Evaluating Latent and Observed Factors in Macroeconomics and Finance

Jushan Bai* Serena Ng †

September 15, 2003

Preliminary: Comments Welcome

Abstract

Common factors play an important role in many disciplines of social science. In economics, the factors are the common shocks that underlie the co-movements of the large number of economic time series. The question of interest is whether some observable economic variables are in fact the underlying unobserved factors. We consider statistics to determine if the observed and the latent factors are exactly the same. We also provide simple to construct statistics that indicate the extent to which the two sets of factors differ. The key to the analysis is that the space spanned by the latent factors can be consistently estimated when the sample size is large in both the cross-section and the time series dimensions. The tests are used to assess how well the Fama and French factors as well as several business cycle indicators approximate the factors in portfolio and individual stock returns, and a large panel of macroeconomic data.

*Department of Economics, 269 Mercer St, NYU, New York, NY 10022 Email: Jushan.Bai@nyu.edu.
†Department of Economics, University of Michigan, Ann Arbor, MI 48109 Email: Serena.Ng@umich.edu

The authors acknowledge financial support from the NSF (grants SBR-9896329, SES-0137084, SES-0136923).

1 Introduction

Many economic theories have found it useful to explain the behavior of the observed data by means of a small number of fundamental factors. For example, the arbitrage pricing theory (APT) of Ross (1976) is built upon the existence of a set of common factors underlying all asset returns. In the capital asset pricing theory (CAPM) of Sharpe (1964), Lintner (1965), and Merton (1973), the 'market return' is the common risk factor that has pervasive effects on all assets. In consumption based CCAPM models of Breeden (1989) and Lucas (1978), aggregate consumption is the source of systematic risk in one-good exchange economies. Because interest rate of different maturities are highly correlated, most models of interest rates have a factor structure. Indeed, Stambaugh (1988) showed the conditions under which an affine-yield model implies a latent-variable structure for bond returns. A fundamental characteristic of business cycles is the comovement of a large number of series, which is possible only if economic fluctuations are driven by common sources.

While the development of theory can proceed without a complete specification of how many and what the factors are, empirical testing does not have this luxury. To test the multifactor APT, one has to specify the empirical counterpart of the theoretical risk factors. To test CAPM, one must specify what the 'market' is, and to test CCAPM, one must specifically define consumption. To test term structure models, the common practice is to identify instruments that are correlated with the underlying state variables, which amounts to finding proxies for the latent factors. To be able to asses the sources of business cycle fluctuations, one has to identify the candidates for the primitive shocks.

A small number of applications have proceeded by replacing the unobserved factors with statistically estimated ones. For example, Lehmann and Modest (1988) used factor analysis, while Connor and Korajzcyk (1998) adopt the method of principal components. The drawback is that the statistical factors do not have immediate economic interpretation. A more popular approach is to rely on intuition and theory as guide to come up with a list of observed variables as 'proxy' or 'mimicking' factors in place of the theoretical factors that are unobserved. Variables such as the unemployment rate in excess of the natural rate, the deviation of output from its potential, are all popular candidates for the state of the economy. In CAPM analysis, the equal-weighted and value-weighted market returns are often used in place of the theoretical market return. In CCAPM analysis, non-durable consumption is frequently used as the systematic risk factor. By regressing asset returns on a large number of financial and macroeconomic variables and analyzing their explanatory power, Chen, Roll and Ross (1986) found that the factors in APT are related to expected and unexpected inflation, interest rate risk, term structure risk, and industrial production. Perhaps the most well-known of observable risk factors are the three Fama and French factors:- the market excess return, the small minus big factor, and the high minus low factor.

There is a certain appeal in associating the latent factors with observed variables as this facilitates economic interpretation. But empirical testing using the observed factors is valid only if the observed and the fundamental factors span the same space. To date, there does not exist a formal test that provides information about the adequacy of the observed variables as proxies for the unobserved factors. The problem is not so much that the fundamental factors are unobserved, but that consistent estimates of them cannot be obtained under the traditional assumption that T is large and N is fixed, or vice versa.

In this paper, we develop several procedures to compare the (individual or set of) observed with the unobserved factors. The point of departure is that we work with large dimensional panels. That is, datasets with a large number of cross-section units (N) and time series observations (T). By allowing both N and T to tend to infinity, the space spanned by the common factors can be estimated consistently. Our analysis thus combines the statistical approach of Lehmann and Modest (1988) and Connor and Korajzcyk (1998), with the economic approach of using observed variables as proxies. We begin in Section 2 by considering estimation of the factors. Because the estimated factors can be treated as though they are known, Section 4 presents tests to compare the observed variables with the estimated factors. In section 4, we use the procedures to compare observed variables with factors estimated from portfolio returns, individual stock returns, and a large set of macroeconomic time series. Proofs are given in the Appendix.

2 Preliminaries

Consider the factor representation for a panel of data x_{it}, $(i = 1, \dots N, t = 1, \dots T)$

$$x_{it} = \lambda_i' F_t + e_{it},$$

where F_t $(r \times 1)$ is the factor process, and λ_i $(r \times 1)$ is the factor loading for unit i. In classical factor analysis, the number of units, N, is fixed and the number of observations, T, tends to infinity. With macroeconomic and financial applications, this assumption is not fully satisfactory because data for a large number of cross-section units are often available over a long time span, and in some cases, N can be much larger than T (and vice versa). For example, daily data on returns for well over one hundred stocks are available since 1960, and a large number of price and interest rate series are available for over forty years on a monthly basis. Classical analysis also assumes e_{it} is iid over t and independent over i. This implies a diagonal variance covariance matrix for $e_t = (e_{1t}, ..., e_{Nt})$, an assumption that is rather restrictive. To overcome these two limitations, we work with high dimensional approximate factor models that allow both N and T to tend to infinity, and in which e_{it} may be serially and cross-sectionally correlated so that the covariance matrix of e_t does not have to be a diagonal matrix. In fact, all that is needed is that the largest eigenvalue of the covariance matrix of e_{it} is bounded as N tends to infinity.

Suppose we observe G_t, an $(m \times 1)$ vector of economic variables. We are ultimately interested in the relationship between G_t and F_t, but we do not observe F_t. It would seem natural to proceed by regressing x_{it} on G_t, and then use some metric to assess the explanatory power of G_t[1]. The idea is that G_t is a good proxy for F_t, it should explain x_{it}. This, however, is not a satisfactory test because even if G_t equals F_t exactly, G_t might still only be weakly correlated with x_{it} if the variance of the idiosyncratic error e_{it} is large. In other words, low explanatory power of G_t by itself may not be the proper criterion for deciding if G_t corresponds to the true factors.

In CAPM analysis, several approaches have been used to check if inference is sensitive to the use of a proxy in place of the market portfolio. Stambaugh (1982) considered returns from a number of broadly defined markets and conclude that inference is not sensitive to the choice of proxies. This does not, however, suggest that the unobservability of the market portfolio has no implication for inference, an issue raised by Roll (1977). Another approach, used in Kandel and Stambaugh (1987) and Shanken (1987), is to obtain the cut-off correlation between the market proxy return and the true market return that would change the conclusion on the hypothesis being tested. They find that if the correlation between G_t and F_t is at or above .7, inference will remain intact. However, this begs the question of whether the correlation between the unobserved market return and the proxy variable is high. While these approaches provide for more cautions inference, the basic problem remains that we do not know how correlated are F_t and G_t. To be able to test F_t and G_t directly, we must first confront the problem that F_t is not observed.

We use the method of principal components to estimate the factors. Let X be the T by N matrix of observations such that the ith column is the time series of the ith cross section. Let V_{NT} be a $r \times r$ diagonal matrix consisting of the r largest eigenvalues of XX'/NT. Let $\widetilde{F} = (\widetilde{F}_1, ..., \widetilde{F}_T)'$ be the principal component estimates of F under the normalization that

[1]Such an approach was adopted in Chen et al. (1986), for example

$\frac{F'F}{T} = I_r$. Then $\widetilde{F}$ is comprised of the r eigenvectors (multiplied by $\sqrt{T}$) associated with the r largest eigenvalues of the matrix $XX'/(NT)$ in decreasing order. Let $\Lambda = (\lambda_1, ..., \lambda_N)'$ be the matrix of factor loadings. The principal components estimator of Λ is $\widetilde{\Lambda} = X'\widetilde{F}/T$. Define $\widetilde{e}_{it} = x_{it} - \widetilde{\lambda}_i'\widetilde{F}_t$.

Denote the norm of a matrix A by $\|A\| = [tr(A'A)]^{1/2}$. The notation M stands for a finite positive constant, not depending on N and T. The following assumptions are needed for consistent estimation of the factors by the method of principal components.

Assumptions: *Assumption A: Common factors*

1. $E\|F_t\|^4 \leq M$ and $\frac{1}{T}\sum_{t=1}^{T} F_t F_t' \xrightarrow{p} \Sigma_F$ for a $r \times r$ positive definite matrix Σ_F.

Assumption B: Heterogeneous factor loadings

The loading λ_i is either deterministic such that $\|\lambda_i\| \leq M$ or it is stochastic such that $E\|\lambda_i\|^4 \leq M$. In either case, $\Lambda'\Lambda/N \xrightarrow{p} \Sigma_\Lambda$ as $N \to \infty$ for some $r \times r$ positive definite non-random matrix Σ_Λ.

Assumption C: Time and cross-section dependence and heteroskedasticity

1. $E(e_{it}) = 0$, $E|e_{it}|^8 \leq M$;

2. $E(e_{it}e_{js}) = \tau_{ij,ts}$, $|\tau_{ij,ts}| \leq \tau_{ij}$ for all (t, s) and $|\tau_{ij,ts}| \leq \gamma_{ts}$ for all (i, j) such that

$$\frac{1}{N}\sum_{i,j=1}^{N} \tau_{ij} \leq M, \ \frac{1}{T}\sum_{t,s=1}^{T} \gamma_{ts} \leq M, \text{ and } \frac{1}{NT}\sum_{i,j,t,s=1} |\tau_{ij,ts}| \leq M$$

3. For every (t, s), $E|N^{-1/2}\sum_{i=1}^{N}\Big[e_{is}e_{it} - E(e_{is}e_{it})\Big]|^4 \leq M$.

Assumption D: $\{\lambda_i\}, \{u_t\}$, and $\{e_{it}\}$ are three groups of mutually independent stochastic variables.

Assumptions A and B together imply r common factors. Assumption C allows for limited time series and cross section dependence in the idiosyncratic component. Heteroskedasticity in both the time and cross section dimensions is also allowed. Under stationarity in the time dimension, $\gamma_N(s, t) = \gamma_N(s - t)$, though the condition is not necessary. Given Assumption C1, the remaining assumptions in C are easily satisfied if the e_{it} are independent for all i and t. The allowance for weak cross-section correlation in the idiosyncratic components leads to the *approximate factor structure* of Chamberlain and Rothschild (1983). It is more general than a *strict factor model* which assumes e_{it} is uncorrelated across i. Assumption D is standard in factor analysis.

