Essays in Econometrics and Random Matrix Theory

by

Matthew C. Harding

B.A. (Hons), Philosophy and Economics, University College London (2000)
M.Phil., Economics, University of Oxford (2002)

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2007

© Matthew C. Harding, MMVII. All rights reserved.

The author hereby grants to MIT permission to reproduce and distribute
publicly paper and electronic copies of this thesis document in whole or in part.

Author . . . .
Department of Economics
May 15, 2007

Certified by . . . .
Jerry Hausman
John and Jennie S. MacDonald Professor
Thesis Supervisor

Certified by . . . .
Victor Chernozhukov
Castle Krob Career Development Associate Professor
Thesis Supervisor

Certified by . . . .
Whitney Newey
Jane Berkowitz Carlton and Dennis William Carlton Professor
Thesis Supervisor

Accepted by . . . .
Peter Temin
Elisha Gray II Professor of Economics
Chairman, Department Committee on Graduate Students

Essays in Econometrics and Random Matrix Theory

by

Matthew C. Harding

Submitted to the Department of Economics
on May 15, 2007, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This dissertation develops new econometric procedures for the analysis of high-dimensional datasets commonly encountered in finance, macroeconomics or industrial organization. First, I show that traditional approaches to the estimation of latent factors in financial data underestimate the number of risk factors. They are also biased towards a single market factor, the importance of which is overestimated in samples. In Chapter 3, I derive a new consistent procedure for the estimation of the number of latent factors by examining the effect of the idiosyncratic noise in a factor model. Furthermore, I show that the estimation of factor loadings by Principal Components Analysis is inconsistent for weak factors and suggest alternative Instrumental Variables procedures. Chapter 4 uses the theoretical results of the earlier chapters to estimate the stochastic dimension of the US economy and shows that global risk factors may obfuscate the relationship between inflation and unemployment. Chapter 5 (co-authored with Jerry Hausman) suggests a new procedure for the estimation of discrete choice models with random coefficients and shows that ignoring individual taste heterogeneity can lead to misleading policy counterfactuals.

Thesis Supervisor: Jerry Hausman
Title: John and Jennie S. MacDonald Professor

Thesis Supervisor: Victor Chernozhukov
Title: Castle Krob Career Development Associate Professor

Thesis Supervisor: Whitney Newey
Title: Jane Berkowitz Carlton and Dennis William Carlton Professor

Acknowledgments

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

Modern economics has at its disposal large quantities of data, from high-frequency financial observations to detailed consumer level scanner data. Very often however the econometric methods employed to analyze such data are based on assumptions of bygone times when data was sparse and computers slow. This dissertation addresses a number of important economic questions by developing econometric procedures aimed to take advantage of the high-dimensional data available to us today. It is written in the hope that we can break the chains of convention and convenience and develop complex large scale statistical models of economic processes. Keynes ones remarked that "the difficulty lies not so much in developing new ideas as in escaping from old ones."

The first part of this dissertation addresses the problem of employing factor analysis in financial and macroeconomic applications using large firm level datasets. Data abundance presents both opportunities and challenges. Random Matrix Theory proves to be a powerful tool for understanding the limitations of traditional approaches to factor analysis and for developing new estimators more finely tuned to high-dimensional financial data.

In Chapter 2 I explain the finite sample behavior of factor models estimated by principal components. I show that in samples of dimensions commonly found in empirical applications it is not possible to distinguish all the factors from the idiosyncratic noise and that this leads to a bias towards the identification of a single factor, as documented by numerous studies. Moreover, I find that the quantities estimated are severely biased compared to their population values, even when correctly identified, and provide an approximation to their finite sample bias and their sampling distribution. Overall, these results challenge the use of PCA, as commonly employed, as a suitable tool for the identification of the underlying factor structure of asset returns in high-dimensional datasets where the number

of time periods is not large relative to the number of assets.

Chapters 3 and 4 develop econometric theory for the estimation of large N, T factor models in structural macro-finance. I employ non-commutative probability theory to derive a new estimator for the number of latent factors based on the moments of the eigenvalue distribution of the empirical covariance matrix. The proposed test combines a minimum distance procedure for the estimation of structural model parameters with a specification test on the empirical eigenvalues to solve the problem of separating the factors from the noise. I also relate the second order unbiased estimation of factor loadings to instrumental variable methods where the number of instruments is large relative to the sample size, and derive a number of alternatives to principal components with excellent finite sample properties. Using a large dataset of international stock returns, I then estimate global supply and demand shocks in a structural New Keynesian macro-finance model of the US economy. I uncover 23 global factors over the period 1973-2006, many of which impact the supply side of the US economy. Chapter 4 also shows that omitting these factors masks the role of unemployment in the Phillips curve and of the real interest rate in aggregate demand.

The second part of this dissertation addresses the estimation of discrete choice models in the presence of individual level heterogeneity. Chapter 5 is co-authored with Jerry Hausman.

Current methods of estimating the random coefficients logit model employ simulations of the distribution of the taste parameters through pseudo-random sequences. These methods suffer from difficulties in estimating correlations between parameters and computational limitations such as the curse of dimensionality. Chapter 5 provides a solution to these problems by approximating the integral expression of the expected choice probability using a multivariate extension of the Laplace approximation. Simulation results reveal that this method performs very well, both in terms of accuracy and computational time.

CHAPTER 2

The Single Factor Bias of Arbitrage Pricing Models in Finite Samples

The Arbitrage Pricing Theory (APT) of Ross (1976) assumes the lack of arbitrage opportunities in capital markets and postulates a linear relationship between actual returns and a set of K common factors, with the implication that the expected returns will be linear functions of the common factor weights. This suggests the use of Factor Analysis (FA) developed by Spearman and Hotelling at the beginning of the last century as a potential tool for the extraction of the K common factors from a sample of returns.

Since factor analysis only identifies the factor loadings up to a non-trivial rotation, the task of extracting the K common factors and determining if they are priced by the market can be difficult. Determining the right rotation is a potentially very complicated task since the number of relevant options can be very large. In practice it is often preferred to fix the rotation on *a priori* grounds and this task is implicitly performed by the use of Principal Components Analysis (PCA) as a substitute for the more laborious Factor Analysis.

Asymptotic conditions for PCA to produce results which are close to FA are provided by Chamberlain and Rothschild (1983). They require for the first K eigenvalues of the covariance matrix of factor returns to grow without bound as the number of securities N in the portfolio increases, while the remaining $N - K$ eigenvalues to remain bounded. This ensures that the returns are linearly related to the underlying K factors and even allows for weak forms of correlation between idiosyncratic shocks.

In practice one would consider the first K largest eigenvalues of the empirical covariance (or correlation) matrix of a panel of returns for a portfolio of N securities over T time periods, where both N and T are large. Since the number of underlying factors is unknown it is necessary to estimate a cut-off point

which separates the K eigenvalues corresponding to the underlying factors from the remaining $N - K$ eigenvalues due to the idiosyncratic noise component.

Over the years numerous studies (Trzcinka, 1986; Connor and Korajczyk, 1993; Geweke and Zhou, 1996; Jones, 2001; Merville and Xu, 2001) have documented the dominance of one factor, labeled as the market factor, which explains most of the sample variation. More limited and inconclusive results have been obtained for identifying other factors such as industry specific factors. This has usually been attributed to the lack of formal criteria for choosing the number of factors from an empirical distribution of eigenvalues of the sample covariance matrix. In practice it is common to choose the number K by visual inspection of the scree plot or by the use of ad-hoc cut-off points of the distribution of eigenvalues.

More recently random matrix theory was employed to describe the distribution of the idiosyncratic noise component, which has a bounded support. Therefore, it is possible to choose the number of factors as the number of eigenvalues outside the finite support of the eigenvalues due to noise, as formalized by Onatski (2005). This approach is also found in the growing econophysics literature where numerous empirical investigations of different asset markets have been performed using this method (Plerou et. al., 2002; Bouchard and Potters, 2003). These studies also document the dominance of the market factor and remain inconclusive on the identification of further factors.

Brown (1989) provides Monte-Carlo evidence of an economy with K factors, each of which is priced and contributes equally to the returns. Moreover, the economy is by construction admissible under the framework of Chamberlain and Rothschild (1983) and calibrated to actual data from the NYSE. Nevertheless, he finds evidence that PCA is biased towards a single factor model. Thus, we cannot conclude that the empirical evidence presented in the numerous studies, some of which were mentioned above, is not also consistent with a setup where more than one factor is present in the economy, but where PCA fails to identify the entire set of relevant factors.

In this paper we use recent results from stochastic eigen-analysis to quantify the intuition of Brown (1989) and explain the single factor bias of arbitrage pricing models through the finite sample behavior of the factors estimated by PCA. We show that in finite samples of dimensions commonly found in these empirical investigations it is not possible to distinguish some of the factors from the idiosyncratic noise element either by heuristic methods or by a random matrix approach. This leads to the bias towards the identification of a single factor which is routinely reported in empirical investigations. Moreover, we find that the quantities estimated are severely biased, even when correctly identified, and provide an approximation to the finite sample bias and their sampling distribution. Overall, these results challenge the use of PCA, as it is commonly used in the empirical finance literature, as a suitable tool for the identification of the underlying factor structure of asset returns in situations where the number of assets is large relative

to the number of time periods.

■ 2.1 Limiting Behavior of Sample Eigenvalues

Let us now consider a version of the model introduced by Brown (1989). Assume that the relevant portfolio consists of N assets which are observed for T time periods. The asset returns are generated by an exact linear factor technology with K factors. Thus, the demeaned asset returns are given by:

$$R_N = \Lambda F' + \epsilon, \tag{2.1}$$

where R_N is a matrix of dimensions $N \times T$ of asset returns, F is an $T \times K$ matrix of factor scores, Λ is a $N \times K$ matrix of factor loadings and ϵ is a $N \times T$ matrix of idiosyncratic noise components. The covariance matrix of returns is given by the $N \times N$ matrix $\Sigma_N = (1/T) R_N R'_N$. We also assume that $E(\epsilon) = 0, E(\epsilon\epsilon') = \sigma_\epsilon^2 I_N$ and $E(|\epsilon_{ii}|^4) < \infty$. Notice that we do not assume normality for most of the results in this paper except for Proposition 2.

Furthermore, we assume that $N \to \infty$, $T \to \infty$, and $N/T \to c \in (0, \infty)$, where c is a constant. The asymptotic framework is similar to that used in other panel data studies (e.g. Hahn and Kuersteiner, 2002) and will facilitate the derivation of finite sample results. It corresponds to a setup where the number of cross-sectional units is large relative to the number of available time periods. We will first characterize the distribution of population eigenvalues and then employ recent results from random matrix theory to characterize their limiting distribution under the large N, large T asymptotic framework. This will allow us to explore the conditions under which the factors are identified and also derive the distribution of the correctly identified factors.

It is important to note the large N, large T asymptotic framework is the appropriate framework to model the samples encountered in practical finance. Traditionally PCA results were derived under the assumption that N is fixed while T goes to infinity. In practice however portfolio managers tend to have only a limited number of time periods available while looking to measure risk factors from a large number of securities.

Furthermore, assume that $E[(1/T)F'F] = \sigma_F^2 I_K$ and $E[(1/T)\epsilon F] = 0$. The then population covariance of factor returns is given by

$$\Sigma_N = \sigma_F^2 \Lambda\Lambda' + \sigma_\epsilon^2 I_{N.} \tag{2.2}$$

We now impose additional assumptions of the $N \times K$ matrix of factor loadings, Λ. Let $\Lambda = [\Lambda_1 : \Lambda_2 : ... : \Lambda_K]$, where each of the columns Λ_i of Λ is given by $\Lambda_i = \beta\iota + \sqrt{\sigma_\beta^2} e_i$. Let β be a constant, ι be the $N \times 1$ vector $(1, 1, 1, ..., 1)'$ and e_i be an i.i.d. vector random variable with mean 0 and variance 1 and finite fourth

moments.

First let us consider the eigenvalue behavior of this model under the assumption that $T \to \infty$ much faster than N. The resulting eigenvalues are labeled as population eigenvalues, since they correspond to a setup where the number of time-periods for which the model is observed is very large.

Let us now compute the population eigenvalues of Σ_N from the decomposition $U_N \Sigma_N U_N^{-1} = \text{diag}\{\lambda_1, \lambda_2, ..., \lambda_N\}$, where $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_N$. Since $\Lambda\Lambda'$ is rank deficient it will have K non-zero eigenvalues and $N - K$ zero eigenvalues. But since the non-zero eigenvalues of the $N \times N$ matrix $\Lambda\Lambda'$ are the same as the eigenvalues of the $K \times K$ matrix $\Lambda'\Lambda$, it is sufficient to consider only the latter one. Furthermore, notice that in the large T limit, $\Lambda_i'\Lambda_j = N(\beta^2 + \sigma_b^2)$ if $i = j$ and $\Lambda_i'\Lambda_j = N\beta^2$ is $i \neq j$. Hence, $\Lambda'\Lambda = N\sigma_b^2 I_K + N\beta^2 J_K$, where J_K is the $K \times K$ matrix of ones. It follows by the Sherman-Morrison theorem (Graybill, 1983) that the eigenvalues l_i of $\Lambda'\Lambda$ are given by $l_1 = N(\sigma_b^2 + K\beta^2)$ and $l_j = N\sigma_b^2$, for $j = 2...K$. Furthermore, from this we obtain the population eigenvalues of Σ_N as:

$$\lambda_1 = N\sigma_F^2 \left(\sigma_\beta^2 + K\beta^2\right) + \sigma_\epsilon^2 \tag{2.3}$$

$$\lambda_i = N\sigma_F^2 \sigma_\beta^2 + \sigma_\epsilon^2, \text{ for i} = 2...K \tag{2.4}$$

$$\lambda_j = \sigma_\epsilon^2, \text{ for j} = K+1...N. \tag{2.5}$$

Let us now consider the sample covariance matrix derived from a panel of returns of a portfolio of size N observed over T time periods where both N and T are large and $N/T - c = o(N^{-1/2})$. Recent advances in random matrix theory have made it possible to relate the distribution of sample eigenvalues to that of the population eigenvalues described above. Below we will use a set of results for "spiked covariance matrices" as derived by Baik and Silverstein (2005), Onatski (2005) and Paul (2005). Identical results were derived independently by these authors based on very similar assumptions. The results of Baik and Silverstein (2005) were derived under slightly more general conditions.

