Sources of Systematic Risk*

Igor Makarov[†] and Dimitris Papanikolaou[‡]

Abstract

Using the restrictions implied by the heteroskedasticity of stock returns, we identify four factors in the U.S. industry returns. The first correlates highly with the market portfolio; the second is a portfolio of stocks that produce investment goods minus stocks that produce consumption goods; the third differentiates between cyclical and noncyclical stocks. The fourth, a portfolio of industries that produce input goods minus the rest of the market, is a robust predictor of excess returns on the market portfolio and bond returns. The extracted factors are shown to contain significant information about future macroeconomic and financial variables.

*We are grateful to Vito Gala, Leonid Kogan, Alexei Onatski, Roberto Rigobon, and Stephen Ross for valuable discussions and comments. We thank participants at MIT Finance lunch, LBS Finance lunch and Panagora Asset Management for helpful comments. We thank John Cochrane, Kenneth French and Motohiro Yogo for sharing their data and Alexei Onatski and Serena Ng for sharing their code. Dimitris Papanikolaou thanks the Zell Center for Risk for financial support.

[†]London Business School. E-mail: imakarov@london.edu.

[‡]Kellogg School of Management. E-mail: d-papanikolaou@kellogg.northwestern.edu.

1 Introduction

In this paper we aim to understand the informational content of factors that affect industry returns. If the economy is subject to multiple shocks, then the cross-section of asset returns should be a valuable source of information about the future state of the economy. The major problem is to separate structural shocks from the noise. Factor analysis partially achieves this goal. However, it suffers from a major drawback. Under the standard assumption of a constant factor covariance matrix, one cannot identify individual factors, but only their linear span. This lack of identification makes interpretation of the corresponding economic shocks difficult and often leads to unstable factors, i.e., factors extracted from different subsamples do not resemble each other.

We relax the standard assumptions of factor analysis and allow the factor variance-covariance matrix to change over time. We require factors to be orthogonal at every point in time, which results in exactly identified factors. The assumption of multiple heteroskedastic factors has intuitive appeal. A wealth of evidence suggests that the world is far from homoskedastic: many macroeconomic quantities, stock and bond returns, and even consumption growth display time-varying volatility[1]. Casual empiricism also seems to support this view. For example, the late 1990s was a relatively calm time for oil stocks. In contrast, by 2008 oil prices had reached all-time highs, and the returns of oil stocks had displayed considerably more fluctuation.

Using identification through heteroskedasticity, we isolate four factors from industry returns. The first factor has similar loadings on all industries, explains the highest proportion of variation in realized returns, and is highly correlated with the value-weighted market portfolio. The second factor is highly correlated with a portfolio of stocks that produce investment goods minus stocks that produce consumption goods (IMC), and might capture shocks to the relative productivity of capital versus consumption goods. The third factor differentiates between cyclical and non-cyclical industries, such as auto manufacturers compared to healthcare providers, and is likely to capture information related to the business cycle. The fourth factor loads positively on industries that produce commodities such as oil, gas, and metals, and negatively on the rest of the market. Therefore, it might capture changes in the relative price of input goods.

Our main finding is that our extracted factors are robust predictors of both financial

[1] Several authors have documented and exploited the presence of heteroskedasticity in macroeconomic and financial variables. An incomplete list includes Officer, 1973; Schwert, 1989; Longstaff and Schwartz, 1992; Lauterbach, 1989; Bansal, Khatchatrian and Yaron, 2005; Baele, Bekaert and Inghelbrecht, 2009.

and macroeconomic variables. The fourth factor contains information about future excess returns on the market portfolio and bond returns in addition to well-known predictors such as the short rate, the term spread, and the dividend yield. This is consistent with the results of Hong, Torous and Valkanov (2007), who find that 12 out of 38 industries lead the market. We find that the predictive power of industries is subsumed by the fourth factor: if this factor is included in the predictive regression, no industry remains a statistically significant predictor of the market. Furthermore, we show that following positive realizations of the first factor, aggregate output, consumption, and investment increase, but unemployment falls. The second factor predicts an increase in the real interest rate, an increase in industrial production and a fall in unemployment. The third factor appears to capture a component of the business cycle that is orthogonal to the first factor: it predicts an increase in industrial production and a fall in unemployment. The positive realizations of the fourth factor are followed by an increase in inflation, and a fall in productivity and consumption growth.

We show that accounting for time-varying volatility significantly improves the stability of the factors: analysis applied to different subsamples identifies essentially the same factors. At the same time, factor heteroskedasticity suggests a possible explanation for the instability we observe in previous applications of the method. The inability to separate the right factors from their linear span often results in arranging factors that are ranked according to their ability to explain the amount of variation in stock returns. If the relative volatility of the factors varies over time, this ranking will be time-dependent. Using Monte Carlo simulations, we show that if the underlying factors are heteroskedastic, then our method significantly outperforms static factor analysis and principal components in identifying the true factors. We perform the analysis by using data from both the United States and the United Kingdom and show that our procedure identifies essentially the same factors in the UK as those identified in the US. Since we do not a priori impose any economic structure on our factors, the fact that the extracted factors are similar across different countries suggests that these factors might be identifying structural economic shocks.

Finally, we explore the ability of the extracted factors to price the cross-section of expected returns. Our four-factor model performs on par with the Fama and French (1993) model with estimated prices of risk being consistent across the set of portfolios. The second and fourth factor carry a negative price of risk, whereas the third factor carries a positive risk premium.

The paper is organized as follows. In Section 2 we briefly review the related literature. In Section 3 we present the econometric framework and we explain how factors are estimated and identified. In Section 4 we provide details of the empirical implementation and present our findings. In Section 5 we explore the informational content of the extracted factors for aggregate quantities and prices. In Section 6 we examine the ability of the extracted factors to price the cross-section of expected returns. In Section 7 we repeat the analysis using data from the United Kingdom. Section 8 concludes.

2 Related Literature

Understanding the forces that jointly determine aggregate output and stock prices has always been at the heart of the macro-finance agenda. Chen, Roll and Ross (1986) show that industrial production growth, default and term premiums are priced risk factors. Connor and Korajczyk (1991) project the Chen, Roll, and Ross factors on factors extracted from stock returns to study the performance of mutual funds. Campbell (1996) presents evidence that revisions in forecasts of future labor income growth is a risk factor that helps price the cross-section of stock returns. Vassalou (2003) finds that news related to future GDP growth is related to the HML and SMB factors of Fama and French (1993). Petkova (2006) relates the HML and SMB factors to innovations in the short-term T-bill, term spread, aggregate dividend yield, and default spread. Similar to Petkova (2006), our focus is on the informational content and interpretation of the extracted factors.

Stock and Watson (2003) provide an excellent review of the literature examining the ability of stock returns to predict financial and macroeconomic variables. Lamont (2001) uses bond and industry returns to construct factor-mimicking portfolios for expectations about future economic variables. He concludes that the resulting portfolios can be useful in forecasting macroeconomic variables. We differ from Lamont (2001), in that we are agnostic about macroeconomic shocks and use heteroskedasticity in stock returns to identify the underlying shocks. Ludvigson and Ng (2007) apply the approach of Stock and Watson (2002) to extract factors from a very large cross-section of macroeconomic and financial variables. The authors find that the extracted factors can forecast excess returns and volatility of the market portfolio. In contrast to Ludvigson and Ng (2007), we extract factors from the cross-section of stock returns and identify them through heteroskedasticity.

Our paper is related to Stock and Watson (1999) and Stock and Watson (2003), who explore the informational content of the market portfolio in predicting future inflation movements. These authors conclude that the relation is very fragile. Rather than focusing on the market portfolio, we use the information contained in the cross-section of returns, and we also examine the relation between stock returns and other economic variables. Consistent with these authors, we find that the market portfolio is a weak predictor of inflation, the same however is not true for other portfolios constructed from the cross-section of returns.

In this paper, we show how to exploit the factors' time-varying volatilities to obtain identification in classical factor analysis. Classical factor analysis assumes that N is relatively small and fixed but $T \to \infty$. The standard model has been extended in several ways. Connor and Korajczyk (1988) develop asymptotic principal component analysis to estimate approximate factor models when T is fixed and $N \to \infty$. Jones (2001) extends the Connor and Korajczyk analysis to allow for time series heteroskedasticity in the factor model residuals. Stock and Watson (2002) consider the case when $N, T \to \infty$ jointly. The various methods described above identify the factors' linear span. To obtain the individual factors, researchers usually assume a particular rotation. Our approach is robust, computationally easy to implement, and can be adjusted to any of the methods above.

Several other papers also explore the idea of using heteroskedasticity in the data for purposes of identification. Using four observable and two latent heteroskedastic factors, specified as GARCH(1,1) processes, King, Sentana and Wadhwani (1994) explore the links between international stock markets. They show that changes in correlations between markets are driven primarily by movements in latent factors. Sentana and Fiorentini (2001) provide a theoretical account of identification in factor models using heteroskedasticity in return series. Rigobon (2003) uses heteroskedasticity of the structural shocks to get identification in simultaneous equation models. Our contribution is that we propose a robust and easily implementable two-stage estimation technique. In the first stage, we use the limited information likelihood method to estimate the unconditional version of the model. In the second stage, we find the correct rotation matrix employing a GMM-type estimator.

3 Econometric Framework

We assume that returns obey the following factor model:

$$\begin{aligned} r_t &= C f_t + \varepsilon_t, \quad \text{t=1,..T}, \\ f_t &= \Lambda_t^{\frac{1}{2}} u_t, \quad \mathrm{E}\Lambda_t = I, \\ u_t &\sim \mathrm{N}(0, I), \quad \varepsilon_t \sim \mathrm{N}(0, \Gamma), \end{aligned} \tag{1}$$

where r_t is an n-vector of returns, f_t - is a k-vector of common factors, C is the n-by-k matrix of factor loadings, with $rank(C) = k$, ε_t is an n-vector of residuals conditionally orthogonal to f_t, Λ_t is a k-by-k diagonal matrix of factor variances at time t, I is an identity matrix of size k, and Γ is k-by-k the diagonal matrix of residual variances. The special case of constant factor covariance matrix, $\Lambda_t = I, \forall t$, corresponds to the traditional factor model.

In the traditional factor model, it is well-known that the matrix of residual variances Γ is identified, but factors are identified only up to an orthogonal rotation matrix $\mathcal{O}$. Harman (1976) provides a textbook treatment. To see this, we note that a matrix $\mathcal{O}$ is called a rotation matrix if $\mathcal{O}'\mathcal{O} = I$. If $\tilde{f}_t = \mathcal{O} f_t$ is a new set of factors after an orthogonal transformation, then their variance-covariance matrix is $\mathcal{O}' I \mathcal{O} = I$. The new factor loadings are $\tilde{C} = C\mathcal{O}'$. But then, the new factors $(\tilde{f}_t)$ are observationally equivalent to the old ones (f_t) since $\tilde{C}\,\tilde{f}_t = C\mathcal{O}'\mathcal{O} f_t = C f_t$. Hence, in the traditional factor analysis framework we can identify the linear subspace created by factors, but not the factors themselves, which is a substantial drawback of the procedure.

If the volatility of factors is time-varying, then the model still implies an unconditional k factor structure for returns, since $E\Lambda_t = I$. More importantly, it imposes additional identification restrictions by requiring that factors be conditionally orthogonal as well. To illustrate the additional restrictions, consider a simple example. Suppose we have two orthogonal factors $f_{1,t}$ and $f_{2,t}$ whose variances at time t equal $\lambda_{1,t}^2$ and $\lambda_{2,t}^2$ respectively. If $\tilde{f}_{1,t} = (f_{1,t} + f_{2,t})/2$ and $\tilde{f}_{2,t} = (f_{1,t} - f_{2,t})/2$ is a linear transformation of the original factors then $E_t(\tilde{f}_{1,t}\tilde{f}_{2,t}) = (\lambda_{1,t}^2 - \lambda_{2,t}^2)/4$. The last equation implies that factors are identified, provided there exists a time t such that $\lambda_{1,t}^2 \neq \lambda_{2,t}^2$. In contrast, in the homoscedastic case, the constraint is degenerate, resulting in underidentification. As long as factor volatilities, λ_{it}, $i = 1..k$ are linearly independent, the factors are identified up to row permutations. Since identification does not depend on the exact dynamics of Λ_t, we do not need to specify unnecessary restrictions on its evolution.