As is well known, the factor model is fundamentally unidentified because $\lambda_i' H H^{-1} F_t = \lambda_i' F_t$ for any invertible matrix H. In economics, exact identification of the factors, F_t, may not always be necessary. If the estimated F_t is used for forecasting as in Stock and Watson (2002), the distinction between F_t and HF_t is immaterial because they will give the same forecast. When stationarity or the cointegrating rank of F_t is of interest, knowing HF_t is sufficient, as F_t has the same cointegrating rank as HF_t. In these situations as well as addressing the question we are interested in, namely, determining if F_t is close to G_t, what is important is consistent estimation of the space spanned by the factors.

Lemma 1 *Let $H = V_{NT}^{-1}(\widetilde{F}'F/T)(\Lambda'\Lambda/N)$. Under Assumptions A-D, and as $N, T \to \infty$,*

i $\min[N,T]\left(\frac{1}{T}\sum_{t=1}^{T} \|\widehat{F}_t - HF_t\|^2\right) = O_p(1)$;

ii $\sqrt{N}(\widetilde{F}_t - HF_t) \xrightarrow{d} V^{-1}QN(0,\Gamma_t)$, *where* $\widetilde{F}'F/T \xrightarrow{p} Q$, $V_{NT} \xrightarrow{p} V$, *and* $\Gamma_t = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} E(\lambda_i \lambda_j{}' e_{it} e_{jt})$.

Part (i), shown in Bai and Ng (2002), establishes that the squared difference between the estimated and the scaled true factors vanish as N and T tend to infinity. As shown in Bai (2003), $\sqrt{N}(\widetilde{F}_t - HF_t) = V_{NT}^{-1}(\widetilde{F}'F/T)\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \lambda_i e_{it} + o_p(1)$, from which the sampling distribution of the factor estimates given in (ii) follows. These limiting distributions are asymptotically independent across t if e_{it} are serially uncorrelated. The tests to be developed are built upon Lemma 1 and the fact that Γ_t is consistently estimable.

3 Comparing the Estimated and the Observed Factors

We observe G_t, and want to know if its m elements are generated by (or is a linear combination of) the r latent factors, F_t. In general, r is an unknown parameter. Consider estimating r using one of the two panel information criterion:

$$\begin{aligned}
\widehat{r} &= \underset{k=0,\ldots kmax}{\text{argmax}}\ PCP(k) \quad \text{where} \quad PCP(k) = \widehat{\sigma}^2(k) + \widehat{\sigma}^2(kmax)k \cdot g(N,T), \\
\widehat{r} &= \underset{k=0,\ldots kmax}{\text{argmax}}\ ICP(k) \quad \text{where} \quad ICP(k) = \log \widehat{\sigma}^2(k) + k \cdot g(N,T),
\end{aligned}$$

where $\widehat{\sigma}^2(k) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \widetilde{e}_{it}^2$, $\widetilde{e}_{it} = x_{it} - \widetilde{\lambda}_i' \widetilde{F}_t$, with $\widetilde{\lambda}_i$ and $\widetilde{F}_t$ estimated by the method of principal components. In Bai and Ng (2002), we showed that $\text{prob}(\widehat{r} = r) \to 1$ as $N, T \to \infty$

if $g(N,T)$ is chosen such that $g(N,T) \to 0$ and $\min[N,T]g(N,T) \to \infty$. Because r can be consistently estimated, in what follows, we simply treat it as though it is known.

Obviously, if $m < r$, the m observed variables cannot span the space of the r latent factors. Testing for the adequacy of G as a set is meaningful only if $m \geq r$. Nonetheless, regardless of the dimension of G_t and F_t, it is still of interest to know if a particular G_t is in fact a fundamental factor. Section 3.1 therefore begins with testing the observed variables one by one. Section 3.2 considers test of the observed variables as a whole.

3.1 Testing G_t One at a Time

Let G_{jt} be an element of the m vector G_t. The null hypothesis is that G_{jt} is an exact factor, or more precisely, that there exists a δ_j such that $G_{jt} = \delta_j' F_t$ for all t. Consider the regression $G_{jt} = \gamma_j' \widetilde{F}_t + error$. Let $\widehat{\gamma}_j$ be the least squares estimate of γ_j and let $\widehat{G}_{jt} = \widehat{\gamma}_j' \widetilde{F}_t$. Consider the t-statistic

$$\tau_t(j) = \frac{\sqrt{N}(\widehat{G}_{jt} - G_{jt})}{\left(var(\widehat{G}_{jt})\right)^{1/2}}.$$

Let Φ_α^τ be the α percentage point of the limiting distribution of $\tau_t(j)$. Then $\mathcal{A}(j) = \frac{1}{T}\sum_{t=1}^{T} 1(|\tau_t(j)| > \Phi_\alpha^\tau)$ is the frequency that $\tau_t(j)$ exceeds the α percent critical value in a sample of size T.

The $\mathcal{A}(j)$ statistic allows G_{jt} to deviate from $\widehat{G}_{jt}$ for a pre-specified number of time points as specified by α. A stronger test is to require G_{jt} not to deviate from $\widehat{G}_{jt}$ by more than sampling error at *every* t. To this end, we also consider the statistic $\mathcal{M}(j) = \max_{1\leq t\leq T} |\tau_t(j)|$. The $\mathcal{M}(j)$ statistic is a test of how far is the $\widehat{G}_{jt}$ curve from G_{jt}. It is a stronger test than $A(j)$ which tests G_{jt} point by point.

Proposition 1 *(exact tests) Let G_t be a vector of m observed factors, and F_t be a vector of r latent factors. Let $\widehat{\tau}_t(j)$ be obtained with $var(\widehat{G}_{jt})$ replaced by its consistent estimate, $\widehat{var}(\widehat{G}_{jt})$. Consider the statistics:*

$$A(j) = \frac{1}{T}\sum_{t=1}^{T} 1(|\widehat{\tau}_t(j)| > \Phi_\alpha) \quad (1)$$

$$M(j) = \max_{1\leq t\leq T} |\widehat{\tau}_t(j)|. \quad (2)$$

Under the null hypothesis that $G_{jt} = \delta' F_t$ and as $N,T \to \infty$, (i) $A(j) \xrightarrow{p} 2\alpha$, and (ii) $P(M(j) \leq x) \approx [2\Phi(x) - 1]^T$, where $\Phi(x)$ is the cdf of a standard normal random variable.

If F_t was observed, $\tau_t(j)$ is the usual t-statistic which is approximately normal for large N. Although our tests are based on $\widetilde{F}_t$, it is close to HF_t by Lemma 1. Thus, $\widehat{G}_{jt} = \widehat{\gamma}_j'\widetilde{F}_t$ is close to $G_{jt} = \delta_j' F_t$ with $\gamma_j' = \delta_j' H^{-1}$. For this reason, $\widehat{\tau}_t(j)$ is asymptotically normal. This allows us to use the critical values from the standard normal distribution for inference. Since the probability that a standard normal random variable z_t exceeds Φ_α equals 2α, we have $A(j) \to 2\alpha$ as stated. Furthermore, the probability that the standardized deviation between G_{jt} and $\widehat{G}_{jt}$ should not exceed the maximum value in a sample of T normal random variables. Extreme value distribution can also be used to approximate the distribution of $M(j)$ after proper centering and scaling.

As shown in the Appendix, the asymptotic variance of $\widehat{G}_{jt}$ is

$$\text{var}(\widehat{G}_{jt}) = \gamma_j' V_{NT}^{-1}(\frac{\widetilde{F}'F}{T})\Gamma_t(\frac{\widetilde{F}'F}{T})V_{NT}^{-1}\gamma_j'.$$

An estimate of it can be obtained by first substituting $\widetilde{F}$ for F, and noting that $\widetilde{F}'\widetilde{F}/T$ is an r-dimensional identity matrix by construction. Thus,

$$\widehat{\text{var}}(\widehat{G}_{jt}) = \widehat{\gamma}_j' V_{NT}^{-1}\widehat{\Gamma}_t V_{NT}^{-1}\widehat{\gamma}_j,$$

where $\widehat{\Gamma}_t$ is a consistent estimate of $\Gamma_t = \frac{1}{N}\sum_{i=1}^N \sum_{j=1}^N \widetilde{\lambda}_i \lambda_j' \widetilde{e}_{it}\widetilde{e}_{jt}$. For fixed T, Γ_t is not consistently estimable. The problem is akin to the problem in time series that summing the T autocovariances will not yield a consistent estimate of the spectrum. With cross-section data, the data have no natural ordering, and the time series solution of truncation is neither intuitive nor possible without a precise definition of distance between observations. However, as shown in Bai and Ng (2003), the following estimator is consistent for Γ_t:

$$\widetilde{\Gamma}_t = \frac{1}{N}\sum_{i=1}^N \sum_{j=1}^N \widetilde{\lambda}_i \widetilde{\lambda}_j' \frac{1}{T}\sum_{t=1}^T \widetilde{e}_{it}\widetilde{e}_{jt} \quad \forall\, t. \tag{3}$$

This covariance estimator is robust to cross-correlation, and accordingly, we refer to it as CS-HAC. It makes use of repeated observations over time, so that under covariance stationarity, the time series observations can be used to estimate the cross-section correlation matrix.

When e_{ij} is cross-sectionally uncorrelated, an estimate of Γ_t is

$$\widehat{\Gamma}_t = \frac{1}{N}\sum_{i=1}^N \widetilde{e}_{it}^2 \widetilde{\lambda}_i \widetilde{\lambda}_i'. \tag{4}$$

If, in addition, time series homoskedasticity is assumed, then $\Gamma_t = \Gamma$ does not depend on t, and a consistent estimate of Γ is

$$\widehat{\Gamma} = \widehat{\sigma}_e^2 \frac{\widetilde{\Lambda}'\widetilde{\Lambda}}{N}, \tag{5}$$

where $\widehat{\sigma}_e^2 = \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \widetilde{e}_{it}^2$, and $\widetilde{\Lambda}$ is the $N \times r$ matrix of estimated factor loadings. Testing if G_{jt} is an exact factor is then quite simple. For example, if $\alpha = 0.025$, the fraction of $\tau_t(j)$ that exceeds 1.96 in absolute value should be close to 5% for large N and T. Thus, $A(j)$ should be close to .05. Furthermore, $M(j)$ should not exceed the maximum of a vector of $N(0,1)$ random variables (in absolute values) of length T. This maximum value increases with T, but can be easily tabulated by simulations or theoretical computations. The 5% critical values for the $M(j)$ test when $T = 50, 100, 200, 400$ are, 3.283, 3.474, 3.656, and 3.830, respectively.

Requiring that G_t be an exact linear combination of the latent factors is rather strong. An observed series might match the variations of the latent factors very closely, and yet is not an exact factor in a statistical sense. Measurement error, for example, could be responsible for deviations between the observed variables and the latent factors. In such instances, it would be useful to gauge how large the departures are.