For the purposes of this note we will assume that at least the first factor can be identified. This is guaranteed if the mean factor loading is high enough in relation to the idiosyncratic component. Our main focus is on the identification of the remaining $K - 1$ factors.

Applying the random matrix theory results quoted above we can obtain the expected value of the sample eigenvalues for large N, large T, which are summarized in the following proposition.

Proposition 2.1. *Let S_N be a sample covariance matrix for a set of observations with the population covariance Σ_N described above. Furthermore, let $V_N S_N V_N^{-1} = \text{diag}\{\hat{\lambda}_1, \hat{\lambda}_2, ..., \hat{\lambda}_N\}$, be its eigenvalue decomposition with $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq ... \geq \hat{\lambda}_N$.*

Then if $N/T - c = o(N^{-1/2})$ we have the following almost sure limits:

$$a) \quad \hat{\lambda}_1 \overset{a.s.}{\to} \left\{ N\sigma_F^2 \left(\sigma_\beta^2 + K\beta^2\right) + \sigma_\epsilon^2 \right\} \left\{ 1 + \frac{1}{T} \frac{\sigma_\epsilon^2}{\sigma_F^2 \left(\sigma_\beta^2 + K\beta^2\right)} \right\}, \qquad (2.6)$$

$$b) \quad \hat{\lambda}_i \overset{a.s.}{\to} \left\{ N\sigma_F^2 \ \sigma_\beta^2 + \sigma_\epsilon^2 \right\} \left\{ 1 + \frac{1}{T} \frac{\sigma_\epsilon^2}{\sigma_F^2 \sigma_\beta^2} \right\}, \text{for } i = 2...K \text{ and } N \geq \frac{1}{T}\left(\frac{\sigma_\epsilon^2}{\sigma_F^2 \sigma_\beta^2}\right)^2,$$
$$(2.7)$$

$$c) \quad \hat{\lambda}_i \overset{a.s.}{\to} \sigma_\epsilon^2 \left(1 + \sqrt{N/T}\right)^2, \text{for } i = 2...K \text{ and } N < \frac{1}{T}\left(\frac{\sigma_\epsilon^2}{\sigma_F^2 \sigma_\beta^2}\right)^2, \qquad (2.8)$$

$$d) \quad \hat{\lambda}_j \overset{a.s.}{\to} \sigma_\epsilon^2 \left(1 + \sqrt{N/T}\right)^2, \text{ for } j = K+1...N. \qquad (2.9)$$

Notice that the sample eigenvalues are biased estimates of the corresponding population eigenvalues. Moreover, the bias is always positive and does not disappear as we add more securities to the portfolio. The bias only disappears as $T \to \infty$, that is as we add more time periods to the sample. The quantity $\sigma_\epsilon^2(1 + \sqrt{N/T})$ corresponds to the upper support bound of the Marcenko-Pastur distribution, which characterizes the sample eigenvalues of the sample covariance matrix with mean zero and variance $\sigma_\epsilon^2 I_N$. We shall label this as the Marcenko-Pastur bound.

To exemplify the results of Proposition 2.1 let us consider the calibration of Brown (1989) based on the NYSE. Thus, let $\beta = 1$, $\sigma_\beta^2 = 0.01$, $\sigma_F^2 = 0.000158$ and $\sigma_\epsilon^2 = 0.0045$. Furthermore, we let $T = 80$ and simulate the factor economy for portfolio sizes between $N = 50$ and $N = 200$. For each value of N we simulate the portfolio 300 times and compute the corresponding sample covariance matrix. This procedure is then used to extract the sample eigenvalues. The results for the first 10 eigenvalues are plotted in Figure 2.1, where we report the mean eigenvalue over the simulations for each portfolio size. Furthermore, for the largest eigenvalue we compute the interquartile range of the distribution which is reported in the figure as the shaded area around the mean value of the first eigenvalue for each N. We also report the Marcenko-Pastur bound introduced above which characterizes the largest eigenvalue compatible with a pure idiosyncratic noise model. Additionally, we plot the population eigenvalues for the calibration at each N and the corresponding almost sure limit derived in Proposition 2.1.

Notice that the sample eigenvalues are biased away from the population eigenvalues at each portfolio size and that the extent of the bias is correctly estimated by the limits derived in Proposition 1. The bias is substantial and the population
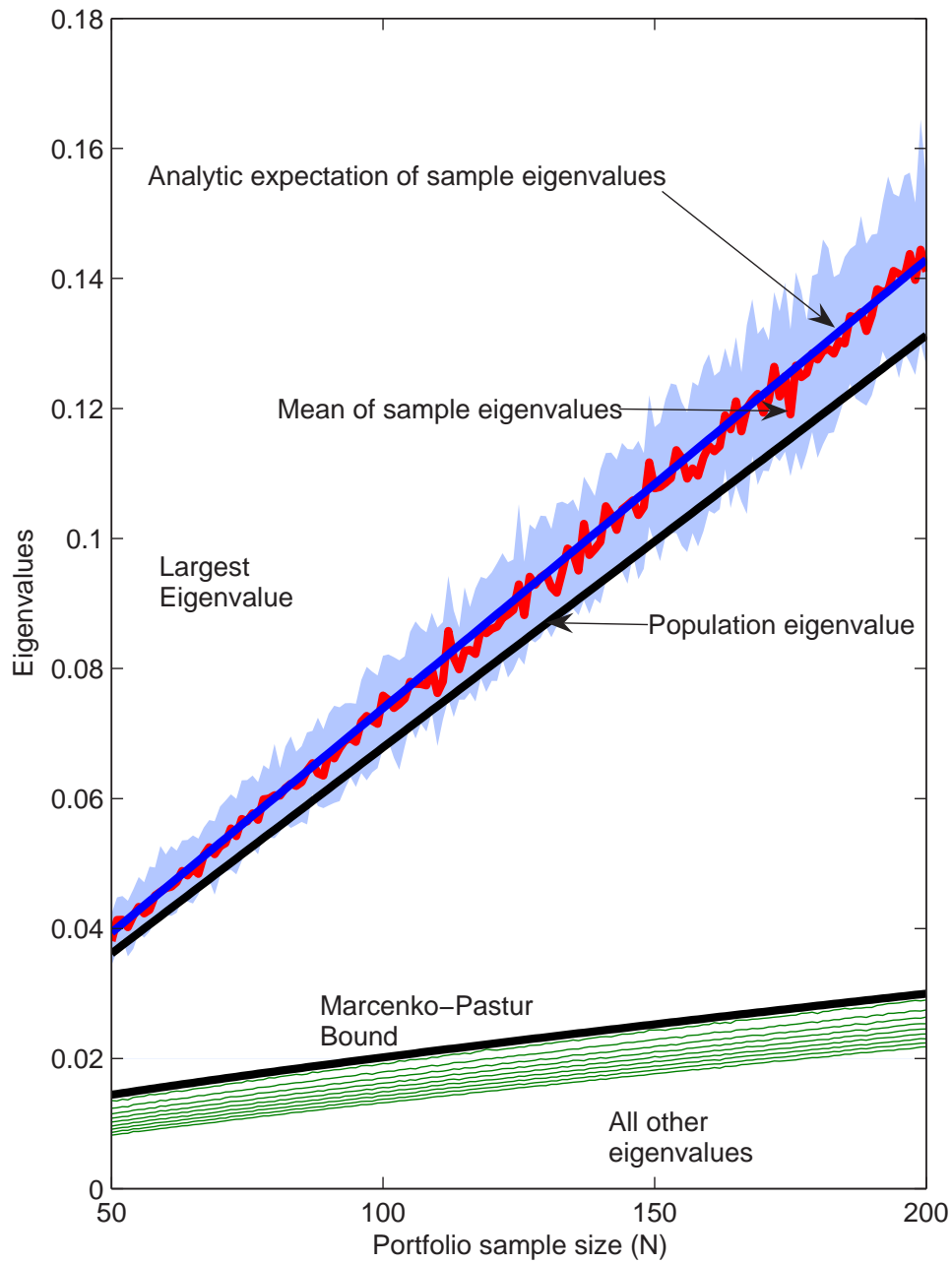
Figure 2-1. Limiting behavior of 10 largest eigenvalues as a function of portfolio size.

eigenvalues correspond the lower 25th percentile of the distribution of sample eigenvalues. Additionally, notice that all the remaining $N - 1$ eigenvalues are bounded from above by the Marcenko-Pastur bound. This implies that in this setting it is not possible to identify any of the $K - 1$ factors even though they contribute equally to the asset returns. No test based on the sample eigenvalues will be able to distinguish the second through $K-$th eigenvalues from the remaining $N - K$ eigenvalues due to the idiosyncratic noise component over the given range of portfolio sizes. Factor estimation based on PCA will only uncover a market factor corresponding to the first eigenvalue which will have high explanatory power, yet it will be a biased estimate of the true first factor.

Although the parameter values used in this simulation are similar to those used in many empirical application we can enquire further as to the minimum portfolio size $N_{\min}$ required to correctly identify all factors of the economy for the given value of T. We obtain $T_{\min} \geq \frac{1}{N}\left(\frac{\sigma_\epsilon^2}{\sigma_F^2 \sigma_\beta^2}\right)^2 \cong 40,000$. This is an extremely large value (over 120 years of daily data) and shows the difficulties involved in identifying the non-diversifiable sources of risk in actual portfolio analysis using PCA. In particular notice that we face the a trade-off between the need for large samples in order to identify all latent factors and model stability over extended periods of time.

Using the results in Onatski (2005) and Paul (2005) we can also give the distribution of the first eigenvalue under the large N, large T asymptotic framework.

Proposition 2.2. *If $\epsilon \sim^d N(0, \sigma_\epsilon^2 I_N)$ and if $N/T - c = o(N^{-1/2})$ we have $\hat{\lambda}_1 \sim^d N(m, q)$, where*

$$m = \left\{N\sigma_F^2 \left(\sigma_\beta^2 + K\beta^2\right) + \sigma_\epsilon^2\right\} \left\{1 + \frac{1}{T}\frac{\sigma_\epsilon^2}{\sigma_F^2 \left(\sigma_\beta^2 + K\beta^2\right)}\right\}, \qquad (2.10)$$

$$q = 2\left\{N\sigma_F^2 \left(\sigma_\beta^2 + K\beta^2\right) + \sigma_\epsilon^2\right\}^2 \left\{1 - \frac{1}{NT}\frac{\sigma_\epsilon^2}{\left[\sigma_F^2 \left(\sigma_\beta^2 + K\beta^2\right)\right]^2}\right\}. \qquad (2.11)$$

■ 2.2 Conclusion

In this chapter we have explained the reasons for the bias of APT models estimated by PCA towards a single factor model. We have shown that unless the period of time over which the portfolio is observed is extremely large, it is not possible to identify all the factors of the economy. This is due to the finite sample bias of the estimated eigenvalues of the sample covariance matrix when the number of periods over which the portfolio is observed is of similar orders of magnitude to that of the number of securities in the portfolio. Using recent results in random matrix theory we have characterized the limiting behavior of the sample eigenvalues and the distribution of the largest eigenvalue.

This chapter challenges the use of PCA in its standard form as a tool for

factor analysis in finite samples. The correct estimation of factors requires the use of bias corrections in finite samples. Moreover, it seems that the need arises to explore the finite sample properties of other factor estimation procedures other than PCA if we are to identify the full set of factors which determine the asset returns in a portfolio observed only for a finite period of time.

Bibliography

BAIK, J., AND J. W. SILVERSTEIN (2005): "Eigenvalues of Large Sample Covariance Matrices of Spiked Population Models," University of Michigan, Ann Arbor.

BOUCHARD, J.-P., AND M. POTTERS (2003): *Theory of Financial Risk and Derivative Pricing*. Cambridge.

BROWN, S. J. (1989): "The Number of Factors in Security Returns," *Journal of Finance*, 54, 1247–1262.

CHAMBERLAIN, G., AND M. ROTHSCHILD (1983): "Arbitrage, Factor Structure, and Mean-Variance Analysis of Large Asset Markets," *Econometrica*, 51, 1281–1304.

CONNOR, G., AND R. A. KORAKJCZYK (1993): "A Test for the Number of Factors in an Approximate Factor Model," *Journal of Finance*, 48, 1263–1291.

GEWEKE, J., AND G. ZHOU (1996): "Measuring the Pricing Error of the Arbitrage Pricing Theory," *Review of Financial Studies*, 9, 557–587.

GRAYBILL, F. A. (1983): *Matrices with Applications in Statistics*. Wadsworth.

HAHN, J., AND G. KUERSTEINER (2002): "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects when Both N and T are Large," *Econometrica*, 70, 1639–1657.

JONES, C. S. (2001): "Extracting Factors from Heteroskedastic Asset Returns," *Journal of Financial Economics*, 62, 293–325.

MERVILLE, L. J., AND Y. XU (2002): "The Changing Factor Structure of Equity Returns," University of Texas.

