol{\iota}_N'\mathbf{\Gamma}_N^{-1}\boldsymbol{\iota}_N)^{-1}\boldsymbol{\iota}_N'\mathbf{\Gamma}_N^{-1}\mathbf{u}_{Nt} \tag{12}$$

with conditional and unconditional MSE given by

$$\begin{aligned} \omega_{Nt}^{UI} &= (1-\breve{\beta}_N)^2\lambda_t + (\boldsymbol{\iota}_N'\mathbf{\Gamma}_N^{-1}\mathbf{\Gamma}_{Nt}\mathbf{\Gamma}_N^{-1}\boldsymbol{\iota}_N)/(\boldsymbol{\iota}_N'\mathbf{\Gamma}_N^{-1}\boldsymbol{\iota}_N)^2 \\ \omega_{N}^{UI} &= (1-\breve{\beta}_N)^2 + (\boldsymbol{\iota}_N'\mathbf{\Gamma}_N^{-1}\boldsymbol{\iota}_N)^{-1} \end{aligned}$$

respectively, where $\breve{\beta}_N = \sum_i \beta_i\gamma_{ii}^{-1}/\sum_i \gamma_{ii}^{-1}$ is the cross-sectional average of the factor loadings, with weights that are inversely proportional to the unconditional idiosyncratic variances.

As expected, the prediction error is correlated with f_{Nt}^{UI}, and again improvements in the predictions can be made. The intuition is that we are now ignoring the information contained in $\mathbf{B}_N$. This could have potentially serious consequences in those empirical applications in which the set of available assets includes arbitrage portfolios (such as returns on foreign exchange forward contracts), for which the sign of the payoff is largely irrelevant (see e.g. Diebold and Nerlove, 1989).

There is also an equally weighted portfolio version of f_t^{UI}, which is simply

$$f_{Nt}^A = (\boldsymbol{\iota}_N'\boldsymbol{\iota}_N)^{-1}\boldsymbol{\iota}_N'\mathbf{u}_N,$$

the cross-sectional average of the returns innovations. In this case, the conditional and unconditional MSE are

$$\begin{aligned}
\omega_{Nt}^A &= (1-\bar{\beta}_N)^2\lambda_t + \bar{\gamma}_{Nt}/N \\
\omega_N^A &= (1-\bar{\beta}_N)^2 + \bar{\gamma}_N/N
\end{aligned}$$

respectively, where $\bar{\gamma}_{Nt}, \bar{\gamma}_N$ and $\bar{\beta}_N$ are the unweighted cross-sectional averages of the conditional and unconditional idiosyncratic variances, and the factor loadings. Again, the prediction error is correlated with f_{Nt}^A, which confirms its suboptimality.

3.4 The relative efficiency of the different basis portfolios

The results in the previous section provide us with an ordering for many of the basis portfolios that we analyse in terms of their MSE. Nevertheless, they do not tell us much about the magnitude of the differences in their MSE as a function of the model parameters. The objective of this subsection is to obtain a few "sufficient" statistics that effectively summarise the information in the model parameters required to analyse the relative efficiency of different mimicking portfolios. To do so, we initially assume that $\mathbf{B}_N, \boldsymbol{\Lambda}_t, \boldsymbol{\Gamma}_{Nt}$, and $\boldsymbol{\tau}_t$, and hence $\boldsymbol{\mu}_{Nt}$ and

$\mathbf{\Sigma}_{Nt}$, are known by the econometrician. Then, in Section 3.6 we shall discuss the effects of not knowing them.

3.4.1 Unconditional GLS vs Unconditional Kalman filter

In order to compare $\mathbf{f}_{Nt}^{UK}$ and $\mathbf{f}_{Nt}^{UG}$, it is convenient to scale the innovations in returns by $\mathbf{\Gamma}_N^{-1/2}$, i.e.:

$$\mathbf{u}_{Nt}^* = \mathbf{B}_N^* \mathbf{f}_t + \boldsymbol{\varepsilon}_{Nt}^*$$

where $\mathbf{u}_{Nt}^* = \mathbf{\Gamma}_N^{-1/2}\mathbf{u}_{Nt}$, $\mathbf{B}_N^* = \mathbf{\Gamma}_N^{-1/2}\mathbf{B}_N$ and $\boldsymbol{\varepsilon}_{Nt}^* = \mathbf{\Gamma}_N^{-1/2}\boldsymbol{\varepsilon}_{Nt}$, so that $V\left(\boldsymbol{\varepsilon}_{Nt}^* \middle| \Phi_{t-1}\right) = \mathbf{\Gamma}_{Nt}^* = \mathbf{\Gamma}_N^{-1/2}\mathbf{\Gamma}_{Nt}\mathbf{\Gamma}_N^{-1/2}$ but $V(\boldsymbol{\varepsilon}_{Nt}^*) = E(\mathbf{\Gamma}_N^*) = \mathbf{I}_N$. Let

$$\mathbf{B}_N^* = \begin{pmatrix} \dot{\mathbf{Q}}_N^* & \ddot{\mathbf{Q}}_N^* \end{pmatrix} \begin{pmatrix} \mathbf{\Psi}_N^{*1/2} \\ 0 \end{pmatrix} \mathbf{V}_N^{*\prime} = \dot{\mathbf{Q}}_N^* \mathbf{\Psi}_N^{*1/2} \mathbf{V}_N^{*\prime}$$

denote the singular value decomposition of the $N \times k$ matrix $\mathbf{B}_N^*$, and define $\dot{\mathbf{u}}_{Nt}^* = \dot{\mathbf{Q}}_N^{*\prime}\mathbf{u}_{Nt}^*$ and $\ddot{\mathbf{u}}_{Nt}^* = \ddot{\mathbf{Q}}_N^{*\prime}\mathbf{u}_{Nt}^*$. The factor structure for these portfolios will be given by

$$\begin{pmatrix} \dot{\mathbf{u}}_{Nt}^* \\ \ddot{\mathbf{u}}_{Nt}^* \end{pmatrix} = \begin{pmatrix} \mathbf{\Psi}_N^{*1/2}\mathbf{V}_N^{*\prime} \\ \mathbf{0} \end{pmatrix} \mathbf{f}_t + \begin{pmatrix} \dot{\boldsymbol{\varepsilon}}_{Nt}^* \\ \ddot{\boldsymbol{\varepsilon}}_{Nt}^* \end{pmatrix},$$

where the unconditional covariance matrix of the idiosyncratic components $\dot{\boldsymbol{\varepsilon}}_{Nt}^* = \dot{\mathbf{Q}}_N^{*\prime}\boldsymbol{\varepsilon}_{Nt}^*$ and $\ddot{\boldsymbol{\varepsilon}}_{Nt}^* = \ddot{\mathbf{Q}}_N^{*\prime}\boldsymbol{\varepsilon}_{Nt}^*$ will be given by

$$E(\mathbf{Q}_N^{*\prime}\mathbf{\Gamma}_{Nt}^*\mathbf{Q}_N^*) = E\begin{pmatrix} \dot{\mathbf{\Gamma}}_{Nt}^* & \dddot{\mathbf{\Gamma}}_{Nt}^* \\ \dddot{\mathbf{\Gamma}}_{Nt}^{\prime *} & \ddot{\mathbf{\Gamma}}_{Nt}^* \end{pmatrix} = \begin{pmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0}' & \mathbf{I}_{N-k} \end{pmatrix}$$

because $\mathbf{Q}^*$ is an orthogonal matrix. In this context, it is straightforward to show that the unconditional GLS basis portfolios can be written as a linear, time-invariant, transformation of $\dot{\mathbf{u}}_{Nt}^*$. Specifically

$$\mathbf{f}_{Nt}^{UG} = \mathbf{V}_N^*\mathbf{\Psi}_N^{*-1/2}\dot{\mathbf{u}}_{Nt}^* = \mathbf{f}_t + \mathbf{V}_N^*\mathbf{\Psi}_N^{*-1/2}\dot{\boldsymbol{\varepsilon}}_{Nt}^*,$$

so that

$$\begin{aligned} \mathbf{\Omega}_{Nt}^{UG} &= \mathbf{V}_N^*\mathbf{\Psi}_N^{*-1/2}\dot{\mathbf{\Gamma}}_{Nt}^*\mathbf{\Psi}_N^{*-1/2}\mathbf{V}_N^{*\prime} = \mathbf{\Omega}_N^{UG}\left(\mathbf{V}_N^*\mathbf{\Psi}_N^{*1/2}\dot{\mathbf{\Gamma}}_{Nt}^*\mathbf{\Psi}_N^{*1/2}\mathbf{V}_N^{*\prime}\right)\mathbf{\Omega}_N^{UG} \\ &= (\mathbf{B}_N'\mathbf{\Gamma}_N\mathbf{B}_N)^{-1}\mathbf{B}_N'\mathbf{\Gamma}_N^{-1}\mathbf{\Gamma}_{Nt}\mathbf{\Gamma}_N^{-1}\mathbf{B}_N(\mathbf{B}_N'\mathbf{\Gamma}_N^{-1}\mathbf{B}_N)^{-1}, \end{aligned}$$

where

$$\mathbf{\Omega}_N^{UG} = \mathbf{V}_N^* \mathbf{\Psi}_N^{*-1} \mathbf{V}_N^{*\prime} = (\mathbf{B}_N' \mathbf{\Gamma}_N^{-1} \mathbf{B}_N)^{-1}.$$

Such a decomposition for $\mathbf{\Omega}_N^{UG}$ is precisely the one we would obtain if we regarded $\mathbf{B}_N^*$ as the design matrix and the $\mathbf{f}_t's$ as the parameters in the cross-sectional regression of $\mathbf{u}_{Nt}^*$ on $\mathbf{B}_N^*$.

On the other hand, $\mathbf{f}_{Nt}^{UK}$ is given by the unconditional least squares projection of $\mathbf{f}_t$ on $\mathbf{f}_{Nt}^{UG}$, which for our chosen scaling of the common factors yields

$$\mathbf{f}_{Nt}^{UK} = (\mathbf{I}_k + \mathbf{V}_N^* \mathbf{\Psi}_N^{*-1} \mathbf{V}_N^{*\prime})^{-1} \mathbf{f}_{Nt}^{UG} = \mathbf{V}_N^* \mathbf{\Psi}_N^* (\mathbf{\Psi}_N^* + \mathbf{I}_k)^{-1} \dot{\mathbf{u}}_{Nt}^*,$$

so that

$$\begin{aligned}
\mathbf{\Omega}_{Nt}^{UK} &= \mathbf{V}_N^* (\mathbf{\Psi}_N^* + \mathbf{I}_k)^{-1} \mathbf{V}_N^{*\prime} \left[\mathbf{\Lambda}_t + \mathbf{V}_N^* \mathbf{\Psi}_N^{*1/2} \dot{\mathbf{\Gamma}}_{Nt}^* \mathbf{\Psi}_N^{*1/2} \mathbf{V}_N^{*\prime} \right] \mathbf{V}_N^* (\mathbf{\Psi}_N^* + \mathbf{I}_k)^{-1} \mathbf{V}_N^{\prime *} \\
&= \mathbf{\Omega}_N^{UK} \left[\mathbf{\Lambda}_t + \mathbf{V}_N^* \mathbf{\Psi}_N^{*1/2} \dot{\mathbf{\Gamma}}_{Nt}^* \mathbf{\Psi}_N^{*1/2} \mathbf{V}_N^{*\prime} \right] \mathbf{\Omega}_N^{UK} \\
&= (\mathbf{B}_N' \mathbf{\Gamma}_N^{-1} \mathbf{B}_N + \mathbf{I}_k)^{-1} (\mathbf{\Lambda}_t + \mathbf{B}_N' \mathbf{\Gamma}_N^{-1} \mathbf{\Gamma}_{Nt} \mathbf{\Gamma}_N^{-1} \mathbf{B}_N) (\mathbf{B}_N' \mathbf{\Gamma}_N^{-1} \mathbf{B}_N + \mathbf{I}_k)^{-1}
\end{aligned}$$

where

$$\mathbf{\Omega}_N^{UK} = \mathbf{V}_N^* (\mathbf{I}_k + \mathbf{\Psi}_N^*)^{-1} \mathbf{V}_N^{\prime *} = (\mathbf{B}_N' \mathbf{\Gamma}_N^{-1} \mathbf{B}_N + \mathbf{I}_k)^{-1}$$