Proposition 2 *(approximate tests) Suppose $G_{jt} = \delta_j' F_t + \varepsilon_{jt}$, with $\varepsilon_{jt} \sim (0, \sigma_\varepsilon^2(j))$. Let $\widehat{G}_{jt} = \widehat{\gamma}_j' \widetilde{F}_t$, where $\widehat{\gamma}_j$ is obtained by least squares from a regression of G_{jt} on $\widetilde{F}_t$. Let $\widehat{\varepsilon}_{jt} = G_{jt} - \widehat{G}_{jt}$. Then (i) $\widehat{\sigma}_\varepsilon^2(j) = \frac{1}{T}\sum_{t=1}^T \widehat{\varepsilon}_{jt}^2 \xrightarrow{p} \sigma_\varepsilon^2(j)$, and (ii)*

$$\frac{\sqrt{T}(\widehat{\varepsilon}_{jt} - \varepsilon_{jt})}{s_{jt}} \xrightarrow{d} N(0,1),$$

where $s_{jt}^2 = F_t' \Sigma_F^{-1} F_t \sigma_\varepsilon^2 + (T/N) var(\widehat{G}_{jt})$.

We generically refer to ε_{jt} as a measurement error, even though it might be due to systematic differences between F_t and G_{jt}. If F_t was observed, the error in fitting G_{jt} would be confined to the first term of s_{jt}^2, as is standard of regression analysis. But because we regress G_t on $\widetilde{F}_t$, $\widehat{\varepsilon}_{jt}$ now consists of the error from estimation of F_t. Proposition 2 is implied by results in Bai and Ng (2003), where prediction intervals for models augmented with factors estimated by the method of principal components are derived. A consistent estimator of s_{jt}^2 is

$$\widehat{s}_{jt}^2 = \widetilde{F}_t' \widetilde{F}_t \widehat{\sigma}_\varepsilon^2(j) + \frac{T}{N}\widehat{\text{var}}(\widehat{G}_{jt}),$$

where $\widehat{\text{var}}(\widehat{G}_{jt})$ can be constructed as discussed above. Notably, the error from having to estimate F_t is decreasing in N.

Explicit consideration of measurement error allows us to construct, for each t, a confidence interval for ε_{jt}. For example, at the 95% level, the confidence interval is

$$(\varepsilon_{jt}^-, \varepsilon_{jt}^+) = (\widehat{\varepsilon}_{jt} - 1.96\frac{\widehat{s}_{jt}}{T}, \quad \widehat{\varepsilon}_{jt} + 1.96\frac{\widehat{s}_{jt}}{T}). \tag{6}$$

If G_{jt} is an exact factor, zero should lie in the confidence interval for each t. It could be of economic interest to know at which time points G_{jt} deviates from $\widehat{G}_{jt}$.

Instead of comparing G_{jt} with $\widehat{G}_{jt}$ at each t, two overall statistics can also be constructed.

$$NS(j) = \frac{\widehat{\text{var}}(\widehat{\varepsilon}(j))}{\widehat{\text{var}}(\widehat{G}(j))} \tag{7}$$

$$R^2(j) = \frac{\widehat{\text{var}}(\widehat{G}(j))}{\widehat{\text{var}}(G(j))}. \tag{8}$$

The $NS(j)$ statistic is simply the noise-to-signal ratio. If G_{jt} is an exact factor, the population value of $NS(j)$ is zero. A large $NS(j)$ thus indicates important departures of G_{jt} from the latent factors. A limitation of this statistic is that it leaves open the question of what is small and what is large. For this reason, we also consider $R^2(j)$, which should be unity if G_{jt} is an exact factor, and zero if the observed variable is irrelevant. The practitioner can draw inference as to how good is G_{jt} as a proxy factor by picking cut off points for $NS(j)$ and $R^2(j)$, in the same way we select the size of the test. These statistics are useful because instead of asking how large is the measurement error in the proxy variable that would overturn hypothesis, we now have consistent estimates for the size of the measurement error.

3.2 Testing G_t as a set

Suppose there are r latent and m observed factors. Whether r exceeds m or vice versa, it is useful to gauge the general coherency between F_t and G_t. To this end, we consider the canonical correlations between F_t and G_t. Let S_{FF} and S_{GG} be the sample variance-covariance matrix of F and G respectively. The sample squared-canonical correlations, denoted by $\widehat{\rho}_k^2, k = 1, \ldots \min[m, r]$, are the largest eigenvalues of the $r \times r$ matrix $S_{FF}^{-1} S_{GF} S_{GG}^{-1} S_{FG}$. It is well known that if F and G are observed and are normally distributed, $z_k = \frac{\sqrt{T}(\widehat{\rho}_k^2 - \rho_k^2)}{2\rho_k(1-\rho_k^2)} \xrightarrow{d} N(0, 1)$ for $k = 1, \ldots \min[m, r]$, see Anderson (1984). Muirhead and Waternaux (1980) provide results for non-normal distributions. For elliptical distributions, they showed that $z_k = \frac{1}{1+\kappa/3} \frac{\sqrt{T}(\widehat{\rho}_k^2 - \rho_k^2)}{2\rho_k(1-\rho_k^2)} \xrightarrow{d} N(0, 1)$, where κ is the excess kurtosis of the distribution.[2] Their results cover the multivariate normal, some contaminated normal (mixture normal), and the multivariate t, which are all elliptical distributions. Our analysis is complicated by the fact that F is not observed but estimated. Nevertheless, the following holds.

[2] A random vector Y is said to have an elliptical distribution if its density is of the form $c|\Omega^{-1/2}|g(y'\Omega^{-1}y)$ for some constant c, positive-definite matrix Ω, and nonnegative function g.

Proposition 3 *Let $\widetilde{\rho}_1^2 \ldots, \widetilde{\rho}_p^2$ be the largest $p = \min[m, r]$ sample squared canonical correlations between $\widetilde{F}$ and G, where $\widetilde{F}_t$ is the principal components estimate of F_t. Let $N, T \to \infty$ with $\sqrt{T}/N \to 0$.*

(i) Suppose that (F_t, G_t) are iid normally distributed,

$$\widetilde{z}_k = \frac{\sqrt{T}(\widetilde{\rho}_k^2 - \rho_k^2)}{2\widetilde{\rho}_k(1-\widetilde{\rho}_k^2)} \xrightarrow{d} N(0,1), \qquad k = 1, \ldots, \min[m, r]. \tag{9}$$

(ii) Suppose that (F_t, G_t) are iid and elliptically distributed. Then

$$\widetilde{z}_k = \frac{1}{(1+\kappa/3)} \frac{\sqrt{T}(\widetilde{\rho}_k^2 - \rho_k^2)}{2\widetilde{\rho}_k(1-\widetilde{\rho}_k^2)} \xrightarrow{d} N(0,1), \qquad k = 1, \ldots, \min[m, r]. \tag{10}$$

where κ is the excess kurtosis.

Proposition 3 establishes that $\widetilde{z}_k$ is asymptotically the same as z_k, so that having to estimate F has no effect on the sampling distribution of the canonical correlations. This allows us to construct $(1-\alpha)$ percent confidence intervals for the population canonical correlations as follows. For $k = 1, \ldots \min[m, r]$,

$$\left(\rho_k^-, \rho_k^+\right) = \left(\widetilde{\rho}_k^2 - \Phi_\alpha \frac{\widetilde{\rho}_i(1-\widetilde{\rho}_i^2)}{\sqrt{T}}, \qquad \widetilde{\rho}_k^2 + \Phi_\alpha \frac{\widetilde{\rho}_i(1-\widetilde{\rho}_i^2)}{\sqrt{T}}\right). \tag{11}$$

If every element of G_t is an exact factor, all the non-zero population canonical correlations should be unity. The confidence interval for the smallest non-zero canonical correlation is thus a bound for the weakest correlation between F_t and G_t.

The only non-zero canonical correlation between a single series, say, G_{jt} and $\widetilde{F}_t$ is $\widetilde{\rho}_1^2$. But this is simply the coefficient of determination from a projection of G_{jt} onto $\widetilde{F}_t$, and thus coincide with $R^2(j)$ as defined in (8). The formula in (11) can therefore be used to obtain a confidence interval for $R^2(j)$ also.

4 Simulations

We use simulations to assess the finite sample properties of the tests. Throughout, we assume $F_{kt} \sim N(0,1), k = 1, \ldots r$, and $e_{it} \sim N(0, \sigma_e^2(i))$, where e_{it} is uncorrelated with e_{jt} for $i \neq j, i, j = 1, \ldots N$. When $\sigma_e^2(i) = \sigma_e^2$ for all i, we have the case of homogeneous data. The factor loadings are standard normal, i.e. $\lambda_{ij} \sim N(0,1), j = 1, \ldots r, i = 1, \ldots N$. The data are generated as $x_{it} = \lambda_i' F_t + e_{it}$. In the experiments, we assume that there are $r = 2$ factors. The data are standardized to have mean zero and unit variance prior to estimation of the factors by the method of principal components.

The observed factors are generated as $G_{jt} = \delta_j' F_t + \varepsilon_{jt}$, where δ_j is a $r \times 1$ vector of weights, and $\varepsilon_{jt} \sim \sigma_\varepsilon(j) N(0, \text{var}(\delta_j' F_t))$. We test $m = 7$ observed variables parameterized as follows:

Parameters for G_{jt}

j	1	2	3	4	5	6	7
δ_{j1}	1	1	1	1	1	1	0
δ_{j2}	1	0	1	0	1	0	0
σ_ε	0	0	.2	.2	2	2	1

The first two factors, G_{1t} and G_{2t} are exact factors since $\sigma_\varepsilon = 0$. Factors three to six are linear combinations of the two latent factors but are contaminated by errors. The variance of this error is small relative to the variations of the factors for G_{3t} and G_{4t}, but is large for G_{5t} and G_{6t}. Finally, G_{7t} is an irrelevant factor as it is simply a $N(0, 1)$ random variable. Prior to testing, the G_ts are also standardized to have mean zero and unit variance. We conduct 1000 replications using Matlab 6.5.

We report results with α set to 0.025. Results for homoskedastic and heteroskedastic errors with $\text{var}(\widehat{G}_t)$ defined as in (5), (4) and (3) are given in Table 1a-c, respectively. According to theory, $A(j)$ should both be 2α if G_{jt} is a true factor, and unity if the factor is irrelevant. Furthermore, $M(j)$ should exceed the the critical value at percentage point α with probability 2α. Columns four and five report the properties of these tests averaged over 1000 replications. Indeed, for G_{1t} and G_{2t}, the rejection rates are close to the nominal size of 5%, even for small samples. For the irrelevant factor G_{7t}, the tests reject the null hypothesis with high probabilities showing the tests have power. The power of the $M(j)$ test is especially impressive. Even when heteroskedasticity in the errors has to be accounted for, the rejection rate is 100 percent, for all sample sizes considered. This means that even with $(N, T) = (50, 50)$, the test can very precisely determine if an observed variable is an exact factor. The $NS(j)$ test reinforces the conclusion that G_{1t} and G_{2t} are exact factors, and that G_{7t} has no coherence with the latent factors. Table 1 also reports the point estimate of $R^2(j)$. The two exact factors have estimates of $R^2(j)$ well above 0.95, while uninformative factors have $R^2(j)$ well below 0.05, with tight confidence intervals.[3]

Because we know the data generating process, we can assess whether or not the confidence intervals around ε_{jt} give correct inference. In the column labelled $CI_\varepsilon(j)$ in Table 1 , we report the probability that the true ε_{jt} lies inside the two-tailed 95% confidence interval

[3]Since $R^2(j)$ is bounded between zero and one, the estimated lower bound should be interpreted in this light.

defined by (6). Evidently, the coverage is excellent for ε_{1t}, ε_{2t}, and ε_{7t}. The result for ε_{7t} might seem surprising at first, but this is in fact showing that the measurement error can be precisely estimated even when F_t and G_{jt} are totally unrelated.