We estimate the model in two stages. In the first stage, we use the limited information likelihood method to estimate the unconditional version of the model. This step does not require any knowledge of Λ_t, and it gives consistent estimates of idiosyncratic variances Γ_k, factors and factor loadings C, up to rotation.

The second stage involves estimating the rotation matrix $\mathcal{O}$ such that the extracted factors are close to being conditionally uncorrelated. We start with the factors f_{it}, $i = 1..K$, which we obtained in the first stage. Since $E\Lambda_t = I$, the factors are by construction unconditionally orthogonal. However, these factors need not be conditionally orthogonal.

Identification boils down to finding the "right" rotation matrix such that the rotated factors are conditionally orthogonal, i.e. the matrix $\mathcal{O}'\Lambda_t\mathcal{O}$ is diagonal for all $t = 1, 2, \ldots, T$.

Any rotation matrix $\mathcal{O} \in SO(k)$ can be represented as a product of $L = k \times (k-1)/2$ elementary rotation matrices T_{ij},

$$\mathcal{O} = \prod_{j>i} T_{ij}\,, \tag{2}$$

where

$$T_{ij} = \begin{array}{c} \\ i \\ \\ j \\ \\ \\ \end{array} \begin{bmatrix} 1 & 0 & \ldots & 0 & 0 \\ 0 & \cos\alpha_l & 0 & \sin\alpha_l & 0 \\ \vdots & 0 & 0 & 0 & \vdots \\ 0 & -\sin\alpha_l & 0 & \cos\alpha_l & 0 \\ 0 & 0 & \ldots & 0 & 1 \end{bmatrix}_{\textstyle\;\;i \qquad\quad j}.$$

The matrix T_{ij} represents a two-dimensional rotation on the plane created by axis i and j. The angle α_l, where $l = 1..L$, is restricted to lie in $[0, 2\pi]$. This representation of the rotation matrix, $\mathcal{O}$, is unique up to row permutations.

In practice, we need to estimate the factor covariance matrix Λ_t. We do not impose any particular restrictions on the volatility process and estimate Λ_t in monthly increments using a rolling window of daily factor realizations over the past six months[2]:

$$\hat{\Lambda}_t = \sum_{s \in S_t} \sum_{j=0}^{w} B(j)\, f_s\, f_{s-j}, \tag{3}$$

where S_t are days between months t and $t-5$. We use the Bartlett window, $B(j)$ of width $w = 2$ to adjust for possible (cross)auto-correlations. As a robustness check, we vary both parameters: window length (three months, one year) and the length of the Bartlett window (3,5,10). In all cases, the results are qualitatively and quantitatively similar.

Due to estimation noise, the matrix $\mathcal{O}'\hat{\Lambda}_t\mathcal{O}$ is unlikely to be diagonal for any choice of the rotation matrix $\mathcal{O}$. Therefore, we use a GMM-type estimator. We choose the $\mathcal{O}^*$ that minimizes the squared off-diagonal elements of the factor correlation matrix. Therefore, $\mathcal{O}^*$

[2]A similar estimator of volatility is also used in French, Schwert and Stambaugh (1987) and Schwert (1989). Andersen, Bollerslev, Diebold and Labys (2003) provides a rigorous treatment of the realized volatility.

is the solution to

$$\mathcal{O}^* = \arg \min_{\alpha_1,\ldots,\alpha_L} \sum_t \sum_{i \neq j} (\hat{\mathcal{P}}_t)^2_{ij}, \tag{4}$$

where the factor correlation matrix, $\hat{\mathcal{P}}_t$, is given by

$$\hat{\mathcal{P}}_t \equiv diag(\mathcal{O}'\hat{\Lambda}_t\mathcal{O})^{-1/2}(\mathcal{O}'\hat{\Lambda}_t\mathcal{O})diag(\mathcal{O}'\hat{\Lambda}_t\mathcal{O})^{-1/2}. \tag{5}$$

The minimization is over the axes of rotation, $\alpha_1, \ldots, \alpha_L$ that characterize the rotation matrix, as shown in Equation 2.

Once the rotation matrix $\mathcal{O}^*$ is estimated, we construct estimates of the factors and the factor loadings as

$$\hat{f}_t = \mathcal{O}^* f_t \tag{6}$$

and

$$\hat{C} = C\mathcal{O}^{*\prime} \tag{7}$$

respectively, where f_t and C are the factor realizations and factor loadings estimated in stage 1.

To test the efficiency of the proposed identification, we perform a number of Monte Carlo experiments. We simulate Model (1) using parameter values obtained from fitting the model to the data. For each simulation we compare the performance of our method to three benchmarks. In the first case, we assume that the matrix of factor loadings C and the matrix of residual variances Γ are known. This case yields the most efficient estimates of the factors given by GLS:

$$\hat{f}_t = (C'\Gamma^{-1}C)^{-1}C'\Gamma^{-1}\, r_t. \tag{8}$$

It provides an upper bound on how well the factors can be recovered in the presence of noise. In the second case, we use traditional factor analysis (FA) with a popular rotation technique "varimax". For the third case, we utilize the estimates given by principal component analysis (PCA). In all cases, we set the true number of factors to four.

Figure 1 shows the distribution of the correlation coefficient of the estimated factors with the true ones derived by our methods. We observe that factors identified using our methodology are remarkably close to the most efficient estimates. At the same time, estimates produced by FA with the varimax rotation, and by PCA, are significantly worse. The difference is especially noticeable for the PCA method: PCA can only successfully separate the first factor (market), which accounts for the highest variability in the data. For FA and

PCA, respectively, the estimate of the first factor has an average correlation with the true factor of 72.4% and 91.7%, while our technique produces a correlation of 96.7%. If the true matrices C and Γ were known, then the average correlation of the GLS estimate would be 97.4%. However, if we use both FA and PCA, we have difficulty separating the remaining factors. For instance, the average correlation between the FA and PCA estimates and the remaining factors are 78.6% and 59.1%, compared to 86.7% for our method, and 88.2% if C and Γ were known. The bottom line is that our method significantly outperforms both FA and PCA in identifying the true factors. Our method is in fact close to the upper bound, i.e., the GLS estimates we would obtain if the rotation were known.

An important question is to what extent these results are sensitive to particular parameter values. Simulation results indicate that the critical parameter is Γ, the matrix of residual variances, and not the specific dynamics of the factors. This is good news, since estimation of the residual variances Γ does not depend on the particular form of identification. This result is intuitive because, given the finite number of assets, the amount of idiosyncratic noise determines how well factor analysis can identify the factors, even when the rotation (matrix C) is known.

4 Empirical Implementation and Results

We apply our analysis to the set of 30 industry portfolios created by Fama and French (1997). The authors classify all stocks into industries based on SIC codes. In our estimation we omit the "Other" industry, which comprises unsorted firms. The industry portfolios are good candidates as basis assets, since the industry loadings on macroeconomic factors are likely to be significantly more stable than are those of individual stocks. While it is true that the structure of an industry, and thus its response to systematic risk factors, may change over time, changes at the industry level are likely to occur far less often than changes at the firm level. Bansal, Fang and Yaron (2005) demonstrate that even though the relative shares of industries change over time, this shift has little impact on the moments of stock returns.

We compute portfolio returns by using CRSP data at both daily and monthly frequencies. We use data from January 1963 to December 2004. We use the 30 industry portfolios because they are well diversified, even in the beginning of our sample. Our results are similar when we extract factors from the Fama and French (1997) 17 or 49 industry portfolios.

4.1 The Number of Factors

The number of common factors in the data is the main parameter of any factor model. The econometrics literature has proposed a number of tests to determine the number of common factors. We utilize the test developed in Onatski (2006), which uses results from the distribution theory of random matrices, and the model selection criteria developed in Bai and Ng (2002). Both approaches allow for heteroskedasticity, and for weak serial and cross-sectional dependence.[3] The results are given in Table 1. Onatski's test suggests that the number of factors in the data is between three and four. This is consistent with the selection criteria in Bai and Ng. The most conservative criterion, BIC3, gives the number of factors as one or two. The least conservative criteria, PC_i, assign five to eight factors.

Our identification technique provides another way to check the number of factors in the data. For example, consider the case where the true number of factors in the data is N, and we extract M factors. If $M < N$, then the M extracted factors will be linear combinations of the true N factors. On the other hand, if $M > N$, then, given that the model is identified, we expect to see the N true factors among the extracted factors, with the remaining $M - N$ being noise. This reasoning suggests that we should start the estimation of the model with one factor, and then gradually increase the number of factors by one, until the additional factor is not correlated with the previously estimated factors.

In Table 2 we present correlations between factors that we obtain when we estimate the model with three, four, and five factors. When we extract three factors, the first and the third factors are linear combinations of the first, third, and fourth factors in a four-factor version of the model. When we estimate five factors, four factors are almost exactly the same as when we estimate the model with four factors, but the fifth extracted factor has low correlation with the previously estimated factors. This finding is consistent with the results provided by the Bai and Ng (2002) and Onatski (2006) tests. If the fifth extracted factor is truly a common factor and not noise, then it should be correlated with the previously extracted factors. Therefore, we conclude that there are four common factors in the data.

Our identification of factors relies on the assumption that the variance-covariance matrix of the factors is time varying. Therefore, it is important to verify that the extracted factors are indeed heteroskedastic. Table 3 shows the parameter estimates when the GARCH(1,1) model is fitted to the extracted factors. We find that both ARCH and GARCH coefficients are significantly different from zero for all factors, which implies the presence of substantial

[3] Both tests require an upper bound on the true number of factors $r_{\max}$. We use the values of $r_{\max} = 8,\ 9$ and 10 suggested in these two papers for our dimension of the data.

time-variation in factor volatilities. Figure 2 displays the conditional variances obtained from GARCH(1,1) estimation. The figure further confirms the heteroskedastic nature of the factors, and demonstrates that changes in their volatilities are not perfectly correlated.

4.2 Estimation Results and Interpretation

Table 4 shows the correlation of industry returns with factor returns. We report our results for estimating the model over the whole sample for the period 1963-2004.

When we look at correlations, we see that the first factor correlates highly with all industry returns, and thus can be interpreted as a common market shock. We also see that the factor correlates more with industries, such as Food, Beer, Consumer Goods, and Health, that are less sensitive to business cycles. When we look at how the second factor correlates with industry portfolios, we see that final consumption good industries, such as Food, Beer, Tobacco, and Utilities, have negative correlations. At the same time, industries such as Fabricated Products, Electrical Equipment, and Business Equipment, which produce capital goods, have positive correlations. These results suggest that we can interpret the second factor as investment minus consumption industries. Looking at the third factor, we see that, unlike the first factor, it correlates more with industries that have returns sensitive to business cycle conditions, such as Auto and Apparel, and less with acyclical industries, such as Food, Health, and Services. The fourth factor correlates positively with industries whose outputs are raw inputs for the production of other goods, such as Chemicals, Steel, Coal and Petroleum, and negatively with other industries. Our interpretation is that the first and third factors capture information about the business cycle, the second factor contains information about the relative price of capital, and the fourth factor contains information about factor input prices.