A summary measure of their relative unconditional efficiency can then be obtained as the ratio of the generalised variances of the prediction errors in $\mathbf{f}_{Nt}^{UK}$ and $\mathbf{f}_{Nt}^{UG}$. In particular,

$$\begin{aligned}
\frac{\left|\mathbf{\Omega}_N^{UK}\right|}{\left|\mathbf{\Omega}_N^{UG}\right|} &= \frac{\left|\mathbf{V}_N^* (\mathbf{I}_k + \mathbf{\Psi}_N^*)^{-1} \mathbf{V}_N^{\prime *}\right|}{\left|\mathbf{V}_N^* \mathbf{\Psi}_N^{*-1} \mathbf{V}_N^{\prime *}\right|} = \prod_{j=1}^{k} \frac{\psi_{jjN}^*}{1 + \psi_{jjN}^*} \\
&= R_c^2 \left(\mathbf{f}_{Nt}^{UG}, \mathbf{f}_t \right) = \prod_{j=1}^{k} \rho_j^2 \left(\mathbf{f}_{Nt}^{UG}, \mathbf{f}_t \right)
\end{aligned}$$

where $R_c^2 \left(\mathbf{f}_{Nt}^{UG}, \mathbf{f}_t \right)$ is the so-called unconditional coefficient of multiple correlation between $\mathbf{f}_t$ and $\mathbf{f}_{Nt}^{UG}$, which provides an overall goodness of fit measure for the multivariate unconditional regression of $\mathbf{f}_t$ on $\mathbf{f}_{Nt}^{UG}$, and $\rho_j \left(\mathbf{f}_{Nt}^{UG}, \mathbf{f}_t \right)$ the j^{th} unconditional canonical correlation between $\mathbf{f}_t$ and $\mathbf{f}_{Nt}^{UG}$ (see e.g. Dhrymes, 1970).

This result confirms that $\omega_{jjN}^{UK} = [\mathbf{\Omega}_N^{UK}]_{jj}$ is always smaller than $\omega_{jjN}^{UG} = [\mathbf{\Omega}_N^{UG}]_{jj}$, especially so when there is almost perfect "collinearity" between the columns of $\mathbf{B}_N^*$, as indicated by its singular values ψ_{jjN}^* being close to 0. Given that $\mathbf{f}_{Nt}^{UG} = \mathbf{f}_t + (\mathbf{B}_N'\mathbf{\Gamma}_N^{-1}\mathbf{B}_N)^{-1}\mathbf{B}_N'\mathbf{\Gamma}_N^{-1}\boldsymbol{\varepsilon}_{Nt}$, and that $\mathbf{f}_{Nt}^{UK}$ is the fitted value from the unconditional least squares regression of $\mathbf{f}_t$ on $\mathbf{f}_{Nt}^{UG}$, the result also says that $\mathbf{f}_{Nt}^{UK}$ is relatively more efficient than $\mathbf{f}_{Nt}^{UG}$ the smaller the signal to noise ratio, as measured by $R_c^2\left(\mathbf{f}_{Nt}^{UG}, \mathbf{f}_t\right)$. Notice that this is exactly when the precision of $\mathbf{f}_{Nt}^{UG}$ is relatively low (see Lehmann and Modest, 1985).

However, when we compare the conditional efficiency of $\mathbf{f}_{Nt}^{UK}$ and $\mathbf{f}_{Nt}^{UG}$, we obtain that

$$\frac{\left|\mathbf{\Omega}_{Nt}^{UK}\right|}{\left|\mathbf{\Omega}_{Nt}^{UG}\right|} = \frac{\left|\mathbf{\Omega}_{N}^{UK}\right|}{\left|\mathbf{\Omega}_{N}^{UG}\right|} \frac{\left|\mathbf{\Lambda}_t + \mathbf{V}_N^*\mathbf{\Psi}_N^{*1/2}\dot{\mathbf{\Gamma}}_{Nt}^*\mathbf{\Psi}_N^{*1/2}\mathbf{V}_N^{*\prime}\right|}{\left|\mathbf{V}_N^*\mathbf{\Psi}_N^{*1/2}\dot{\mathbf{\Gamma}}_{Nt}^*\mathbf{\Psi}_N^{*1/2}\mathbf{V}_N^{*\prime}\right|} \frac{\left|\mathbf{\Omega}_{N}^{UK}\right|}{\left|\mathbf{\Omega}_{N}^{UG}\right|} = \frac{\left[R_c^2\left(\mathbf{f}_{Nt}^{UG}, \mathbf{f}_t\right)\right]^2}{R_{at}\left(\mathbf{f}_{Nt}^{UH}, \mathbf{f}_t\right)} \tag{13}$$

where $R_{at}\left(\mathbf{f}_{Nt}^{UH}, \mathbf{f}_t\right)$ is the so-called conditional coefficient of alienation between $\mathbf{f}_{Nt}^{UH} = \mathbf{f}_t + \mathbf{B}_N\mathbf{\Gamma}_N^{-1}\boldsymbol{\varepsilon}_{Nt}$ and $\mathbf{f}_t$, which is such that $1 - R_{at}^2(\mathbf{f}_{Nt}^{UH}, \mathbf{f}_t)$ provides an alternative overall goodness of fit measure for the multivariate conditional regression of $\mathbf{f}_{Nt}^{UH}$ on $\mathbf{f}_t$ (see e.g. Dhrymes, 1970). Given that this expression cannot be guaranteed to be smaller than 1, nothing can been said in general about $\mathbf{\Omega}_{Nt}^{UG}$ versus $\mathbf{\Omega}_{Nt}^{UK}$, other than their average difference $\mathbf{\Omega}_N^{UG} - \mathbf{\Omega}_N^{UK} = \mathbf{V}_N^*(\mathbf{\Psi}_N^* + \mathbf{\Psi}_N^{*2})^{-1}\mathbf{V}_N^{\prime *}$ is a p.s.d. matrix. In fact, it is fairly easy to find instances in which $\mathbf{\Omega}_{Nt}^{UK} - \mathbf{\Omega}_{Nt}^{UG}$ will be p.s.d. (e.g. when $\dot{\mathbf{\Gamma}}_{Nt}^* = \mathbf{I}_k$ and $\mathbf{\Lambda}_t - (2\mathbf{I}_k + \mathbf{V}_N^*\mathbf{\Psi}_N^{*-1}\mathbf{V}_N^{*\prime})$ is p.s.d.). This apparent paradox is a direct consequence of the fact that neither $\mathbf{f}_{Nt}^{UG}$ nor $\mathbf{f}_{Nt}^{UK}$ use the information in Φ_{t-1} to form portfolios.

3.4.2 Unconditional vs Conditional GLS

But given that the main innovation of this paper is the introduction of basis portfolios that use the information in Φ_{t-1} to combine the underlying assets, an important question that we have to address is how much more efficient such dynamic investment strategies are relative to their passive counterparts in replicating the underlying factors. In this respect, it is possible to show using the same framework as before that

$$\mathbf{f}_{Nt}^{G} = \mathbf{f}_t + \mathbf{V}_N^{*}\mathbf{\Psi}_N^{*-1/2}\dot{\boldsymbol{\eta}}_{Nt}^{*}$$

where $\dot{\boldsymbol{\eta}}_{Nt}^{*} = \dot{\boldsymbol{\varepsilon}}_{Nt}^{*} - \dddot{\mathbf{\Gamma}}_{Nt}^{*}\ddot{\mathbf{\Gamma}}_{Nt}^{*-1}\ddot{\boldsymbol{\varepsilon}}_{Nt}^{*}$ is the residual from the conditional regression of $\dot{\boldsymbol{\varepsilon}}_{Nt}^{*}$ on $\ddot{\boldsymbol{\varepsilon}}_{Nt}^{*}$. Hence,

$$\mathbf{\Omega}_{Nt}^{G} = \mathbf{V}_N^{*}\mathbf{\Psi}_N^{*-1}\dot{\mathbf{\Upsilon}}_{Nt}^{*}\mathbf{\Psi}_N^{*-1/2}\mathbf{V}_N^{*\prime}$$

where

$$\dot{\mathbf{\Upsilon}}_{Nt}^{*} = V\left(\dot{\boldsymbol{\eta}}_{Nt}^{*}\middle|\,\Phi_{t-1}\right) = \dot{\mathbf{\Gamma}}_{Nt}^{*} - \dddot{\mathbf{\Gamma}}_{Nt}^{*}\ddot{\mathbf{\Gamma}}_{Nt}^{*-1}\dddot{\mathbf{\Gamma}}_{Nt}^{*}$$

and

$$\mathbf{\Omega}_{N}^{G} = \mathbf{V}_N^{*}\mathbf{\Psi}_N^{*-1}\dot{\mathbf{\Upsilon}}_{N}^{*}\mathbf{\Psi}_N^{*-1}\mathbf{V}_N^{*\prime}$$

with $\dot{\mathbf{\Upsilon}}_{N}^{*} = V\left(\dot{\boldsymbol{\eta}}_{Nt}^{*}\right) = E(\dot{\mathbf{\Upsilon}}_{Nt}^{*})$. Hence, both basis portfolios coincide iff $\dddot{\mathbf{\Gamma}}_{Nt}^{*} = \mathbf{0}$. Therefore, the difference between $\mathbf{f}_{Nt}^{G}$ and $\mathbf{f}_{Nt}^{UG}$ stands from the fact that while $\dot{\boldsymbol{\varepsilon}}_{Nt}^{*}$ and $\ddot{\boldsymbol{\varepsilon}}_{Nt}^{*}$ are unconditionally uncorrelated, they will generally be conditionally correlated. That means that there is often useful information in $\ddot{\mathbf{u}}_{Nt}^{*}$ about $\mathbf{f}_t$ even though the corresponding factor loadings are zero, but that information cannot be exploited without using time-varying weights. Moreover, since $\dot{\mathbf{\Gamma}}_{Nt}^{*} - \dot{\mathbf{\Upsilon}}_{Nt}^{*} = \dddot{\mathbf{\Gamma}}_{Nt}^{*}\ddot{\mathbf{\Gamma}}_{Nt}^{*-1}\dddot{\mathbf{\Gamma}}_{Nt}^{*}$ is a p.s.d. matrix by construction, our analysis confirms that so is $\mathbf{\Omega}_{Nt}^{UG} - \mathbf{\Omega}_{Nt}^{G}$.

Once more, we can compute an overall summary measure of relative conditional efficiency as

$$\frac{\left|\mathbf{\Omega}_{Nt}^{G}\right|}{\left|\mathbf{\Omega}_{Nt}^{UG}\right|} = \frac{\left|\dot{\mathbf{\Upsilon}}_{Nt}^{*}\right|}{\left|\dot{\mathbf{\Gamma}}_{Nt}^{*}\right|} = R_{at}^2(\dot{\boldsymbol{\varepsilon}}_{Nt}^{*}, \ddot{\boldsymbol{\varepsilon}}_{Nt}^{*}) = \prod_{j=1}^{k}\left[1 - \rho_{jt}^2(\dot{\boldsymbol{\varepsilon}}_{Nt}^{*}, \ddot{\boldsymbol{\varepsilon}}_{Nt}^{*})\right]$$

where $R_{at}^2(\dot{\boldsymbol{\varepsilon}}_{Nt}^{*}, \ddot{\boldsymbol{\varepsilon}}_{Nt}^{*})$ is the conditional coefficient of alienation between $\dot{\boldsymbol{\varepsilon}}_{Nt}^{*}$ and $\ddot{\boldsymbol{\varepsilon}}_{Nt}^{*}$, and $\rho_{jt}(\dot{\boldsymbol{\varepsilon}}_{Nt}^{*}, \ddot{\boldsymbol{\varepsilon}}_{Nt}^{*})$ the j^{th} conditional canonical correlation between $\dot{\boldsymbol{\varepsilon}}_{Nt}^{*}$ and $\ddot{\boldsymbol{\varepsilon}}_{Nt}^{*}$.