ONATSKI, A. (2005): "Asymptotic Distribution of Principal Components Estimators of Large Spherical Factor Models," Columbia University.

PAUL, D. (2005): "Asymptotics of the Leading Sample Eigenvalues for a Spiked Covariance Model," Stanford University.

PLEROU, V., V. GOPIKRISHNAN, B. ROSENAU, L. AMARAL, T. GUHR, AND E. STANLEY (2002): "Random Matrix Approach to Cross-Correlations in Financial Data," *Physical Review E*, 65, 1–17.

ROSS, S. (1976): "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory*, 3, 343–362.

TRZCINKA, C. (1986): "On the Number of Factors in the Arbitrage Pricing Model," *Journal of Finance*, 41, 347–368.

CHAPTER 3

Structural Estimation of High-Dimensional Factor Models

This chapter develops new techniques for the estimation of factor models in large datasets where the number of observations grows with the time dimension. Factor models relate observed data to a small set of unobserved variables which are then estimated. These models underlie many important tools of modern economics and finance, but no definitive econometric theory exists for the case of large panel datasets commonly encountered today.

We relate the identification and estimation of factor models to the asymptotic behavior of estimated eigenvalues of large random matrices, providing a connection between economics and the new mathematical field of Random Matrix Theory. In this chapter we take a structural approach and show how the estimation of factor models can be improved by incorporating the economic assumptions of the model into the estimation procedure. In particular, we allow for arbitrary parametric forms of heteroskedasticity and autocorrelation.

We show that the key to identifying the number of latent factors lies in correctly understanding the structure of the noise (idiosyncratic effects) in the data. Once we can separate the estimated eigenvalues of a large factor model into those due to the latent structure and those due to the noise, we can construct a procedure that will consistently estimate the number of factors. Furthermore, we show that Principal Components Analysis (PCA) becomes unreliable for weak factors and relate the second order unbiased estimation of the factor loadings to recent advances in the estimation of models with instrumental variables when the number of instruments is large.

In the next chapter we apply these econometric procedures to a structural New Keynesian macro-finance model, we uncover a substantial number of global factors which act as supply shocks to the US economy. We characterize the

nature of these factors and show that they can be used to explain the inflationary experience of the US over the past few decades. Moreover, we show that ignoring the effect of global factors on the US economy masks the role of unemployment in the Phillips curve and of the real interest rate in aggregate demand.

While factor models have been used for almost a century, standard econometric methods were developed under the assumption that the time dimension grows large while the cross-section dimension is small and bounded. In applications where both the number of individuals and the number of time periods is large, standard econometric theory fails and becomes an unreliable statistical guide to data analysis. New econometric procedures for the estimation of high-dimensional factor models are the subject of active research (Bai and Ng, 2002; Stock and Watson, 2005; Onatski, 2006).

The application of econometric methods which take into account the special nature of large panel datasets leads us to reconsider stylized facts which we have taken for granted, such as the number of factors explaining most of the variation in financial returns. Chapter 2 shows that in a large panel data setup, the estimated eigenvalues corresponding to strong factors are severely upward biased in finite samples. Since the ratio of the largest eigenvalue to the sum of all eigenvalues has been traditionally used to measure the effect of the factors, there is a bias towards accepting only a few (3-5) factors as explaining most of the data. In fact, many other factors may exist in the data and contain potentially valuable economic information.

The identification of the number of factors is central to the estimation of factor models and in Section 3.1 we show that it is possible to separate the identification of the number of factors from the estimation of the factor loadings and factor scores, and estimate the number of factors consistently in large factor models. In Section 3.2 we explain why PCA estimation is inconsistent for weak factors and develop a number of instrumental variable approaches to the estimation of factor loadings with excellent finite sample properties. The next chapter develops a structural macroeconomic model and estimates global factors and their effect on the US economy.

■ 3.1 Determining the Number of Factors

We are interested in the following large (N, T) panel data model with latent factors:

$$R_t = \Lambda F_t + U_t, \tag{3.1}$$

for $t = 1 \ldots T$. R_t is an $N \times 1$ vector of observations, F_t is a $p \times 1$ vector of latent factors, Λ is an $N \times p$ matrix of coefficients (factor loadings) and U_t is an $N \times 1$ vector of idiosyncratic errors. In this model only R_t is observed while Λ, F_t and U_t are unobserved for all t.

The aim of this model is to explain the variation in R_t with reference to a

small dimensional set of latent factors F_t by decomposing the observed variation into a common component ΛF_t and an idiosyncratic component U_t. In order to simplify the discussion we shall refer to the cross-sectional dimension N as "individuals", while we let the time-series dimension T denote "periods". Note that the coefficients Λ correspond to loadings or weights of the common factors F_t for each individual.

This particular statistical model originates in the work of Spearman and Hotelling and has been incorporated in many economic models of interest. The traditional econometric approach to solving this model was derived under the assumption that N is a fixed small number while T is large (Goldberger, 1972; Robinson, 1974; Zellner, 1970). With the availability of large dimensional panel data where both N and T are large, this model has received renewed attention and is currently an active area of research (Amengual and Watson, 2006; Bai and Ng, 2002; Onatski, 2006).

In finance the factor model of equation 3.1 corresponds to the Arbitrage Pricing Theory (APT) of Ross (1976), which explains the returns $R_{i,t}$ on $i = 1 \ldots N$ assets observed over $t = 1 \ldots T$ time periods by reference to a small set of risk factors F_t and asset specific shocks $U_{i,t}$. These multi-factor asset pricing models represent a major improvement over simpler single-factor CAPM models in evaluating the risk-return trade-off. While observable proxies have been used for the unobserved factors F_t in many applications, practitioners tend to agree that statistical factors derived from the econometric estimation of the model tend to outperform models evaluated by factor proxies (Miller, 2006).

More recently, the standard factor model above has been incorporated in more complex hybrid models involving both observed and latent factors:

$$BY_t + \Gamma Z_t + \Lambda F_t + U_t = 0, \tag{3.2}$$

where Y_t corresponds to a set of $N \times 1$ dimensional endogenous variable and Z_t is a set of $N \times 1$ dimensional observed exogenous variables with coefficients B and $Gamma$ respectively.

In microeconomics, such a model was first used by Gorman (1980) to analyze the characteristics of demand. In particular it is convenient to interpret the term $\Lambda F_t + U_t$ as a multifactor error structure or interactive fixed effects (Bai, 2005; Pesaran, 2006). This model has recently received considerable attention in labor economics in the study of the relationship between education and earnings (Heckman and Navarro, 2006).

Factor models are also very popular in macroeconomics, where a forecasting model that uses factors constructed from numerous macroeconomic time series can substantially improve forecasting (Stock and Watson, 2006). The recent field of macro-finance has also relied on the estimation of factors from bond yields in order to improve the performance of small-scale structural macroeconomic models

(Ang and Piazzesi, 2003). In the next chapter we introduce a new approach to the structural estimation of New Keynesian models by employing a multifactor error structure to identify global supply and demand shocks to the US economy through the extraction of factors from international stock markets.

The primary focus of our chapter, however, is determining the number of factors and the estimation of the factor loadings in equation 3.1 for high-dimensional models where both N and T are large. This is captured by the following assumption:

Assumption 3.1 (Asymptotics): *The number of individuals increases with the sample size. Thus, $N \to \infty$ and $T \to \infty$ and $N/T \to c \in (0, \infty)$.*

This assumption is familiar to the literature on large N and T panel data (Hahn and Kuersteiner, 2002). The constant c, representing the limiting ratio of rows to columns in our panel, will play a very important role in the subsequent discussion.

The traditional statistical approach of Anderson and Rubin (1956) for solving factor models involves the assumption that the errors $U_{i,t}$ are uncorrelated both across individuals and across time. In many economics and finance applications, this assumptions has proved to be too restrictive. In particular, Chamberlain and Rothschild (1983) show that if we allow for weak heteroskedaticity and time dependence of the error terms in the APT model, the mean returns are approximately linear in the factor loadings. Thus, the model remains correct under weak departures from the strict factor structure in the large N and T limit.

A major contribution of this chapter is the development of an approach that allows us to deal with the approximate factor model where both heteroskedasticity and autocorrelations are possible. In order to do so, we need to describe the precise nature of the weak heteroskedasticity and autocorrelations compatible with the approximate factor model.

Let $U = [U_1, U_2, ..., U_T]$ be the $N \times T$ matrix of errors in equation 3.1, with elements $U_{i,j}$ for $i = 1 \ldots N$ and $t = 1 \ldots T$, and where each column t corresponds to a realization of the errors at time t for the N individuals in the sample. Let vec(U) be the $NT \times 1$ vector obtained by stacking the columns of U, i.e. vec$(U) = (U_1 U_2 ... U_T)'$.

Assumption 3.2 (Cross-sectional and Time Dependence) *There is an $N \times N$ matrix A_N and a $T \times T$ matrix B_T such that $E(\text{vec}(U)) = 0, \text{Cov}(\text{vec}(U)) = A_N \otimes B_T$ and $E((U_{i,j})^4) < \infty$ for all N, T and A_N is unrelated to B_T.*

This assumption states that the errors U are mean zero and have finite fourth order moments. The most important assumption lies in the (N, T) separability of the covariance matrix into an $N \times N$ component A_N and a $T \times T$ component B_T. The matrix A_N captures the cross-sectional dependence between individuals, while B_T captures the form of time dependence. Note that this allows for very general forms of cross-sectional and time dependence, such as full correlation matrices. It does, however, limit the number of unknown parameters by assuming

that the cross-sectional dependence and the time dependence are unrelated to each other.

In particular, note that one familiar distribution satisfying Assumption 3.2 is the matrix variate normal distribution (Arnold, 1981), where $\text{vec}(U)$ $\text{Normal}_{(N,T)}(0, A_N \otimes B_T)$, where the distribution function is given by:

$$f(U) = (2\pi)^{-\frac{NT}{2}} \det(A_N)^{-T/2} \det(B_T)^{-N/2} \exp\left\{\text{tr}[-\frac{1}{2}U'A_N^{-1}UB_T^{-1}]\right\}. \quad (3.3)$$

The framework of approximate factor structure derives identification by requiring that the covariance matrix of the observations R_t has p unbounded eigenvalues corresponding to the latent factors and $N - p$ bounded eigenvalues corresponding to the idiosyncratic noise component. In order to account for these additional constraints, we first require some additional definitions. Consider the spectral decomposition of an arbitrary symmetric $n \times n$ matrix C_n. Since the matrix C_n is symmetric, we can find a matrix V with columns that are orthogonal to each other such that $A_n = V'DV$. The matrix $D = \text{diag}\{\lambda_1, \lambda_2, ..., \lambda_n\}$ contains the set of eigenvalues of the matrix C_n, while the columns of V are the eigenvectors of C_n. We can now define the following proper cummulative distribution function $F^{C_n}(\lambda)$ on the spectrum $\lambda_i \in \{\lambda_1, \lambda_2, ..., \lambda_n\}$ of C_n:

$$F^{C_n}(\lambda) = \frac{1}{n}\left\{\text{Number of Eigenvalues of } C_n \leq \lambda\right\} = \frac{1}{n}\sum_{\lambda_i \leq \lambda} 1. \quad (3.4)$$

Note that the spectrum does count multiplicities of eigenvalues.

Assumption 3.3 (Bounded Spectrum): *Denote by F^{A_N} and F^{B_T} the eigenvalue distribution of the matrix A_N and B_T respectively. Then as $N \to \infty$ and $T \to \infty$, $F^{A_N} \to F^A$ and $F^{B_T} \to F^B$, the eigenvalue distributions converge to nonrandom limiting distributions F^A and F^B. Moreover, let $\|\text{Sp}(A_N)\|$ and $\|\text{Sp}(B_T)\|$ denote the spectral norms of A_N and B_T. Assume that both spectral norms are bounded in N and T respectively.*

This assumption is required in order to guarantee that the (scaled) eigenvalue distribution of $N^{-1}UU'$ converges to a non-random distribution as $N \to \infty$, and moreover that the spectrum of $N^{-1}UU'$ is also bounded (Silverstein and Bai, 1995; Bai and Silverstein, 1998). If $A_N = I_N$ and $B_T = I_T$ then the limiting distribution of $N^{-1}UU'$ converges to the limiting distribution of Marcenko and Pastur (1967) with support bounded on $\left[(1 - \sqrt{c})^2, (1 + \sqrt{c})^2\right]$ as discussed in Chapter 2.

Assumption 3.2 and 3.3 rule out certain types of explosive behavior of the error terms and require that the variances converge for each time series to a finite value. Additionally, this assumption ensures that the moments of the empirical eigenvalue distribution of $N^{-1}UU'$ converge almost surely since the limiting distribution has bounded support. These moments will play an essential role in our procedure to estimate the number of factors.

Notice that the assumptions on the idiosyncratic error terms imposed above allow for a large range of empirically relevant models. Many forms of heteroskedasticity and autocorrelation are consistent with the assumptions on weak dependence. While some stochastic processes in the time-series literature are naturally excluded, such a unit roots, others such as structural breaks are allowed. In particular, notice that the assumptions above are consistent with structural breaks in the variance occurring at some unknown change points. It seems that these models have not been discussed so far in the literature on factor models, but they are undoubtedly important in the context of factor models based on large N, T panel data where we have the temptation to include data going back for many years and thus potentially covering more than one variance regime. To illustrate this point let us assume that we are interested in constructing a model using data before and after a financial crisis such as the Asian crisis of the late 1990s, but we are unsure as to the exact change point in the time series of the variances. If we assume τ periods to have been in the variance regime σ_1 and $T - \tau$ periods in the variance regime σ_2, a simple model constructed along the lines of Assumptions 2 and 3 would be $A_N = I$ and $B_T = [\sigma_1 \otimes I_\tau] \oplus [\sigma_2 \otimes I_{T-\tau}]$, where $\oplus$ stands for the direct sum of the two matrix spaces. Hence the spectrum of A_N consists of $\lambda_i = 1$ for $i = 1 \ldots N$, while the spectrum of B_T consists of $\lambda_i = \sigma_1$ for $i = 1 \ldots \tau$ and $\lambda_i = \sigma_2$ for $i = \tau + 1 \ldots T$. These two spectra satisfy the boundedness condition of Assumption 3.2 and are thus admissible. Moreover, τ does not have to be determined a priori and can be a model parameter that is estimated at the same time as the number of factors. This illustrates the flexibility of our approach in the estimation of factor models by using the empirical eigenvalue distribution.