To verify this initial interpretation, we construct several auxiliary portfolios. First, we use the 1997 Bureau of Economic Analysis (BEA) Input-Output tables to construct a portfolio of investment minus consumption industries (IMC). As in Papanikolaou (2008), we classify industries according to their contribution to Consumption Expenditures or Private Investment, and we then construct the portfolio as the difference of two value-weighted portfolios for firms that belong to each sector.

We project the four extracted factors on the market portfolio, the IMC portfolio and a small number of industry portfolios. We project our third factor on two cyclical goods industries (Fabricated Products, Apparel) and two acyclical industries (Beer, Health), and our fourth factor on the market portfolio and four input industries (Mines, Steel, Oil, Coal).

Table 5 presents the results. The first factor is highly correlated (85%) with the value-weighted market portfolio. The second factor is highly correlated (87%) with the investment minus consumption portfolio. The third factor is related to the business cycle. When projecting the third factor on selected industries, we see that the industries enter with the right signs, i.e., positive for cyclical, and negative for acyclical industries, and the correlation of the extracted factor with the fitted values is over 90%. The fourth factor is an input-price factor. The four input industries enter with a positive sign, but the market portfolio enters with a negative sign. The correlation of the fourth factor with its projection is about 85%. These results confirm our initial interpretation regarding the economic content of the extracted factors.

4.3 Sample Dependence

One of the main concerns of factor analysis is that the resulting factors are very sample-dependent, and so results estimated in-sample may not hold out-of-sample. To address this issue, we divide our sample in two sub-samples, 1963 to 1983 and 1984 to 2004. For each subperiod, we use the matrix of factor loadings (C) and idiosyncratic variances (Γ) estimated using data within that period to provide GLS estimates of the factor realizations for the entire sample (1963-2004) using Equation 8. We then compute the correlation between factors extracted from the whole sample ($\hat{f}$) and the factors extracted from each sub-sample ($\hat{f}^A, \hat{f}^B$). We show the results in Table 6. In all cases the correlation is consistently higher than 90%. In what follows, we report our results for the factors extracted from the full sample since utilizing the full sample is the most efficient means of summarizing the covariation in the data.

As an additional robustness check, we take the following steps to mitigate sample dependence. We substitute the second factor for the IMC portfolio, whose construction does not depend on a particular sample. For the other factors we replace their estimated loadings with their signs. The resulting factors are very close to the extracted factors, with pairwise correlations of over 90%. Our empirical results using these alternative factors are qualitatively and quantitatively similar and available upon request.

5 Predictive Regressions

In this section we explore the informational content of the extracted factors for aggregate economic and financial variables. For financial variables we focus on the CRSP value-weighted market portfolio, the Fama and Bliss (1987) zero-coupon bonds with maturities from one to five years, the one-month Treasury-bill rate, and the default premia measured as the difference in Moody's seasoned Aaa and Baa corporate bond yields. The economic variables of interest are inflation, industrial production and unemployment.

Given our interpretation of the first and the third factors as capturing business cycle risk, these factors should predict variables which vary across the business cycle, such as output, unemployment, and default spreads.

The second factor is highly correlated with the investment minus consumption portfolio, and as such it likely proxies investment-specific shocks as in Papanikolaou (2008). If this is the case, then it should predict an increase in output and employment as the economy allocates increased resources to the investment sector. Moreover, this reallocation of resources implies that consumption is expected to be higher in the future, and thus the second factor should also predict increases in the real interest rate.

Finally, the fourth factor, which might proxy for commodity prices, should predict an increase in inflation, and thus nominal interest rates. Pollet (2005) and Driesprong, Jacobsen and Maat (2008) find that changes in oil prices predict future market returns. Thus, we explore whether the same holds for our fourth factor, and also examine if it can predict bond returns, as in most general equilibrium models equity and bond premia are positively correlated.

Whenever possible, we report results for the period January 1951 to December 2004. Our factors are estimated in the January 1963 to December 2004 sample due to the need for daily return data. We obtain estimates of realizations of the factors for the period 1951-1963 by $\hat{f} = (C'\Gamma^{-1}C)^{-1}C'\Gamma^{-1}\, r_t$, where the matrix of coefficients (C) and idiosyncratic variance (Γ) are estimated using the 1963-2004 sample period.

5.1 Asset Prices

Here, we wish to determine whether our factors capture information about the investment opportunity set faced by investors. If the factors can predict the evolution of moments of asset returns, then they could be used to hedge investor's exposure to such risks, in the spirit of Merton (1973).

5.1.1 Equity Risk Premium

We examine the ability of the input-price factor to predict future returns on the market portfolio. We estimate

$$mkt_{t+1} = \alpha_0 + \beta\, f_{4,t} + \alpha_1\, r_{f,t} + \alpha_2\, yield_t + \alpha_3\, spread_t + \alpha_4\, ldp_t + \varepsilon_{t+1}. \tag{9}$$

The dependent variable (mkt) refers to the simple return on the CRSP value-weighted stock market index over the one-month Treasury-bill rate, r_f refers to the one-month Treasury-bill rate, *yield* refers to the Moody's seasoned Aaa corporate bond yield minus one-month Treasury-bill rate, *spread* is the difference in Moody's seasoned Aaa and Baa corporate bond yields, and *ldp* is the log dividend-price ratio. We compute the dividend-price ratio as dividends over the past year divided by the current price. We compute heteroskedasticity and autocorrelation consistent (HAC) standard errors using the Newey and West (1987) procedure.

Table 7 presents the results of predictive regressions of the excess market return by the fourth factor. The regression results show that the fourth factor does have the ability to predict the market, and that the coefficient of the factor is very stable across different subsamples. The predictive power is stronger in the second half of the sample (1978-2004). Given the increased importance of oil shocks during this period, this finding is consistent with our intuition of the fourth factor being a proxy for shocks to input prices. We note that, as documented in Goyal and Welch (2008), it is in this sample that most of the known predictors do not perform well.

The reported predictive regressions are in-sample. Goyal and Welch (2008), among others, demonstrate that many successful predictors in-sample fail to deliver the same performance out-of-sample. One of the main indicators for the poor out-of-sample performance is the instability of coefficients estimated over different subsamples. However, as evident from Table 7, the coefficient at the fourth factor is reasonably stable over time, which suggests that its ability to predict the market is not sample-driven. As a further robustness check, we estimate the parameters only on the data available at the time τ, and we use the estimated parameters to predict the market return one month ahead. Thus, the parameters are re-estimated each month, and the best forecast of the market return at $t = \tau + 1$ is given by

$$\hat{m}kt_{\tau+1} = \hat{a}_\tau + \hat{b}_\tau\, f_{4,\tau}, \tag{10}$$

where $\hat{a}_\tau$ and $\hat{b}_\tau$ denote the parameter estimates at time τ. We compute the out-of-sample

R^2 as

$$R^2 = 1 - \frac{\sum_{t=1}^{T}(mkt_t - \widehat{mkt}_t)^2}{\sum_{t=1}^{T}(mkt_t - \overline{mkt}_t)^2}, \tag{11}$$

where mkt_t is the realized value of the market return, $\widehat{mkt}_t$ is its prediction and $\overline{mkt}_t$ is the prevailing historical mean. Finally, we use the first half of the sample (1951-1977) to obtain the initial estimates, and compute the out-of-sample R^2 for the second half (1978-2004).

We find that the input factor improves on the historical average. Using the input factor alone results in an out-of-sample R^2 statistic of 0.5% at monthly frequency. In contrast, 38 out of the 39 predictive regressors considered by Goyal and Welch (2008) underperformed the prevailing historical mean, resulting in negative out-of-sample R^2.

Our results are related to the findings of Hong et al. (2007), who document that some industries lead the market. They find that in the U.S. stock market, 12 out of 38 industries are statistically significant predictors of market portfolio returns one month ahead. They also show that similar results hold for the eight largest stock markets outside the U.S. Since they include lagged market portfolio returns in their predictive regressions, they essentially show that it is not the industries per se that predict the market, but rather some industry components which are orthogonal to the market.

Given the ability of the fourth factor to predict the market, we verify whether industry portfolios have residual predictive power once the fourth factor is taken into account. Therefore, we estimate

$$mkt_{t+1} = \alpha_0 + \beta\, f_{4,t} + \alpha_1\, ind_{it} + \alpha_2\, mkt_{it} + \varepsilon_{t+1}. \tag{12}$$

As before, the dependent variable mkt is the simple return on the CRSP value-weighted stock market index over the one-month Treasury-bill rate, $f_{4,t}$ refers to our fourth factor, and ind_{it} refers to returns of an industry portfolio constructed by Fama and French (1997). The regressions are similar to those in Hong et al. (2007) except that we run a horse race between the fourth factor and the 49 Fama and French industry returns. Due to data availability, the sample period is July 1963 to December 2004.

Table 8 shows the result of estimating Equation (12). Once we include the fourth factor as a control variable, none of the 49 industry coefficients is significant. At the same time, the results for the fourth factor are virtually unchanged with regard to inclusion of any specific industry return. The only exception is a lower t-statistic when we include the Mines industry.

Hong et al. (2007) interpret the finding that industries lead the market as evidence in favor of the limited information-processing capacity of investors. We show that this predictability comes from a systematic component of returns. Therefore, it is possible that the source of this predictability is systematic movements in risk premiums, rather than underreaction to industry-specific news.

5.1.2 Bond Risk Premiums

In many general equilibrium models equity and bond risk premiums move together. If that is the case, then the input factor should be able to predict both equity returns and bond returns. We use the data of Fama and Bliss (1987) to investigate whether the fourth factor has the ability to predict excess bond returns. We estimate the regression:

$$r^{(k)}_{t,t+12} - y^{(1)}_t = a_k + b_k f_{4,t} + c_k Y_t + e_t, \qquad k = 2..5 \tag{13}$$

where $r^{(k)}_{t,t+12}$ refers to the return of a bond with maturity k between months t and $t+12$ and Y_t refers to a vector of controls that includes the short rate, slope, default spread, log dividend yield, and the Cochrane and Piazzesi (2005) factor, which is a linear combination of forward rates.

Table 9 shows the results. The fourth factor is a statistically significant predictor of excess bond returns for all maturities from two to five years. When we include the vector of controls, the fourth factor is still statistically significant for all bond maturities. This finding suggests that the input price factor captures information about expected returns on bonds as well as on equity; and that this information is somehow not fully incorporated into the dividend yield, the Cochrane and Piazzesi (2005) factor, the term and default spreads, or the short rate. The source of this predictability is not obvious, but the fact that it predicts both excess bond and stock returns with the same sign provides additional evidence that is consistent with a risk-based story.

5.1.3 Interest Rates

Here, we estimate

$$X_{t+k} - X_t = a_0 + A(k)F_t + B(L)\,X_t + c_k Y_t + e_t, \tag{14}$$

where X_t is a vector containing the short term rate (one-month Treasury bill); the slope of the yield curve, which we define as the difference between the yield on 10-year Aaa bonds and the short rate; and the default premium, which we define as the difference between the Baa and Aaa rated bonds.

The results appear in Table 10. The second factor predicts increases in the level of nominal interest rates, which partly verifies our hypothesis. What remains to be seen is that the second factor has no ability to predict movements in inflation, which implies it is predicting movements in real interest rates. The first and third factors predict a tightening of default spreads, which is consistent with their interpretation as business cycle factors. Finally, the fourth factor is a statistically significant predictor of the short rate at horizons of up to one year.

5.2 Macroeconomic Series

Here, we explore the informational content of our extracted factors for aggregate quantities and prices. We focus on inflation, aggregate output, and unemployment. In particular, given that the second and fourth factors predict future interest rates, their ability to predict inflation may shed some light on whether they predict real or nominal rates.