The differential effect of using conditional information in forming basis portfolios can be seen by noticing that while the coefficient of alienation in the corresponding unconditional regression is 1, $R^2_{at}(\dot{\boldsymbol{\varepsilon}}^*_{Nt}, \ddot{\boldsymbol{\varepsilon}}^*_{Nt})$ is between 0 and 1. It is conceivable, though, that due to the implicit averaging in our efficiency criterion, the unconditional MSE of $\mathbf{f}^{UG}_{Nt}$, $\boldsymbol{\Omega}^{UG}_N$, might not be much larger than the unconditional MSE of $\mathbf{f}^G_{Nt}$, $\boldsymbol{\Omega}^G_N$. In particular, $\left|\boldsymbol{\Omega}^G_N\right| / \left|\boldsymbol{\Omega}^{UG}_N\right| = \left|\dot{\boldsymbol{\Upsilon}}^*_N\right|$ because $E(\dot{\boldsymbol{\Gamma}}^*_{Nt}) = \mathbf{I}_k$. It turns out that under certain assumptions commonly made regarding random matrices, closed form expressions are available to assess how large these unconditional efficiency gains could be in practice. In particular, let's assume that the conditional covariance matrix of $(\dot{\boldsymbol{\varepsilon}}^{*\prime}_{Nt}, \ddot{\boldsymbol{\varepsilon}}^{*\prime}_{Nt})'$ follows a central Wishart distribution with parameters $(G^{-1}\mathbf{I}_N, G)$, $(G \geq N)$. Then, from Theorems 7.3.5, and 7.3.6 in Anderson (1984) we will have that $\dot{\boldsymbol{\Gamma}}^*_{Nt}$ follows a central Wishart distribution of order k with parameters $(G^{-1}\mathbf{I}_k, G)$, and furthermore, that $\dot{\boldsymbol{\Upsilon}}^*_{Nt}$ and $\ddot{\boldsymbol{\Gamma}}^*_{Nt}\ddot{\boldsymbol{\Gamma}}^{*-1}_{Nt}\ddot{\boldsymbol{\Gamma}}^*_{Nt}$ will follow independent central Wishart distributions of the same order with parameters $(G^{-1}\mathbf{I}_k, G-N+k)$ and $(G^{-1}\mathbf{I}_k, N-k)$ respectively. As a result, $\boldsymbol{\Omega}^G_N = G^{-1}(G-N+k)(\mathbf{V}^*_N\boldsymbol{\Psi}^{*-1}_N\mathbf{V}^{*\prime}_N)$, and $\left|\boldsymbol{\Omega}^G_N\right| / \left|\boldsymbol{\Omega}^{UG}_N\right| = (G-N+k)^k/G^k$, which can be significantly smaller than 1 when N is large relative to k. Under the same distributional assumptions, Theorem 7.5.1 in Anderson (1984) also implies that both $\left|\boldsymbol{\Omega}^{UG}_{Nt}\right|$ and $\left|\boldsymbol{\Omega}^G_{Nt}\right|$ can be written as (proportional to) the product of k mutually independent chi-square distributions with different degrees of freedom, from where the exact distribution for $R^2_{at}(\dot{\boldsymbol{\varepsilon}}^*_{Nt}, \ddot{\boldsymbol{\varepsilon}}^*_{Nt})$ can be obtained. The single factor case, in which $R^2_{at}(\dot{\varepsilon}^*_{Nt}, \ddot{\boldsymbol{\varepsilon}}^*_{Nt})$ is simply one minus the coefficient of determination in the conditional regression of $\dot{\varepsilon}^*_{Nt}$ on $\ddot{\boldsymbol{\varepsilon}}^*_{Nt}$, is particularly illustrative. The differential effect of using the conditional information in forming basis portfolios can be seen by noticing that while the coefficient of determination of the corresponding unconditional regression is 0, $R^2_{at}(\dot{\varepsilon}^*_{Nt}, \ddot{\boldsymbol{\varepsilon}}^*_{Nt})$ follows a beta distribution with parameters $(G-N+1)/2$ and $(G-1)/2$ under the same distributional assumptions, with a mean value of $(G - N + 1)/M$, which somewhat remarkably, turns out to be the same as $\omega^{G}_{11N}/\omega^{UG}_{11N}$.

3.4.3 Other basis portfolios

Analogous expressions can be obtained for many other pairwise comparisons (see Sentana, 2000a). For the sake of brevity, we shall only mention here that in the case of a single factor, it is not possible to prove that ω^{UI}_{Nt} is always larger than ω^{UG}_{Nt}. However, a sharper result is available for the unconditional correlation of f^{UI}_{Nt} with f_t, $R^2_c\left(f^{UI}_{Nt}, f_t\right)$, which is always less than or equal to $R^2_c\left(f^{UG}_{Nt}, f_t\right) = R^2_c\left(f^{UK}_{Nt}, f_t\right)$ in view of Theorem 4, with equality iff $\mathbf{B}_N$ is proportional to $\boldsymbol{\iota}_N$ by the Cauchy-Schwartz inequality. Hence, how good a predictor f^{UI}_{Nt} is will depend of the "correlation" between $\mathbf{B}'_N$ and $\boldsymbol{\iota}'_N$ in the metric of $\mathbf{\Gamma}^{-1}_N$. Notice also that no monotonicity result regarding $R^2_c\left(f^{UI}_{Nt}, f_t\right)$ can be established.

In principle, one would expect f^{A}_{Nt} to be less efficient than f^{UI}_{Nt}, since we are not only ignoring the information contained in $\mathbf{B}_N$, but also in $\mathbf{\Gamma}_N$. It is possible, though, to find counterexamples in which ω^{A}_{N} is smaller than ω^{UI}_{N}. But note once more that the unconditional correlation of f^{A}_{Nt} with f_t, $R^2_c\left(f^{A}_{Nt}, f_t\right)$ can never exceed $R^2_c\left(f^{UK}_{Nt}, f_t\right) = R^2_c\left(f^{UG}_{Nt}, f_t\right)$ by virtue of Theorem 4.

3.5 Large N Sample Results

A minimum requirement for any factor representing portfolio is that it converges in unconditional MS to $\mathbf{f}_t$ as $N \to \infty$. But as all the basis portfolios that we are considering have the same mean as the factors, we only need to show that the unconditional variance of forecasting errors vanishes asymptotically.

Let's start with the OLS estimates $\mathbf{f}^{O}_{Nt}$, and define $\gamma_1 = \sup_N \delta_1(\mathbf{\Gamma}_N) < \infty$ and $\gamma_\infty = \inf_N \delta_N(\mathbf{\Gamma}_N) > 0$, where $\delta_1(\mathbf{A})$ and $\delta_n(\mathbf{A})$ denote the largest and smallest eigenvalues of the $n \times n$ matrix $\mathbf{A}$. Given that $\mathbf{\Omega}^{O}_{N} = (\mathbf{B}'_N\mathbf{B}_N)^{-1}\ (\mathbf{B}'_N\mathbf{\Gamma}_N\mathbf{B}_N)\ (\mathbf{B}'_N\mathbf{B}_N)^{-1} \leq (\mathbf{B}'_N\mathbf{B}_N)^{-1}\gamma_1$, then the norm of $\mathbf{\Omega}^{O}_{N}$ goes to 0 because the norm of $(\mathbf{B}_N'\mathbf{B}_N)$ goes to ∞ in view of our assumption that the common factors are pervasive. As the unconditional forecast error variances of $\mathbf{f}_{Nt}^K$, $\mathbf{f}_{Nt}^G$, $\mathbf{f}_{Nt}^{UK}$, $\mathbf{f}_{Nt}^{UG}$, $\mathbf{f}_{Nt}^{UKG}$ and $\mathbf{f}_{Nt}^{KUG}$ are at most as large (in the matrix sense) as that of $\mathbf{f}_{Nt}^O$ for any N, these estimators are also MS consistent.

Let's now consider $f_{Nt}^A = \bar{\beta}_N f_t + \bar{\varepsilon}_{Nt}$. Since $V(\bar{\varepsilon}_{Nt}) \leq \gamma_1/N$, it is clear that f_N^A converges to a multiple of f_t, provided that $\bar{\beta}_N$ does not converge to 0, in which case f_{Nt}^A will converge to 0 too.[12] A similar situation arises with f_{Nt}^{UI}, which will also converge to a multiple of f_t unless $\bar{\beta}_N \to 0$, in which case $\breve{\beta}_N$ will converge to 0 too, because $(\gamma_\infty/\gamma_1)\bar{\beta}_N \leq \breve{\beta}_N \leq (\gamma_1/\gamma_\infty)\bar{\beta}_N$.

If $\mathbf{u}_{Nt}$ were (conditionally) normally distributed for all N, so would be the different factor representing portfolios and the corresponding forecast errors. But even if this is not the case, under suitable regularity conditions on the dependence of the idiosyncratic components of returns, it would be possible to prove that for large N, the forecasting errors scaled by $\sqrt{N}$ would be approximately normally distributed with zero mean and the corresponding MSEs as variances.[13] Notice, though, that this is totally compatible with the distribution of the actual predictors being very close to that of $\mathbf{f}_t$, and hence possibly highly non-normal. Such results would allow us to give confidence intervals of approximately correct size for N large enough. Once more, the superiority of $\mathbf{f}_{Nt}^K$ is confirmed by the fact that it would provide the narrowest intervals of all for any given size.

[12] Note that although a zero average factor loading coefficient is an unlikely event, there is nothing in the assumptions made in section 2 which prevents its ocurrence. In particular, since the requirement for the pervasiveness of the common factor (i.e. $\psi_N^2 = N(\bar{\beta}_N^2 + \bar{\sigma}_N^2) \to \infty$ as $N \to \infty$), is ensured by the necessary and sufficient condition for the existence of a riskless limit asset (i.e. $N\bar{\sigma}_N^2 \to \infty$), we may well have situations in which $\bar{\beta}_N \to 0$, and even $N\bar{\beta}_N^2 \to 0$ (see Sentana, 1997).

[13] For instance, if the idiosyncratic errors were not only unconditionally uncorrelated, but also independent across assets, one could apply a standard central limit theorem for heterogenously distributed random variables to $\sqrt{N}\bar{\varepsilon}_{Nt}$, and show that $\sqrt{N}(f_{Nt}^A - \bar{\beta} f_t) \xrightarrow{d} N(0, \bar{\gamma})$ as $N \to \infty$, where $\bar{\beta} = \lim_{N\to\infty} N^{-1}\sum_{i=1}^N \beta_i$ and $\bar{\gamma} = \lim_{N\to\infty} N^{-1}\sum_{i=1}^N \gamma_{ii}$ (see Zaffaroni, 2000).

3.6 The effects of parameter uncertainty

It is important to stress that the efficiency analysis in Section 3.4 is predicated on all the model parameters, $\boldsymbol{\theta}$ say, being known. For our purposes, they can be divided into the following groups:

1. Static factor model parameters:
   (a) Factor loadings $\mathbf{b}_N = vec(\mathbf{B}_N)$
   (b) Unconditional idiosyncratic variances $\boldsymbol{\gamma}_N = vecd(\boldsymbol{\Gamma}_N)$
2. Risk prices $\boldsymbol{\tau}$
3. Conditional variance parameters:
   (a) $\boldsymbol{\psi}$ and $\boldsymbol{\delta}$, which only enter through the common factor variances $\boldsymbol{\Lambda}_t$
   (b) $\boldsymbol{\phi}$ and $\boldsymbol{\rho}$, which only enter through the standardised idiosyncratic variances $\boldsymbol{\Gamma}^*_{Nt} = \boldsymbol{\Gamma}_N^{-1/2}\boldsymbol{\Gamma}_{Nt}\boldsymbol{\Gamma}_N^{-1/2}$.

In reality, of course, $\boldsymbol{\theta}$ will be estimated, and therefore subject to measurement error. The issue of robustness becomes then relevant, because the more efficient basis portfolios typically require knowledge of more parameters than the less efficient ones. For instance, Lehmann and Modest (1985) present examples in which for finite N, f_t^{UI} is more robust than f_t^{UG} when the factor loadings are unknown. The purpose of this subsection is to analyse the effects of parameter uncertainty on our results. In view of our motivation, though, we shall only do so in a sampling framework in which the number of risky assets N is fixed, while the number of time series observations T grows without bounds.