Assumption 3.4 (Pervasive factors): *Assume that the factors F_t are independent of U_t. Denote by $\mu_0 = \min\{\mathrm{Sp}(\frac{1}{T}\sum_{t=1}^{T} \Lambda F_t F_t' \Lambda')\}$, the smallest eigenvalue of the covariance of the factors weighted by the factor loadings. Then for all $N \to \infty$, there is some $M > 0$ and $M \to \infty$ such that $\mu_0 \geq M$.*

Notice that this is consistent with both random and fixed factor loadings Λ. This assumption requires that the latent factors impact at least a fraction of the individuals, where the fraction increases with the sample size. Thus even factors which are relatively weak will be revealed in large samples due to their effect on a large number of individuals. The pervasive factors reflect the structural part of an economic model. We can think of them as the systemic component of our model to be distinguished from the idiosyncratic noise perturbations. In the context of the application in the next chapter, we think of pervasive factors as global supply and demand shocks which impact a large number of firms simultaneously. Note, however, that pervasive factors are not the only kind of factors one might be interested in. While not the subject of this chapter, we could adjust the current methodology to estimate small scale factors that are related to a small number of firms but where the number of firms affected by them does not increase with the

sample size. For example, several firms might rely on the same supply network and thus have correlated fluctuations. In certain circumstances, we might be able to distinguish these factors from the background idiosyncratic noise. These small scale factors are, however, excluded by Assumption 3.4 since for them $\mu_0 \to 0$ as $N \to \infty$.

Small scale factors might be particularly interesting if we wish to detect unusual correlations in large samples and correspond to multicollinearities in the data, a topic that has received some attention in the statistics literature. Assumption 3.4, however, corresponds to the current economic practice and mirrors the assumptions in Chamberlain and Rothschild (1983).

■ 3.1.1 Factor Identification Strategy

The classical statistics literature on factor models recognizes that we can write the covariance matrix of the observations as:

$$\Sigma_N = \frac{1}{T} \sum_{t=1}^{T} R_t R_t' = \frac{1}{T} \sum_{t=1}^{T} \Lambda F_t F_t' \Lambda' + \frac{1}{T} \sum_{t=1}^{T} U_t U_t' = \Xi_N + \Omega_N, \qquad (3.5)$$

where $\mathrm{rank}(\Xi_N) = p$ and $N \to \infty$. Thus, the covariance matrix of the observations can be thought of as a finite (p) rank perturbation Ξ_N of the idiosyncratic noice covariance Ω_N. If we were to observe the population equivalents of our matrices Ξ_N and Ω_N, which we denote by Ξ_0 and Ω_0, under the assumptions of our factor model, we would observe an infinite number of small bounded eigenvalues for Ω_0 and a small number of infinite eigenvalues for Ξ_0. In finite samples, however, it has been noted that even for the simplest case of the strict factor model with $A_N = I_N$ and $B_T = I_T$ both the eigenvalues of Ξ_N and the eigenvalues of Ω_N increase with N (Brown, 1989). Until very recently this has prompted economists and statisticians to believe that factor identification based on the empirical distribution of eigenvalues is impossible (Bai and Ng, 2002). In this chaper we show how factor identification based on the empirical eigenvalue distribution is in fact possible and provides a very powerful new approach to factor analysis in a large class of models.

In order to identify the number of factors we rely on recent advances in the field of random matrix theory which provides mathematical tools that enable us to characterize the empirical eigenvalue distribution for many symmetric matrices (Edelman and Rao, 2005). Our approach is structural in that it relies on explicit assumptions about the form of cross-sectional and time dependence. In particular, Assumption 3.2 states that the covariance of the idiosyncratic terms is separable between a cross-sectional correlation matrix A_N and a time dependence matrix B_T. For many cases of interest it is sufficient to impose a specific structural form on these two matrices and parameterize these matrices as $A_N(\theta_A)$ and $B_T(\theta_B)$, where $\theta = (\theta_A, \theta_B)$ is a low dimensional vector of unknown structural covariance

parameters. For example, the model with two variance regimes discussed above depends on $\theta = (\sigma_1, \sigma_2, \tau)$, where σ_1 and σ_2 are the two variances and τ/T is the probability of being in the first regime. Our procedure estimates the unknown parameter vector θ at the same time as the number of factors p. The main question in identifying the finite rank p perturbation due to the pervasive factors is how does the empirical eigenvalue distribution of Ω_N depend on (A_N, B_T)? While in general we cannot analytically characterize the empirical distribution of Ω_N, in Section 3.1.2 we show that we can compute the moments of the empirical eigenvalue distribution of Ω_N in terms of A_N and B_T, which gives rise to a minimum distance estimation procedure and a downward testing procedure of the moment conditions that correctly identifies the number of factors.

In order to simplify mathematical notation we restrict our attention to the case where $B_T = I_T$ and discuss the remaining cases in Section 3.1.4. Furthermore, notice that without loss of generality, we let $0 < c \leq 1$, since the non-zero eigenvalues of CC', for some $N \times T$ dimensional matrix C are the same as the eigenvalues of $C'C$. The remaining $T - N$ eigenvalues of $C'C$ are all zero.

Define the Cauchy Transform of an eigenvalue distribution function F^C for some matrix C as:

$$G_C(w) = \frac{1}{N} \lim_{N \to \infty} E \left\{ \text{tr}[\frac{1}{wI_N - C}] \right\} = \int \frac{1}{w - \lambda} F^C(\lambda) \qquad (3.6)$$

for $w \in \mathbb{C}^+$ with $\mathfrak{Im}(w) > 0$. This analytic function plays an important role in many random matrix theory results where it serves as an analogue to the Fourier transform in traditional probability theory. In particular, it allows us to recover the eigenvalue probability density function from the Stieltjes-Perron Inversion:

$$\frac{dF^C(\lambda)}{d\lambda} = -\frac{1}{\pi} \lim_{\xi \to 0} \mathfrak{Im}[G_C(w + i\xi)]. \qquad (3.7)$$

First consider the limit distribution of the eigenvalues of the noise covariance matrix Ω_N.

Proposition 3.1: As $N \to \infty$, $T \to \infty$, $N/T \to c$, $A_N \to A$ and $F^{A_N} \to F^A$, *the empirical eigenvalue distribution $F^{\Omega_N}(\lambda)$ converges to a non-random asymptotic distribution function $F^\Omega(\lambda; F^A, c)$ with bounded support.*

Proof: See Silverstein (1995), Rao and Edelman (2006).

Notice that the asymptotic distribution depends on c and also on the asymptotic eigenvalue distributions of the matrices A (and B in the more general case) corresponding to the population values of the cross-sectional and time-series correlations. The resulting asymptotic distribution function can be derived implicitly for certain types of matrices A in terms of its Cauchy transformation, but requires numerical methods to evaluate (Rao and Edelman, 2006). Therefore, we focus our attention on a set of linear spectral statistics corresponding to the moments

of the eigenvalue distribution, which have more convenient properties. For an arbitrary covariance matrix C we define

$$m_C(\lambda) = \int g(\lambda)dF^C(\lambda) = \frac{1}{N}\sum_{j=1}^{N}f(\lambda_j), \qquad (3.8)$$

where $\lambda_j \in \mathrm{Sp}(C)$. We are especially interested in the monomials, $g(\lambda) \in \{\lambda, \lambda^2, ..., \lambda^s\}$ and denote the corresponding moments by $\{m_C^1, m_C^2, ..., m_C^3\}$. Notice that these monomials define the raw moments of the eigenvalue distribution F^C.

If C is some empirical covariance matrix C_N, then

$$m_{C_N}^s(\lambda) = N^{-1}\,\mathrm{tr}[(C_N)^s]. \qquad (3.9)$$

Moreover, standard results on bounded moment convergence and continuous mapping (e.g. Billingsley, 1995) imply that if $C_N \to C$ and $F^{C_N} \to F^C$, a proper probability distribution with bounded support, then:

$$\lim_{N\to\infty} m_{C_N}^s(\lambda) = m_C^s(\lambda) = \int \lambda^s dF^C(\lambda) < \infty. \qquad (3.10)$$

In particular note that for the covariance model introduced in Assumption 3.2, the moments of the eigenvalue distribution of the error term in our factor model exist and are finite as a consequence of Proposition 3.1. The challenge consists of being able to compute the limiting moments of the eigenvalue distribution. Below we introduce a procedure based on free probability theory that relates the moments of the limiting eigenvalue distribution to the moments of the eigenvalue distribution of the cross-sectional and time-series correlation matrices.

Moreover, it can be shown that the moments of the eigenvalue distribution of a random covariance matrix C_N satisfy a Central Limit Theorem (Bai and Silverstein, 2004):

Proposition 3.2 (CLT): *Let $\overline{g}(w) = -(1-c)/w + cG_C(w)$. Then*

$$N^{-1}\begin{pmatrix} m_{C_N}^1 & - & m_C^1(\lambda) \\ & ... & \\ m_{C_N}^s & - & m_C^s(\lambda) \end{pmatrix} \sim N(\Delta, V), \qquad (3.11)$$

where for $j = 1 \ldots s$ and $k = 1 \ldots s$

$$\Delta_j = -\frac{1}{2\pi i}\int w^j \frac{c\int \overline{g}(w)^3 v^2(1+v\overline{g}(w))^{-3}dF^A}{\left(1 - c\int \overline{g}(w)^2 v^2(1+v\overline{g}(w))^{-2}F^A\right)^2}dw \qquad (3.12)$$

$$V_{jk} = -\frac{1}{2\pi^2} \int \int \frac{w_1^j w_2^k}{(\overline{g}(w_1) - \overline{g}(w_2))^2} \frac{d\overline{g}(w_1)}{dw_1} \frac{d\overline{g}(w_2)}{dw_2} dw_1 dw_2, \tag{3.13}$$

where the contours are assumed to be non-overlapping, closed, taken in the positive direction in the complex plane, each enclosing the support of F^C.

In general however it is not possible to compute these integrals over the complex plane analytically. For $A = I_N$ the answer is known and was derived by Jonsson (1982) using a combinatoric proof. In this case the expressions above reduce to:

$$\Delta_j = \frac{1}{4} \left(\left(1 - \sqrt{c}\right)^{2j} + \left(1 + \sqrt{c}\right)^{2j} \right) - \frac{1}{2} \sum_{r=0}^{j} \binom{j}{r}^2 c^r \tag{3.14}$$

$$(V)_{j,k} = 2c^{j+k} \sum_{r_1=0}^{j-1} \sum_{r_2=0}^{k} \binom{j}{r_1} \binom{k}{r_2} \left(\frac{1-c}{c}\right)^{j+k} \sum_{l=1}^{j-k} l \binom{2j-1-(r_1+l)}{j-1} \binom{2k-1-(r_2+l)}{k-1}$$
$$\tag{3.15}$$

Recall that our strategy for identifying the number of latent factors involves performing an eigenvalue decomposition of the covariance matrix between the observed time series, Σ_N. Moreover, we have assumed that the covariance matrix between the unobserved error terms is Ω_N, where the error terms where drawn from a matrix variate distribution with separable cross-sectional and time series correlation (Assumption 3.2). Note that by Proposition 3.1, the empirical eigenvalue distribution of Ω_N converges to some non-random proper distribution function $F^{\Omega}(\lambda; \theta, c)$ as $N \to \infty$, $T \to \infty$ and $N/T \to c$ where θ is the unknown vector of covariance parameters.

Let us assume for the moment that our data does not contain unobserved latent factors, that is $\Xi_N = 0$. In this case, Proposition 3.2 suggests a minimum distance procedure for estimating the vector of unknown covariance parameters θ using the moments of the empirical eigenvalue distribution. Let

$$\Pi(\hat{\Omega}_N) = \left[N^{-1} \operatorname{tr}(\Omega_N^1), N^{-1} \operatorname{tr}(\Omega_N^2), ..., N^{-1} \operatorname{tr}(\Omega_N^s) \right]', \tag{3.16}$$

be the vector of the first s moments of the empirical eigenvalue distribution of Ω_N and denote by

$$\Pi(\theta) = \left[m_{\Omega}^1(\theta, c), m_{\Omega}^2(\theta, c), ..., m_{\Omega}^s(\theta, c) \right]', \tag{3.17}$$

the corresponding vector of limiting moments as $N \to \infty$, $T \to \infty$ and $N/T \to c$. In Section 3.1.2 we show how to derive expressions for these limiting moments analytically. In order to estimate the vector of unknown parameters, we could

apply the following minimum distance procedure:

$$\hat{\theta} = \text{argmin}_{\theta}\left(\Pi(\theta) - \Pi(\hat{\Omega}_N)\right)'\hat{V}^{-1}(\Pi(\theta) - \Pi(\hat{\Omega}_N)), \qquad (3.18)$$

where $(\hat{V})_{j,k}$ is a consistent estimate of equation 3.13.