5.2.1 Inflation, Output and Employment

To examine the ability of the extracted factors to predict inflation, we estimate

$$\pi_{t+k} - \pi_t = a_0 + A(k)F_t + B(L)\Delta\pi_t + c_k Y_t + e_t, \tag{15}$$

where $\pi_t = \ln(CPI_t) - \ln(CPI_{t-1})$ is inflation, F_t are the factor realizations obtained in Section 4.2, and Y_t is a vector of control variables that includes three lags of industrial production growth and unemployment. We follow Stock and Watson (1999) and include three lags of inflation to account for serial correlation in the data. We measure inflation with the CPI-U index, which is available from the Bureau of Labor Statistics (BLS)

To find out whether the extracted factors can predict output and employment, we estimate

$$X_{t+k} - X_t = a_0 + A(k)F_t + B(L)\, X_t + c_k Y_t + e_t, \tag{16}$$

where X_t is log output or log unemployment, F_t are the factor realizations, and Y_t is a vector of control variables that includes lags of first differences in inflation. We include three lags of the dependant variable to account for serial correlation in the data. We measure output with the Industrial Production Index, which is available from the Federal Reserve Board of

Governors (G17 Release). We measure employment with the Civilian Unemployment Rate, which is available from the BLS. We estimate Equations 15 and 16 using monthly data.

We present the results in Table 11. The fourth factor is a statistically significant predictor of inflation for horizons of up to one year. This result is in sharp contrast with Stock and Watson (2003) who document that asset prices have no predictive power for inflation. The market portfolio indeed has little predictive ability. However, the forth factor appears to be a valuable addition to other predictors.

As a robustness check, we explore the ability of the fourth factor to predict inflation out of sample. We estimate the parameters only on the data available at the time τ, and we use the estimated parameters to predict inflation one month ahead. We compare the out-of-sample performance of a reference model that includes three lags of inflation versus a model that also includes the fourth factor. The parameters are re-estimated each month, and the best forecast of inflation at $t = \tau + 1$ is given by

$$\hat{\pi}_{\tau+1} = \pi_\tau + \hat{a}_\tau + \hat{b}_\tau \, f_{4,\tau} + \hat{c}_\tau(L)\Delta\pi_\tau, \tag{17}$$

where $\hat{a}_\tau$, $\hat{b}_\tau$ and $\hat{c}_\tau(L)$ denote the parameter estimates at time τ. We compare the performance of this forecast to the model that excludes the fourth factor, i.e.

$$\overline{\pi}_{\tau+1} = \pi_\tau + \overline{a}_\tau + \overline{c}_\tau(L)\Delta\pi_\tau, \tag{18}$$

where $\overline{a}_\tau$ and $\overline{c}_\tau(L)$ denote the parameter estimates at time τ. We compute the out-of-sample R^2 as

$$R^2 = 1 - \frac{\sum_{t=1}^{T}(\pi_t - \widehat{\pi}_t)^2}{\sum_{t=1}^{T}(mkt_t - \overline{\pi}_t)^2}. \tag{19}$$

Finally, we use the first half of the sample (1951-1977) to obtain the initial estimates, and compute the out-of-sample R^2 for the second half (1978-2004). We find that including the fourth factor has a lower out-of-sample mean squared error and the fourth factor alone has an out-of-sample R^2 of 0.4%.

We see that the first factor predicts an increase in industrial production. The first factor also predicts an increase in employment, as demonstrated by the fall in the unemployment index. Most of the responses are statistically significant at horizons beyond three months. The third factor exhibits similar responses. Industrial production and employment rise sharply following positive realizations of the third factor, and the coefficients are statistically significant for horizons from one month to a year. The second factor shows some statistically significant short-run responses: both industrial production and employment rise in the short run (up to six months). Given that the second factor has no ability to predict inflation, coupled with the observation that it predicts changes in the level of nominal interest rates, suggests that it captures movements in real interest rates. This is consistent with the interpretation of the second factor capturing investment-specific shocks and thus movements in the expected growth of consumption. We perform additional robustness checks (not reported) and find that these results are robust to controlling for the short rate, the slope of the term structure and the default spread, as well as measuring output by GDP and measuring employment by non-farm payrolls.

6 Asset Pricing Tests

In this section, we briefly explore the ability of extracted factors to price different cross-sections of stocks. We focus on portfolios of stocks sorted first on size and then on Book-to-Market, Momentum, Short- or Long-Term Reversal[4]. These portfolios are known to have high spreads in mean returns and are now standard test assets. We estimate each cross-section separately, and also in a pooled cross-section comprised of 6 Size/Book-to-Market, 6 Size/Momentum, 6 Size/Short-Term Reversal and 6 Size/Long-Term Reversal portfolios. Along with our four-factor model, we present the results for the CAPM, the Fama and French (1993) three-factor model, and the Carhart (1997) model that includes the momentum factor.

We use Hansen and Singleton (1982) two-step GMM procedure with the Stochastic Discount Factor of the form:

$$m = a - b(f - \mu_f),$$

and the following moment restrictions:

$$E[mR^e] = 0, \tag{20}$$

$$E[f] = \mu_f, \tag{21}$$

where f are the candidate asset pricing factors and R^e are the excess returns of the test assets. The vector of unknown parameters contains the risk prices and the mean realizations of the candidate asset pricing factors, $\theta = [b, \mu_f]$.[5] Following Yogo (2006), we use the first

[4] We use the data from Kenneth French's web site:
http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

[5] By estimating the factor means jointly with the parameters of the SDF yields standard errors that account for the fact that factor means are estimated. See Cochrane (2001) for a textbook treatment.

stage weighting matrix,

$$W = \begin{bmatrix} kI_N & 0 \\ 0 & \hat{\Sigma}_{ff}^{-1} \end{bmatrix},$$

where $k = \det(\hat{\Sigma}_{R^e})^{-1/N}$ and N is the number of test assets. This first stage weighting matrix puts equal weight on each of the N asset pricing restrictions. We compute the spectral density matrix for the second stage using the Newey and West (1987) estimator with three lags of returns.

We show the estimation results in Table 12 and plot the in-sample average versus the predicted risk premia in Table 13. Along with the estimated prices of risk (b), we also report the mean absolute pricing errors (MAPE), the sum of squared errors (SSQE) and the J-test for the validity of the over-identifying restrictions.

In most cases, the performance of our four-factor model falls between that of the Fama-French and the Carhart models. The price of risk of the first factor is not significant, which is consistent with the findings of Chen et al. (1986). The second (investment-goods price) factor carries a negative risk premium and helps price the value cross-section, as it is negatively correlated with the HML portfolio[6]. The third (business cycle) and fourth (input-goods price) factors have a positive and negative price of risk respectively. Moreover, they seem to help in explaining the value, momentum and short-term reversal cross-section of expected returns, since risk of these portfolios, as measured by the loadings on our factors, is not the same.

7 Factors in the United Kingdom

To find out if our results can be replicated in a different sample, we perform an additional robustness check, using the stock market data for the United Kingdom that is available through the London Share Price Database (LSPD). We use the Financial Times Stock Exchange (FTSE) index as the market portfolio. We replicate the Fama and French (1997) industry classification to map stocks into 30 industry portfolios. Because of data availability, we have to drop the coal and textile industries. We exclude the Other industry, and we extract factors using the method we describe in Section 3.

We show the results in Tables 14 and 15. The UK factors are similar to the US factors. The first factor corresponds to the market portfolio, the second factor loads positively on industries that produce investment goods, and negatively on industries that produce consump-

[6]See Papanikolaou (2008) for a theoretical justification.

tion goods. The third factor loads positively on cyclical industries such as manufacturing and negatively on acyclical industries such as utilities. The fourth factor is positively correlated with industries that produce commodities and negatively correlated with the market portfolio.

We also examine whether our main predictability results hold in this sample. We repeat the exercise in Section 5.2.1 to investigate if the extracted factors predict inflation in the United Kingdom. We show the results in Table 16. The results are somewhat sensitive to the inclusion of bond yields, so we also report results with (Panel B) and without (Panel A), controlling for the three-month and ten-year bond yield. The fourth factor is a statistically significant predictor of inflation at all horizons, regardless of whether we include bond yields in the specification. The third factor is also a statistically significant predictor of inflation at horizons greater than six months, but only if we also include bond yields.

In addition, we explore the ability of the fourth factor to predict excess stock market (FTSE Index) and bond returns (the 10-year Gilt) in the United Kingdom. We present the results in Table 17. The fourth factor is a statistically significant predictor of one-month ahead returns on the market portfolio and ten-year government bonds. In addition to the sign, the magnitude of coefficients is also comparable to those in Tables 7 and 9.

Based on our results, we conclude that our UK analysis delivers results that are qualitatively and quantitatively similar to those obtained for the U.S. This similarity reinforces our view that these factors represent macroeconomic shocks. Moreover, even though we do not take a stand as to the mechanism through which these shocks affect asset prices and real quantities, the results suggest that it is robust across different samples.

8 Conclusion

The volatility of many economic and financial variables changes over time. Rather than treating this heteroskedasticity as an econometric problem, we exploit it in the data, thus achieving a complete identification of factors in factor analysis. We develop a two-stage estimation procedure that has a GMM interpretation and is easy to implement. This method is essentially model-free, since it does not rely on a structural model to achieve identification.

We demonstrate the value of this method by studying the returns of the US industry portfolios. Our analysis reveals four factors that should correspond to different sources of systematic risk in the economy. These factors contain information about future macroeconomic quantities, and variables that characterize the investment opportunity set in the economy. These extracted factors and their responses can serve as a rough guide for the types of shocks that are included in production-based models. We investigate the robustness of these results by repeating the analysis in a different sample, the United Kingdom, and we find that they are qualitatively and quantitatively similar to the results we obtain for the United States.

A multi-factor model with heteroskedastic factors can be a parsimonious way to characterize the time-variation in comovement across assets. For instance, Baele et al. (2009) use a heteroskedastic factor model to study the determinants of stock and bond correlations. Our method is quite general and can be applied equally well to other settings. For example, it can be used to study the comovement in exchange rates across countries or to analyze the factors that affect the term structure of interest rates.

References

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2003). Modeling and forecasting realized volatility, *Econometrica* **71**(2): 579–625.

Baele, L., Bekaert, G. and Inghelbrecht, K. (2009). The determinants of stock and bond return comovements, *working paper*, Columbia University.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models, *Econometrica* **70**(1): 191–221.

Bansal, R., Fang, E. and Yaron, A. (2005). Equity capital: A puzzle?, *mimeo*.

Bansal, R., Khatchatrian, V. and Yaron, A. (2005). Interpretable asset markets?, *European Economic Review* **49**(3): 531–560.

Campbell, J. Y. (1996). Understanding risk and return, *Journal of Political Economy* **104**(2): 298–345.

Carhart, M. M. (1997). On persistence in mutual fund performance, *Journal of Finance* **52**(1): 57–82.

Chen, N.-F., Roll, R. and Ross, S. A. (1986). Economic forces and the stock market, *Journal of Business* **59**(3): 383–403.

Cochrane, J. H. (2001). *Asset Pricing*, Princeton University Press.

Cochrane, J. H. and Piazzesi, M. (2005). Bond risk premia, *American Economic Review* **95**(1): 138–160.

Connor, G. and Korajczyk, R. A. (1988). Risk and return in an equilibrium apt : Application of a new test methodology, *Journal of Financial Economics* **21**(2): 255–289.

Connor, G. and Korajczyk, R. A. (1991). The attributes, behavior and performance of u.s. mutual funds, *Review of Quantitative Finance and Accounting* **1**: 5–26.

Driesprong, G., Jacobsen, B. and Maat, B. (2008). Striking oil: Another puzzle?, *Journal of Financial Economics* **89**(2): 307–327.

Fama, E. F. and Bliss, R. R. (1987). The information in long-maturity forward rates, *American Economic Review* **77**(4): 680–92.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* **33**(1): 3–56.

Fama, E. F. and French, K. R. (1997). Industry costs of equity, *Journal of Financial Economics* **43**(2): 153–193.