There are two standard ways to handle parameter uncertainty in our signal extraction set up. The first possibility is to take a Bayesian perspective, and repeat the analysis in Section 3.1.1 conditioning on the available sample observations, but not on the unknown values of the parameters. Specifically, we would need to find the portfolio weights $\mathbf{w}_{Nt}$ that minimise

$$E\left[(\mathbf{f}_t - \mathbf{w}'_{Nt}\mathbf{u}_{Nt})(\mathbf{f}_t - \mathbf{w}'_{Nt}\mathbf{u}_{Nt})' |\, \Phi^{t-T}_{Nt-1}\right]$$

in a matrix sense, where Φ^{t-T}_{Nt-1} is the information set generated by $\mathbf{r}_{Nt-1}, \ldots, \mathbf{r}_{Nt-T}$. In view of Theorem 1, it is easy to see that such a Bayesian least squares projection will be achieved by using:

$$\mathbf{w}^K_{Nt} = E^{-1}(\mathbf{u}_{Nt}\mathbf{u}'_{Nt} |\, \Phi^{t-T}_{Nt-1})E(\mathbf{u}_{Nt}\mathbf{f}'_t |\, \Phi^{t-T}_{Nt-1})$$

which are the weights of the conditional Kalman filter estimates $\mathbf{f}^K_{Nt}$, as long as we interpret the different moments involved as unconditional on the parameters. To find out the required expressions, we can either use predictive densities or the law of iterated expectations. The second route involves integrating the parameters out with respect to their posterior distributions, which depend on the observed sample and the prior information. In order to implement such an approach, though, we would need to strengthen the distributional assumptions for $\mathbf{u}_{Nt}|\Phi^{t-T}_{Nt-1}$ (see Fiorentini et al. (2001) for an application of this analysis to the smoothed estimates of $\mathbf{f}_t$).

The second approach, which is classical in nature, consists in replacing the true parameters by their estimators in the different expressions for the factor representing portfolios. If the number of periods, T, on which we have observations on asset returns is rather large, it is possible to prove that replacing true values with consistent estimates would not seriously affect the unconditional MSEs of the different basis portfolios. In not so large data panels, though, the fact that parameter estimators are used would obviously affect our results, especially if we cannot put much confidence on them because their standard errors are large. Unfortunately, it is generally impossible to obtain closed form expressions for the exact unconditional MSEs except in very simple examples (cf. Magnus and Pesaran, 1991).

For that reason, in the remaining of this section we shall obtain asymptotic expansions up to order T^{-1} of the unconditional MSEs of the estimated basis portfolios, and relate them to the finite sample bias and variance of the parameter estimators. The main advantage of this second approach is that it can be applied to any root-T consistent estimation method. In particular, let $\tilde{\boldsymbol{\theta}}_{t,T}$ denote some extremum estimator that maximises the objective function $L_{t,T}(\boldsymbol{\theta}) = \sum_{s=t-T}^{t-1} l_s(\boldsymbol{\theta})$, and assume that

$$P\left[\lim_{T\to\infty} \sup_{\boldsymbol{\theta}} \left\| \tilde{\boldsymbol{\theta}}_{t,T} - \boldsymbol{\theta}_0 \right\| = 0\right] = 1$$

$$\sqrt{T}\frac{1}{T}\sum_s \frac{\partial l_s(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \to N(\mathbf{0}, \mathcal{I}_0)$$

$$P\left[\lim_{T\to\infty} \left\| \frac{1}{T}\sum_s \frac{\partial l_s^2(\boldsymbol{\theta}_{t,T}^*)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} - \mathcal{J}_0 \right\| = 0\right] = 1$$

where $\boldsymbol{\theta}_0$ are the true values of the parameters, $\boldsymbol{\theta}_{t,T}^*$ is any sequence that converges in probability to $\boldsymbol{\theta}^0$, and $\mathcal{J}_0, \mathcal{I}_0$ are non-stochastic square matrices of order $\dim(\boldsymbol{\theta})$, with $\mathcal{I}_0$ p.d., so that

$$\sqrt{T}\left(\tilde{\boldsymbol{\theta}}_{t,T} - \boldsymbol{\theta}_0\right) \to N(\mathbf{0}, \mathcal{J}_0^{-1}\mathcal{I}_0\mathcal{J}_0^{-1})$$

For instance, $\tilde{\boldsymbol{\theta}}_{t,T}$ will be the maximum likelihood estimator when $l_s(\boldsymbol{\theta})$ is the log of the true conditional density of the observed returns, in which case $\mathcal{I}_0 + \mathcal{J}_0 = 0$ by the information matrix equality.

We are interested in obtaining

$$E_{\theta_0}\left[\left(\tilde{\mathbf{f}}_{Nt}^L - \mathbf{f}_t\right)\left(\tilde{\mathbf{f}}_{Nt}^L - \mathbf{f}_t\right)'\right]$$

where $\tilde{\mathbf{f}}_{Nt}^L$ is a generic mimicking portfolio $\mathbf{f}_{Nt}^L$ evaluated at $\tilde{\boldsymbol{\theta}}_{t,T}$, and E_{θ_0} means an expectation taken with respect to the true distribution of the data. To do so, it is convenient to carry out a second order Taylor expansion of $vech\left[\left(\tilde{\mathbf{f}}_{Nt}^{L}-\mathbf{f}_t\right)\left(\tilde{\mathbf{f}}_{Nt}^{L}-\mathbf{f}_t\right)'\right]$ around $\boldsymbol{\theta}_0$ as follows:

$$vech\left\{\left(\tilde{\mathbf{f}}_{Nt}^{L}-\mathbf{f}_t\right)\left(\tilde{\mathbf{f}}_{Nt}^{L}-\mathbf{f}_t\right)'\right\}=vech\left\{\left(\mathbf{f}_{Nt}^{L}-\mathbf{f}_t\right)\left(\mathbf{f}_{Nt}^{L}-\mathbf{f}_t\right)'\right\}$$
$$+\mathbf{J}_t(\boldsymbol{\theta}_0)(\tilde{\boldsymbol{\theta}}_{t,T}-\boldsymbol{\theta}_0)+\left[\mathbf{I}_{k^2}\otimes(\tilde{\boldsymbol{\theta}}_{t,T}-\boldsymbol{\theta}_0)\right]\mathbf{H}_t(\boldsymbol{\theta}_0)(\tilde{\boldsymbol{\theta}}_{t,T}-\boldsymbol{\theta}_0)+o\left(\left\|(\tilde{\boldsymbol{\theta}}_{t,T}-\boldsymbol{\theta}_0)\right\|^2\right)$$

where $\mathbf{J}_t(\boldsymbol{\theta}_0)$ and $\mathbf{H}_t(\boldsymbol{\theta}_0)$ are the Jacobian and the Hessian matrices respectively, of the vector function above with respect to all the parameters, evaluated at $\boldsymbol{\theta}_0$. Then, under suitable regularity conditions, we obtain

$$\begin{aligned}
& E\left[\left(\tilde{f}_{iNt}^{L}-f_{it}\right)\cdot\left(\tilde{f}_{jNt}^{L}-f_{jt}\right)\Big|\,\Phi_{t-1}\right]-\omega_{ijNt}^{K}=\tilde{\omega}_{ijNt}^{K}-\omega_{ijNt}^{K}\\
= & E\left\{\left[\left(f_{jNt}^{L}-f_{jt}\right)\cdot\partial f_{iNt}^{L}/\partial\boldsymbol{\theta}'+\left(f_{iNt}^{L}-f_{it}\right)\cdot\partial f_{jNt}^{L}/\partial\boldsymbol{\theta}'\right]\Big|\,\Phi_{t-1}\right\}(\tilde{\boldsymbol{\theta}}_{t,T}-\boldsymbol{\theta}_0)\\
& +(\tilde{\boldsymbol{\theta}}_{t,T}-\boldsymbol{\theta}_0)'E\left\{\left[\begin{array}{c}\partial f_{iNt}^{L}/\partial\boldsymbol{\theta}\cdot\partial f_{jNt}^{L}/\partial\boldsymbol{\theta}'+\partial f_{jNt}^{L}/\partial\boldsymbol{\theta}\cdot\partial f_{iNt}^{L}/\partial\boldsymbol{\theta}'\\ +\left(f_{jNt}^{L}-f_{jt}\right)\cdot\partial^2 f_{iNt}^{L}/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'\\ +\left(f_{iNt}^{L}-f_{it}\right)\cdot\partial^2 f_{jNt}^{L}/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'\end{array}\right]\middle|\,\Phi_{t-1}\right\}\\
& \times(\tilde{\boldsymbol{\theta}}_{t,T}-\boldsymbol{\theta}_0)+o_p\left(T^{-1}\right)
\end{aligned}$$

after taking conditional expectations element by element.

Given their pre-eminence in our discussion, we shall concentrate on the conditional Kalman filter scores. A significant advantage of these scores in this context derives from (8), which implies that all the above terms vanish, except those involving the outer product of the first derivatives of $\mathbf{f}_{Nt}^{K}$, whose analytical expressions can be found in Appendix 3. Moreover, since under our assumptions:

$$\left(\tilde{\boldsymbol{\theta}}_{t,T}-\boldsymbol{\theta}_0\right)=\mathcal{J}_0^{-1}\frac{1}{T}\sum_s\frac{\partial l_s(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}+o_p(T^{-1/2})$$

it is clear that the covariances of the cross-products of the estimation errors in $\boldsymbol{\theta}$ with the conditional expected values of the Hessian terms $\mathbf{h}_{ijt}(\boldsymbol{\theta}_0)$ are $o(T^{-1})$, which means that they are negligible relative to the product of the respective expected values. Therefore, we can finally write

$$\begin{aligned}
& E\left[\left(\tilde{f}_{iNt}^{K}-f_{it}\right)\left(\tilde{f}_{jNt}^{K}-f_{jt}\right)\right]-\omega_{ijN}^{K} \\
= \; & \frac{1}{T}tr\left[E\left(\frac{\partial f_{iNt}^{K}}{\partial \boldsymbol{\theta}}\frac{\partial f_{jNt}^{K}}{\partial \boldsymbol{\theta}'}+\frac{\partial f_{jNt}^{K}}{\partial \boldsymbol{\theta}}\frac{\partial f_{iNt}^{K}}{\partial \boldsymbol{\theta}'}\right)\left(\mathcal{J}_{0}^{-1}\mathcal{I}_{0}\mathcal{J}_{0}^{-1}\right)\right]+o\left(T^{-1}\right)
\end{aligned}$$

3.7 An illustrative example

In order to gauge the different quantities discussed in the previous sections, we have computed their values for the following simplified version of the model considered by King et al. (1994):

$$r_{it}=\beta_{i}(\lambda_{t}\tau+f_{t})+\varepsilon_{it} \quad i=1,2,3$$

where

$$\begin{aligned}
f_{t} &= \lambda_{t}^{1/2}f_{t}^{*} \\
\lambda_{t} &= 1+\psi_{1}\left[(f_{t-1}^{k})^{2}+\omega_{t-1}^{k}-1\right]+\delta_{1}(\lambda_{t-1}-1)
\end{aligned} \tag{14}$$

$$\left.\begin{aligned}
\varepsilon_{it} &= \gamma_{iit}^{1/2}\varepsilon_{it}^{*} \\
\gamma_{iit} &= \gamma_{ii}+\phi_{i1}\left[(u_{it-1}-\beta_{i}f_{t-1}^{k})^{2}+\beta_{i}^{2}\omega_{t-1}^{k}-\gamma_{i}\right]+\rho_{i}(\gamma_{iit-1}-\gamma_{ii})
\end{aligned}\right\} \quad i=1,2,3$$

$$\left(\begin{array}{cccc} f_{t}^{*} & \varepsilon_{1t}^{*} & \varepsilon_{2t}^{*} & \varepsilon_{3t}^{*}\end{array}\right)'|\Phi_{t-1}\sim N(\mathbf{0},\mathbf{I}_{4}) \tag{15}$$

and f_{t}^{k} and ω_{t}^{k} are the conditional Kalman filter estimate of f_{t} and its conditional MSE respectively.