Our focus however, is on estimating the rank of the matrix Ξ_N when rank($\Xi_N > 0$) subject to the constraint that we only observe $\Sigma_N = \Xi_N + \Omega_N$. This implies that we have to use the spectral decomposition of the covariance matrix Σ_N to estimate both the rank of Ξ_N and any additional covariance parameters θ that Ω_N depends on.

By the assumptions of our factor model, rank(Ξ_N) $= p \ll N$. Moreover, we know that the p eigenvalues capturing the effect of the latent factors F_t diverge to infinity as $N \to \infty$ while the N eigenvalues corresponding to the noise term U_t remain bounded. In large enough samples, this produces a separation of the spectrum of Σ_N into two parts, a first part with mass $(N - p)/N$ located to the left but bounded by zero from below and a second part with mass p/N to the right which diverges as $N \to \infty$ (Bai and Silverstein, 1998, 2004; Baik and Silverstein, 2006; Dozier and Silverstein, 2004). In particular note that the p eigenvalues do not "pull" the remaining $N-p$ eigenvalues to the right. Identifying the number of latent factors thus requires us to estimate the number of eigenvalues to the right of this spectral gap which separates the eigenvalues due to the noise term U_t from those due to the latent factors. In finite samples however this gap is not evident due to the presence of weak factors for which the corresponding eigenvalues are close to the upper bound of the distribution of eigenvalues due to the idiosyncratic terms. Thus, we require statistical techniques in order to separate the eigenvalues due to the factors from those due to the noise.

If we were to compare the asymptotic moment expressions of equation 3.17 for $\Pi(\theta)$, with the empirical moments of the empirical eigenvalue distribution of the observed covariance matrix Σ_N,

$$\Pi(\Sigma_N) = \left[N^{-1}\,\text{tr}(\Sigma_N^1), N^{-1}\,\text{tr}(\Sigma_N^2), ..., N^{-1}\,\text{tr}(\Sigma_N^s)\right]', \qquad (3.19)$$

we would find a poor match. The asymptotic moment expressions are correct for the unobserved covariance matrix Ω_N but not for the observed covariance matrix $\Sigma_N = \Xi_N + \Omega_N$. We can exploit this inconsistency in the moment conditions of the eigenvalue distribution which occurs in the presence of latent factors to specify a downward testing moment selection procedure (Andrews, 1999).

Let Sp(Σ_N) denote the spectrum of the covariance matrix of the observations, Σ_N, where we have ordered the eigenvalues in decreasing order. That is, Sp(Σ_N) $= \{\lambda_1, \lambda_2, ..., \lambda_N\}$, with λ_1 being the largest eigenvalue and including multiplicities.

Note that expression 3.19 can be re-written as:

$$\Pi(\Sigma_N) = \left[N^{-1} \sum_{\lambda_j \in \mathrm{Sp}(\Sigma_N)} \lambda_j^1, N^{-1} \sum_{\lambda_j \in \mathrm{Sp}(\Sigma_N)} \lambda_j^2, ..., N^{-1} \sum_{\lambda_j \in \mathrm{Sp}(\Sigma_N)} \lambda_j^s \right]'. \quad (3.20)$$

Now let $\mathrm{Sp}_p(\Sigma_N)$ be the truncated spectrum where we have removed the first p largest eigenvalues, $\mathrm{Sp}_p(\Sigma_N) = \mathrm{Sp}(\Sigma_N)\backslash\{\lambda_1, \lambda_2, ..., \lambda_p\}$. Let $\Pi_p(\Sigma_N)$ be the vector of the first s empirical moment conditions evaluated using the truncated spectrum $\mathrm{Sp}_p(\Sigma_N)$,

$$\Pi_p(\Sigma_N) = \left[N^{-1} \sum_{\lambda_j \in \mathrm{Sp}_p(\Sigma_N)} \lambda_j^1, N^{-1} \sum_{\lambda_j \in \mathrm{Sp}_p(\Sigma_N)} \lambda_j^2, ..., N^{-1} \sum_{\lambda_j \in \mathrm{Sp}_p(\Sigma_N)} \lambda_j^s \right]'. \quad (3.21)$$

Notice that if p is the true number of latent factors then the moment conditions $\Pi_p(\Sigma_N)$ will match with the asymptotic moments for the covariance of error terms $\Pi(\theta)$ from expression 3.20 above. By evaluating the moment conditions on the truncated spectrum, we have removed the effect of the latent factors and we expect the distance between $(\Pi_p(\Sigma_N) - \Pi(\theta))$ to be small if the correct number of factors p has been identified and large otherwise. This suggests a testing procedure based on the minimized objective function, commonly referred to as the J-test (Hansen, 1982). The number of unobserved factors, $\hat{p}$ is estimated by:

$$\hat{p} = \mathrm{argmin}_{p=0,1,2,...} \hat{J}(\hat{\theta}; \mathrm{Sp}_p(\Sigma_N)), \quad (3.22)$$

where the vector of unknown covariance parameters is estimated using the moment conditions computed from the truncated spectrum $\mathrm{Sp}_p(\Sigma_N)$. The procedure is applied recursively by truncating the spectrum of the observed covariance matrix Σ_N from the right and re-estimating the vector of parameters θ until the corresponding J-test is minimized. In Section 3.1.3 we show that the J-test is minimized after the spectrum of the observed covariance matrix Σ_N has been truncated by the true number of factors, thereby estimating $\hat{p}$ consistently.

Notice that the the true number of factors is only revealed asymptotically as $N \to \infty$. Identifying the true number of pervasive factors requires the sample to be large enough such that the spectrum separates between a set of $N-p$ eigenvalues corresponding to the error terms U_t and a set of p eigenvalues corresponding to the factors F_t. Unfortunately, in small samples the eigenvalues corresponding to the p factors may not always separate and exhibit a phase transition phenomenon where eigenvalues corresponding to weak factors do not detach from the spectrum of the error terms and converge in probability to the upper bound of the spectrum of Ω_N rather than to their true asymptotic limits which diverge with N. Chapter 2 shows how this leads to a single factor bias in estimated arbitrage

pricing models commonly used in finance. Note that this is not a feature of the estimation procedure but rather of the data; in small samples, some weak factors are obfuscated by the error terms and cannot be identified. Our procedure, however, does guarantee that if a factor is strong enough to overcome the phase transition phenomenon and emerge from the shadow of the error terms, it will be picked up by our algorithm and will be correctly identified as a latent factor. This allows us to estimate all the weak factors which can be identified from the data in addition to the strong factors which have traditionally been estimated.

The identification procedure described above can be applied to a broader set of latent structures than linear factor models. Harding (2006) extends this procedure to identify latent structures in large networks by exploiting a similar eigenvalue separation process that occurs for the spectrum of the adjacency matrix of a large random graph with random and structural components.

■ 3.1.2 Free Probability Derivation of Moments

In order to implement the identification strategy outlined above, we need to compute the limiting moments of the eigenvalue distribution of the covariance of the error terms U_t as a function of the covariance model parameters θ,

$$m_\Omega^s = \lim_{N \to \infty} (1/N) \, E\left\{\text{tr}(\Omega_N^s)\right\}. \tag{3.23}$$

Ignoring time series correlations for the moment, one of the implications of Assumption 3.2 is that we can write $\Omega_N = (1/T)UU' = (1/T)A_N^{1/2}\epsilon\epsilon'A_N^{1/2}$, where $U = A_N^{1/2}\epsilon$ and $(\epsilon)_{i,j}$ is iid mean zero with finite fourth order moments. Moreover, we assume that A_N and ϵ are independent. Denote the covariance of the iid terms $(\epsilon)_{i,j}$ by $\Psi_N = (1/T)\epsilon\epsilon'$ and notice that:

$$\begin{aligned} m_\Omega^s &= \lim_{N \to \infty} (1/N) \, E\left\{\text{tr}\left(\left[(1/T)\,A_N^{1/2}\epsilon\epsilon'A_N^{1/2}\right]^s\right)\right\} \\ &= \lim_{N \to \infty} (1/N) \, E\left\{\text{tr}((1/T)\left[A_N\Psi_N\right]^s)\right\}. \end{aligned} \tag{3.24}$$

The focus of this section is to introduce a procedure to analytically derive the large N limiting eigenvalue distribution moments m_Ω^s based on our knowledge of the limiting eigenvalue distribution of A_N and Ψ_N. Note, however, that even though by assumption A_N and Ψ_N have limiting eigenvalue distributions with bounded support and A_N is independent of Ψ_N this is not sufficient to guarantee that the eigenvalue distribution of $(A_N\Psi_N)^s$ will depend only on the underlying limiting distributions. The free probability approach developed below will provide additional conditions under which it is possible to relate the moments of the limiting eigenvalue distribution of mixed moments of products of random matrices to the limiting moments of the eigenvalues of their constituent matrices.

The computation of expressions such as those of the moments m_Ω^s can be

prohibitive analytically if we start from the individual elements of the random matrix due to the combinatoric complexity of the resulting traces of powers of mixed products between matrices. Instead we prefer to think of the whole random matrix as a random variable on a non-commutative probability space defined below.

Consider a probability space Θ and the Hilbert space $L^\infty(\Theta)$ of bounded measurable functions h defined on Θ, corresponding to the random variables on Θ. The functions h are allowed to be complex valued but for the purpose of estimating a factor model we can restrict our attention to real random variables on Θ. Furthermore, there exists a probability law $\mathfrak{P}$ on h which measures the probability that the value of h lies in a certain sub-interval in the (real or complex) image of h. The space of bounded linear functionals ϕ on the Hilbert space $L^\infty(\Theta)$ given by $\phi(h) = \int_\Theta h d\mathfrak{P}$ defines a von Neumann algebra, $\mathcal{A}$, on Θ.

Definition 3.1: *A non-commutative probability space is a pair $(\mathcal{A}, \phi)$, where $\mathcal{A}$ is an algebra endowed with a unit (1) and ϕ a linear functional on $\mathcal{A}$ such that $\phi(1) = 1$.*

Note that classical probability spaces also satisfy the above definition but that we are interested in relaxing the commutativeness assumption imposed by classical probability on scalar random variables. Furthermore, it is convenient to also assume that $\mathcal{A}$ is a von Neumann algebra as discussed above.

For the non-commutative probability space $(\mathcal{A}, \phi)$ and a random variable $X \in \mathcal{A}$, we can define the s-th moment of X as $m_X^s = \phi(X^s)$. Computing expectations over random variables in classical probability is often simplified when we can assume independence between the random variables. We now extend the notion of independence in classical probability by employing the concept of freeness from operator algebras (Voiculescu, 1985; Speicher, 2005).

Definition 3.2 (Freeness): *Let $(\mathcal{A}, \phi)$ be the non-commutative probability space of Definition 3.1 and $\{\mathcal{A}_1, \mathcal{A}_2, ..., \mathcal{A}_J\} \subset \mathcal{A}$ subalgebras of $\mathcal{A}$ with the a unit (1). Then the algebras $\mathcal{A}_1, \mathcal{A}_2, ..., \mathcal{A}_J$ are free with respect to ϕ if*

$$\phi(a_1 a_2 ... a_K) = 0 \tag{3.25}$$

if $a_1 \in \mathcal{A}_{j(1)}$, $a_2 \in \mathcal{A}_{j(2)}$, ... , $a_K \in \mathcal{A}_{j(K)}$, where $j(k)$ is an index function on the set $\{1, 2, ..., J\}$ and $j(k) \neq j(k+1)$ for all $k = 1..(K-1)$, and $\phi(a_k) = 0$ for all $k = 1 \ldots K$.

By extension, the random variables $X_1, X_2, ..., X_K \in \mathcal{A}$ are freely independent if the subalgebras generated by them are free with respect to ϕ. Thus, the operator concept of freeness is a particular generalization of the classical probability concept of independence, where freeness with respect to ϕ corresponds to independence of σ-algebras, and free independence of random variables corresponds to the classical notion of independence of random variables. Note, however, that this particular extension of independence to non-commutative probability spaces

is but one of the possible non-commutative extensions with convenient properties for the analysis of covariance matrices.

Freeness is a convenient property of random variables since it amounts to an iterative procedure for computing mixed moments of products of random variables from the moments of the constituent random variables. Notice that we can re-write equation 3.25 for the case when $\phi(a_k) \neq 0$ by subtracting the individual means:

$$\phi((a_1 - \phi(a_1)1)(a_2 - \phi(a_2)1)\ldots(a_k - \phi(a_k)1)) = 0. \tag{3.26}$$

Our primary focus is on the computation of mixed moments $m_{ab}^s = \phi((ab)^s)$. If a and b are free then,

$$\phi((a - \phi(a)1)(b - \phi(b)1)) = 0, \tag{3.27}$$

and expanding,

$$\phi(ab - \phi(a)b - a\phi(b) + \phi(a)\phi(b)) = 0, \tag{3.28}$$

$$\phi(ab) - \phi(a)\phi(b) - \phi(a)\phi(b) + \phi(a)\phi(b)) = 0, \tag{3.29}$$

$$m_{ab}^1 = \phi(ab) = \phi(a)\phi(b). \tag{3.30}$$

Notice that this expression is the same as the one we would obtain if a and b were independent random variables in a classical probability space. Now consider, $m_{ab}^2 = \phi((ab)^2)$. Since a and b are non-commutative, $m_{ab}^s = \phi(abab)$, we start by expanding the expression:

$$\phi((a - \phi(a)1)(b - \phi(b)1)(a - \phi(a)1)(b - \phi(b)1)) = 0. \tag{3.31}$$

In the Appendix we show that this leads to the following expression for the mixed second moment in a and b:

$$\phi(abab) = \phi^2(a)\phi(bb) + \phi^2(b)\phi(aa) - \phi^2(a)\phi^2(b). \tag{3.32}$$

This expression however does not reduce to the same expression one would obtain if a and b were independent commutative random variables in a classical probability space, since

$$\phi(abab) \neq \phi(a^2b^2) = \phi(a^2)\phi(b^2). \tag{3.33}$$

The definition of free independence can thus be applied recursively to obtain the mixed higher order moments of ab and other similar products.