French, K. R., Schwert, G. W. and Stambaugh, R. F. (1987). Expected stock returns and volatility, *Journal of Financial Economics* **19**(1): 3–29.

Goyal, A. and Welch, I. (2008). A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies* **21**(4): 1455–1508.

Hansen, L. P. and Hodrick, R. J. (1980). Forward exchange rates as optimal predictors of future spot rates: An econometric analysis, *Journal of Political Economy* **88**(5): 829–53.

Hansen, L. P. and Singleton, K. J. (1982). Generalized instrumental variables estimation of nonlinear rational expectations models, *Econometrica* **50**(5): 1269–86.

Harman, H. H. (1976). *Modern Factor Analysis*, 3rd edn, University of Chicago Press.

Hong, H., Torous, W. and Valkanov, R. (2007). Do industries lead stock markets?, *Journal of Financial Economics* **83**(2): 367–396.

Jones, C. S. (2001). Extracting factors from heteroskedastic asset returns, *Journal of Financial Economics* **62**(2): 293–325.

King, M., Sentana, E. and Wadhwani, S. (1994). Volatility and links between national stock markets, *Econometrica* **62**(4): 901–33.

Lamont, O. A. (2001). Economic tracking portfolios, *Journal of Econometrics* **105**(1): 161–184.

Lauterbach, B. (1989). Consumption volatility, production volatility, spot-rate volatility, and the returns on treasury bills and bonds, *Journal of Financial Economics* **24**(1): 155–179.

Longstaff, F. A. and Schwartz, E. S. (1992). Interest rate volatility and the term structure: A two-factor general equilibrium model, *Journal of Finance* **47**(4): 1259–82.

Ludvigson, S. C. and Ng, S. (2007). The empirical risk-return relation: A factor analysis approach, *Journal of Financial Economics* **83**(1): 171–222.

Merton, R. C. (1973). An intertemporal capital asset pricing model, *Econometrica* **41**(5): 867–87.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* **55**(3): 703–08.

Officer, R. R. (1973). The variability of the market factor of the new york stock exchange, *Journal of Business* **46**(3): 434–53.

Onatski, A. (2006). Determining the number of factors from empirical distribution of eigenvalues, *mimeo*, Columbia University, Department of Economics.

Papanikolaou, D. (2008). Investment-specific technological change and asset prices, *mimeo*, Northwestern University.

Petkova, R. (2006). Do the fama-french factors proxy for innovations in predictive variables?, *Journal of Finance* **61**(2): 581–612.

Pollet, J. (2005). Predicting asset returns with expected oil price changes, *working paper*, University of Illinois at Urbana-Champaign - Department of Finance.

Rigobon, R. (2003). Identification through heteroskedasticity, *The Review of Economics and Statistics* **85**(4): 777–792.

Schwert, G. W. (1989). Why does stock market volatility change over time?, *Journal of Finance* **44**(5): 1115–53.

Sentana, E. and Fiorentini, G. (2001). Identification, estimation and testing of conditionally heteroskedastic factor models, *Journal of Econometrics* **102**(2): 143–164.

Stock, J. H. and Watson, M. W. (1999). Forecasting inflation, *Journal of Monetary Economics* **44**(2): 293–335.

Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors, *Journal of the American Statistical Association* **97**: 1167–1179.

Stock, J. H. and Watson, M. W. (2003). Forecasting output and inflation: The role of asset prices, *Journal of Economic Literature* **41**(3): 788–829.

Vassalou, M. (2003). News related to future gdp growth as a risk factor in equity returns, *Journal of Financial Economics* **68**(1): 47–73.

Yogo, M. (2006). A consumption-based explanation of expected stock returns, *Journal of Finance* **61**(2): 539–580.

Tables and Graphs

Figure 1: Correlation between extracted and true factors, comparison across methodologies

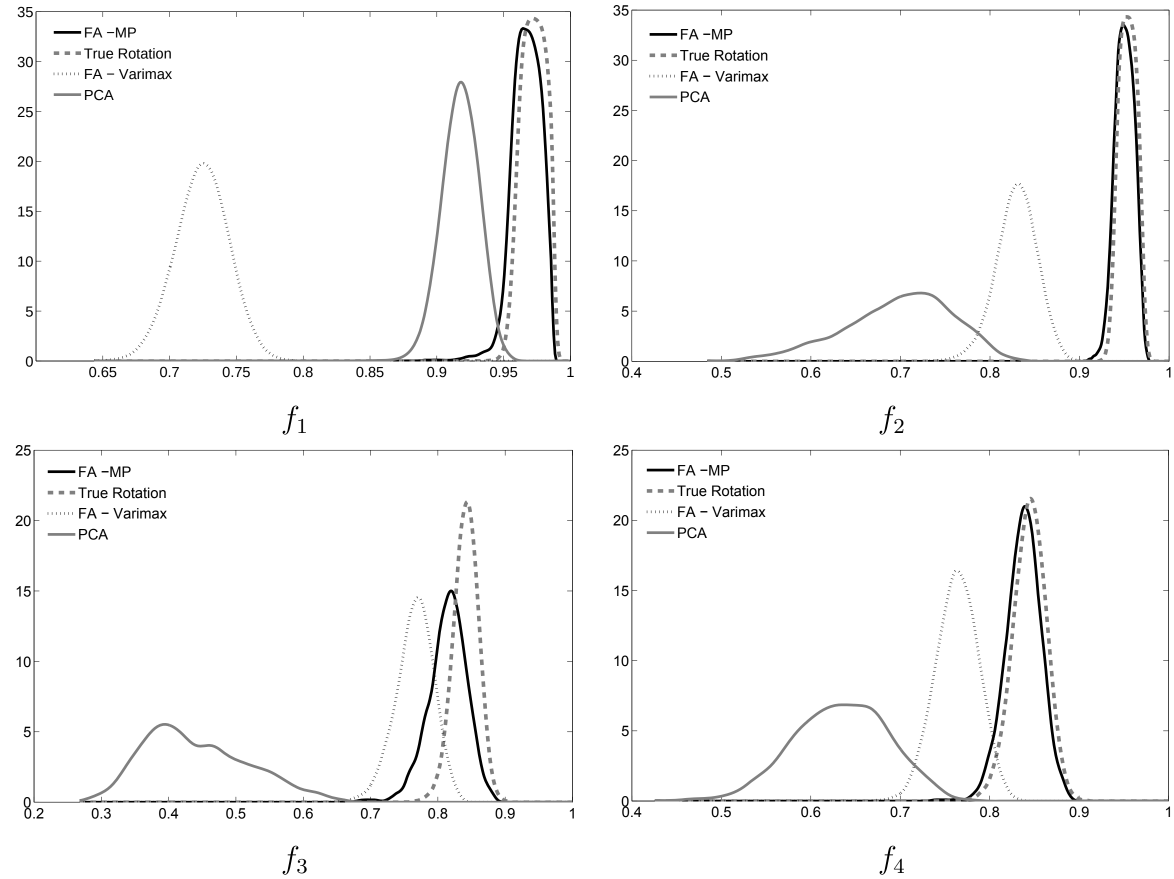


Figure 1 compares the performance of our method (black solid line) to three benchmarks. The first is the GLS estimate when the matrix of factor loadings is known (grey dotted line). The second is principal component analysis (solid grey) and the third is static factor analysis using the 'varimax' criterion (dotted black line). The graphs plot the distribution of correlation coefficients between the extracted and true factors. Each simulation sample has a length of 600 months. The data generating process for each factor is assumed to be univariate GARCH(1,1), using the parameter estimates shown in Table 3. We simulate a cross-section of 29 assets using the point estimates of the factor loadings estimated in Section 4. We repeat the procedure 4,000 times.

Table 1: The number of factors in the 29 Industry Portfolios

Data Frequency	Onatski		Bai and Ng				
	rmax	N	rmax	$PC_{1,2}$	PC_3	IC_i	BIC_3
Daily	8,9	3	8	5	5	2	1
	10	4					
Monthly	8	3	8	7	8	5	2
	9	4					

Table 2: Specification check

$\rho(\%)$	f_1	f_2	f_3	f_4
f_1	90.60	-2.20	27.63	-29.18
f_2	2.66	99.70	0.87	6.65
f_3	3.78	-4.83	63.67	75.70
f_1	99.07	-0.40	-5.07	1.07
f_2	0.27	98.37	-10.76	1.73
f_3	2.92	9.57	98.16	-7.65
f_4	8.80	-8.72	-3.26	82.96
f_5	8.82	-12.12	-13.06	-55.76

Table 1 reports the number of factors in 29 industry portfolios according to the tests developed in Onatski (2006a) and Bai and Ng (2002). There are seven information criteria in Bai and Ng (2002): PCi, ICi, $i = 1, 2, 3$, and BIC3. Their results are reported separately only if they give different number of factors. All tests require an upper bound on the true number of factors $r_{\max}$. We follow Onatski (2006) and Bai and Ng (2002) and use values of $r_{\max} = 8, 9,$ and 10 in the tests. The sample period is January 1963 to December 2004. Table 2 reports correlation between extracted factors when the model in Section 3 is estimated with three, four, and five factors. The data are returns on 29 industry portfolios constructed in Fama and French (1997). We use the 30 industry classification and drop the 'Other' industry. The industry return data are from Kenneth French's website. The sample period is January 1963 to December 2004.

Table 3: GARCH(1,1) estimates of f_i

Garch(1,1)	f_1	f_2	f_3	f_4
κ	0.091	0.023	0.049	0.040
	(0.037)	(0.012)	(0.017)	(0.014)
a	0.099	0.131	0.129	0.114
	(0.023)	(0.031)	(0.032)	(0.025)
b	0.810	0.838	0.825	0.854
	(0.047)	(0.041)	(0.035)	(0.030)

Figure 2: Conditional Variance of f_i

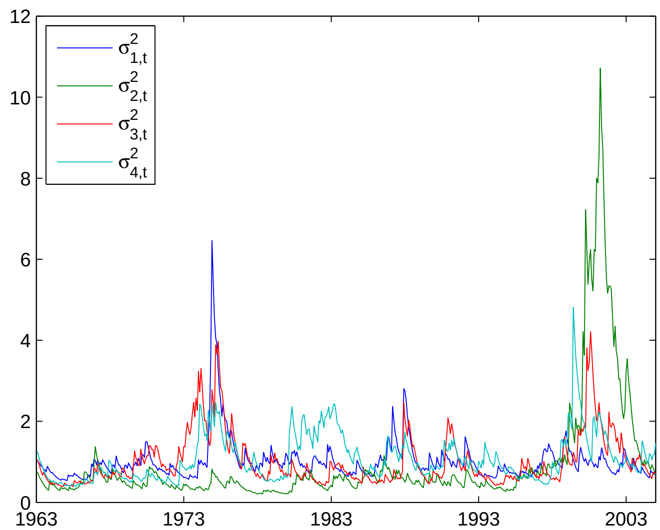


Table 3 reports estimates of GARCH(1,1) parameters for the factors extracted from 29 Fama-French industry portfolios using the two step estimation method described in Section 3. We demean each factor and fit the following model: $f_{i,t} = \sigma_{i,t}\varepsilon_{i,t}$, where $\varepsilon_{i,t}$ is an independent, identically distributed (i.i.d.) sequence $\sim$ N(0,1). The time-conditional variance, $\sigma^2_{i,t}$, process is modeled as $\sigma^2_{i,t} = \kappa + b\sigma^2_{i,t-1} + a\varepsilon^2_{i,t}$. Standard errors are provided in parenthesis. Figure 2 plots the time series of estimated conditional volatilities.