In theory, $\tau(j)$ and $M(j)$ should always reject the null hypothesis when G_3, G_4, G_5 and G_6 are being tested since none of these are exact factors. Table 1 shows that this is the case when N and T both exceed 100. For smaller N and/or T, the power of the tests depend on how large are the measurement errors. For G_{5t} and G_{6t}, which have a high noise-to-signal ratio of 4, $M(j)$ still rejects with probability one when N and T are small, while the $A(j)$ has a respectable rejection rate of 0.85. However, when the signal-to-noise ratio is only .04 as in G_{3t} and G_{4t}, both the $A(j)$ and the $M(j)$ under-reject the null hypothesis, and the problem is much more severe for $A(j)$.

Notice that when the noise-to-signal ratio is .04, $R^2(j)$ remains at around .95. This would be judged high in a typical regression analysis, and yet we would reject the null hypothesis that G_{jt} is an exact factor. It is for cases such as this that having a sense of how big is the measurement error is useful. In our experience, an $NS(j)$ above .5, and/or a $R^2(j)$ below 0.95 is symptomatic of non-negligible measurement errors. By these guides, G_{3t} and G_{4t} are strong proxies for the latent factors.

It is of interest to remark that whether the measurement error has large or small variance, the confidence intervals constructed according to (6) bracket the true error quite precisely. This is useful in empirical work since we can learn if the discrepancy between G_{jt} and the latent factors are systematic or occasional.

Table 2 reports results for testing four sets of observed factors using canonical correlations. These are formed from G_{1t} to G_{6t} as defined above, plus four $N(0,1)$ random variables unrelated to F_t, labelled G_{7t} to G_{10t}. The four sets of factors are defined as follows:

Set 1: G_{3t}, G_{4t}, and G_{7t} Set 2: G_{5t}, G_{6t}, and G_{7t}
Set 3: G_{1t}, G_{2t}, and G_{7t} Set 4: G_{7t}, G_{8t}, G_{9t}, and G_{10t}.

Because $\min[m,r]$ is 2 in each of the four case, the smaller two of the four canonical correlations should be zero, while the largest two should be unity if every G_t is an exact factor. We use (11) to construct confidence intervals for ρ_2^2, i.e. the smaller of the two non-zero canonical correlations. Table 2 shows that Set 3 has maximal correlation with $\widetilde{F}_t$ as should be the case since G_{1t} and G_{3t} are exact factors, and a weight of zero on G_{7t} would indeed maximize the correlation between G and $\widetilde{F}$. When Set 4 is being tested, zero is in the confidence interval as should be the case since this is a set of irrelevant factors. For Set 2 which has factors contaminated by large measurement errors, the test also correctly detects a very small canonical correlation. When measurement errors are small but non-zero, the sample correlations are non-zero but also not unity. The practitioner again has to take a stand on whether a set of factors is useful. The values of $\widehat{\rho}_2^2$ for Set 1 are around .9, below our cut-off point of .95. We would thus be more concerned with accepting Set 1 as a valid set than accepting its elements (i.e. G_{3t} and G_{4t}) as individually valid factors.

Finally, Figures 1 and 2 depict the performance of exact factor testing and measurement error testing. Data are generated with two factors. Figure 1 displays the confidence intervals for the factor processes along with true factor processes under exact factor testing. The confidence intervals are

$$([\widehat{G}_{jt} - 1.96N^{-1/2}var(\widehat{G}_{jt})^{1/2}, \, \widehat{G}_{jt} + 1.96N^{-1/2}var(\widehat{G}_{jt})^{1/2}]$$

for $t = 1, ..., T$ and $j = 1, 2$. The left panel is for the first factor with $N = 50, 100$, and the right panel is for the second factor with different Ns. The confidence intervals become narrower for larger N. Next, we do not assume exact factors are observable, instead, $G_{jt} = \delta_j' F_t + \varepsilon_{jt}$ for $j = 1, 2$ are observable. The measurement errors ε_{jt} are estimated for $t = 1, ..., T$ and $j = 1, 2$. Confidence intervals are constructed as, according to Proposition 2:

$$[\widehat{\varepsilon}_{jt} - 1.96T^{-1/2}\, s_{jt}, \, \widehat{\varepsilon}_{jt} + 1.96T^{-1/2}\, s_{jt}].$$

These confidence intervals are plotted in Figure 1. In simulations, the true error processes ε_{jt} are known, and they are also plotted in Figure 2. It is clear that the confidence intervals cover the true process. In contrast with exact factor testing, the confidence intervals do not become narrow as N increases. This is due to parameter uncertainty about δ, which is estimated with T observations. When T increases, the confidence band will be narrower.

5 Empirical Applications

In this section, we take our tests to the data. Factors estimated from portfolios, stock returns, and a large set of economic variables will be tested against various G_{jt}s. The base factors are the three (FF) Fama and French factors, denoted 'Market', 'SMB', and 'HML'.[4] For annual data, we also consider aggregate consumption growth (DC) and the 'CAY' variable suggested by Lettau and Ludvigson (2001). For monthly data, we include, in addition to the FF factors,

[4]Small minus big is the difference between the average return of three small portfolios and three big portfolios. High minus low is the average return on two value and two growth portfolios. See Fama and French (1993). $Market = R_m - R_f$ is value weighted return on all NYSE, AMEX, and NASDAQ minus the one month treasury bill rate from Ibbotson Associates.

variables considered in Chen et al. (1986). These are the first lag of annual consumption growth 'DC', inflation 'DP', the growth rate of industrial production 'DIP', a term premia 'TERM', and a risk premia 'RISK''.[5] In each case, we analyze the data from 1960-1996. We also split the sample at various points to look for changes in relations between the observed and the latent factors over the forty years. The data are standardized to be mean zero with unit variances prior to estimation by the method of principal components. The G_ts are likewise standardized prior to implementation of the tests. In view of the properties of the data, we only report results for heteroskedastic errors with $\widehat{\text{var}}(\widehat{G}_t)$ defined as in (4).

We determine the number of factors with $g(N,T)$ specified as:

$$g_1(N,T) = \log\left(\frac{NT}{N+T}\right)\frac{N+T}{NT}$$
$$g_2(N,T) = \frac{(N+T)}{NT}\log(\min[N,T]).$$

5.1 Portfolios

In this application, x_{it} are monthly or annual observations on 100 portfolios available from Kenneth French's web site.[6] These are the intersections of 10 portfolios formed on size (market equity) and 10 portfolios formed on the ratio of book to market equity. A total of 89 portfolios are continuously available for the full sample. Depending on the sample period, the PCP and ICP select between 4 and 6 factors. We set $r = 6$ in all subsequent tests. The results are reported in Table 3.

For annual data, the $A(j)$ test rejects the null hypothesis of exact factors in more than 5% of the sample. The critical value for the $M(j)$ test is 3.28. It cannot reject the null hypothesis that SMB is an exact factor at the 5% level, and HML and Market at around the 10% level. However, the evidence does not support CAY and aggregate consumption growth as exact factors, as $NS(j)$ and $R^2(j)$ indicate the presence of non-trivial measurement errors. The canonical correlations suggest only three well defined relations between F_t and G_t. The remaining relations are extremely weak. The canonical correlations between the three FF factors alone, and $\widetilde{F}_t$ are .970, .962, and .950, respectively. Compared with the results when

[5] The data are taken from citibase. 'DC' is the growth rate of PCEND, 'DIP' is the growth rate of IP, 'DP' is the growth rate of PUNEW. The risk premium,'RISK', is the BAA rated bond rate (FYBAAC) minus the 10 year government bond rate (FYGT10). The term premia, 'TERM' is 10 year government bond FYGT10 minus the three month treasury bill rate FYGM3.

[6] The url is `mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_100_port_sz.html`.

'CAY' and 'DC' are included, little is gained by including 'CAY' and 'DC'. This suggests that the FF factors underlie the three non-zero canonical correlations in the five variable set.

Results for testing monthly data are given in Table 4. Several features are noteworthy. First, the FF factors continue to be strong proxies for systematic risks, and there is some weak evidence that the SMB is an exact factor in the 1988-1996 sub-sample. Of the three FF factors, 'Market' has the highest $R^2(j)$ with the unobserved factors. Second, the macroeconomic variables have unstable relations with the unobserved factors over time, but the relations have become stronger since the 1980s. Inflation was a good proxy factor between 1960-82, but over the entire sample, RISK has the highest $R^2(j)$ and the lowest $NS(j)$. In contrast, the relation between the FF factors and the latent factors appears to be quite stable, displaying little variation in both $R^2(j)$ and $NS(j)$ over time. Third, three of the eight sample canonical correlations between $\widetilde{F}_t$ and G_t are practically zero, from which we can conclude that the eight observed factors considered cannot span the true factor space. Because the five non-zero canonical correlations are in fact far from unity, we can also conclude that measurement errors are significant enough that the eight observed variables cannot even span a five dimensional subspace of the true factors.

5.2 Stock Returns

We next apply our tests to monthly stock returns. Because portfolios are aggregated from stocks, individual stock returns should have larger idiosyncratic variances than portfolios. Thus, the common components in the individual returns can be expected to be smaller than those in the portfolios. It is thus of interest to see if good proxy variables can be found when the data have larger idiosyncratic noises.

Data for 190 firms are available from CRSP over the entire sample. The two PCP criteria select 6 and 5 factors in this panel of data, while the ICP always selects 4. We set r to 6 in the analysis. The results in Table 5 reveal that 'Market' continues to be a strong proxy factor with a low noise to signal ratio. However, SMB, HML, and especially the macroeconomic variables now appear to be poor proxies for the factors in the returns data. The best of the proxy macroeconomic variable – 'RISK'– still has a noise-to-signal ratio that is ten times larger than the 'Market' factor.

The above analysis indicates that the factors in annual portfolios are better approximated by observed variables than the factors in monthly portfolios, and finding proxies for the factors in the monthly portfolios is in turn a less challenging task than finding observed variables to proxy the factors in individual returns. This is because high frequency and/or disaggregated data are more likely to be contaminated by noise. Thus, even though more data are available at the high frequency and disaggregated levels, they are less reliable proxies for the systematic variations in the data. Inference using observed variables as proxies for the common factors could be inaccurate. Of all the variables considered, the most satisfactory proxy for the latent factors in *both* portfolios and individual stock returns appears to be the 'Market' factor as described in Fama and French (1993). Its signal to noise ratio is systematically high, and its coherence with the latent factors is robust across sample periods.

5.3 Macroeconomic Factors

A fundamental characteristic of business cycles is the co movement of a large number of economic series. Such a phenomenon can be rationalized by common factors being the driving force of economic fluctuations. This has been used as a justification to represent the unobserved state of the economy by variables thought to be dominated by variations arising from common sources. Industrial production, unemployment rate, and various interest rate spreads have been used for this purpose. However, there has been no formal analysis of how good the proxy variables are.