Note that although we are forced to consider a fairly simple design for the sake of tractability, it possesses the two relevant features mentioned in the introduction as motivating our work: the cross-sectional dimension is small, and conditioning information plays a crucial role in deriving asset risk premia.

When the conditional variances are time-varying, there are no closed form expressions for the unconditional MSEs of the dynamic basis portfolios. Nevertheless, we can easily obtain them by Monte Carlo integration in a single but very long simulation of the model of size 250,000, either directly as the sample mean squared differences between the factor scores and the true factors, or as the sample mean of the corresponding conditional MSEs. A similar approach can be used to compute $E\left[\partial f_{Nt}^{K}/\partial\boldsymbol{\theta}\cdot\partial f_{Nt}^{K}/\partial\boldsymbol{\theta}'\right]$, provided the required moments are bounded. As for $\mathcal{J}_0^{-1}\mathcal{I}_0\mathcal{J}_0^{-1}$, we consider the maximum likelihood estimator that uses the true conditional density of observed returns, and compute $\mathcal{I}_0 = -\mathcal{J}_0$ as the sample variance of $\partial l_s(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}$ (see Appendix 4 for analytical expressions for the score).

Two sets of values for $\mathbf{b}_N$ have been selected, $\mathbf{b}_N = (1,1,1)'$ and $\mathbf{b}_N = (\sqrt{2}/2, \sqrt{2}/2, -\sqrt{2})'$, corresponding to unit mean and zero dispersion in the $\beta_i' s$, and zero mean and unit dispersion respectively. For each value of $\mathbf{b}_N$, two values of $\boldsymbol{\gamma}_N$ have been selected: $\boldsymbol{\gamma}_N = 3\boldsymbol{\iota}$ and $\boldsymbol{\gamma}_N = .75\boldsymbol{\iota}$, representing low and high signal to noise ratios. Then, for each of the four combinations, we consider two pairs of values for (ψ,δ) and (ϕ_i,ρ_i), namely (0,0) and (.1,.85) in order to compare the classical framework of constant variances with the more empirically realistic GARCH context.[14] In this respect, it is important to emphasise that the parameters that influence the relative performance of the dynamic portfolios vis-a-vis the static ones are the unconditional variances of the conditional variances, and not merely the sum of the ARCH and GARCH coefficients.[15] Finally, we have maintained the risk price τ at .1 throughout.

Note that since we are mainly interested in comparing the relative merits of actively versus passively managed basis portfolios, we have assumed that $\boldsymbol{\Gamma}_N$ is

[14]When an ARCH parameter such as ψ or ϕ_i is 0, then the information matrix is singular because the corresponding GARCH parameter δ or ρ_i is not identified. For that reason, we set the latter parameters to 0 whenever the former are 0, and exclude them from the computation of the information matrix. A formal treatment of the asymptotic distribution of the ML estimator in this special case is beyond the scope of the present paper (see Andrews, 1999).

[15]In the case of a strong Gaussian GARCH(1,1) model with ARCH parameter ψ, GARCH parameter δ, and unit unconditional variance, the unconditional variance of the conditional variance is given by $2\psi^2/(1-3\psi^2-2\psi\delta-\delta^2)$, provided the fourth moment is bounded.

the scalar matrix $\gamma \mathbf{I}_3$ for the sake of conciseness.[16] As a result, $f_t^O = f_t^{UG}$ and $f_t^{UI} = f_t^A$ in all our designs. Note also that when $\mathbf{b}_N = (1,1,1)'$, it is also true that $f_t^A = f_t^{UG}$, while when $\mathbf{b}_N = (\sqrt{2}/2, \sqrt{2}/2, -\sqrt{2})'$, f_t^A is simply the cross-sectional average of the idiosyncratic terms, and therefore, completely uncorrelated with f_t. In this case, in fact, the null portfolio is better than f_t^A in the conditional and unconditional MSE sense, with the same correlation.

Table 1 contains the unconditional MSEs for the different designs, and provides a clear illustration of the analysis in Section 3.4. In particular, the ranking is always

$$\omega^K \leq \omega^{KUG} \leq \omega^{UK} < \omega^G \leq \omega^{UG} \leq \omega^A$$

In addition, the single most important determinant of the performance of f_t^K, f_t^G, f_t^{KUG}, f_t^{UK}, and f_t^{UG} seems to be the unconditional signal to noise ratio, as measured by $R_c^2\left(f_t^{UG}, f_t\right) = 3/(3+\gamma)$, while the crucial parameter for f_t^A is $\bar{\beta}_N$. The unconditional MSE reductions achieved by using time-varying weightings are certainly noticeable, but the effect is relatively small. The same conclusion can be obtained from Table 2, which contains the unconditional correlations between the different basis portfolios and the underlying common factor f_t. The main difference is that the ranking is now

$$R(f_t^K, f_t) \geq R(f_t^G, f_t) \geq R(f_t^{KUG}, f_t) \geq R(f_t^{UK}, f_t) = R(f_t^{UG}, f_t) \geq R(f_t^A, f_t)$$

Importantly, the deterioration in performance due to estimation error shown in the last column of Table 1, is roughly of the same order of magnitude across designs, although it worsens the lower the signal to noise ratio, and the higher the dispersion in the $\beta' s$. In any case, the picture that emerges from these results is that for the sort of (temporal) sample sizes typically encountered in practice, the

[16] However, in computing the effects of parameter uncertainty, we do not use the fact that γ_{ii}, or indeed ϕ_i and ρ_i are the same across assets.

conditional Kalman filter factor representing portfolios do not lose their attractiveness when the parameters of the model have to be estimated.

As we have already mentioned, though, unconditional MSEs mask potentially important features that differentiate the performance of the various basis portfolios over time. For that reason, Table 3 contains the three quartiles, and the smallest and largest values of the distributions of ratios of conditional MSEs, as summary statistics of the conditional efficiency of f_t^K relative to the other mimicking portfolios. A noticeable result is that in all the cases that we analyse ($f_t^G, f_t^{KUG}, f_t^{UK}$, and f_t^{UG}), relative performance seems to depend on the signal to noise ratio, but not on the mean or the dispersion of the $\beta's$. The two other evident features are that the different Kalman filter-based portfolios clearly outperform both GLS portfolios, and that within each group, dynamic portfolios are significantly better than static ones, not only in terms of average performance, but also in terms of dispersion.

4 Conclusions

In this paper, we study the statistical properties of several ways of constructing factor representing portfolios in the context of the dynamic version of the APT developed by King et al. (1994), in which changing information implies a changing risk perception by the agents. Methodologically, our major contributions are to focus the problem in its natural signal extraction framework, and to use conditional information to form better mimicking portfolios.

In this context, we show that although many basis portfolios converge in MS to the factors as the number of assets increases, there are some that may be substantially more efficient in relatively small (cross-sectional) samples, in the sense of having lower variability in their "tracking errors". In particular, we prove that the dynamic basis portfolios generated by the conditional Kalman filter updating recursions, are the *best* (in the conditional and unconditional MSE sense) mimicking portfolios possible for any conditional distribution of returns. In addition, we prove that these basis portfolios maximise conditional correlation with the common factors.

Our results also suggest that the efficiency gains of using these dynamic portfolios instead of the static ones considered so far in the literature, could be substantial over time, although the average difference is relatively small. Asymptotic expansions of unconditional MSEs show that for the sort of (temporal) sample sizes typically encountered in practice, these conditional Kalman filter factor representing portfolios do not lose their attractiveness when the parameters of the model have to be estimated.

Finally, given that a maintained assumption of our analysis is that the first and second conditional moments of asset returns are correctly specified, the study of the effects of model misspecification on the consistency and efficiency of the different procedures constitutes a fruitful avenue for further research.

Appendix

A Proofs of results

A.1 Theorem 1

First, it is clear that because $E(\mathbf{f}_t|\Phi_{t-1}) = \mathbf{0}$ and $E(\mathbf{u}_{Nt}|\Phi_{t-1}) = \mathbf{0}$, we obtain a conditionally unbiased predictor of $\mathbf{W}'(\Phi_{t-1})\mathbf{f}_t$ iff $\mathbf{c}(\Phi_{Nt-1}) = \mathbf{0}$, so that we can effectively restrict ourselves to the class of "conditionally linear" predictors $\mathbf{D}'(\Phi_{t-1})\mathbf{u}_{Nt}$. But then, the solution to our problem is simply the conditionally linear least squares projection of $\mathbf{W}'(\Phi_{t-1})\mathbf{f}_t$ on the conditionally (closed) linear subspace generated by $\mathbf{u}_{Nt}$, $\mathcal{U}_{Nt}$ (see Hansen and Richard, 1987). Since such a projection will be given by $\mathbf{W}'(\Phi_{t-1})\mathbf{\Lambda}_t\mathbf{B}'_N\mathbf{\Sigma}^{-1}_{Nt}\mathbf{u}_{Nt}$, our conditional result follows. In addition, from the properties of projections, $E\left[\mathbf{D}'(\Phi_{t-1})\mathbf{u}_{Nt}\left(\mathbf{f}_t - \mathbf{f}^K_{Nt}\right)'\middle|\,\Phi_{t-1}\right] = \mathbf{0}$, which together with $E(\mathbf{f}_t|\Phi_{t-1}) = \mathbf{0}$ and $E(\mathbf{u}_{Nt}|\Phi_{t-1}) = \mathbf{0}$ prove (8). Finally, note that by the law of iterated expectations

$$\mathbf{\Omega}^{CL}_N - \mathbf{\Omega}^K_N = E\left(\mathbf{\Omega}^{CL}_{Nt} - \mathbf{\Omega}^K_{Nt}\right)$$

where $\mathbf{\Omega}^{CL}_N$ and $\mathbf{\Omega}^{CL}_{Nt}$ are the unconditional and conditional MSE of an arbitrary conditionally linear factor representing portfolio, $\mathbf{\Omega}^K_N$ is the unconditional MSE of $\mathbf{f}^K_{Nt}$, and the right hand side expectation is taken with respect to the unconditional distribution of Φ_{t-1}. But since a scalar weighted average of p.s.d. matrices with non-negative weights is a p.s.d. matrix, our unconditional result follows. □

A.2 Theorem 2:

First of all, it is clear that since the conditional correlation will be invariant to conditionally affine transformations, we can take $c(\Phi_{t-1}) = 0$ without loss of generality. Then, we can formally characterise the portfolio weights that maximise the (squared) conditional correlation between $\mathbf{d}'(\Phi_{t-1})\mathbf{u}_{Nt}$ and $\mathbf{w}'(\Phi_{t-1})\mathbf{f}_t$ as

$$\mathbf{d}^*(\Phi_{t-1}) = \arg\max_{\mathbf{d}(\Phi_{t-1})} \frac{\mathbf{d}'(\Phi_{t-1})\mathbf{B}\mathbf{\Lambda}_t\mathbf{w}(\Phi_{t-1})\mathbf{w}'(\Phi_{t-1})\mathbf{\Lambda}_t\mathbf{B}\mathbf{d}(\Phi_{t-1})}{\mathbf{d}'(\Phi_{t-1})\mathbf{\Sigma}_t\mathbf{d}(\Phi_{t-1})\cdot\mathbf{w}'(\Phi_{t-1})\mathbf{\Lambda}_t\mathbf{w}(\Phi_{t-1})}$$

For each value of the conditioning variables, this is a standard algebraic problem with a well-known solution. In particular, $\mathbf{d}^*(\Phi_{t-1})$ is given by the eigenvector associated with the maximum eigenvalue of the rank 1 matrix $\mathbf{B}\mathbf{\Lambda}_t\mathbf{w}(\Phi_{t-1})$ $\mathbf{w}'(\Phi_{t-1})\mathbf{\Lambda}_t\mathbf{B}$ in the metric of $\mathbf{\Sigma}_t$. That is,

$$\mathbf{d}^*(\Phi_{t-1}) = \frac{\mathbf{\Sigma}_t^{-1}\mathbf{B}\mathbf{\Lambda}_t\mathbf{w}(\Phi_{t-1})}{\sqrt{\mathbf{w}'(\Phi_{t-1})\mathbf{\Lambda}_t\mathbf{B}'\mathbf{\Sigma}_t^{-1}\mathbf{B}\mathbf{\Lambda}_t\mathbf{w}(\Phi_{t-1})}}$$

as required. $\square$

A.3 Theorem 3:

For each value of the conditioning variables, this problem has the well-known solution $\mathbf{D}^{**}(\Phi_{t-1}) = \mathbf{\Gamma}_{Nt}^{-1}\mathbf{B}_N(\mathbf{B}_N'\mathbf{\Gamma}_{Nt}^{-1}\mathbf{B}_N)^{-1}\mathbf{W}(\Phi_{t-1})$ (see Magnus and Neudecker, 1988). Then, by the law of iterated expectations, we can again prove that $\mathbf{W}'(\Phi_{t-1})\mathbf{f}_{Nt}^G$ also minimises the unconditional MSE within the same class. $\square$

B Relation between conditional GLS factors and principal components

The following lemma establishes the relation between the spectral decomposition of $\mathbf{\Gamma}_{Nt}^{-1/2}\mathbf{\Sigma}_{Nt}\mathbf{\Gamma}_{Nt}^{-1/2}$ and the spectral decomposition of $\mathbf{\Lambda}_t^{1/2}\mathbf{B}_N'\mathbf{\Gamma}_{Nt}^{-1}\mathbf{B}_N\mathbf{\Lambda}_t^{1/2}$.