Table 4: Correlations of Industry Portfolios with extracted factors

Industry	Correlations			
	f_1	f_2	f_3	f_4
Food Products	0.80	-0.27	0.26	-0.16
Beer & Liquor	0.81	-0.19	0.04	-0.23
Tobacco Products	0.63	-0.18	-0.03	0.01
Recreation	0.59	0.26	0.57	-0.13
Printing & Publishing	0.68	0.11	0.52	-0.16
Consumer Goods	0.85	-0.02	0.13	-0.19
Apparel	0.55	0.04	0.72	-0.17
Healthcare	0.89	0.03	-0.03	-0.13
Chemicals	0.75	0.02	0.46	0.25
Textiles	0.52	0.04	0.68	-0.07
Construction and Construction Materials	0.78	0.05	0.50	0.12
Steel Works	0.58	0.39	0.46	0.37
Fabricated Products and Machinery	0.69	0.35	0.50	0.22
Electrical Equipment	0.73	0.23	0.42	0.02
Automobiles and Trucks	0.53	0.17	0.55	0.06
Aircraft, Ships, & Railroad equipment	0.67	0.11	0.47	0.15
Mining	0.39	0.11	0.32	0.47
Coal	0.40	0.04	0.19	0.36
Petroleum and Natural Gas	0.57	-0.06	0.22	0.48
Utilities	0.59	-0.30	0.20	0.13
Communication	0.62	0.15	0.22	-0.10
Personal & Business Services	0.66	0.54	0.17	-0.12
Business Equipment	0.65	0.69	0.25	-0.03
Business Supplies & Shipping Containers	0.78	0.01	0.43	0.18
Transportation	0.68	0.10	0.54	0.13
Wholesale	0.71	0.14	0.51	0.08
Retail	0.66	0.09	0.54	-0.32
Restaurants, Hotels & Motels	0.69	0.05	0.45	-0.20
Banking, Insurance, Real Estate & Trading	0.78	-0.05	0.41	-0.03

Table 4 presents the correlations of the 29 industry portfolios with the four factors f_i. The factors are extracted from a cross-section of 29 industry portfolios using the two step estimation method described in Section 3. We use the 30 industry classification of Fama and French (1997) and drop the 'Other' industry. The industry return data are from Kenneth French's website. Correlations are computed using monthly data over the period January 1963 to December 2004.

Table 5: Factor Interpretation

Factor	Portfolio					$R^2(\%)$
	MKT					
f_1	0.851					73.0
	(36.91)					
	IMC					
f_2	0.87					75.8
	(39.73)					
		Fab. Prod	Clothes	Beer	Health	
f_3		3.590	0.969	-0.204	-0.542	81.7
		(-3.99)	(-11.46)	(9.85)	(10.98)	
	MKT	Steel	Mines	Coal	Oil	
f_4	-1.173	0.895	0.241	0.168	0.795	70.1
	(-24.98)	(15.75)	(7.52)	(5.85)	(19.47)	

Table 5 reports coefficients from a projection of factors f_i on the market portfolio, the IMC portfolio constructed in Papanikolaou (2008) and a subset of industries, for the period 1963-2004. The factors f_i are extracted from 29 industry portfolios using the two step estimation method described in section 3. We use the 30 industry classification of Fama and French (1997) and drop the 'Other' industry. The industry return data are from Kenneth French's website. All portfolio returns have been normalized to zero mean and unit variance. t statistics are reported in parenthesis.

Table 6: Alternative Estimation Period

$\rho(\%)$	f_1	f_2	f_3	f_4
f_1^A	99.40	1.72	3.92	5.85
f_2^A	-0.56	91.15	25.61	-17.47
f_3^A	-0.06	-24.04	93.54	10.21
f_4^A	3.38	8.18	-13.44	96.65
f_1^B	98.31	-7.82	-4.45	-0.14
f_2^B	2.96	92.65	-28.93	-0.14
f_3^B	2.32	33.48	90.68	-2.66
f_4^B	-2.21	8.47	2.87	97.45

Table 6 reports correlation between factors extracted over different sub-samples. We estimate Equation (1) over the whole sample, the 1963-1983 and the 1984-2004 period. For each subperiod, we use the estimated matrix of factor loadings (C) and idiosyncratic variances (Γ) to provide estimates of the factor realizations for the entire sample, as $\hat{f}^j = (C'\Gamma^{-1}C)^{-1}C'\Gamma^{-1}\, r_t$, where r_t is a vector of returns on the industry portfolios. Thus we obtain estimates of the factors estimated using the entire sample ($\hat{f}$), the first half ($\hat{f}^A$) and the second half ($\hat{f}^B$). We use the 30 industry classification of Fama and French (1997) and drop the 'Other' industry. The industry return data are from Kenneth French's website.

Table 7: Predictability of Equity Returns

Period	const	f_4(%)	r_f	yield	spead	ldp	R^2(%)	T
1951-2004	0.01	-0.46					0.87	647
	(3.32)	(-2.41)						
	70.03	-0.46	-3.82	-1.46	21.02	0.01	3.24	647
	(1.94)	(-2.49)	(-3.46)	(-0.77)	(2.52)	(1.50)		
1951-1977	0.01	-0.38					0.24	323
	(2.21)	(-1.31)						
	0.03	-0.28	-8.00	-7.14	45.05	0.01	6.32	323
	(0.98)	(-1.18)	(-3.25)	(-1.59)	(2.83)	(0.62)		
1978-2004	0.01	-0.49					1.08	323
	(2.56)	(-2.00)						
	0.18	-0.59	-10.35	-10.54	21.06	0.03	5.19	323
	(4.17)	(-2.45)	(-4.06)	(-2.93)	(2.00)	(3.68)		

Table 7 reports estimates from OLS predictive regressions of the market excess return across different subsamples

$$mkt_{t+1} = \alpha_0 + \beta\, f_{4,t} + \alpha_1\, r_{f,t} + \alpha_2\, yield_t + \alpha_3\, spread_t + \alpha_4\, ldp_t + \varepsilon_{t+1}.$$

The dependent variable (mkt) refers to the simple return on the CRSP value-weighted stock market index over the one-month Treasury-bill rate. r_f refers to the one-month Treasure bill rate, *yield* refers to the Moody's seasoned Aaa corporate bond yield minus one month T-bill rate, *spread* is the difference in Moody's seasoned Aaa and Baa corporate bond yields and *ldp* is the log dividend-price ratio. The dividend-price ratio is computed as dividends over the past year divided by the current price. Newey and West (1987) corrected t-statistics are reported in parentheses. Adjusted R^2 and number of observations in each case are displayed. We report results for the entire sample period (January 1951-December 2004) and the first and second half separately. We obtain estimates of realizations of the fourth factor for the period 1951-1963 by $\hat{f} = (C'\Gamma^{-1}C)^{-1}C'\Gamma^{-1}\, r_t$, where the matrix of coefficients (C) and idiosyncratic variance (Γ) are estimated using the 1963-2004 sample period. Newey and West (1987) (with three lags) corrected t-statistics are reported in parentheses.

Table 8: Predictability of Equity Returns (controlling for Industry Returns)

Industry	f_4		Ind		Industry	f_4		Ind	
	coef(%)	t-stat	coef(%)	t-stat		coef(%)	t-stat	coef(%)	t-stat
Agriculture	-0.56	-2.76	2.76	0.69	Shipbuilding & Railroad Equip.	-0.54	-2.56	-2.23	-0.59
Food Products	-0.58	-2.63	-2.14	-0.33	Defense	-0.55	-2.66	-0.64	-0.19
Candy & Soda	-0.57	-2.68	-1.00	-0.25	Precious Metals	-0.49	-2.24	-1.56	-0.59
Beer & Liquor	-0.58	-2.75	-2.12	-0.47	Mining	-0.38	-1.63	-7.17	-1.72
Tobacco	-0.55	-2.74	-4.64	-1.50	Coal	-0.53	-2.35	-0.80	-0.30
Recreation	-0.60	-2.91	-6.34	-1.85	Petroleum and Natural Gas	-0.47	-1.94	-3.77	-0.61
Entertainment	-0.55	-2.68	0.29	0.07	Utilities	-0.61	-3.08	10.74	1.72
Printing & Publishing	-0.47	-2.22	7.99	1.28	Communication	-0.58	-2.87	-4.39	-0.68
Consumer Goods	-0.63	-2.88	-8.16	-1.16	Personal Services	-0.52	-2.58	5.02	1.21
Apparel	-0.55	-2.43	0.51	0.09	Business Services	-0.52	-2.43	3.54	0.44
Healthcare	-0.56	-2.76	-0.41	-0.20	Computers	-0.57	-2.84	-2.36	-0.52
Medical Equipment	-0.61	-2.92	-6.45	-1.09	Electronic Equipment	-0.54	-2.70	2.86	0.58
Pharmaceuticals	-0.60	-2.75	-5.36	-0.87	Measuring & Control Equipment	-0.56	-2.78	1.30	0.25
Chemicals	-0.54	-2.43	-1.73	-0.23	Business Supplies	-0.56	-2.73	0.01	0.00
Rubber & Plastic	-0.56	-2.75	-1.31	-0.26	Shipping Containers	-0.55	-2.74	-1.13	-0.20
Textiles	-0.56	-2.73	-0.33	-0.08	Transportation	-0.56	-2.75	1.08	0.20
Construction Materials	-0.56	-2.74	-3.54	-0.55	Wholesale	-0.55	-2.75	-1.05	-0.15
Construction	-0.56	-2.76	0.54	0.13	Retail	-0.57	-2.28	-0.76	-0.09
Steel Works	-0.45	-1.95	-4.69	-1.06	Restaraunts, Hotels & Motels	-0.60	-2.64	-2.54	-0.52
Fabricated Products	-0.54	-2.58	-1.05	-0.30	Banking	-0.52	-2.67	8.53	1.43
Machinery	-0.59	-2.61	2.71	0.36	Insurance	-0.51	-2.59	10.70	1.74
Electrical Equipment	-0.56	-2.76	-2.86	-0.45	Real Estate	-0.55	-2.71	0.75	0.19
Automobiles	-0.56	-2.74	-6.37	-1.33	Trading	-0.54	-2.74	7.97	0.82
Aircraft	-0.56	-2.74	1.79	0.43	Other	-0.56	-2.76	1.27	0.41

The table reports estimates of $mkt_{t+1} = \alpha_0 + \beta\, f_{4,t} + \alpha_1\, ind_{it} + \alpha_2\, mkt_{it} + \varepsilon_{t+1}$. The dependent variable (mkt) refers to the simple return on the CRSP value-weighted stock market index over the one-month Treasury-bill rate, while ind refers to returns of the corresponding industry portfolio. We use the 49 industry classification of Fama and French (1997). The data on portfolio returns are from Kenneth French's website. The sample period is July 1963 to December 2004. Newey and West (1987) (with three lags) corrected t-statistics are reported in parentheses.

Table 9: Predictability of Bond Returns

Maturity	f_4(%)	r_f	*yield*	*spread*	*ldp*	*cp*	R^2(%)	T
2 Years	-0.33						3.11	468
	(-3.07)							
	-0.22	3.32	1.92	-2.50	-1.59	0.41	41.06	468
	(-2.99)	(2.00)	(0.85)	(-0.33)	(-2.31)	(5.59)		
3 Years	-0.59						2.94	468
	(-3.00)							
	-0.40	4.59	2.46	-7.07	-2.79	0.84	42.06	468
	(-2.95)	(1.56)	(0.61)	(-0.53)	(-2.31)	(6.36)		
4 Years	-0.81						2.90	468
	(-3.00)							
	-0.55	6.47	4.71	-18.72	-3.44	1.23	44.78	468
	(-3.02)	(1.65)	(0.88)	(-1.05)	(-2.14)	(6.90)		
5 Years	-0.99						2.90	468
	(-3.01)							
	-0.70	8.65	9.03	-27.45	-3.94	1.37	42.21	468
	(-3.07)	(1.80)	(1.38)	(-1.26)	(-2.01)	(6.22)		

Table 9 presents estimates of:

$$r^{(k)}_{t,t+12} - y^{(1)}_t = a_k + b_k f_{4,t} + c_k Y_t + e_t, \qquad k = 2..5$$

where $r^{(k)}_{t,t+12}$ refers to the return of a bond with maturity k between months t and $t+12$ and Y_t refers to a vector of controls that includes the short rate (r_f), the 10-year yield on Aaa rated bonds minus the short rate (*yield*), the Aaa minus Baa credit spread (*spread*), the log dividend yield (*ldp*) and the Cochrane and Piazzesi (2005) hump-shaped factor (*cp*). $f_{4,t}$ refers to the fourth factor extracted in Section 4. The data on discount bonds are from Fama and Bliss (1987) and are available through WRDS. The data on the hump-shaped factor are from John Cochrane's website. The sample period is January 1964 to December 2002. t statistics with Hansen and Hodrick (1980) errors are reported in parenthesis. The coefficients on *ldp* and *cp* are multiplied by 100.