We estimate the latent factors from 150 monthly series considered in Stock and Watson (2002). This data consist of series from output, consumption, employment, investment, prices, wages, interest rate, and other financial series such as exchange rates. In addition to the five macroeconomic variables considered throughout, we also tested if unemployment rate (LHUR) is a common factor. For this application, we also consider a post Bretton Woods sub-sample. The results are reported in Table 6.

Given the noise in monthly data, rejecting the null hypothesis that the variables are exact factors is hardly surprising. Although industrial production is a widely used indicator of economic activity, it is systematically dominated by DC, DP, RISK, and UR. As in the results for asset returns, the relations between the macroeconomic variables and the factors also appear to be unstable. While inflation was the best proxy factor before 1983, the unemployment rate has done extremely well since. The non-zero canonical correlations also reveal instability in the relations, as they are higher in the later than the earlier sub-samples. One interpretation of this result is that idiosyncratic shocks were more important in the sixties and seventies, but common shocks have become more important in recent years as sources of economic fluctuations.

To the extent that the FF factors are good proxies for the factors in portfolios, one might wonder if the common shocks to economic activity are also the common shocks to portfolio returns. To shed some light on this issue, we also tested if the FF factors are related to the factors in macroeconomic data. As seen from Table 5, there is hardly any evidence for a relation between the FF factors and the panel of macroeconomic variables.

6 Conclusion

It is common practice in empirical work to proxy unobserved common factors by observed variables, yet hardly any formal procedures exist to assess if the observed variables equal, or are close to the factors. This paper exploits the fact that the space spanned by the common factors can be consistently estimated from large dimensional panels. We develop several tests that can serve as guides as to which variables are close to the factors. The tests have good properties in simulations. We estimate the common factors in portfolios, individual returns, as well as a large set of macroeconomic data. The Fama and French factors approximate the factors in portfolios and individual stock returns much better than any single macroeconomic variable. Although industrial production is a widely used indicator of economic activity, inflation, unemployment rate, and the risk premium have stronger coherence with the macroeconomic factors. Inflation was a good proxy of the factors in portfolios and macroeconomic data prior to the 1980s, but its importance has diminished since.

Proof of Proposition 2

Rewrite $G_t = \delta' F_t + \varepsilon_t$ as $G_t = \delta' H^{-1}\widetilde{F}_t + \varepsilon_t + \delta' H^{-1}(HF_t - \widetilde{F}_t)$, or

$$G_t = \gamma' \widetilde{F}_t + \varepsilon_t + \gamma'(HF_t - \widetilde{F}_t),$$

where $\gamma = H^{-1\prime}\delta$. In matrix notation,

$$G = \widetilde{F}\gamma + \varepsilon + (FH' - \widetilde{F})\gamma. \tag{12}$$

The least squares estimator of γ is

$$\widehat{\gamma} = (\widetilde{F}'\widetilde{F}/T)^{-1}(\widetilde{F}'G/T) = \frac{1}{T}\widetilde{F}'G.$$

Substituting G of (12) into $\widehat{\gamma}$, we have

$$\begin{aligned} \widehat{\gamma} &= \gamma + \frac{1}{T}\widetilde{F}'\varepsilon + \frac{1}{T}\widetilde{F}'(FH' - \widetilde{F})\gamma, \\ \sqrt{T}(\widehat{\gamma} - \gamma) &= \frac{1}{\sqrt{T}}\widetilde{F}'\varepsilon + \sqrt{T}\frac{1}{T}\widetilde{F}'(FH' - \widetilde{F})\gamma. \end{aligned}$$

From Lemma B.3 of Bai (2003), the last term is $\sqrt{T}\cdot O_p(\min[N,T]^{-1}) \to 0$ provided $\sqrt{T}/N \to 0$. Thus, $\sqrt{T}(\widehat{\gamma} - \gamma) = \frac{1}{\sqrt{T}}\widetilde{F}'\varepsilon + o_p(1)$. From

$$\frac{1}{\sqrt{T}}\widetilde{F}'\varepsilon = H\frac{1}{\sqrt{T}}F'\varepsilon + \frac{1}{\sqrt{T}}(\widetilde{F} - FH')'\varepsilon = H\frac{1}{\sqrt{T}}F'\varepsilon + o_p(1)$$

where we use the fact that $\frac{1}{\sqrt{T}}(\widetilde{F} - FH')'\varepsilon = o_p(1)$ if $\sqrt{T}/N \to 0$. By the CLT, $T^{-1/2}F'\varepsilon = T^{-1/2}\sum_{t=1}^{T} F_t\varepsilon_t$ is asymptotically normal. The asymptotic variance is the probability limit of $\sigma_\varepsilon^2 \frac{F'F}{T}$. So the asymptotic variance of $HT^{-1/2}F'\varepsilon$ is the limit of $H\frac{F'F}{T}H'$. But $H\frac{F'F}{T}H' = I + o_p(1)$, see Bai (2003). Thus $\sqrt{T}(\widehat{\gamma} - \gamma)\xrightarrow{d} N(0, \sigma_\varepsilon^2 I)$.

Now $G_t = \widehat{\gamma}'\widetilde{F}_t + \widehat{\varepsilon}_t$, and we also have $G_t = \gamma'\widetilde{F}_t + \varepsilon_t + \gamma'(HF_t - \widetilde{F}_t)$. Equating, we have

$$\widehat{\gamma}'\widetilde{F}_t + \widehat{\varepsilon}_t = \gamma'\widetilde{F}_t + \varepsilon_t + \gamma'(HF_t - \widetilde{F}_t),$$

which implies

$$\widehat{\varepsilon}_t - \varepsilon_t = (\gamma' - \widehat{\gamma}')\widetilde{F}_t + \gamma'(HF_t - \widetilde{F}_t).$$

Thus, $\sqrt{T}(\widehat{\varepsilon}_t - \varepsilon_t) = -\widetilde{F}_t'\sqrt{T}(\widehat{\gamma} - \gamma) - \gamma'\sqrt{T/N}\sqrt{N}(\widetilde{F}_t - HF_t)$. But $\sqrt{T}(\widehat{\gamma} - \gamma)\xrightarrow{d} N(0, \sigma_\varepsilon^2 I)$ and $\gamma'\sqrt{N}(HF_t - \widetilde{F}_t)\xrightarrow{d} N(0, \gamma' V^{-1}Q\Gamma_t Q' V^{-1}\gamma)$, and $\widetilde{F}'\widetilde{F}_t = F_t'H'HF_t + o_p(1) = F_t'(F'F/T)^{-1}F_t + o_p(1)$, the stated result follows.

Proof of Proposition 3

Let $\widetilde{B} = S^{-1}_{\widetilde{F}\widetilde{F}} S_{\widetilde{F}G} S^{-1}_{GG} S_{G\widetilde{F}}$, and $B = S^{-1}_{FF} S_{FG} S^{-1}_{GG} S_{GF}$. Let $\widetilde{\rho}_k$ and $\widehat{\rho}_k$ be the canonical correlations of $\widetilde{B}$ and B, respectively. Because eigenvalues are continuous functions, we will have $\sqrt{T}(\widetilde{\rho}^2_k - \widehat{\rho}^2_k) \stackrel{p}{\longrightarrow} 0$ provided that $\sqrt{T}(\widetilde{B} - B) \stackrel{p}{\longrightarrow} 0$. That is, the canonical correlations of $\widetilde{B}$ have the same limiting distributions as those of B when $\widetilde{B}$ and B are asymptotically equivalent. We next establish $\sqrt{T}(\widetilde{B} - B) \stackrel{p}{\longrightarrow} 0$. First note that because H is full rank, the canonical correlations of B is the same as the canonical correlations of B^*, where

$$B^* = S^{-1}_{HF\,HF} S_{HF\,G} S^{-1}_{GG} S_{G\,HF}.$$

Thus, it suffices to show that $\sqrt{T}\widetilde{B} - B^*) \stackrel{p}{\longrightarrow} 0$. But this is implied by

$$\sqrt{T}(S^{-1}_{\widetilde{F}\widetilde{F}} - S^{-1}_{HF\,HF}) \stackrel{p}{\longrightarrow} 0 \quad (13)$$

$$\sqrt{T}(S_{\widetilde{F}G} - S_{HF\,G}) \stackrel{p}{\longrightarrow} 0. \quad (14)$$

Consider (13). Now

$$(S^{-1}_{\widetilde{F}\widetilde{F}} - S^{-1} S_{HF\,HF}) = S^{-1}_{\widetilde{F}\widetilde{F}} (S_{HF\,HF} - S_{\widetilde{F}\widetilde{F}}) S^{-1}_{HF\,HF}.$$

Thus, (13) is implied by $\sqrt{T}(S_{HF\,HF} - S_{\widetilde{F}\widetilde{F}}) \stackrel{p}{\longrightarrow} 0$, or that

$$\sqrt{T}\left(\frac{HF'FH}{T} - \frac{\widetilde{F}'\widetilde{F}}{T}\right) \stackrel{p}{\longrightarrow} 0. \quad (15)$$

By Lemma B.2 and B.3 of Bai (2003),

$$\begin{aligned} \frac{\widetilde{F}'(\widetilde{F} - FH')}{T} &= O_p(\min[N,T]^{-1}) \\ \frac{F'(\widetilde{F} - FH')}{T} &= O_p(\min[N,T]^{-1}). \end{aligned}$$

Adding and subtracting terms, (15) becomes

$$\frac{-\sqrt{T}(\widetilde{F} - FH')'\widetilde{F}}{T} - \frac{-\sqrt{T}HF'(\widetilde{F} - FH')}{T} = \sqrt{T} O_p(\min[N,T]^{-1}) \to 0$$

if $\sqrt{T}/N \to 0$. For (14),

$$\sqrt{T}\left(\frac{\widetilde{F}'G}{T} - \frac{HF'G}{T}\right) = \sqrt{T}\left(\frac{(\widetilde{F} - FH')'G}{T}\right).$$

But $\frac{1}{T}(\widetilde{F} - FH')'G = O_p(\min[N,T]^{-1})$, see Lemma B.2 of Bai (2003). Thus, (14) is $O_p(\sqrt{T}/\min[N,T]) \stackrel{p}{\longrightarrow} 0$, establishing $\sqrt{T}(\widetilde{B} - B^*) = o_p(1)$ and thus the proposition.