Consider the space of $N \times N$ real matrices $\mathcal{M}_N(\mathbb{R})$ and X a random matrix on this space whose elements $(X)_{i,j}$ are random variables on a classical probability

space $(\Theta, \mathfrak{P})$. Define the algebra of functions

$$\mathcal{A}_N = \bigcap_{1 \leq s < \infty} L^s(X, \mathcal{M}_N) \qquad (3.34)$$

for the s-integrable random matrices of dimension $N \times N$ for $1 \leq s < \infty$. Note that this implies that all elements $(X)_{i,j}$ have finite moments since $(X)_{i,j} \in \mathcal{A}_N$. Furthermore, let $\phi_N : \mathcal{A} \to \mathbb{R}$ be an operator defined as:

$$\phi_N(Y) = \frac{1}{N} E_X \operatorname{tr}(Y) = \frac{1}{N} \sum_{j=1}^{N} E(Y_{j,j}) = \frac{1}{N} \int_X \operatorname{tr}(Y) d\mathfrak{P}. \qquad (3.35)$$

Corollary 3.1: $(\mathcal{A}_N, \phi_N)$ *is a non-commutative probability space.*

This result follows immediately as a particular instance of Definition 3.1. It implies that we can think of random matrices both in terms of the usual commutative probability spaces on which each element of the matrix is defined but also in terms of the whole matrix as a random variable defined on a non-commutative probability space. In particular given the connection between the trace and the eigenvalues of the matrix, it turns out to be more convenient to think of the covariance matrices in our factor model in terms of the non-commutative probability space.

Recall that our interest in using non-commutative probability is mainly due to the necessity of computing moments of the type $\phi_N[(A_N \Psi_N)^s]$. In order to employ the moment expansion procedure above, we would first need to show that (A_N, Ψ_N) are freely independent. In general, however, two arbitrary matrices are not freely independent since their eigenspaces may satisfy a particular relationship to each other, even if the elements of the matrices are independent.

One of the main insights of random matrix theory is that certain matrices become freely independent asymptotically as $N \to \infty$ (Voiculescu, 1998). Note that asymptotic freeness for large random matrices Y requires both the convergence of the probability law $\mathfrak{P}$ as $N \to \infty$ and Definition 3.2 to be satisfied for $\phi_N(Y) = \lim_{N \to \infty} \frac{1}{N} E_X \operatorname{tr}(Y)$. The convergence of the probability law implies the convergence of all moments of the eigenvalue distribution in the large N limit.

Consider the set of matrices S distributed uniformly on the Stiefel manifold (Anderson, 2003, Definition 4.5.1).

Definition 3.3: *Let S_N be an $N \times N$ matrix satisfying $S'_N S_N = I_N$ and $S_N H_N \overset{d}{=} S_N$ for all orthogonal matrices H_N. Then S_N is uniformly distributed on the group of square orthogonal matrices $\mathcal{O}(N)$.*

Let μ be the probability measure on the random matrices S_N in Definition 3.3. Then μ is the unique probability measure on $\mathcal{O}(N)$ such that for some $D \subset \mathcal{O}(N)$, $\mu(\Gamma D) = \mu(D\Gamma) = \mu(D)$ for al $\Gamma \in \mathcal{O}(N)$. The distribution μ is referred to as the Haar (invariant) distribution on $\mathcal{O}(N)$.

Proposition 3.3: *Let S_N be an $N \times N$ matrix with the Haar distribution and X_N and Y_N two sequences of random $N \times N$ symmetric matrices such that their empirical eigenvalue distributions converge to proper non-random distributions with bounded support. If S_N is independent of X_N and Y_N then X_N and $S_N Y_N S'_N$ are asymptotically free as $N \rightarrow \infty$.*

Proof: See Speicher (2005).

Note that by the spectral decomposition of the the matrix Y_N, the effect of the Haar distributed random matrix S_N is to introduce a random rotation in the eigenvectors of Y_N. Asymptotic freeness requires us to identify matrices that are rotationally invariant and thus preserve the information on the eigenvalue distribution independently of the eigenvectors which are now randomly rotated. Hence we are particularly interested in matrices Y_N which are unitarily invariant, that is the spectrum of Y_N and that of $S_N Y_N S'_N$ is the same for S_N on the orthogonal group. More formally, we can use the following lemma (Anderson, 2003, Lemma 13.3.2) for normalized matrices.

Lemma 3.1: *Let C_N be an arbitrary matrix of order N and define the following normalization:*

$$J(C_N) = \mathrm{diag}\left\{ (C_N)_{1,1} \Big/ |(C_N)_{1,1}|, (C_N)_{2,2} \Big/ |(C_N)_{2,2}|, ..., (C_N)_{N,N} \Big/ |(C_N)_{N,N}| \right\}.$$
(3.36)

If the orthogonal matrix S_N of order N has a distribution such that $(S_N)_{i,1} \geq 0$ and if $S^*_N = J(S_N H_N) S_N H_N$ has the same distribution for every orthogonal matrix H_N, then S_N has the conditional Haar invariant distribution.

Proof: See Anderson (2003, pp. 542).

The conditional Haar invariant distribution is the conditional distribution of a normalized orthogonal matrix S_N with the Haar distribution, where we let the $(S_N)_{i,1} \geq 0$. It is equal to 2^N times the Haar distribution.

If Y_N is a covariance matrix, unitary invariance requires that the normalized eigenvectors W_N from the spectral decomposition of $Y_N = W_N D_N W'_N$ be distributed conditionally Haar and independent of D_N, the diagonal matrix of eigenvalues. This ensures that further rotations by Haar distributed orthogonal matrices S_N do not change the eigenvalue distribution. Covariance matrices satisfying this requirement include those derived from matrices with iid Normal elements, $Y_N = (1/T)\epsilon\epsilon'$, where $(\epsilon)_{i,j}$ are distributed iid N(0,1). This is captured by the following result:

Proposition 3.4: *Let $W_N = (w_1, w_2, ..., w_N)'$ be the matrix of normalized eigenvectors of a covariance matrix Y_N, where $(w)_{1,i} \geq 0$ and where Y_N is distributed according to a Wishart distribution with mean I_N, then W_N has the conditional Haar invariant distribution and W_N is distributed independently of the eigenvalues of Y_N.*

Proof: See Anderson (2003, Theorem 13.3.3).

Returning to the moment expressions 2.23 and 2.24, we see that A_N and Ψ_N

are asymptotically free if Ψ_N is unitarily invariant. An important special case is given by Proposition 3.4 where Ψ_N is drawn from a standard Wishart distribution, that is $\Psi_N = (1/T)\epsilon\epsilon'$, where $(\epsilon)_{i,j}$ are distributed iid N(0,1) . This implies that we can apply Definition 3.2 and compute mixed moments of $(A_N\Psi_N)^s$ using the recursive procedure outlined above. Since the sequence of matrices A_N is given by our parametric model (or estimated by some alternative procedure), it has known moments $m_{A_N}^s$ for all N as $N \to \infty$. The moments of Ψ_N are known to converge to the moments of the Marcenko-Pastur distribution under a general set of assumptions.

Proposition 3.5: *Let ϵ be an $N \times T$ random matrix with elements which are iid with mean 0, variance 1 and finite fourth order moments. Then as $N \to \infty$, $T \to \infty$ and $N/T \to c$, the empirical eigenvalue distribution of $\Psi_N = (1/T)\epsilon\epsilon'$ converges almost surely to the non-random Marcenko-Pastur distribution whose moments are given by:*

$$m_\Psi^s = \lim_{N \to \infty} \frac{1}{N} E\mathrm{tr}\left\{ [(1/T)\,\epsilon\epsilon']^s \right\} = \sum_{r=1}^s \frac{1}{s}\binom{s}{r}\binom{s}{r-1}c^{s-1}. \tag{3.37}$$

Proof: Jonsson (1982). Note that the moments of $\Psi_N = (1/T)\epsilon\epsilon'$ are given by the Narayana polynomials in $c = N/T$. The first few moments are:

$$m_\Psi^1 = 1 \tag{3.38}$$

$$m_\Psi^2 = 1 + c \tag{3.39}$$

$$m_\Psi^3 = 1 + 3c + c^2 \tag{3.40}$$

$$m_\Psi^4 = 1 + 6c + 6c^2 + c^3 \tag{3.41}$$

$$m_\Psi^5 = 1 + 10c + 20c^2 + 10c^3 + c^4 \tag{3.42}$$

$$m_\Psi^6 = 1 + 15c + 50c^2 + 50c^3 + 15c^4 + c^5. \tag{3.43}$$

We can now use the moments given by equation 3.37 and the expressions in equation 3.30 and 3.32 to compute the mixed moments of $A_N\Psi_N$ in the large N limit using the property of free independence between A_N and Ψ_N. From equation 2.30 we know that $m^1(A\Psi) = m_A^1 m_\Psi^1$. But since $m_\Psi^1 = 1$, we have

$$m^1(A\Psi) = m_A^1. \tag{3.44}$$

Similarly by equation 3.32 we know that

$$m^2(A\Psi) = \left(m_A^1\right)^2 m_\Psi^2 + \left(m_\Psi^1\right)^2 m_A^2 - \left(m_A^1\right)^2 \left(m_\Psi^1\right)^2. \tag{3.45}$$

Substituting $m_\Psi^2 = 1 + c$ and $m_\Psi^1 = 1$ in the expression above we obtain:

$$m^2(A\Psi) = \left(m_A^1\right)^2 (1 + c) + m_A^2 - \left(m_A^1\right)^2 \tag{3.46}$$

$$m^2(A\Psi) = m_A^2 + c(m_A^1)^2. \tag{3.47}$$

We can continue this process to obtain:

$$m^3(A\Psi) = m_A^3 + 3cm_A^1 m_A^2 + c^2(m_A^1)^3 \tag{3.48}$$

$$m^4(A\Psi) = m_A^4 + 2c\left(\left(m_A^2\right)^2 + 2m_A^1 m_A^3\right) + 6c^2(m_A^1)^2 m_A^2 + c^3((m_A^1)^4). \tag{3.49}$$

Thus, we have shown how to compute the moments of the eigenvalue distribution of the covariance of the error terms U_t, $\Omega_N = (1/T)UU' = (1/T)A_N^{1/2}\epsilon\epsilon'A_N^{1/2}$ in terms of the moments of the eigenvalue distribution of A_N in the large N limit. Since these moments will be functions of the unknown parameters θ, the expressions derived above will also be functions of θ once we substitute a precise covariance model for A_N. To illustrate, consider the model with $A_N = \sigma I_N$. We have only one unknown parameter, the variance scale coefficient σ, $\theta = \{\sigma\}$. Substituting the moments of A in equations 3.44, 3.47, 3.48 and 3.49, we obtain the first four spectral moments of the white noise covariance matrix to be:

$$m_\Omega^1 = \sigma \tag{3.50}$$

$$m_\Omega^2 = (1 + c)\sigma^2 \tag{3.51}$$

$$m_\Omega^3 = \left(c^2 + 3c + 1\right)\sigma^3 \tag{3.52}$$

$$m_\Omega^4 = (1 + c)\left(c^2 + 5c + 1\right)\sigma^4. \tag{3.53}$$

The free probability framework introduced above allows us to compute the moments of the empirical eigenvalue distribution of the noise covariance matrix in terms of the population covariance matrix assumed by our factor model by a number of algebraic operations on free moments. While these computations are relatively straightforward and only involve basic algebra, higher order moments may involve a substantial number of terms and thus it may be more convenient to use a mathematical software package such as Maple or Mathematica to derive the moment expressions (Rao and Edelman, 2006).

It is also possible to derive the moment expressions using the Cauchy transform

defined in equation 3.6. In the Appendix we show that the moment expressions derived above are also more generally given by the following implicit relationship:

Proposition 3.6: *Let m_Ω^s be the limiting moments of the empirical noise covariance Ω and m_A^s be the limiting moments of the correlation matrix A. Then for $w \in \mathbb{C}^+$ with $\mathfrak{Im}(w) > 0$ we have that*

$$\sum_{s=1}^{\infty} \frac{m_\Omega^s}{w^s} = \sum_{s=1}^{\infty} \frac{m_A^s}{w^s} \left(1 + c \sum_{r=1}^{\infty} \frac{m_\Omega^r}{w^r} \right)^s. \tag{3.54}$$

The relationship between the first four moments is given by:

$$m_\Omega^1 = m_A^1 \tag{3.55}$$

$$m_\Omega^2 = m_A^2 + c(m_A^1)^2 \tag{3.56}$$

$$m_\Omega^3 = m_A^3 + 3cm_A^1 m_A^2 + c^2(m_A^1)^3 \tag{3.57}$$

$$m_\Omega^4 = m_A^4 + 2c\left((m_A^2)^2 + 2m_A^1 m_A^3 \right) + 6c^2(m_A^1)^2 m_A^2 + c^3((m_A^1)^4). \tag{3.58}$$

Proof: See Appendix. In order to derive the expressions for the moments in equations 3.44, 3.47, 3.48 and 3.49 we can expand this expression in $1/w$ and match the coefficients on $1/(w^s)$.

In the next section we revisit our estimator for the number of factors based on the identification strategy outlined in Section 3.1.1 and the procedure for deriving the limiting moments of the eigenvalue distribution as described in this section and analyze its statistical properties.