Table 10: Interest Rates

Horizon	Short Rate, y^{1m}				Term Structure Slope, $y^{10}-y^{1m}$				Default Spread, $y^{BAA}-y^{AAA}$			
(months)	f_1	f_2	f_3	f_4	f_1	f_2	f_3	f_4	f_1	f_2	f_3	f_4
1	-0.004	0.002	-0.001	0.000	0.000	0.002	0.002	0.004	-0.000	-0.000	**-0.001**	**-0.001**
	(-1.72)	(0.89)	(-0.52)	(0.03)	(0.02)	(0.68)	(0.85)	(1.66)	(-0.75)	(-1.25)	(**-2.65**)	(**-2.10**)
2	-0.001	**0.009**	0.004	**0.007**	-0.002	-0.003	-0.001	0.000	**-0.001**	-0.000	**-0.001**	**-0.001**
	(-0.23)	(**2.52**)	(1.08)	(**2.05**)	(-0.79)	(-1.06)	(-0.31)	(0.05)	(**-2.78**)	(-0.27)	(**-2.35**)	(**-2.09**)
3	0.003	**0.014**	**0.008**	**0.012**	-0.004	**-0.008**	-0.004	-0.006	**-0.002**	0.000	**-0.002**	-0.000
	(0.92)	(**3.63**)	(**2.10**)	(**3.07**)	(-1.46)	(**-2.55**)	(-1.12)	(-1.60)	(**-4.35**)	(0.28)	(**-3.49**)	(-0.03)
6	0.003	**0.018**	0.004	0.010	-0.004	**-0.010**	0.000	-0.000	**-0.003**	-0.001	**-0.002**	0.000
	(0.76)	(**3.61**)	(0.83)	(1.87)	(-1.23)	(**-2.55**)	(0.12)	(-0.11)	(**-4.96**)	(-0.76)	(**-2.77**)	(0.01)
12	0.008	**0.022**	0.010	**0.022**	-0.008	-0.010	-0.003	-0.006	**-0.002**	0.000	0.001	0.001
	(1.45)	(**3.08**)	(1.56)	(**3.04**)	(-1.78)	(-1.87)	(-0.69)	(-1.20)	(**-2.69**)	(0.13)	(0.61)	(1.30)

Table 10 presents estimation results for

$$X_{t+k} - X_t = a_0 + A(k)F_t + B(L)X_t + c_k Y_t + e_t$$

Standard errors are adjusted using the Hansen-Hodrick estimator. The interest rate data is from Datastream. F_t refers to the factors extracted in Section 3. The riskless rate is the yield on a 1m T-Bill, the 10-year yield is the yield on Aaa bonds and the default premium is the yield on Baa minus the yield on Aaa bonds. The vector of controls includes three lags of the riskless rate, the slope and the default spread. The coefficients $A(k)$ on the factor realizations are multiplied by 100. t-statistics are reported in parenthesis. Sample includes data from January 1951 to December 2004. We obtain estimates of factor realizations for the period 1951-1963 by $\hat{f} = (C'\Gamma^{-1}C)^{-1}C'\Gamma^{-1}\, r_t$, where the matrix of coefficients (C) and idiosyncratic variance (Γ) are estimated using the 1963-2004 sample period.

Table 11: Inflation, Output and Employment

Horizon	CPI, SA				Industrial Production, SA				Unemployment, SA			
(months)	f_1	f_2	f_3	f_4	f_1	f_2	f_3	f_4	f_1	f_2	f_3	f_4
1	-0.089	0.166	0.146	**0.294**	-0.017	**0.077**	**0.084**	0.049	-0.023	-0.211	**-0.359**	-0.176
	(-0.77)	(1.35)	(1.20)	**(2.49)**	(-0.49)	(**2.06**)	(**2.29**)	(1.36)	(-0.16)	(-1.38)	(**-2.39**)	(-1.20)
2	-0.035	0.056	0.146	**0.299**	0.070	**0.185**	**0.170**	0.018	-0.157	**-0.546**	**-0.607**	-0.079
	(-0.36)	(0.55)	(1.32)	**(2.83)**	(1.22)	(**2.98**)	(**2.77**)	(0.30)	(-0.76)	(**-2.47**)	(**-2.79**)	(-0.37)
3	0.007	0.102	0.154	**0.256**	**0.163**	**0.252**	**0.290**	0.030	-0.373	**-0.753**	**-0.915**	-0.044
	(0.08)	(1.12)	(1.51)	(**2.65**)	(**2.08**)	(**2.96**)	(**3.16**)	(0.33)	(-1.37)	(**-2.55**)	(**-2.88**)	(-0.14)
6	0.049	0.141	**0.176**	**0.168**	**0.472**	**0.294**	**0.600**	-0.030	**-1.326**	**-1.050**	**-1.644**	0.016
	(0.69)	(1.87)	(**2.02**)	(**1.97**)	(**3.65**)	(**1.99**)	(**3.93**)	(-0.19)	(**-2.92**)	(**-2.00**)	(**-3.07**)	(0.03)
12	**0.176**	**0.175**	0.152	**0.178**	**0.856**	0.389	**0.751**	0.031	**-2.776**	-1.639	**-2.844**	-0.342
	(**2.49**)	(**2.17**)	(1.89)	(**2.12**)	(**4.10**)	(1.51)	(**3.10**)	(0.12)	(**-3.52**)	(-1.64)	(**-3.10**)	(-0.34)

Table 11 presents estimation results for

$$\pi_{t+k} - \pi_t = a_0 + A(k)F_t + B(L)\Delta\pi_t + c_k Y_t + e_t$$

and

$$X_{t+k} - X_t = a_0 + A(k)F_t + B(L)X_t + c_k Y_t + e_t$$

where $\pi_t \equiv \ln(CPI_t) - \ln(CPI_{t-1})$ and $\pi_{t+k} \equiv \frac{1}{k}(\ln(CPI_{t+k}) - \ln(CPI_t))$. The data on inflation, industrial production and unemployment is from the St. Louis Federal Reserve and is seasonally adjusted. F_t refers to the factors extracted in Section 3. For the inflation regression, the vector of controls, Y_t, includes three lags of industrial production growth and unemployment. For industrial production and unemployment we only control for own lags. The coefficients $A(k)$ on the factor realizations are multiplied by 100. t-statistics computed using the Hansen and Hodrick (1980) estimator are reported in parenthesis. Sample includes data from January 1951 to December 2004. We obtain estimates of factor realizations for the period 1951-1963 by $\hat{f} = (C'\Gamma^{-1}C)^{-1}C'\Gamma^{-1} r_t$, where the matrix of coefficients (C) and idiosyncratic variance (Γ) are estimated using the 1963-2004 sample period.

Table 12: Asset Pricing Tests

C/S	MKT	SMB	HML	MOM	f_1	f_2	f_3	f_4	$MAPE$ (%)	$SSQE$ (%)	J
25 ME/BM	2.78								2.99	3.37	79.0
	(2.60)										(0.000)
	4.08	3.22	8.81						1.23	0.70	62.2
	(3.36)	(2.19)	(4.93)								(0.000)
	7.16	3.13	14.71	0.19					0.97	0.38	29.2
	(3.60)	(1.32)	(4.80)	(3.91)							(0.110)
					0.09	-0.30	0.29	-0.32	1.21	0.55	45.8
					(1.32)	(-3.23)	(3.18)	(-2.37)			(0.001)
25 ME/MOM	3.18								4.20	6.95	107.9
	(3.04)										(0.000)
	1.82	1.84	-4.81						3.67	5.57	111.7
	(1.37)	(1.22)	(-1.72)								(0.000)
	5.86	4.58	10.68	0.07					1.69	1.18	73.5
	(3.73)	(3.04)	(3.81)	(4.25)							(0.000)
					0.11	-0.06	0.12	-0.90	1.67	1.12	31.0
					(1.03)	(-0.61)	(0.87)	(-2.28)			(0.073)
25 ME/SR	3.20								2.82	3.09	78.1
	(2.98)										(0.000)
	4.88	2.26	9.92						2.33	2.20	64.6
	(3.43)	(1.40)	(3.29)								(0.000)
	3.15	1.93	4.30	-0.09					1.05	0.59	48.2
	(1.79)	(1.00)	(0.92)	(-2.87)							(0.001)
					0.10	-0.09	0.21	-0.74	1.53	0.80	27.4
					(1.26)	(-1.04)	(1.98)	(-2.67)			(0.158)
25 ME/LR	2.16								2.18	1.80	58.5
	(1.94)										(0.000)
	3.47	3.00	9.34						0.83	0.40	41.1
	(2.60)	(1.82)	(4.00)								(0.008)
	4.18	5.35	10.91	0.10					0.56	0.12	17.8
	(2.46)	(2.85)	(3.41)	(3.54)							(0.663)
					0.01	-0.07	0.19	0.23	1.21	0.62	40.0
					(0.21)	(-0.90)	(2.21)	(1.33)			(0.007)
6 ME/BM 6 ME/MOM 6 ME/SR 6 ME/LR	3.82								3.00	3.02	132.6
	(3.71)										(0.000)
	4.54	2.99	9.11						2.09	2.07	126.2
	(4.02)	(2.15)	(5.55)								(0.000)
	6.09	3.62	12.02	0.07					1.57	1.12	100.1
	(4.81)	(2.66)	(6.96)	(5.10)							(0.000)
					-0.11	-0.11	0.29	-0.50	1.97	1.35	120.8
					(-2.06)	(-1.88)	(3.83)	(-4.38)			(0.000)

Table 12 presents results of asset pricing tests. The test assets include 25 portfolios sorted on Market-Equity (ME) and Book-to-Market, 25 portfolios sorted on Market-Equity (ME) and past return between $t-2$ and $t-12$ (MOM), 25 portfolios sorted Market-Equity (ME) and past return between t and $t-1$ (SR) and 25 portfolios sorted on Market-Equity (ME) and past 5-year return (LR). The last row contains results of a pooled estimation on a total of 24 portfolios, 6 of every sort. See main text for the estimation details.