Table 1a: Tests for G_{jt}: Homoskedasticity Variance

N	T	j	$A(j)$	$M(j)$	$NS(j)$	$CI_{\varepsilon}(j)$	$R2(j)$	$R2^{-}(j)$	$R2^{+}(j)$
50	50	1	0.03	0.02	0.03	0.97	0.97	0.96	0.98
50	50	2	0.03	0.03	0.03	0.97	0.97	0.96	0.98
50	50	3	0.17	0.64	0.07	0.97	0.94	0.92	0.95
50	50	4	0.16	0.60	0.07	0.97	0.94	0.92	0.95
50	50	5	0.85	1.00	5.05	0.95	0.22	0.13	0.32
50	50	6	0.85	1.00	5.08	0.95	0.22	0.12	0.32
50	50	7	0.95	1.00	339.24	0.96	0.04	-0.01	0.09
100	50	1	0.03	0.01	0.01	0.97	0.99	0.98	0.99
100	50	2	0.03	0.01	0.01	0.97	0.99	0.98	0.99
100	50	3	0.27	0.94	0.05	0.97	0.95	0.94	0.96
100	50	4	0.27	0.94	0.05	0.97	0.95	0.94	0.96
100	50	5	0.89	1.00	4.79	0.95	0.23	0.13	0.33
100	50	6	0.89	1.00	4.77	0.95	0.22	0.13	0.32
100	50	7	0.96	1.00	290.41	0.94	0.04	-0.00	0.09
50	100	1	0.03	0.01	0.03	0.97	0.97	0.97	0.98
50	100	2	0.03	0.01	0.03	0.97	0.97	0.97	0.98
50	100	3	0.17	0.74	0.07	0.97	0.94	0.92	0.95
50	100	4	0.17	0.73	0.07	0.97	0.94	0.92	0.95
50	100	5	0.85	1.00	4.52	0.96	0.21	0.14	0.28
50	100	6	0.86	1.00	4.49	0.96	0.21	0.14	0.28
50	100	7	0.96	1.00	1780.38	0.97	0.02	-0.00	0.04
200	100	1	0.03	0.01	0.01	0.97	0.99	0.99	0.99
200	100	2	0.03	0.01	0.01	0.97	0.99	0.99	0.99
200	100	3	0.40	1.00	0.05	0.97	0.96	0.95	0.96
200	100	4	0.40	1.00	0.05	0.97	0.96	0.95	0.96
200	100	5	0.93	1.00	4.32	0.95	0.21	0.14	0.28
200	100	6	0.92	1.00	4.27	0.95	0.21	0.14	0.28
200	100	7	0.98	1.00	431.39	0.96	0.02	-0.00	0.04
100	200	1	0.03	0.01	0.01	0.97	0.99	0.98	0.99
100	200	2	0.03	0.00	0.01	0.97	0.99	0.98	0.99
100	200	3	0.27	1.00	0.05	0.97	0.95	0.94	0.96
100	200	4	0.27	1.00	0.05	0.97	0.95	0.94	0.96
100	200	5	0.90	1.00	4.23	0.96	0.20	0.15	0.25
100	200	6	0.90	1.00	4.15	0.96	0.21	0.16	0.25
100	200	7	0.98	1.00	946.21	0.96	0.01	-0.00	0.02

The 5% critical values for the $M(j)$ test when $T = 50, 100, 200, 400$ are, 3.283, 3.474, 3.656, and 3.830, respectively. The $A(j)$ and the confidence intervals are based on the critival value 1.96. The tests are based on $\widehat{\text{var}(\widehat{G}_t)}$ defined in (5).

Table 1b: Tests for G_{jt}: Heteroskedastic Variance

N	T	j	$A(j)$	$M(j)$	$NS(j)$	$CI_\varepsilon(j)$	$R2(j)$	$R2^-(j)$	$R2^+(j)$
50	50	1	0.05	0.07	0.01	0.95	0.99	0.99	0.99
50	50	2	0.05	0.07	0.01	0.95	0.99	0.99	0.99
50	50	3	0.41	1.00	0.05	0.95	0.96	0.94	0.97
50	50	4	0.42	1.00	0.05	0.95	0.95	0.94	0.97
50	50	5	0.92	1.00	4.92	0.94	0.23	0.13	0.32
50	50	6	0.92	1.00	4.89	0.94	0.23	0.13	0.32
50	50	7	0.97	1.00	193.91	0.94	0.04	-0.01	0.09
100	50	1	0.05	0.05	0.00	0.95	1.00	0.99	1.00
100	50	2	0.05	0.05	0.00	0.96	1.00	0.99	1.00
100	50	3	0.55	1.00	0.04	0.95	0.96	0.95	0.97
100	50	4	0.55	1.00	0.04	0.95	0.96	0.95	0.97
100	50	5	0.95	1.00	4.75	0.94	0.23	0.13	0.33
100	50	6	0.95	1.00	4.69	0.95	0.23	0.13	0.32
100	50	7	0.98	1.00	177.17	0.93	0.04	-0.00	0.09
50	100	1	0.05	0.07	0.01	0.95	0.99	0.99	0.99
50	100	2	0.05	0.06	0.01	0.95	0.99	0.99	0.99
50	100	3	0.42	1.00	0.05	0.95	0.95	0.95	0.96
50	100	4	0.42	1.00	0.05	0.95	0.95	0.95	0.96
50	100	5	0.93	1.00	4.41	0.95	0.21	0.14	0.28
50	100	6	0.93	1.00	4.39	0.95	0.21	0.14	0.28
50	100	7	0.98	1.00	408.56	0.95	0.02	-0.00	0.04
200	100	1	0.05	0.05	0.00	0.95	1.00	1.00	1.00
200	100	2	0.05	0.04	0.00	0.95	1.00	1.00	1.00
200	100	3	0.67	1.00	0.04	0.95	0.96	0.95	0.97
200	100	4	0.67	1.00	0.04	0.95	0.96	0.95	0.97
200	100	5	0.96	1.00	4.29	0.95	0.21	0.14	0.28
200	100	6	0.96	1.00	4.24	0.94	0.22	0.15	0.28
200	100	7	0.99	1.00	465.28	0.95	0.02	-0.00	0.04
100	200	1	0.05	0.07	0.00	0.95	1.00	1.00	1.00
100	200	2	0.05	0.07	0.00	0.95	1.00	1.00	1.00
100	200	3	0.55	1.00	0.04	0.95	0.96	0.95	0.96
100	200	4	0.56	1.00	0.04	0.95	0.96	0.95	0.96
100	200	5	0.95	1.00	4.17	0.95	0.21	0.16	0.25
100	200	6	0.95	1.00	4.09	0.95	0.21	0.16	0.26
100	200	7	0.99	1.00	751.77	0.95	0.01	-0.00	0.02

The 5% critical values for the $M(j)$ test when $T = 50, 100, 200, 400$ are, 3.283, 3.474, 3.656, and 3.830, respectively. The $A(j)$ and the confidence intervals are based on the critical value 1.96.

Table 1c: Tests for G_{jt}: Cross Correlated Errors

N	T	j	$A(j)$	$M(j)$	$NS(j)$	$CI_{\varepsilon}(j)$	$R2(j)$	$R2^{-}(j)$	$R2^{+}(j)$
50	50	1	0.01	0.02	0.03	0.99	0.97	0.96	0.98
50	50	2	0.01	0.02	0.03	0.99	0.97	0.96	0.98
50	50	3	0.08	0.25	0.07	0.99	0.94	0.92	0.95
50	50	4	0.08	0.25	0.07	0.99	0.94	0.92	0.95
50	50	5	0.66	1.00	5.05	0.98	0.22	0.13	0.32
50	50	6	0.66	1.00	5.08	0.98	0.22	0.12	0.32
50	50	7	0.72	1.00	339.24	0.98	0.04	-0.01	0.09
100	50	1	0.04	0.10	0.01	0.96	0.99	0.98	0.99
100	50	2	0.04	0.09	0.01	0.96	0.99	0.98	0.99
100	50	3	0.22	0.63	0.05	0.96	0.95	0.94	0.96
100	50	4	0.23	0.63	0.05	0.97	0.95	0.94	0.96
100	50	5	0.77	1.00	4.79	0.98	0.23	0.13	0.33
100	50	6	0.77	1.00	4.77	0.98	0.22	0.13	0.32
100	50	7	0.80	1.00	290.41	0.97	0.04	-0.00	0.09
50	100	1	0.00	0.00	0.03	1.00	0.97	0.97	0.98
50	100	2	0.00	0.00	0.03	1.00	0.97	0.97	0.98
50	100	3	0.04	0.05	0.07	1.00	0.94	0.92	0.95
50	100	4	0.04	0.05	0.07	1.00	0.94	0.92	0.95
50	100	5	0.64	1.00	4.52	0.99	0.21	0.14	0.28
50	100	6	0.64	1.00	4.49	0.99	0.21	0.14	0.28
50	100	7	0.72	1.00	1780.38	0.99	0.02	-0.00	0.04
200	100	1	0.04	0.10	0.01	0.96	0.99	0.99	0.99
200	100	2	0.04	0.10	0.01	0.96	0.99	0.99	0.99
200	100	3	0.34	0.89	0.05	0.96	0.96	0.95	0.96
200	100	4	0.35	0.89	0.05	0.96	0.96	0.95	0.96
200	100	5	0.84	1.00	4.32	0.98	0.21	0.14	0.28
200	100	6	0.83	1.00	4.27	0.97	0.21	0.14	0.28
200	100	7	0.86	1.00	431.39	0.98	0.02	-0.00	0.04
100	200	1	0.00	0.00	0.01	1.00	0.99	0.98	0.99
100	200	2	0.00	0.00	0.01	1.00	0.99	0.98	0.99
100	200	3	0.09	0.35	0.05	1.00	0.95	0.94	0.96
100	200	4	0.09	0.33	0.05	1.00	0.95	0.94	0.96
100	200	5	0.75	1.00	4.23	0.99	0.20	0.15	0.25
100	200	6	0.74	1.00	4.15	1.00	0.21	0.16	0.25
100	200	7	0.81	1.00	946.21	0.99	0.01	-0.00	0.02

The 5% critical values for the $M(j)$ test when $T = 50, 100, 200, 400$ are, 3.283, 3.474, 3.656, and 3.830, respectively. The $A(j)$ and the confidence intervals are based on the critical value 1.96. The tests are based on $\widehat{\text{var}(\widehat{G}_t)}$ defined in (5).

Table 2: Testing G_t Jointly:

	T	Set	$\widehat{\rho}_s^2$	$\widehat{\rho}_s^2(-)$	$\widehat{\rho}_s^2(+)$
50	50	1	0.85	0.82	0.89
50	50	2	0.09	0.02	0.16
50	50	3	0.96	0.95	0.97
50	50	4	0.03	-0.01	0.08
100	50	1	0.87	0.84	0.90
100	50	2	0.10	0.03	0.17
100	50	3	0.98	0.98	0.99
100	50	4	0.04	-0.01	0.08
50	100	1	0.85	0.83	0.88
50	100	2	0.08	0.03	0.13
50	100	3	0.97	0.96	0.97
50	100	4	0.02	-0.01	0.04
200	100	1	0.88	0.85	0.90
200	100	2	0.08	0.03	0.13
200	100	3	0.99	0.99	0.99
200	100	4	0.02	-0.01	0.04
100	200	1	0.87	0.85	0.89
100	200	2	0.07	0.04	0.11
100	200	3	0.98	0.98	0.99
100	200	4	0.01	-0.00	0.02

ρ_s^2 is the smallest non-zero canonical correlation between $\widetilde{F}$ and G.