Lemma 1 *Let $\mathbf{\Delta}_{Nt}$ and $\mathbf{E}_{Nt}$ denote the eigenvalues and eigenvectors of the $k \times k$ matrix $\mathbf{\Lambda}_t^{1/2}\mathbf{B}_N'\mathbf{\Gamma}_{Nt}^{-1}\mathbf{B}_N\mathbf{\Lambda}_t^{1/2}$. Then the k largest eigenvalues and associated eigenvectors of the matrix $\mathbf{\Gamma}_{Nt}^{-1/2}\mathbf{\Sigma}_{Nt}\mathbf{\Gamma}_{Nt}^{-1/2}$ are given by $(\mathbf{I}_k + \mathbf{\Delta}_{Nt})$ and $\mathbf{\Gamma}_{Nt}^{-1/2}\mathbf{B}_N\mathbf{\Lambda}_t^{1/2}$ $\mathbf{E}_{Nt}\mathbf{\Delta}_{Nt}^{-1/2}$ respectively.*

Proof. Since

$$\mathbf{\Gamma}_{Nt}^{-1/2}\mathbf{\Sigma}_{Nt}\mathbf{\Gamma}_{Nt}^{-1/2} = (\mathbf{\Gamma}_{Nt}^{-1/2}\mathbf{B}_N\mathbf{\Lambda}_t^{1/2})(\mathbf{\Gamma}_{Nt}^{-1/2}\mathbf{B}_N\mathbf{\Lambda}_t^{1/2})' + \mathbf{I}_N$$

the eigenvalues of $\mathbf{\Gamma}_{Nt}^{-1/2}\mathbf{\Sigma}_{Nt}\mathbf{\Gamma}_{Nt}^{-1/2}$ will be one plus the eigenvalues of $(\mathbf{\Gamma}_{Nt}^{-1/2}\mathbf{B}_N\mathbf{\Lambda}_t^{1/2})$ $(\mathbf{\Gamma}_{Nt}^{-1/2}\mathbf{B}_N\mathbf{\Lambda}_t^{1/2})'$, which in turn are the eigenvalues of $\mathbf{\Lambda}_t^{1/2}\mathbf{B}_N'\mathbf{\Gamma}_{Nt}^{-1}\mathbf{B}_N\mathbf{\Lambda}_t^{1/2}$ and $N-k$ zeros. As for the eigenvectors, we just need to orthonormalise the columns of the $N \times k$ matrix $\mathbf{\Gamma}_{Nt}^{-1/2}\mathbf{B}_N\mathbf{\Lambda}_t^{1/2}$, which can be achieved by postmultiplying it by the $k \times k$ matrix $\mathbf{E}_{Nt}\mathbf{\Delta}_{Nt}^{-1/2}$. □

If we then form the conditional least squares projection of $\mathbf{u}_{Nt}^{\dagger} = \mathbf{\Gamma}_{Nt}^{-1/2}\mathbf{u}_{Nt}$ on the space generated by its k first conditionally orthogonal principal components $\mathbf{\Delta}_{Nt}^{-1/2}\mathbf{E}_{Nt}'\mathbf{\Lambda}_t^{1/2}\mathbf{B}_N'\mathbf{\Gamma}_{Nt}^{-1}\mathbf{u}_{Nt}$, we get

$$(\mathbf{\Gamma}_{Nt}^{-1/2}\mathbf{B}_N\mathbf{\Lambda}_t^{1/2}\mathbf{E}_{Nt}\mathbf{\Delta}_{Nt}^{-1/2})(\mathbf{\Delta}_{Nt}^{-1/2}\mathbf{E}_{Nt}'\mathbf{\Lambda}_t^{1/2}\mathbf{B}_N'\mathbf{\Gamma}_{Nt}^{-1}\mathbf{u}_{Nt})$$

where the matrix of conditional projection coefficients are the corresponding eigenvectors. Straightforward algebraic manipulations then show that

$$\mathbf{f}_{Nt}^{G} = (\mathbf{\Lambda}_t^{1/2}\mathbf{E}_{Nt}\mathbf{\Delta}_{Nt}^{-1/2})(\mathbf{\Delta}_{Nt}^{-1/2}\mathbf{E}_{Nt}'\mathbf{\Lambda}_t^{1/2}\mathbf{B}_N'\mathbf{\Gamma}_{Nt}^{-1}\mathbf{u}_{Nt})$$

In particular, if $\mathbf{B}_N'\mathbf{\Gamma}_{Nt}^{-1}\mathbf{B}_N$ is a diagonal matrix, then $\mathbf{f}_{Nt}^{G}$ and the k first conditionally orthogonal principal components of $\mathbf{\Gamma}_{Nt}^{-1/2}\mathbf{u}_{Nt}$ are proportional to each other.

C Derivatives of the conditional Kalman filter basis portfolios

Since we can write $\mathbf{f}_{Nt}^{K}$ as $\mathbf{\Omega}_{Nt}^{K}\mathbf{B}_N'\mathbf{\Gamma}_{Nt}^{-1}\mathbf{u}_{Nt}$, its differential will be given by

$$\begin{aligned} d\mathbf{f}_{Nt}^{K} &= (d\mathbf{\Omega}_{Nt}^{K})\mathbf{B}_N'\mathbf{\Gamma}_{Nt}^{-1}\mathbf{u}_{Nt} + \mathbf{\Omega}_{Nt}^{K}(d\mathbf{B}_N')\mathbf{\Gamma}_{Nt}^{-1}\mathbf{u}_{Nt} \\ &\quad -\mathbf{\Omega}_{Nt}^{K}\mathbf{B}_N'\mathbf{\Gamma}_{Nt}^{-1}(d\mathbf{\Gamma}_{Nt})\mathbf{\Gamma}_{Nt}^{-1}\mathbf{u}_{Nt} + \mathbf{\Omega}_{Nt}^{K}\mathbf{B}_N'\mathbf{\Gamma}_{Nt}^{-1}(d\mathbf{u}_{Nt}) \end{aligned}$$

Hence, we will have that

$$\begin{aligned} \frac{\partial \mathbf{f}_{Nt}^{K}}{\partial \boldsymbol{\theta}'} &= (\mathbf{u}_{Nt}'\mathbf{\Gamma}_{Nt}^{-1}\mathbf{B}_N \otimes \mathbf{I}_k)\mathbf{D}_k\frac{\partial \boldsymbol{\omega}_{Nt}^{K}}{\partial \boldsymbol{\theta}'} + (\mathbf{\Omega}_{Nt}^{K} \otimes \mathbf{u}_{Nt}'\mathbf{\Gamma}_{Nt}^{-1})\frac{\partial \mathbf{b}_N}{\partial \boldsymbol{\theta}'} \\ &\quad -(\mathbf{u}_{Nt}'\mathbf{\Gamma}_{Nt}^{-1} \otimes \mathbf{\Omega}_{Nt}^{K}\mathbf{B}_N'\mathbf{\Gamma}_{Nt}^{-1})\mathbf{E}_N\frac{\partial \boldsymbol{\gamma}_{Nt}}{\partial \boldsymbol{\theta}'} + \mathbf{\Omega}_{Nt}^{K}\mathbf{B}_N'\mathbf{\Gamma}_{Nt}^{-1}\frac{\partial \mathbf{u}_{Nt}}{\partial \boldsymbol{\theta}'} \end{aligned}$$

where $\boldsymbol{\omega}_{Nt}^K = vech(\boldsymbol{\Omega}_{Nt}^K) = \mathbf{D}_k^+ vec(\boldsymbol{\Omega}_{Nt}^K)$, $\mathbf{D}_k$ is the duplication matrix of order k and $\mathbf{D}_k^+$ its Moore-Penrose inverse (see Magnus and Neudecker, 1988). Now, since the differential of $\boldsymbol{\Omega}_{Nt}^K$ is $-\boldsymbol{\Omega}_{Nt}^K d\left(\mathbf{B}_N'\boldsymbol{\Gamma}_{Nt}^{-1}\mathbf{B}_N + \boldsymbol{\Lambda}_t^{-1}\right)\boldsymbol{\Omega}_{Nt}^K$ and

$$\begin{aligned} d\left(\mathbf{B}_N'\boldsymbol{\Gamma}_{Nt}^{-1}\mathbf{B}_N + \boldsymbol{\Lambda}_t^{-1}\right) &= (d\mathbf{B}_N')\boldsymbol{\Gamma}_{Nt}^{-1}\mathbf{B}_N + \mathbf{B}_N'\boldsymbol{\Gamma}_{Nt}^{-1}(d\mathbf{B}_N) \\ &\quad -\mathbf{B}_N'\boldsymbol{\Gamma}_{Nt}^{-1}(d\boldsymbol{\Gamma}_{Nt})\boldsymbol{\Gamma}_{Nt}^{-1}\mathbf{B}_N - \boldsymbol{\Lambda}_t^{-1}(d\boldsymbol{\Lambda}_t)\boldsymbol{\Lambda}_t^{-1} \end{aligned}$$

after some algebraic manipulations, we can show that

$$\begin{aligned} \frac{\partial \boldsymbol{\omega}_{Nt}^K}{\partial \boldsymbol{\theta}'} &= \mathbf{D}_k^+\left[-2(\boldsymbol{\Omega}_{Nt}^K \otimes \boldsymbol{\Omega}_{Nt}^K\mathbf{B}_N'\boldsymbol{\Gamma}_{Nt}^{-1})\frac{\partial \mathbf{b}_N}{\partial \boldsymbol{\theta}'}\right. \\ &\quad +(\boldsymbol{\Omega}_{Nt}^K\mathbf{B}_N'\boldsymbol{\Gamma}_{Nt}^{-1} \otimes \boldsymbol{\Omega}_{Nt}^K\mathbf{B}_N'\boldsymbol{\Gamma}_{Nt}^{-1})\mathbf{E}_N\frac{\partial \boldsymbol{\gamma}_{Nt}}{\partial \boldsymbol{\theta}'} \\ &\quad \left.+\,(\boldsymbol{\Omega}_{Nt}^K\boldsymbol{\Lambda}_t^{-1} \otimes \boldsymbol{\Omega}_{Nt}^K\boldsymbol{\Lambda}_t^{-1})\mathbf{E}_k\frac{\partial \boldsymbol{\lambda}_t}{\partial \boldsymbol{\theta}'}\right] \end{aligned}$$

Finally, note that since

$$d\mathbf{u}_{Nt} = d(\mathbf{r}_{Nt} - \mathbf{B}_N\boldsymbol{\Lambda}_t\boldsymbol{\tau}) = -(d\mathbf{B}_N)\boldsymbol{\Lambda}_t\boldsymbol{\tau} - \mathbf{B}_N(d\boldsymbol{\Lambda}_t)\boldsymbol{\tau} - \mathbf{B}_N\boldsymbol{\Lambda}_t(d\boldsymbol{\tau})$$

we can easily see that

$$\frac{\partial \mathbf{u}_{Nt}}{\partial \boldsymbol{\theta}'} = -(\boldsymbol{\tau}'\boldsymbol{\Lambda}_t \otimes \mathbf{I}_N)\frac{\partial \mathbf{b}_N}{\partial \boldsymbol{\theta}'} - (\boldsymbol{\tau}' \otimes \mathbf{B}_N)\mathbf{E}_k\frac{\partial \boldsymbol{\lambda}_t}{\partial \boldsymbol{\theta}'} - \mathbf{B}_N\boldsymbol{\Lambda}_t\frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\theta}'} \tag{A1}$$