■ 3.1.3 Implementation and Finite Sample Performance

Recall the basic model setup of our factor model, $R_t = \Lambda F_t + U_t$, for $t = 1 \dots T$ where R_t is an $N \times 1$ vector of observations, F_t is a $p \times 1$ vector of latent factors, Λ is an $N \times p$ matrix of coefficients (factor loadings) and U_t is an $N \times 1$ vector of idiosyncratic errors. The identification strategy outlined above implies a computational procedure that leads to the consistent estimation of the number of factors p in the factor model with $N \to \infty$, $T \to \infty$ and $N/T \to c$. The advantage of this procedure consists is that it does not require the estimation of the unknown factor loadings Λ and factor scores F_t first and is therefore unaffected by complications resulting from the estimation of weak factor scores, which will be discussed in Section 3.2.

We can now summarize the steps required for the implementation of our estimator for the number of factors. For simplicity we continue to assume that

$B_T = I_T$. This assumption will be relaxed in the next section.

First, we choose a parametric model for the idiosyncratic error terms U_t in terms of the population covariance matrix $A_N(\theta)$, where the correlations are described in terms of the low dimensional parameter vector θ. Second, we compute the moments of the eigenvalue distribution of $A_N(\theta)$ for large N, $\{m_A^1, m_A^2, m_A^3, ...\}$. Third, we apply the free probability result of Proposition 3.6 to compute the moments of the asymptotic eigenvalue distribution of the covariance matrix Ω of U_t for a large (N, T) sample drawn from a distribution with covariance matrix A_N. We label these moments as $\Pi(\theta) = [m_\Omega^1, m_\Omega^2, m_\Omega^3, ...]'$. Fourth, we use a minimum distance approach to estimate the unknown covariance parameters θ by minimizing a weighted distance between $\Pi(\theta)$ and its sample equivalent $\Pi(\Sigma_N) = [N^{-1} \operatorname{tr}(\Sigma_N^1), N^{-1} \operatorname{tr}(\Sigma_N^2), ...]$ applied to the covariance matrix of observations R_t denoted by Σ_N. Fifth, we remove the largest eigenvalue of the spectrum of Σ_N and re-estimate the parameters θ using the minimum distance procedure. We repeat step 5 by progressively removing large eigenvalues until an (arbitrary) upper bound on the number of factors has been reached. Sixth, we compare the minimized objective functions, $\hat{J}(\hat{\theta})$ obtained by removing large eigenvalues and choose the one which is smallest within the set of minimized objective functions. The number of eigenvalues which had been removed for the computation of that objective function is our consistent estimate of the number of factors.

Proposition 3.7: *Let Σ_N be the covariance matrix of observations R_t in a large N, T factor model with $N \to \infty$, $T \to \infty$ and $N/T \to c$. Let $\operatorname{Sp}_p(\Sigma_N)$ be the spectrum of the matrix Σ_N where we removed the largest p eigenvalues and $\hat{J}(\hat{\theta}; \operatorname{Sp}_p(\Sigma_N))$ the (scaled) minimized objective function of the minimum distance procedure for the estimation of θ outlined above. Then, for*

$$\hat{p} = \operatorname{argmin}_{p=0,1,2,...} \hat{J}(\hat{\theta}; \operatorname{Sp}_p(\Sigma_N)), \tag{3.59}$$

is a consistent estimate of the number of factors p_0 of the factor model $R_t = \Lambda F_t + U_t$.

Proof: See Appendix.

In order to implement the estimation procedure described above we compute the minimum distance estimates of θ using the Nelder-Mead algorithm available in most common software packages such as Matlab or Gauss. The optimal weighting matrix for the moment conditions Π is difficult to compute analytically for the approximate factor model with arbitrary A_N. By Proposition 3.2, however, we can use the bootstrap or the jackknife to estimate the weighting matrix, since the moments are asymptotically Normal .

For the strict factor model with iid errors, the optimal weighting matrix is easy to implement and we can use an efficient two-step procedure which uses the estimated $\hat{\theta}$ from a first step estimation that employs equal weighting of the moment conditions. We label the one-step minimum distance procedure MD and

the two-step weighted estimator MDW.

In simulations we have also found that the performance of our estimator can be improved by adding a panel information criterion which penalizes the objective function in equation 3.84 if the selected number of factors is too large. The intuition is that in some cases the difference between the estimated $\hat{J}$ at p_0 and at $p_0 + 1$ may be very small. In such cases it is beneficial to augment equation 3.84 with a penalty function of the form $p\hat{\sigma}^2 g(N, T)$, where p is the number of excluded eigenvalues, $\hat{\sigma}^2$ is the estimated (average) variance at step p and $g(N, T)$ a function such that $g(N, T) \to 0$ in large samples. In simulations we have found the following choice due to Bai and Ng (2002) to perform very well:

$$g(N, T) = \left(\frac{N + T}{NT} \right) \log \left(\frac{NT}{N + T} \right). \tag{3.60}$$

We can augment the estimators MD and MDW defined above with the additional penalty function $p\hat{\sigma}^2 g(N, T)$ to obtain two alternative estimators which we label MD-IC and MDW-IC.

In Figure 3.1 we plot the objective function given by equation 2.59 for a particular simulation of the exact factor model with 5 factors using the design given in the Appendix. The objective function is minimized for all four choices of estimators of the number of factors (MD, MDW, MD-IC, MDW-IC) at the correct number of factors. In Table 3.1 and 3.2 we explore the finite sample properties of our estimators for different choices of N and T such that $c \in \{0.3, 0.5, 0.7, 0.9\}$. We use two simulation designs, one with strong factors and the other one with weak factors and a strict factor model. We report the mean number of chosen factors over 5000 simulations. While both the use of optimal weighting and of the panel information criterion improve the performance of the estimator our estimators appear to work well in all cases. Furthermore, we can estimate the unknown variance parameter $\theta = \sigma^2$ accurately and with low MSE without having to estimate the unobserved factors first. A particular advantage of our approach is that it works very well irrespective of whether the factors are weak or strong. This makes it especially useful when trying to estimate the weak factors in a model and not just the few strong ones.

We are currently expanding this approach to the estimation of the number of factors in order to construct confidence intervals around the estimated value of $\hat{(p)}$. Proposition 3.2 implies that the minimized objective function J is distributed as a non-central chi-square random variable. This fact can be employed to construct confidence intervals and will be explored in future research.

■ 3.1.4 Time Series Correlations

In some applications we may wish to allow for weak time series correlations of the idiosyncratic errors. If the true model is such that the idiosyncratic errors

are correlated over time but we wrongly assume that they are independent over time, the number of factors that is estimated using a misspecified model will be biased. We investigate this further in Table 3.3 using the simulation design given in the Appendix with five strong factors and autocorrelated idiosyncratic errors. We use the moments of the eigenvalue distribution for the misspecified model to construct the moment conditions and employ the minimum distance procedure described above.

We notice that the number of estimated factors is upward-biased. This is due to the fact that if the true model has autocorrelated idiosyncratic errors the resulting eigenvalue distribution will have a larger spectral radius than the distribution for uncorrelated idiosyncratic errors. This leads to eigenvalues in the right tail of the distribution of eigenvalues due to the noise in the factor model to be falsely categorized as factors. The results in Table 3.3 are for a relatively weak degree of autocorrelation (0.1). We have found that, for the case where the true model has autocorrelations in excess of 0.3 and where the misspecified model is estimated, our estimator will fail to converge. Table 3.3 also shows that if the degree of misspecification is small our estimators will have fairly small bias. This suggests that our approach is robust to minor deviations from the assumed parametric model but will fail if the degree of misspecification is large. If the true model is one where the idiosyncratic errors are autocorrelated, we can construct the correct estimator if we impose the correct assumptions on the parametric form of the autocorrelations.

Recall that by Assumption 3.2 the case of time-correlated idiosyncratic errors implies a model where $B_T \neq I_T$ such that the spectral density of B_T converges to a non-random distribution with bounded spectrum. Consider, for example, a model where the idiosyncratic errors follow an AR(1) process $U_{j,t} = \rho \ U_{j,t-1} + \epsilon_{j,t}$, for $\epsilon_{j,t}$ white noise such that $E(\epsilon_{j,t}) = 0$ and $E(\epsilon_{j,t}^2) = \sigma^2$. Recall that $E(U_{j,t}^2) = \sigma^2/(1 - \rho^2)$ and $E(U_{j,t}U_{j,t-k}) = (\sigma^2 \rho^k)/(1 - \rho^2)$. This implies a separable covariance model with $A_N = (\sigma^2/(1 - \rho^2))I_N$ and $(B_T)_{m,n} = \rho^{|m-n|}$. Thus the model for the time series correlations B_T corresponds to a Toeplitz matrix where the first (main) diagonal is 1, the second (upper and lower) diagonals are ρ, the third (upper and lower) diagonals are ρ^2 etc. Note that in asset return factor models such as APT the degree of autocorrelation would typically be small.

In order to guarantee that the spectrum of B_T is bounded we need to assume absolute summability, i.e.

$$\sum_{k=0}^{\infty} |\rho|^k = \frac{1}{1 - |\rho|} < \infty, \tag{3.61}$$

which implies $|\rho| < 1$. In order to compute the eigenvalue distribution of the matrix B_T as $T \to \infty$ we define the Fourier series $f(\zeta)$ such that

$$f(\zeta) = \lim_{T\to\infty} \sum_{k=-\infty}^{+\infty} \rho^{|k|} \exp(ik\zeta) = \frac{1 - \rho^2}{1 - 2\rho\cos(\zeta) + \rho^2}. \tag{3.62}$$

Let λ_k, $k = 1 \ldots T$ be the eigenvalues of the matrix B_T for $T \to \infty$. Then, by a classic theorem of Grenander and Szego (1958), we have that for any positive integer s

$$m_B^s = \lim_{T\to\infty} \frac{1}{T} \sum_{k=1}^{T} \lambda_k^s = \frac{1}{2\pi} \int_0^{2\pi} [f(\zeta)]^s d\zeta. \tag{3.63}$$

This expression gives the moments of the eigenvalue distribution of the population time-covariance matrix B_T for large T. The univariate integral above can easily be evaluated using numerical integration techniques. The above approach can be employed to compute the moments of the population spectrum F^B for other choices of B_T in the large T limit.

In order to apply the procedure outlined in the section above we have to compute the moments of the asymptotic eigenvalue distribution of the covariance matrix of observations for $N \to \infty$ and $T \to \infty$ and $N/T \to c$:

$$m_\Omega^s = \lim_{N\to\infty} (1/N) E \left\{ \text{tr}(\Omega_N^s) \right\}. \tag{3.64}$$

If we assume that cross-sectional correlations are given by a scale factor times I_N then we can write $\Omega_N = (1/T)UU' = (1/T)\epsilon B_T \epsilon'$, where $U = \epsilon B_T^{1/2}$ and $(\epsilon)_{i,j}$ is iid mean zero with finite fourth order moments. Then,

$$m_\Omega^s = \lim_{N\to\infty} (1/N) E \left\{ \text{tr}([(1/T)\,\epsilon B_T \epsilon']^s) \right\}. \tag{3.65}$$

Notice, however, that the non-zero eigenvalues of $(1/T)\epsilon B_T \epsilon'$ and $(1/T)\epsilon'\epsilon B_T$ are the same. Hence we can apply the free probability procedure presented in Section 3.1.2 to compute the mixed moments of $(1/T)\epsilon'\epsilon$ and B_T. For the case where B_T is a Toeplitz matrix corresponding to an AR(1) process, the moments of the eigenvalue distribution of B_T were given above. The relationship between the moments of the eigenvalue distribution of the covariance matrix of the observations and the moments of the eigenvalue distribution of B_T can be summarized by the following result:

Proposition 3.8: *Let m_Ω^s be the limiting moments of the empirical noise covariance Ω and m_B^s be the limiting moments of the correlation matrix B. Then for $w \in \mathbb{C}^+$ with $\mathfrak{Im}(w) > 0$ we have that*

$$\sum_{s=1}^{\infty} \frac{m_\Omega^s}{w^s} = \frac{1}{c} \left\{ \sum_{s=1}^{\infty} \frac{m_B^s}{w^s} \left[c \left(1 + \sum_{r=1}^{\infty} \frac{m_\Omega^r}{w^r} \right) \right]^s \right\}. \tag{3.66}$$

The relationship between the first four moments is given by:

$$m_\Omega^1 = m_B^1 \tag{3.67}$$

$$m_\Omega^2 = cm_B^2 + \left(m_B^1\right)^2 \tag{3.68}$$

$$m_\Omega^3 = c^2 m_B^3 + 3cm_B^2 m_B^1 + \left(m_B^1\right)^3 \tag{3.69}$$

$$m_\Omega^4 = c^3 m_B^4 + 4c^2(m_B^3 m_B^1 + \frac{1}{2}\left(m_B^2\right)^2) + 6cm_B^2(m_B^1)^2 + \left(m_B^1\right)^4. \tag{3.70}$$

Proof: See Appendix.

In Table 3.4 we illustrate the performance of our estimator by Monte-Carlo simulations for a model with 5 factors constructed according to the design given in the Appendix and with the additional requirement that the idiosyncratic errors are autocorrelated with coefficient $\rho = 0.3$. We apply two procedures to estimate the unknown parameters (p, σ^2, ρ), corresponding to the number of factors, the variance scale and the degree of autocorrelation respectively. The first procedure is the unweighted minimum distance method, while the second procedure augments the minimum distance objective function with the panel information criterion in order to estimate p. Both methods were described in the previous section. We notice that the estimator of p based on the minimum distance procedure augmented with the panel information criterion performs extremely well in choosing the correct number of factors. Similarly the unknown covariance parameters σ^2 and ρ are also estimated precisely with low MSE.