Table 3: Testing the Factors in 100 FF Portfolios: Annual Data

Sample	j	$A(j)$	$M_\tau(j)$	$R^2(j)$	$NS(j)$	$\widehat{\rho}(k)^2$
60-96	Market	0.270	4.343	0.984 (0.979, 0.989)	0.016	0.995(0.993, 0.997)
T=37	SMB	0.162	2.865	0.971 (0.962, 0.980)	0.030	0.970(0.961, 0.980)
N=94	HML	0.081	2.479	0.962 (0.950, 0.974)	0.039	0.933(0.912, 0.954)
	DC	0.757	43.309	0.098 (0.007, 0.188)	9.252	0.034(-0.023, 0.092)
	CAY	0.892	49.005	0.088 (0.001, 0.175)	10.392	0.015(-0.024, 0.054)

The $A(j)$ test is the frequency that $|\widehat{\tau}_t(j)|$ exceeds the critical value of 1.96 in the sample of size T, where the heteroskedasticity robust variance-covariance matrix (4) is used. The $M(j)$ test is the frequency $\max_{1\le t\le T}|\widehat{\tau}_t(j)|$ exceeds the critical value for a sample of size T. The 5% critical values when $T = 50, 100, 200, 400$ are, 3.283, 3.474, 3.656, and 3.830, respectively. R^2, defined in (8), is the ratio of the fraction of the variance of G_t that is explained by $\widetilde{F}_t$. $NS(j)$, defined in (7), is the noise-to-signal ratio. $\widehat{\rho}(k)^2$ is the vector of squared canonical correlations of G_t with respect to $\widetilde{F}_t$.

Table 4: Testing the Factors in 100 FF Portfolios: Monthly Data

Sample	j	$A(j)$	$M(j)$	$R^2(j)$	$NS(j)$	$\widehat{\rho}(k)^2$
60-96	Market	0.288	9.847	0.973 (0.971, 0.976)	0.028	0.992(0.991, 0.993)
T=444	SMB	0.259	7.132	0.926 (0.920, 0.933)	0.079	0.918(0.911, 0.926)
N=95	HML	0.182	5.351	0.886 (0.876, 0.896)	0.128	0.832(0.818, 0.846)
	DC	0.948	130.976	0.010 (0.001, 0.020)	95.525	0.342(0.306, 0.378)
	DIP	0.917	51.805	0.059 (0.038, 0.080)	15.938	0.024(0.010, 0.038)
	DP	0.955	153.342	0.031 (0.015, 0.047)	31.463	0.006(-0.001, 0.014)
	TERM	0.901	69.627	0.171 (0.139, 0.202)	4.865	0.000(-0.000, 0.000)
	RISK	0.840	74.642	0.277 (0.241, 0.312)	2.613	0.000(-0.000, 0.000)
60-82	Market	0.380	9.041	0.965 (0.961, 0.969)	0.036	0.993(0.992, 0.994)
T=276	SMB	0.341	6.488	0.913 (0.903, 0.923)	0.095	0.907(0.896, 0.917)
N=95	HML	0.178	4.556	0.887 (0.874, 0.899)	0.128	0.843(0.826, 0.860)
	DC	0.717	33.495	0.414 (0.369, 0.458)	1.417	0.792(0.770, 0.814)
	DIP	0.928	38.171	0.107 (0.073, 0.142)	8.313	0.043(0.020, 0.067)
	DP	0.569	13.704	0.729 (0.702, 0.757)	0.371	0.007(-0.003, 0.017)
	TERM	0.920	79.347	0.026 (0.007, 0.044)	37.627	0.000(-0.000, 0.000)
	RISK	0.779	34.169	0.411 (0.366, 0.455)	1.435	-0.000(-0.000,-0.000)
83-96	Market	0.238	5.098	0.982 (0.979, 0.985)	0.018	0.993(0.992, 0.994)
T=168	SMB	0.208	4.034	0.940 (0.932, 0.949)	0.063	0.921(0.909, 0.932)
N=101	HML	0.190	4.340	0.917 (0.905, 0.929)	0.090	0.897(0.882, 0.912)
	DC	0.839	37.060	0.132 (0.084, 0.180)	6.576	0.523(0.471, 0.575)
	DIP	0.744	23.411	0.159 (0.108, 0.210)	5.283	0.127(0.080, 0.174)
	DP	0.881	47.689	0.071 (0.034, 0.109)	13.033	0.013(-0.004, 0.030)
	TERM	0.893	31.936	0.104 (0.060, 0.148)	8.594	0.000(-0.000, 0.000)
	RISK	0.685	24.456	0.338 (0.280, 0.397)	1.955	-0.000(-0.000,-0.000)
73-87	Market	0.383	9.075	0.970 (0.966, 0.975)	0.030	0.992(0.991, 0.993)
T=180	SMB	0.428	6.840	0.903 (0.889, 0.916)	0.108	0.923(0.912, 0.934)
N=98	HML	0.200	4.287	0.912 (0.900, 0.924)	0.096	0.794(0.768, 0.821)
	DC	0.733	22.633	0.488 (0.436, 0.540)	1.048	0.506(0.455, 0.558)
	DIP	0.944	88.219	0.058 (0.025, 0.091)	16.164	0.030(0.006, 0.055)
	DP	0.806	33.120	0.310 (0.254, 0.366)	2.226	0.002(-0.004, 0.008)
	TERM	0.900	40.481	0.168 (0.118, 0.218)	4.949	0.000(-0.000, 0.000)
	RISK	0.900	44.992	0.158 (0.109, 0.207)	5.339	-0.000(-0.000,-0.000)
88-96	Market	0.176	4.581	0.983 (0.980, 0.986)	0.017	0.991(0.990, 0.993)
T=108	SMB	0.046	3.568	0.962 (0.955, 0.969)	0.039	0.957(0.949, 0.965)
N=101	HML	0.148	4.075	0.909 (0.893, 0.926)	0.100	0.908(0.891, 0.924)
	DC	0.806	23.760	0.169 (0.105, 0.234)	4.908	0.475(0.407, 0.543)
	DIP	0.676	13.737	0.233 (0.163, 0.303)	3.289	0.113(0.057, 0.170)
	DP	0.546	9.324	0.434 (0.363, 0.504)	1.307	0.020(-0.006, 0.046)
	TERM	0.898	23.007	0.133 (0.074, 0.193)	6.510	-0.000(-0.000, 0.000)
	RISK	0.769	23.595	0.282 (0.210, 0.354)	2.548	-0.000(-0.000, 0.000)

Table 5: Testing the Factors in Monthly CRSP Returns

Sample	j	$A(j)$	$M(j)$	$R^2(j)$	$NS(j)$	$\widehat{\rho}(k)^2$
60-96	Market	0.255	6.685	0.967 (0.964, 0.970)	0.034	0.972(0.970, 0.975)
T=444	SMB	0.468	9.677	0.603 (0.574, 0.631)	0.660	0.588(0.559, 0.618)
N=190	HML	0.628	14.925	0.436 (0.401, 0.471)	1.294	0.363(0.327, 0.399)
	DC	0.917	46.358	0.048 (0.029, 0.068)	19.755	0.066(0.044, 0.088)
	DIP	0.876	46.788	0.052 (0.032, 0.072)	18.346	0.025(0.011, 0.039)
	DP	0.908	72.374	0.020 (0.007, 0.033)	49.119	0.010(0.001, 0.020)
	TERM	0.941	111.105	0.015 (0.004, 0.027)	63.647	0.000(-0.000, 0.000)
	RISK	0.876	39.177	0.056 (0.035, 0.076)	17.018	0.000(-0.000, 0.000)
60-82	Market	0.264	6.299	0.969 (0.965, 0.972)	0.032	0.975(0.972, 0.978)
T=276	SMB	0.471	7.261	0.664 (0.632, 0.697)	0.505	0.639(0.604, 0.673)
N=190	HML	0.562	11.003	0.525 (0.485, 0.566)	0.904	0.453(0.410, 0.497)
	DC	0.866	34.175	0.094 (0.061, 0.127)	9.613	0.122(0.086, 0.158)
	DIP	0.953	43.521	0.066 (0.037, 0.094)	14.219	0.029(0.009, 0.048)
	DP	0.920	56.533	0.026 (0.008, 0.045)	37.305	0.008(-0.003, 0.018)
	TERM	0.964	112.982	0.020 (0.004, 0.037)	48.203	0.000(-0.000, 0.000)
	RISK	0.957	44.627	0.048 (0.023, 0.073)	19.774	-0.000(-0.000,-0.000)
83-96	Market	0.286	4.747	0.967 (0.962, 0.972)	0.034	0.971(0.967, 0.976)
T=168	SMB	0.530	10.600	0.556 (0.506, 0.606)	0.799	0.578(0.530, 0.627)
N=190	HML	0.619	17.831	0.464 (0.409, 0.519)	1.155	0.382(0.324, 0.439)
	DC	0.935	75.460	0.022 (-0.000, 0.043)	45.479	0.055(0.021, 0.088)
	DIP	0.958	160.489	0.013 (-0.004, 0.029)	78.175	0.019(-0.002, 0.039)
	DP	0.958	66.233	0.020 (-0.001, 0.041)	48.263	0.000(-0.003, 0.003)
	TERM	0.857	25.031	0.073 (0.035, 0.111)	12.623	-0.000(-0.000,-0.000)
	RISK	0.810	27.875	0.096 (0.054, 0.139)	9.373	-0.000(-0.000,-0.000)
73-87	Market	0.256	6.538	0.973 (0.969, 0.977)	0.028	0.976(0.972, 0.979)
T=180	SMB	0.467	9.064	0.662 (0.622, 0.702)	0.510	0.662(0.622, 0.702)
N=190	HML	0.656	16.801	0.553 (0.504, 0.602)	0.808	0.419(0.364, 0.474)
	DC	0.906	43.853	0.093 (0.053, 0.134)	9.710	0.153(0.105, 0.202)
	DIP	0.928	60.124	0.066 (0.031, 0.102)	14.042	0.048(0.017, 0.078)
	DP	0.911	37.918	0.049 (0.018, 0.080)	19.349	0.017(-0.002, 0.035)
	TERM	0.944	106.380	0.025 (0.003, 0.048)	38.253	0.000(-0.000, 0.000)
	RISK	0.900	42.361	0.093 (0.052, 0.133)	9.768	-0.000(-0.000,-0.000)
88-96	Market	0.250	4.272	0.963 (0.957, 0.970)	0.038	0.969(0.963, 0.975)
T=108	SMB	0.583	10.659	0.551 (0.488, 0.614)	0.815	0.626(0.570, 0.682)
N=190	HML	0.648	18.149	0.437 (0.366, 0.507)	1.290	0.365(0.292, 0.437)
	DC	0.917	53.033	0.032 (-0.000, 0.065)	29.829	0.062(0.018, 0.107)
	DIP	0.889	44.873	0.040 (0.004, 0.077)	23.791	0.020(-0.006, 0.046)
	DP	0.991	61.828	0.039 (0.003, 0.074)	24.910	0.005(-0.008, 0.018)
	TERM	0.889	23.788	0.086 (0.036, 0.137)	10.587	0.000(-0.000, 0.000)
	RISK	0.750	19.579	0.208 (0.140, 0.276)	3.814	0.000(-0.000, 0.000)