To compute these expressions recursively, though, we will need to know the values of $\partial\boldsymbol{\lambda}_t/\partial\boldsymbol{\theta}$ and $\partial\boldsymbol{\gamma}_{Nt}/\partial\boldsymbol{\theta}$. For the particular example in Section 3.7 in particular, it is straightforward to show from (14) and (15) that the required expressions will be given by:

$$\begin{aligned} \frac{\partial \lambda_t}{\partial \boldsymbol{\theta}} &= \psi_1\left(2f_{t-1}^K\frac{\partial f_{t-1}^K}{\partial \boldsymbol{\theta}} + \frac{\partial \omega_{t-1}^K}{\partial \boldsymbol{\theta}}\right) + \delta_1\frac{\partial \lambda_{t-1}}{\partial \boldsymbol{\theta}} \\ &\quad + \left[(f_{t-1}^K)^2 + \omega_{t-1}^K - 1)\right]\frac{\partial \psi_1}{\partial \boldsymbol{\theta}} + (\lambda_{t-1} - 1)\frac{\partial \delta_1}{\partial \boldsymbol{\theta}} \end{aligned}$$

and

$$
\begin{aligned}
\frac{\partial \gamma_{it}}{\partial \boldsymbol{\theta}} &= 2\phi_{i1}\left[-(u_{it-1}-\beta_i f_{t-1}^K)f_{t-1}^K+\beta_i\omega_{t-1}^K\right]\frac{\partial \beta_i}{\partial \boldsymbol{\theta}}+(1-\phi_{i1}-\rho_{i1})\frac{\partial \gamma_i}{\partial \boldsymbol{\theta}} \\
&+\phi_{i1}\left[2(u_{it-1}-\beta_i f_{t-1}^K)\left(\frac{\partial u_{it-1}}{\partial \boldsymbol{\theta}}-\beta_i\frac{\partial f_{t-1}^K}{\partial \boldsymbol{\theta}}\right)+\beta_i^2\frac{\partial \omega_{t-1}^K}{\partial \boldsymbol{\theta}}\right] \\
&+\left[(u_{it-1}-\beta_i f_{t-1}^K)^2+\beta_i^2\omega_{t-1}^K-\gamma_i\right]\frac{\partial \phi_{i1}}{\partial \boldsymbol{\theta}} \\
&+\rho_{i1}\frac{\partial \gamma_{it-1}}{\partial \boldsymbol{\theta}}+(\gamma_{it-1}-\gamma_i)\frac{\partial \rho_{i1}}{\partial \boldsymbol{\theta}}
\end{aligned}
$$

respectively.

D The score of a conditionally heteroskedastic in mean factor model

Bollerslev and Wooldridge (1992) show that the score function $\partial l_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ of any multivariate conditionally heteroskedastic dynamic regression model with conditional mean $\boldsymbol{\mu}_{Nt}$ and covariance matrix $\boldsymbol{\Sigma}_{Nt}$ is given by the following expression:

$$
\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}=\frac{\partial \boldsymbol{\mu}_{Nt}'}{\partial \boldsymbol{\theta}}\boldsymbol{\Sigma}_{Nt}^{-1}\mathbf{u}_{Nt}+\frac{1}{2}\frac{\partial vec'\left(\boldsymbol{\Sigma}_{Nt}\right)}{\partial \boldsymbol{\theta}}\left(\boldsymbol{\Sigma}_{Nt}^{-1}\otimes\boldsymbol{\Sigma}_{Nt}^{-1}\right)vec\left(\mathbf{u}_{Nt}\mathbf{u}_{Nt}'-\boldsymbol{\Sigma}_{Nt}\right)
$$

Given that $\mathbf{u}_{Nt}=\mathbf{r}_{Nt}-\boldsymbol{\mu}_{Nt}$, we will have that $\partial\boldsymbol{\mu}_{Nt}/\partial\boldsymbol{\theta}'=-\partial\mathbf{u}_{Nt}/\partial\boldsymbol{\theta}'$, which can be found in (A1). Then, since the differential of $\boldsymbol{\Sigma}_{Nt}$ is

$$
d(\mathbf{B}_N\boldsymbol{\Lambda}_t\mathbf{B}_N'+\boldsymbol{\Gamma}_{Nt})=(d\mathbf{B}_N)\boldsymbol{\Lambda}_t\mathbf{B}_N'+\mathbf{B}_N(d\boldsymbol{\Lambda}_t)\mathbf{B}_N'+\mathbf{B}_N\boldsymbol{\Lambda}_t(d\mathbf{B}_N')+d\boldsymbol{\Gamma}_{Nt}
$$

we have that the corresponding Jacobian will be:

$$
\frac{\partial vec\left(\boldsymbol{\Sigma}_{Nt}\right)}{\partial \boldsymbol{\theta}'}=(\mathbf{I}_{N^2}+\mathbf{K}_{NN})(\mathbf{B}_N\boldsymbol{\Lambda}_t\otimes\mathbf{I}_N)\frac{\partial \mathbf{b}_N}{\partial \boldsymbol{\theta}'}+\mathbf{E}_N\frac{\partial \boldsymbol{\gamma}_{Nt}}{\partial \boldsymbol{\theta}'}+(\mathbf{B}_N\otimes\mathbf{B}_N)\mathbf{E}_k\frac{\partial \boldsymbol{\lambda}_t}{\partial \boldsymbol{\theta}'}
$$

where $\mathbf{E}_n$ is the unique $n^2\times n$ "diagonalisation" matrix which transforms $vec(\mathbf{A})$ into $vecd(\mathbf{A})$ as $vecd(\mathbf{A})=\mathbf{E}_n'vec(\mathbf{A})$, and $\mathbf{K}_{mn}$ is the commutation matrix of orders m and n (see Magnus, 1988).

After some straightforward algebraic manipulations, we get:

$$\begin{aligned}
\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \frac{\partial \mathbf{b}'_N}{\partial \boldsymbol{\theta}} vec(\boldsymbol{\Sigma}_{Nt}^{-1}\mathbf{u}_{Nt}\boldsymbol{\tau}'\boldsymbol{\Lambda}_t + \boldsymbol{\Sigma}_{Nt}^{-1}\mathbf{u}_{Nt}\mathbf{u}'_{Nt}\boldsymbol{\Sigma}_{Nt}^{-1}\mathbf{B}_N\boldsymbol{\Lambda}_t - \boldsymbol{\Sigma}_{Nt}^{-1}\mathbf{B}_N\boldsymbol{\Lambda}_t) \\
&+ \frac{\partial \boldsymbol{\tau}'}{\partial \boldsymbol{\theta}}\boldsymbol{\Lambda}_t\mathbf{B}'_N\boldsymbol{\Sigma}_{Nt}^{-1}\mathbf{u}_{Nt} + \frac{1}{2}\frac{\partial \boldsymbol{\gamma}'_{Nt}}{\partial \boldsymbol{\theta}} vecd\left(\boldsymbol{\Sigma}_{Nt}^{-1}\mathbf{u}_{Nt}\mathbf{u}'_{Nt}\boldsymbol{\Sigma}_{Nt}^{-1} - \boldsymbol{\Sigma}_{Nt}^{-1}\right) \\
&+ \frac{1}{2}\frac{\partial \boldsymbol{\lambda}'_t}{\partial \boldsymbol{\theta}} vecd\left(\mathbf{B}'_N\boldsymbol{\Sigma}_{Nt}^{-1}\mathbf{u}_{Nt}\boldsymbol{\tau}' + \mathbf{B}'_N\boldsymbol{\Sigma}_{Nt}^{-1}\mathbf{u}_{Nt}\mathbf{u}'_{Nt}\boldsymbol{\Sigma}_{Nt}^{-1}\mathbf{B}_N - \mathbf{B}'_N\boldsymbol{\Sigma}_{Nt}^{-1}\mathbf{B}_N\right)
\end{aligned}$$

Given that $\boldsymbol{\gamma}_{Nt} > \mathbf{0}$ in view of our assumptions, we can use the standard Woodbury formula to prove that

$$\begin{aligned}
\boldsymbol{\Sigma}_{Nt}^{-1}\mathbf{u}_{Nt} &= \boldsymbol{\Gamma}_{Nt}^{-1}\left[\mathbf{u}_{Nt} - \mathbf{B}_N\mathbf{f}_{Nt}^K\right] \\
\mathbf{B}'_N\boldsymbol{\Sigma}_{Nt}^{-1}\mathbf{u}_{Nt} &= \boldsymbol{\Lambda}_t^{-1}\mathbf{f}_{Nt}^K \\
\boldsymbol{\Lambda}_t\mathbf{B}'_N\boldsymbol{\Sigma}_{Nt}^{-1}\mathbf{u}_{Nt} &= \mathbf{f}_{Nt}^K \\
\boldsymbol{\Sigma}_{Nt}^{-1}\mathbf{u}_{Nt}\mathbf{u}'_{Nt}\boldsymbol{\Sigma}_{Nt}^{-1}\mathbf{B}_N\boldsymbol{\Lambda}_t - \boldsymbol{\Sigma}_{Nt}^{-1}\mathbf{B}_N\boldsymbol{\Lambda}_t &= \boldsymbol{\Gamma}_{Nt}^{-1}\left[\mathbf{u}_{Nt}\mathbf{f}_{Nt}^K - \mathbf{B}_N\left(\mathbf{f}_{Nt}^K\mathbf{f}_{Nt}^K + \boldsymbol{\Omega}_{Nt}^K\right)\right] \\
\boldsymbol{\Sigma}_{Nt}^{-1}\mathbf{u}_{Nt}\mathbf{u}'_{Nt}\boldsymbol{\Sigma}_{Nt}^{-1} - \boldsymbol{\Sigma}_{Nt}^{-1} &= \boldsymbol{\Gamma}_{Nt}^{-1}\begin{bmatrix} (\mathbf{u}_{Nt} - \mathbf{B}_N\mathbf{f}_{Nt}^K)(\mathbf{u}_{Nt} - \mathbf{B}_N\mathbf{f}_{Nt}^K)' \\ +\mathbf{B}_N\boldsymbol{\Omega}_{Nt}^K\mathbf{B}'_N - \boldsymbol{\Gamma}_{Nt} \end{bmatrix}\boldsymbol{\Gamma}_{Nt}^{-1} \\
\mathbf{B}'_N\boldsymbol{\Sigma}_{Nt}^{-1}\mathbf{u}_{Nt}\mathbf{u}'_{Nt}\boldsymbol{\Sigma}_{Nt}^{-1}\mathbf{B}_N - \mathbf{B}'_N\boldsymbol{\Sigma}_{Nt}^{-1}\mathbf{B}_N &= \boldsymbol{\Lambda}_t^{-1}\left[\left(\mathbf{f}_{Nt}^K\mathbf{f}_{Nt}^K + \boldsymbol{\Omega}_{Nt}^K\right) - \boldsymbol{\Lambda}_t\right]\boldsymbol{\Lambda}_t^{-1}
\end{aligned}$$

which greatly simplifies the computations (see Sentana, 2000b).

References

Anderson, T.W., 1984, An introduction to multivariate statistical analysis, 2nd ed. (Wiley, New York).

Andrews, D.W.K., 1999, Estimation when a parameter in on the boundary, Econometrica 67, 1341-1383.

Barlett, M.S., 1938, Methods of estimating mental factors, Nature 141, 609-610.

Bartholomew, D. J., 1987, Latent variable models and factor analysis (Charles Griffin, London).

Bollerslev, T., R.Y. Chou and K.F. Kroner, 1992, ARCH modeling in finance: a review of the theory and empirical evidence, Journal of Econometrics 52, 5-59.

Bollerslev, T. and J. Wooldridge, 1992, Quasi-ML estimation and inference in dynamic models with time-varying variances, Econometric Reviews 11, 143-172.