Table 13: Asset Pricing Tests - Actual vs Predicted Expected Returns

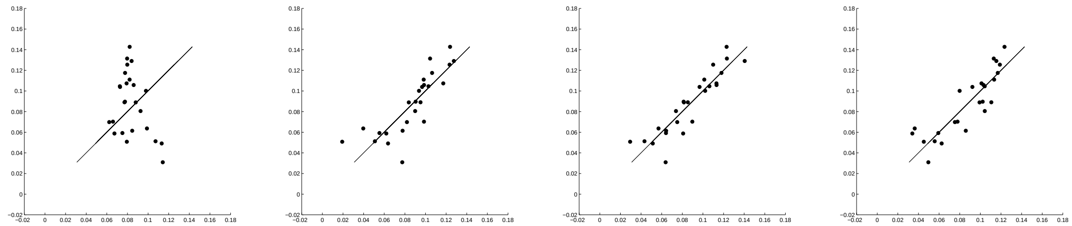
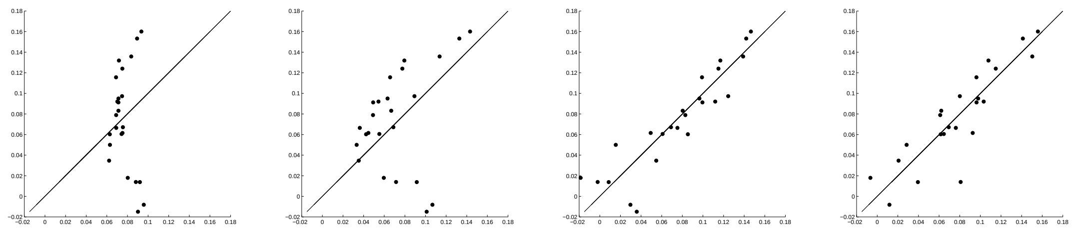
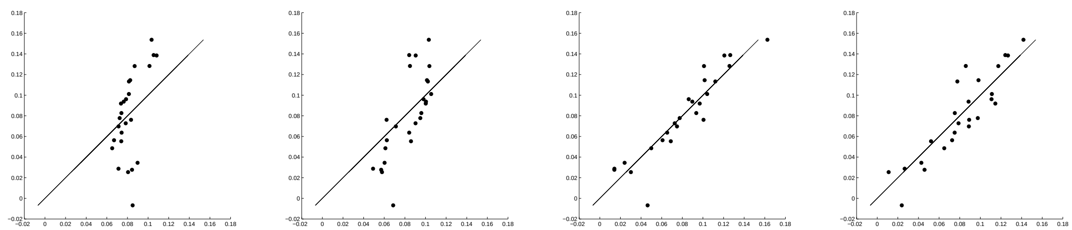
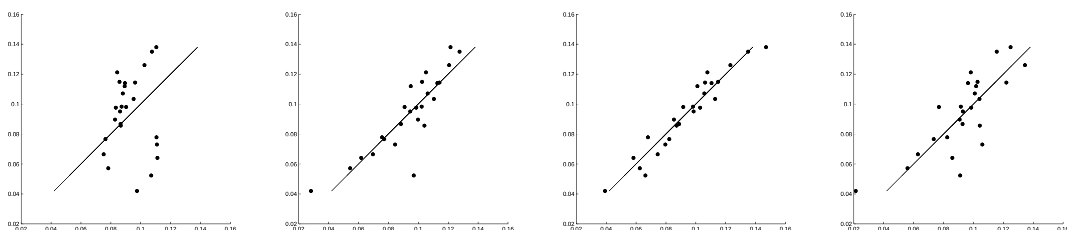
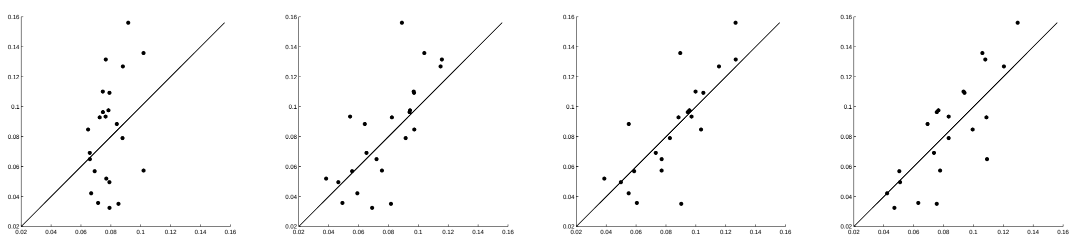
CS	CAPM	FF3	Carhart	4f
25 ME/BM				
25 ME/MOM				
25 ME/SR				
25 ME/LR				
6 ME/BM 6 ME/MOM 6 ME/SR 6 ME/LR				

Table 13 plots visually the results of asset pricing tests. We plot historical average excess returns for the test assets versus predicted average excess returns. The test assets include 25 portfolios sorted on Market-Equity (ME) and Book-to-Market, 25 portfolios sorted on Market-Equity (ME) and past return between $t-2$ and $t-12$ (MOM), 25 portfolios sorted Market-Equity (ME) and past return between t and $t-1$ (SR) and 25 portfolios sorted on Market-Equity (ME) and past 5-year return (LR). The last row contains results of a pooled estimation on a total of 24 portfolios, 6 of every sort.

Table 14: UK: Correlations of Industry portfolios with extracted factors

Industry	Correlations			
	f_1^{UK}	f_2^{UK}	f_3^{UK}	f_4^{UK}
Food Products	0.84	-0.20	0.06	-0.22
Beer & Liquor	0.80	-0.22	0.09	-0.21
Tobacco Products	0.63	-0.20	-0.18	0.02
Recreation	0.76	0.16	0.25	0.14
Printing & Publishing	0.70	0.45	0.24	0.03
Consumer Goods	0.75	-0.05	0.24	-0.13
Apparel	0.57	0.05	0.24	0.11
Healthcare	0.67	0.01	-0.23	-0.20
Chemicals	0.75	0.07	0.043	0.09
Construction and Construction Materials	0.79	0.05	0.36	0.24
Steel Works	0.40	0.28	0.33	0.25
Fabricated Products and Machinery	0.72	0.14	0.56	0.15
Electrical Equipment	0.57	0.34	0.62	0.09
Automobiles and Trucks	0.56	0.19	0.41	-0.02
Aircraft, Ships, & Railroad equipment	0.66	0.06	0.36	0.12
Mining	0.61	0.14	0.17	0.52
Petroleum and Natural Gas	0.64	0.00	-0.11	0.63
Utilities	0.63	-0.01	-0.24	-0.08
Communication	0.57	0.50	-0.06	0.04
Personal & Business Services	0.66	0.63	0.21	0.01
Business Equipment	0.44	0.88	0.15	0.03
Business Supplies & Shipping Containers	0.67	0.08	0.42	0.18
Transportation	0.79	0.10	0.35	0.05
Wholesale	0.75	0.20	0.41	0.17
Retail	0.83	-0.05	0.07	0.00
Restaurants, Hotels & Motels	0.82	0.12	0.17	0.03
Banking, Insurance, Real Estate & Trading	0.89	0.19	0.02	0.17

Table 14 presents the correlations of the 27 industry portfolios with the four factors f_i for the United Kingdom. The stock return data are from the London Share Price database (LSPD). The industry portfolios are constructed using the classification scheme of Fama and French (1997), mapping SIC codes to 30 industries. Due to lack of data, we drop the Coal, Textiles and Other industries in the UK sample. The factors are extracted from 27 Fama-French industry portfolios using two step estimation method described in Section 3. Correlations are computed using monthly data over the period January 1985 to December 2004.

Table 15: UK: Factor Interpretation

Factor	Portfolio					$R^2(\%)$
	MKT					
f_1^{UK}	0.936					88.1
	(42.50)					
		Business Equip.	Telecom.	Food	Beer	
f_2^{UK}		0.949	0.040	-0.271	-0.206	95.7
		(53.14)	(2.15)	(-13.83)	(-10.45)	
		Fab. Prod	Autos	Utilities	Health	
f_3^{UK}		0.704	0.207	-0.390	-0.399	66.3
		(-7.35)	(-9.67)	(9.55)	(8.05)	
	MKT	Steel	Mines	Coal	Oil	
f_4^{UK}	-0.965	0.266	0.564	-	0.821	77.9
	(-24.98)	(15.75)	(7.52)	-	(19.47)	

Table 15 reports coefficients from a projection of the UK factors (f_i) on a subset of industries, for the period 1987-2007. The UK stock data is from the London share Price Database (LSPD). The industry portfolios are constructed using the classification scheme of Fama and French (1997), mapping SIC codes to 30 industries. Due to lack of data, we drop the Coal, Textiles and Other industries in the UK sample. The factors are extracted from 27 Fama-French industry portfolios using two step estimation method described in Section 3. All returns have been normalized to zero mean and unit variance.

Table 16: UK: Inflation

Horizon	A (no controls)				B (controlling for yields)			
(months)	f_1^{UK}	f_2^{UK}	f_3^{UK}	f_4^{UK}	f_1^{UK}	f_2^{UK}	f_3^{UK}	f_4^{UK}
1	-0.17	-0.04	0.06	**0.54**	-0.20	-0.11	0.11	**0.45**
	(-1.08)	(-0.26)	(0.42)	(**3.35**)	(-1.38)	(-0.76)	(0.79)	(**3.01**)
2	-0.08	0.00	0.07	**0.41**	-0.10	-0.06	0.13	**0.33**
	(-0.78)	(0.04)	(0.59)	(**3.24**)	(-1.03)	(-0.59)	(1.32)	(**3.16**)
3	0.02	0.01	0.08	**0.31**	-0.00	-0.05	0.14	**0.23**
	(0.22)	(0.06)	(0.72)	(**2.94**)	(-0.03)	(-0.55)	(1.64)	(**2.65**)
6	0.07	-0.09	0.10	**0.26**	0.03	-0.15	**0.17**	**0.16**
	(0.86)	(-0.88)	(1.06)	(**2.56**)	(0.45)	(-1.89)	(**2.83**)	(**2.39**)
12	0.07	-0.03	0.04	**0.27**	0.03	-0.08	**0.11**	**0.16**
	(1.32)	(-0.30)	(0.43)	(**2.73**)	(0.56)	(-1.01)	(**2.37**)	(**2.77**)

Table 16 presents estimation results for

$$\pi_{t+k} - \pi_t = a_0 + A(k)F_t^{UK} + B(L)\Delta\pi_t + c_k Y_t + e_t$$

where $\pi_t \equiv \ln(RPI_t) - \ln(RPI_{t-1})$ and $\pi_{t+k} \equiv \frac{1}{k}\left(\ln(RPI_{t+k}) - \ln(RPI_t)\right)$. RPI refers to the UK Retail Prices Index (RPI), F_t refers to the factors extracted in section 7 using stock market data from the United Kingdom. Standard errors are adjusted using the Hansen and Hodrick (1980) estimator. Sample includes data from 1987 to 2007. In panel A the vector of controls, Y_t, includes three lags of industrial production growth. In Panel B we also control for the 3m and 10y bond yields. The inflation data are from the Office for National Statistics, bond yield data are from the Bank of England.

Table 17: UK: Predictability of Stock Market and Bond Returns.

Asset (ret_{t+1})	const	f_4(%)	r_f	$yield$	ret_t	R^2
FTSE	0.321	-0.487				0.78%
	(1.12)	(-1.87)				
	0.061	-0.559	0.036	0.172	0.084	0.56%
	(0.04)	(-1.94)	(0.23)	(0.81)	(1.34)	
10-y bond	0.008	-0.151				1.13%
	(1.51)	(-2.14)				
	0.014	-0.147	0.113	0.077	0.1530	4.11%
	(0.06)	(-1.99)	(1.93)	(0.53)	(2.56)	

Table 17 reports estimates from OLS predictive regressions of

$$ret_{t+1} - r_{f,t+1} = const + af_{4,t}^{UK} + \alpha_1 r_{f,t} + \alpha_2 yield_t + \alpha_3 ret_t + \varepsilon_{t+1}.$$

The dependent variable, ret, refers to 1-month returns of UK market portfolio and the 10-year bond. The UK market portfolio is the Financial Times Stock Exchange (FTSE) Index. $f_{4,t}^{UK}$ refers to the fourth factor extracted using stock market data from the United Kingdom. r_f is the 1-month LIBOR. The yield is the 10-year yield minus the 3-month rate. Newey and West (1987) corrected t-statistics are reported in parentheses. Adjusted R^2 in each case are displayed. The sample period is January 1987 to December 2007. The data on LIBOR and bond yields are from the Bank of England. The data on stock market (FTSE), interest rate and bond returns are from the Global Financial database.