Table 6: Testing the Factors in Macroeconomic Data

Sample	j	$A(j)$	$M(j)$	$R^2(j)$	$NS(j)$	$\widehat{\rho}(k)^2$
60-96	Market	0.941	364.555	0.008 (-0.000, 0.016)	124.359	0.862(0.850, 0.874)
T=444	SMB	0.917	380.642	0.019 (0.007, 0.032)	50.978	0.677(0.652, 0.701)
N=150	HML	0.926	466.551	0.010 (0.001, 0.019)	97.817	0.292(0.257, 0.328)
	DC	0.635	12.028	0.610 (0.582, 0.638)	0.640	0.164(0.133, 0.196)
	DIP	0.563	15.308	0.585 (0.556, 0.615)	0.708	0.030(0.014, 0.045)
	DP	0.491	10.775	0.744 (0.724, 0.765)	0.344	0.004(-0.002, 0.010)
	TERM	0.782	25.664	0.426 (0.391, 0.461)	1.347	0.000(-0.000, 0.000)
	RISK	0.709	16.283	0.636 (0.609, 0.663)	0.573	-0.000(-0.000,-0.000)
	UR	0.637	22.638	0.624 (0.597, 0.652)	0.602	-0.000(-0.000,-0.000)
60-82	Market	0.924	337.746	0.011 (-0.001, 0.023)	91.476	0.906(0.896, 0.917)
T=276	SMB	0.906	230.985	0.023 (0.006, 0.041)	41.760	0.766(0.742, 0.790)
N=150	HML	0.938	456.227	0.011 (-0.001, 0.023)	92.165	0.300(0.255, 0.345)
	DC	0.692	13.930	0.671 (0.640, 0.703)	0.490	0.055(0.029, 0.081)
	DIP	0.667	19.124	0.527 (0.486, 0.567)	0.898	0.030(0.010, 0.049)
	DP	0.736	19.880	0.830 (0.812, 0.848)	0.205	0.007(-0.003, 0.017)
	TERM	0.710	26.791	0.398 (0.353, 0.443)	1.511	0.000(-0.000, 0.000)
	RISK	0.732	18.976	0.671 (0.639, 0.703)	0.491	0.000(-0.000, 0.000)
	UR	0.612	31.987	0.751 (0.726, 0.777)	0.331	0.000(-0.000, 0.000)
83-96	Market	0.917	78.037	0.075 (0.037, 0.114)	12.277	0.950(0.943, 0.957)
T=168	SMB	0.839	40.687	0.073 (0.035, 0.111)	12.683	0.729(0.694, 0.764)
N=150	HML	0.905	136.574	0.036 (0.008, 0.063)	27.151	0.512(0.459, 0.564)
	DC	0.738	21.232	0.443 (0.387, 0.499)	1.257	0.090(0.048, 0.131)
	DIP	0.571	13.877	0.586 (0.538, 0.634)	0.706	0.059(0.025, 0.094)
	DP	0.720	24.163	0.356 (0.297, 0.414)	1.812	0.027(0.003, 0.051)
	TERM	0.815	20.764	0.370 (0.312, 0.428)	1.702	0.000(-0.000, 0.000)
	RISK	0.839	21.151	0.461 (0.405, 0.516)	1.170	0.000(-0.000, 0.000)
	UR	0.327	4.632	0.936 (0.926, 0.945)	0.069	0.000(-0.000, 0.000)
73-96	Market	0.924	90.926	0.037 (0.015, 0.058)	26.279	0.902(0.891, 0.912)
T=288	SMB	0.903	162.320	0.065 (0.037, 0.092)	14.472	0.751(0.727, 0.776)
N=150	HML	0.944	311.158	0.024 (0.006, 0.041)	41.259	0.542(0.503, 0.581)
	DC	0.722	17.252	0.720 (0.693, 0.748)	0.388	0.157(0.119, 0.196)
	DIP	0.569	19.118	0.586 (0.550, 0.623)	0.705	0.045(0.021, 0.068)
	DP	0.646	17.733	0.715 (0.688, 0.743)	0.398	0.006(-0.003, 0.016)
	TERM	0.750	26.238	0.472 (0.430, 0.514)	1.119	-0.000(-0.000, 0.000)
	RISK	0.698	13.381	0.630 (0.596, 0.663)	0.589	-0.000(-0.000, 0.000)
	UR	0.378	7.895	0.835 (0.817, 0.852)	0.198	0.000(-0.000, 0.000)

References

Anderson, T. W. (1984), *An Introduction to Multivariate Statistical Analysis*, Wiley, New York.

Bai, J. (2003), Inferential Theory for Factor Models of Large Dimensions, *Econometrica* **71:1**, 135–172.

Bai, J. and Ng, S. (2002), Determining the Number of Factors in Approximate Factor Models, *Econometrica* **70:1**, 191–221.

Bai, J. and Ng, S. (2003), Confidence Intervals for Factor Forecasts with Many Predictors. Unpublished manuscript.

Breeden, D. (1989), An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities, *Journal of Financial Economics* **7**, 265–296.

Chamberlain, G. and Rothschild, M. (1983), Arbitrage, Factor Structure and Mean-Variance Analysis in Large Asset Markets, *Econometrica* **51**, 1305–1324.

Chen, N., Roll, R. and Ross, S. (1986), Economic Forces and the Stock Market, *Journal of Business* **59**, 383–403.

Connor, G. and Korajzcyk, R. (1998), Risk and Return in an Equilibrium APT Application of a New Test Methodology, *Journal of Financial Economics* **21**, 225–289.

Fama, E. and French, K. (1993), Common Risk Factors in the Returns on Stocks and Bonds, *Journal of Financial Economics.*

Kandel, S. and Stambaugh, R. (1987), On Correlations and Inference About Mean-Variance Efficiency, *Journal of Financial Economics* **18**, 61–90.

Lehmann, B. and Modest, D. (1988), The Empirical Foundations of the Arbitrage Pricing Theory, *Journal of Financial Economics* **21**, 213–254.

Lettau, M. and Ludvigson, S. (2001), Resurrecting the (C)CAPM: A Cross-Sectional Test When Risk Premia are Time Varying, *Journal of Politcal Economy* **109:6**, 1238–1287.

Lintner, J. (1965), Security Prices, Risk, and Maximal Gains from Diversification, **20**, 587–615.

Lucas, R. E. (1978), Asset Prices in an Exchange Economy, *Econometrica* **46**, 1429–1445.

Merton, R. (1973), An Intertemporal Capital Asset Pricing Model, *Econometrica* **41**, 867–886.

Muirhead, R. J. and Waternaux, C. (1980), Asymptotic distribution in canonical correlation analysis and other multivariate procedures for nonnormal populations, *Bimetrika* **67:1**, 31–43.

Roll, R. (1977), A Critique of the Asset Pricing Theory's Tests: Part I, *Journal of Financial Economics* **4**, 129–176.

Ross, S. (1976), The Arbitrage Theory of Capital Asset Pricing, *Journal of Finance* **13**, 341–360.

Shanken, J. (1987), Multivariate Proxies and Asset Pricing Relations: Living with the Roll Critique, *Journal of Financial Economics* **18**, 91–110.

Sharpe, W. (1964), Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk, *Journal of Finance* **19**, 425–442.

Stambaugh, R. (1982), On the Exclusion of Assets from Tests of the Two Parameter Model, *Journal of Financial Economics* **10**, 235–168.

Stambaugh, R. (1988), The Information in Forward Rates: Implications for Models of the Term Structure, *Journal of Financial Economics* **21**, 41–70.

Stock, J. H. and Watson, M. W. (2002), Macroeconomic Forecasting Using Diffusion Indexes, *Journal of Business and Economic Statistics* **20:2**, 147–162.

Figure 1: Factor Processes and Their Confidence Intervals

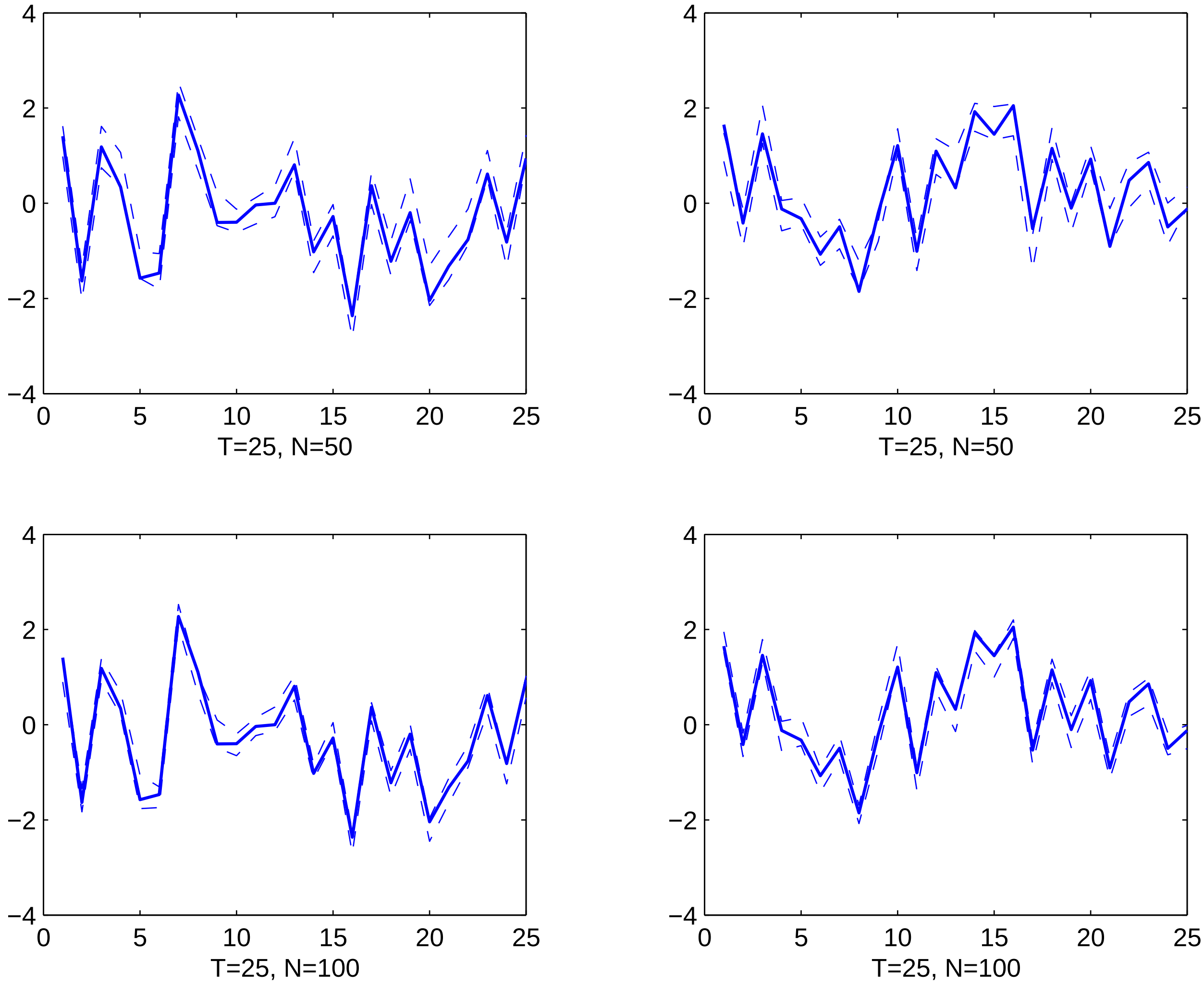

Figure 2: Measurement Errors and Their Confidence intervals