Chamberlain, G., 1983, Funds, factors and diversification in arbitrage pricing models, Econometrica 51, 1305-1324.

Chamberlain, G. and M. Rothschild, 1983, Arbitrage, factor structure, and mean-variance analysis on large asset markets, Econometrica 51, 1281-1304.

Chen, N.F., 1983, Some empirical tests of the theory of arbitrage pricing, Journal of Finance 38, 1393-1414.

Connor, G., 1984, A unified beta pricing theory, Journal of Economic Theory 34, 13-31.

Connor, G. and R.A. Korajczyk, 1988, Risk and return in an equilibrium APT: application of a new test methodology, Journal of Financial Economics 21, 255-289.

Dhrymes, P., 1970, Econometrics: statistical foundations and applications, (Harper and Row, New York).

Diebold, F.X. and M. Nerlove, 1989, The dynamics of exchange rate volatility:

a multivariate latent factor ARCH model, Journal of Applied Econometrics 4, 1-21.

Engle, R.F., V.M. Ng, and M. Rothschild, 1990, Asset pricing with a factor-ARCH structure: empirical estimates for Treasury Bills, Journal of Econometrics 45, 213-237.

Fang, K.T., S. Kotz, and K.W. Ng, 1990, Symmetric multivariate and related distributions (Chapman and Hall, London).

Fiorentini, G., E. Sentana, and N. Shephard, 2001, Likelihood estimation of latent GARCH structures, mimeo, CEMFI.

Gourieroux, C., A. Monfort, and E. Renault, 1991, A general framework for factor models, mimeo, INSEE.

Grinblatt, M. and S. Titman, 1987, The relation between mean-variance efficiency and arbitrage pricing, Journal of Business 60, 97-112.

Hansen, L.P. and S.F. Richard, 1987, The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models, Econometrica 55, 587-613.

Harvey, A.C., 1981, The econometric analysis of time series (Phillip Allan, London).

Harvey, A.C., 1989, Forecasting, structural models and the Kalman filter (Cambridge University Press, Cambridge).

Harvey, A.C., E. Ruiz, and E. Sentana, 1992, Unobservable component time series models with ARCH disturbances, Journal of Econometrics 52, 129-157.

Huberman, G., 1982, A simple approach to the arbitrage pricing theory, Journal of Economic Theory 28, 183-191.

Huberman, G., S. Kandel, and R.F. Stambaugh, 1987, Mimicking portfolios and exact arbitrage pricing, Journal of Finance 42, 1-9.

Ingersoll, J., 1984, Some results in the theory of arbitrage pricing, Journal of

Finance 39, 1021-1039.

King, M.A., E. Sentana, and S.B. Wadhwani, 1994, Volatility and links between national stock markets, Econometrica 62, 901-933.

Lawley, D.N. and A.E. Maxwell, 1971, Factor analysis as a statistical method, 2nd ed. (Butterworths, London).

Lehmann, B.N., 1992, Notes on dynamic factor pricing models, Review of Quantitative Finance and Accounting 2, 69-87.

Lehmann, B.N. and D.M. Modest, 1985, The empirical foundations of the arbitrage pricing theory: the optimal construction of basis portfolios, NBER Working Paper 1726.

Lehmann, B.N. and D.M. Modest, 1988, The empirical foundations of the arbitrage pricing theory, Journal of Financial Economics 21, 213-254.

Magnus, J.R., 1988, Linear structures (Oxford University Press, New York).

Magnus, J. and H. Neudecker, 1988, Matrix differential calculus with applications in Statistics and Econometrics (Wiley, Chichester).

Magnus, J. and B. Pesaran, 1991, The bias of forecasts from a first-order autoregression, Econometric Theory 7, 222-35.

Roll, R.W., 1977, A critique of the asset pricing theory's tests - part I: on past and potential testability of the theory, Journal of Financial Economics 4, 129-176.

Ross, S.A., 1976, The arbitrage theory of capital asset pricing, Journal of Economic Theory, 13, 341-360.

Ross, S.A., 1978, Mutual fund separation in financial theory - the separating distributions, Journal of Economic Theory 17, 254-286.

Rubin, D.B. and D.T. Thayer, 1982, EM algorithms for ML factor analysis, Psychometrika 57, 69-76.

Sentana, E., 1991, Time-varying volatility and returns on ordinary shares: an empirical investigation, unpublished doctoral dissertation, University of London.

Sentana, E., 1992, Factor representing portfolios in large asset markets, LSE FMG Discussion Paper 135.

Sentana, E., 1997, Risk and return in the Spanish stock market: some evidence from individual assets, Investigaciones Económicas 21, 297-359.

Sentana, E., 2000a, Factor representing portfolios in large asset markets, CEMFI Working Paper 0001.

Sentana, E., 2000b, The likelihood function of conditionally heteroskedastic factor models, Annales d'Economie et de Statistique 58, 1-19.

Sentana, E. and G. Fiorentini, 2001, Identification, estimation and testing of conditionally heteroskedastic factor models, Journal of Econometrics 102, 143-164.

Sheikh, A.M., 1996, BARRA's risk models, BARRA Research Insights, BARRA.

Sorensen, E.H., J.J. Mezrich, and D.N. Thadani, 1993, Currency hedging through portfolio optimization, Journal of Portfolio Management, 19, 78-85.

Stambaugh, R. F., 1984, Asset pricing with information, Journal of Financial Economics 12, 357-369.

Thomson, G.H., 1951, The factorial analysis of human ability, 5th ed. (Houghton-Mifflin, New York).

Zaffaroni, P., 2000, Contemporaneous aggregation of GARCH processes, mimeo, University of Cambridge.

Table 1

Basis portfolios unconditional mean square errors

$\bar{\beta}_N$	$\bar{\sigma}^2_N$	γ	ψ, ϕ_i	δ, ρ_i	f_t^K	f_t^G	f_t^{KUG}	f_t^{UK}	f_t^{UG}	f_t^A	$\tilde{\omega}^K - \omega^K$
1	**0**	**3**	**0**	**0**	*.5*	*1*	*.5*	*.5*	*1*	*1*	16.748/T
			.1	**.85**	.475	.923	.492	*.5*	*1*	*1*	17.824/T
		.75	**0**	**0**	*.2*	*.25*	*.2*	*.2*	*.25*	*.25*	7.532/T
			.1	**.85**	.188	.235	.196	*.2*	*.25*	*.25*	12.939/T
0	**1**	**3**	**0**	**0**	*.5*	*1*	*.5*	*.5*	*1*	*2*	37.683/T
			.1	**.85**	.480	.945	.493	*.5*	*1*	*2*	28.394/T
		.75	**0**	**0**	*.2*	*.25*	*.2*	*.2*	*.25*	*1.25*	13.524/T
			.1	**.85**	.190	.240	.196	*.2*	*.25*	*1.25*	16.240/T

Notes: Numbers in *italics* represent theoretical values.

$\bar{\beta}_N$: cross-sectional average of the factor loadings.

$\bar{\sigma}^2_N$: cross-sectional variance of the factor loadings.

γ: unconditional variance of idiosyncratic risks.

ψ, ϕ_i: ARCH parameters for common and idiosyncratic factors.

δ, ρ_i: GARCH parameters for common and idiosyncratic factors.

f_t^K: conditional Kalman filter mimicking portfolio (based on f_t^G).

f_t^G: conditional GLS mimicking portfolio.

f_t^{KUG}: conditional Kalman filter mimicking portfolio based on f_t^{UG}.

f_t^{UK}: unconditional Kalman filter mimicking portfolio.

f_t^{UG}: unconditional GLS mimicking portfolio.

f_t^A: equally-weighted mimicking portfolio.

$\tilde{\omega}^K - \omega^K$: difference between the unconditional MSE of f_t^K evaluated at ML estimators and true values.

Table 2

Basis portfolios unconditional correlations with f_t

$\bar{\beta}_N$	$\bar{\sigma}^2_N$	γ	ψ, ϕ_i	δ, ρ_i	f_t^K	f_t^G	f_t^{KUG}	f_t^{UK}	f_t^A
1	**0**	**3**	**0**	**0**	*.707*	*.707*	*.707*	*.707*	*.707*
			.1	**.85**	.724	.719	.712	*.707*	*.707*
		.75	**0**	**0**	*.894*	*.894*	*.894*	*.894*	*.894*
			.1	**.85**	.901	.899	.897	*.894*	*.894*
0	**1**	**3**	**0**	**0**	*.707*	*.707*	*.707*	*.707*	*0*
			.1	**.85**	.722	.718	.713	*.707*	*0*
		.75	**0**	**0**	*.894*	*.894*	*.894*	*.894*	*0*
			.1	**.85**	.900	.899	.897	*.894*	*0*

Notes: Numbers in *italics* represent theoretical values.

$\bar{\beta}_N$: cross-sectional average of the factor loadings.

$\bar{\sigma}^2_N$: cross-sectional variance of the factor loadings.

γ: unconditional variance of idiosyncratic risks.

ψ, ϕ_i: ARCH parameters for common and idiosyncratic factors.

δ, ρ_i: GARCH parameters for common and idiosyncratic factors.

f_t^K: conditional Kalman filter mimicking portfolio (based on f_t^G).

f_t^G: conditional GLS mimicking portfolio.

f_t^{KUG}: conditional Kalman filter mimicking portfolio based on f_t^{UG}.

f_t^{UK}: unconditional Kalman filter mimicking portfolio.

f_t^A: equally-weighted mimicking portfolio.

Table 3

Basis portfolios relative conditional efficiency

(min, $q_{.25}$, median, $q_{.75}$, max)

$\bar{\beta}_N$	$\bar{\sigma}^2_N$	γ	ψ, ϕ_i	δ, ρ_i	$\omega_t^K/\omega_t^{KUG}$					ω_t^K/ω_t^G				
1	**0**	**3**	**0**	**0**	*1*					*.5*				
			.1	**.85**	.627	.957	.979	.992	1	.286	.479	.515	.552	.855
		.75	**0**	**0**	*1*					*.8*				
			.1	**.85**	.473	.951	.977	.991	1	.568	.763	.797	.832	.978
0	**1**	**3**	**0**	**0**	*1*					*.5*				
			.1	**.85**	.664	.968	.985	.994	1	.266	.480	.511	.544	.816
		.75	**0**	**0**	*1*					*.8*				
			.1	**.85**	.526	.964	.982	.993	1	.551	.762	.793	.827	.977

$\bar{\beta}_N$	$\bar{\sigma}^2_N$	γ	ψ, ϕ_i	δ, ρ_i	ω_t^K/ω_t^{UK}					ω_t^K/ω_t^{UG}				
1	**0**	**3**	**0**	**0**	*1*					*.5*				
			.1	**.85**	.278	.938	.967	.983	1	.079	.444	.488	.529	.797
		.75	**0**	**0**	*1*					*.8*				
			.1	**.85**	.245	.926	.961	.981	1	.162	.722	.767	.807	.964
0	**1**	**3**	**0**	**0**	*1*					*.5*				
			.1	**.85**	.206	.953	.975	.988	1	.056	.453	.491	.527	.812
		.75	**0**	**0**	*1*					*.8*				
			.1	**.85**	.373	.943	.969	.984	1	.253	.733	.770	.806	.968

Notes: Numbers in *italics* represent ratios of theoretical unconditional MSEs.

$\bar{\beta}_N$: cross-sectional average of the factor loadings.

$\bar{\sigma}^2_N$: cross-sectional variance of the factor loadings.

γ: unconditional variance of idiosyncratic risks.

ψ, ϕ_i: ARCH parameters for common and idiosyncratic factors.

δ, ρ_i: GARCH parameters for common and idiosyncratic factors.

$\omega_t^K/\omega_t^{KUG}$: relative conditional efficiency of the conditional Kalman filter mimicking portfolios based on f_t^G and f_t^{UG}.

ω_t^K/ω_t^G: relative conditional efficiency of conditional Kalman filter and conditional GLS mimicking portfolios.

ω_t^K/ω_t^{UK}: relative conditional efficiency of conditional and unconditional Kalman filter mimicking portfolios.

ω_t^K/ω_t^{UG}: relative conditional efficiency of conditional Kalman filter and unconditional GLS mimicking portfolios.