The expressions given in Propositions 6 and 8 above cover the cases where either B_T or A_N are known be the identity matrix up to a scaling factor. It is possible however to estimate models where both $B_T \neq I_T$ and $A_N \neq I_T$. In such cases the covariance matrix of the residuals is given by:

$$\Omega_N = \frac{1}{T}A_N^{1/2}\epsilon B_T \epsilon' A_N^{1/2}. \tag{3.71}$$

The corresponding moment conditions can be applied by first computing the moments of the eigenvalue distribution of $\Psi_N = \frac{1}{T}\epsilon B_T \epsilon'$ using Proposition 3.8 and then computing the moments of the eigenvalue distribution of the product $A_N \Psi_N$ by employing Proposition 3.6.

In some cases we may not be able to specify an exact parametric model for A_N or B_T. It may, however, be possible to derive estimates of A_N and B_T using a consistent method for estimating the residuals $U_{i,t}$ without estimating the number of factors or the factor loadings and factor scores. Such a procedure was recently suggested by Pesaran (2006) and involves augmenting the factor model

with observed factor proxies constructed from the cross-sectional averages of the model. In this situation it may be able to estimate the number of factors by a two-step procedure which first derives consistent estimates of A_N or B_T and then uses the estimated moments to extract the correct number of factors. Such an extension will be a subject of future research. Note that however it will not be possible to estimate both A_N and B_T from a single sample in general.

■ 3.2 Second Order Unbiased Estimation of the Factor Model

■ 3.2.1 Inconsistency of Principal Components for Weak Factors

Consider the classical factor model of equation 3.1 $R_t = \Lambda F_t + U_t$, for $t = 1 \ldots T$. Recall that R_t is an $N \times 1$ vector of observations, F_t is a $p \times 1$ vector of latent factors, Λ is an $N \times p$ matrix of coefficients (factor loadings) and U_t is an $N \times 1$ vector of idiosyncratic errors. In this model only R_t is observed while Λ, F_t and U_t are unobserved for all t. Estimation of the classical factor model requires estimation of the $N \times p$ matrix of factor loadings and predicting the $p \times T$ values of the latent factors F_t (factor scores). In this section we will restrict our attention to the exact factor model and assume that $U_{i,t}$ is iid Normal with mean 0 and variance σ^2.

Traditionally, for small values of N, factor models are estimated by maximum likelihood methods. High dimensional factor models however require the estimation of Np parameters which proves to be computationally infeasible if N is larger than 25. Thus, practitioners often employ Principal Components Analysis (PCA) applied to the covariance matrix of observations:

$$\Sigma_N = \frac{1}{T} \sum_{t=1}^{T} R_t R_t' = \Lambda \left(\frac{1}{T} \sum_{t=1}^{T} F_t F_t' \right) \Lambda' + \frac{1}{T} \sum_{t=1}^{T} U_t U_t' = \Xi_N + \Omega_N, \quad (3.72)$$

in order to estimate the factor loadings (Jolliffe, 2002). The equation above does not allow for the separate identification of both factor loadings and factor scores and we have to impose additional identifying restrictions. In particular it is common to assume that the factors F_t are orthogonal to each other and have unit variance. Moreover, since Ξ_N is invariant to orthogonal rotations S of the factor loadings, $\Xi_N = \Lambda \Lambda' = (\Lambda S')(\Lambda S')'$ for $S'S = I$, we require additional normalizations. Principal Components normalizes the Euclidean distance of the estimated factor loadings, $||\tilde{\Lambda}_j||_2 = 1$ for $j = 1 \ldots p$. The statistics literature often refers to this as "fixing the rotation" of the factors, a process that consists of reporting the normalized factor loadings $\tilde{\Lambda}_j = \hat{\Lambda}_j / ||\hat{\Lambda}_j||_2$ for $j = 1 \ldots p$ and estimated factors that are orthogonal to each other.

The PCA estimator $\tilde{\Lambda}$ is the $N \times p$ matrix of the first p (normalized) eigenvectors from the spectral decomposition $\Sigma_N = V D V'$, associated with the p largest diagonal elements of D, where the columns of V are orthonormal by construction.

The factor loadings are defined only up to a change in sign. Given estimates of the factor loadings, we can estimate factor scores by Generalized Least Squares regressions on the cross-section. Alternatively, the factor scores can be approximated by the first p normalized eigenvectors of the $T \times T$ matrix $(1/N)R'R$, where R is the $N \times T$ matrix with columns $R_1, R_2, ..., R_T$ (Connor and Korajczyk, 1986).

Define

$$\tilde{\mu} = \frac{\min\left(\mathrm{Sp}(\lim_{N\to\infty}(\Lambda\Lambda'))\right)}{\max(\mathrm{Sp}(\lim_{N\to\infty}(\Omega_N)))} \tag{3.73}$$

which corresponds to the ratio of the minimum eigenvalue of the spectrum due to the factors over the maximum eigenvalue due to the noise term. It can be thought of as a measure of the spectral gap discussed in Section 3.1.1. Thus, it also measures the "strength" of the factors. A strong factor corresponds to a factor for which $\tilde{\mu} \gg 0$, while a weak factor leads to a corresponding eigenvalue for which $\tilde{\mu}$ is close to zero. By Assumptions 3 and 4 we have $\tilde{\mu} > 0$ as $N \to \infty$, if the factors are identified. In finite samples, however, we expect to have both weak and strong factors.

Under the asymptotic framework of Assumption 3.1, $N \to \infty$ and $T \to \infty$ and $N/T \to c \in (0, \infty)$ it has recently been noticed that the estimated sample eigenvectors are inconsistent estimators of the corresponding population eigenvectors (Hoyle and Rattray, 2004; Paul, 2005; Onatsky, 2006). Let us assume that for each factor $j = 1 \ldots p$ the corresponding vector of true factor loadings Λ_j and its estimate $\hat{\Lambda}_j$ have been normalized such that $||\Lambda_j||_2 = ||\hat{\Lambda}_j||_2 = 1$. Denote by $\varnothing(x, y)$ the cosine of the angle between two arbitrary vectors x and y, where $\varnothing(x, y) = x'y/(||x||_2 \cdot ||y||_2)$. The proposition below states conditions under which consistency continues to hold for the classical factor model estimated by PCA even though the sample eigenvectors are inconsistent.

Proposition 3.9 (Consistency of PCA): *The degree of inconsistency in the estimates of the factor loadings Λ as $N \to \infty$, $T \to \infty$ and $N/T \to c \in (0, \infty)$ is given by*

$$\sqrt{\frac{1 - \frac{c}{\tilde{\mu}^2}}{1 + \frac{c}{\tilde{\mu}}}} \leq \varnothing(\hat{\Lambda}_j, \Lambda_j) \leq 1. \tag{3.74}$$

If $\tilde{\mu} \to \infty$ as $N \to \infty$ then the PCA estimate of Λ_j is consistent, i.e. $\varnothing(\hat{\Lambda}_j, \Lambda_j) \to 1$.

For random factor loadings Λ we can think of $\varnothing$ as a measure of the correlation between the two vectors. It is perhaps surprising that the PCA estimator of a factor model is consistent, but it is important to stress that it follows as a result of the more specialized assumptions imposed by an economic factor model and it does not hold true for an arbitrary application of PCA. Notice that in a factor model the inconsistency of the estimated factor loadings (and factor scores) depends on the ratio between c and $\tilde{\mu}$ only and, under the asymptotic framework of Assumption 3.1, c converges to a constant, while under Assumption 3.4 on

pervasiveness of economic factors, $\tilde{\mu}$ diverges as $N \to \infty$. Since the ratio tends to zero the inconsistency disappears.

In finite samples, however, it might be the case that the measure of the spectral gap, $\tilde{\mu}$, is close to zero for the weak factors. In such cases the estimation of the factor loadings may suffer substantial biases, presenting challenges for the estimation of weak economic factors. In order to investigate this effect, we use Monte Carlo to simulate a factor model with 5 weak factors using the simulation design described in the Appendix. We calibrate the simulations such that $\tilde{\mu} < 2$. In Table 3.5 we present the results for the estimated coefficients on the factor loadings using PCA. We notice that PCA performs very poorly in this case. It is interesting to note that, by construction, the model has both weak and strong factors. The PCA estimates, however, are poor for all factor loadings, not just the ones on the weak factors. In the section below we discuss alternative estimation procedures employing instrumental variables. Table 3.5 shows that, by contrast, an instrumental variables approach continues to provide satisfactory estimates even though PCA fails. This is due to the very different approach to estimation of the two methods and will be discussed in more detail below. Note that the results of Table 3.5 also seem to indicate that the PC estimator is likely not to have moments for the case of weak factors. This is similar to the problems encountered in the estimation of equations with weak instruments (Hahn, Kuersteiner and Hausman, 2004).

PCA does not only offer a poor approach to estimation in the presence of weak factors, it also suffers two more serious short-comings which are easily corrected by alternative instrumental variables based procedures. First, PCA estimation of the factor loadings does not allow us to impose economic restrictions on the estimated coefficients. Such restrictions are common in the macroeconomics literature and are particularly important in Factor Augmented VAR models (Stock and Watson, 2005). Even small departures from the standard framework, such as imposing exclusion restrictions on the factors in some equations, present major challenges. Restrictions severely limit the use of standard eigenvector techniques in the estimations.

Second, performing inference on the estimated factor loadings is very difficult due to the complicated distributions of eigenvectors (Bai, 2003; Paul, 2005; Onatski, 2006). In particular, the asymptotic distributions depend on the large eigenvalues but the eigenvalues themselves are only observed with bias in the sample (Paul, 2005; Onatski, 2006; Harding, Chapter 2).

■ 3.2.2 Estimation by Instrumental Variables

Factor analysis can also be thought of as a generalization of multivariate linear regression analysis with measurement error (Madansky, 1964). Although this fact has been recognized for a very long time, the application of instrumental variables (IV) procedures to the estimation of factor models is generally regarded

as inferior to estimation by PCA. In this section we show that high-dimensional factor models estimated by IV suffer from a finite sample bias problem similar to that encountered in the recent econometrics literature on estimation with many instruments (Hansen, Hausman and Newey, 2006). While this explains the practitioners' reluctance to apply IV methods in large datasets, it also provides a solution to the second order unbiased estimation by using IV estimators with improved finite sample performance.

Above we have seen how PCA imposes a normalization on the factor loadings $\tilde{\Lambda}_j = \hat{\Lambda}_j / ||\hat{\Lambda}_j||_2$. Other normalizations are also possible without loss of generality. In particular partition the matrix of factor coefficients as follows:

$$\Lambda = \begin{pmatrix} \hat{\Lambda}_1 \\ \hat{\Lambda}_2 \end{pmatrix}, \tag{3.75}$$

where $\hat{\Lambda}_1$ is a $p \times p$ submatrix of $\hat{\Lambda}$. We can now define the normalized factor loadings to be:

$$\tilde{\Lambda} = \begin{pmatrix} \hat{\Lambda}_1 \\ \hat{\Lambda}_2 \end{pmatrix} \hat{\Lambda}_1^{-1} = \begin{pmatrix} I_p \\ \hat{\Lambda}_2 \hat{\Lambda}_1^{-1} \end{pmatrix} = \begin{pmatrix} I_p \\ \tilde{\Lambda}_2 \end{pmatrix}. \tag{3.76}$$

Under this normalization we have, $R_{j,t} = F_{j,t} + u_{j,t}$, for $j = 1 \ldots p$. This means that, without loss of generality, we can choose the first p observations to act as proxies measured with error of the underlying latent factors. Now choose any observation $R_{p+k,t}$, for $k \geq 1$. Substituting the first p vectors of observations for the p factors in the equation for observation $p + k$ we obtain:

$$R_{p+k,t} = \sum_{j=1}^{p} \tilde{\Lambda}_{j,2}^k R_{j,t} + u_{p+k,t} - \sum_{j=1}^{p} \tilde{\Lambda}_{j,2}^k u_{j,t}, \tag{3.77}$$

where $\tilde{\Lambda}_{j,2}^k$ corresponds to the row k and column j entry of the normalized loadings matrix $\tilde{\Lambda}_2$.

If we assume an exact factor structure with $u_{i,t}$ iid then we have

$$E \left[R_{p+m,t} \left(R_{p+k,t} - \sum_{j=1}^{p} \tilde{\Lambda}_2^{p+k} R_{j,t} \right) \right] = 0, \tag{3.78}$$

for $m \geq 1$ and $m \neq k$. Therefore we can use all observations other than the first p observations and observation $p + k$ as instruments for the p factor proxies used in the equation for R_{p+k}.

Notice that for the exact factor model with $A_N = I_N$ and $B_T = I_T$ we have $N - p - 1$ moment conditions and the observations on the remaining $N - p - 1$ variables $(R_{p+1}, ..., R_{p+k-1}, R_{p+k+1}, ..., R_N)$ are valid instruments which can be used in order to estimate the coefficients $\tilde{\Lambda}_2^j$ for $j = 1..p$.

For other choices of A_N and B_T not all moment conditions will be available. The identification of equation 3.6 under such conditions will follow the usual procedures for the identification of equations with endogenous right hand side variables. This framework, however, is sufficiently flexible and allows us to complement the set of equations with additional linear restrictions or moment conditions in order to guarantee identification.

Notice that for the exact factor model, each $N - p$ equation $R_{p+k,t}$ can be written as in equation 3.6. Each equation is identified with a degree of over-identification equal to $N - 2p - 1$. Therefore we can estimate all the factor loadings $\tilde{\Lambda}_2$ using instrumental variables by repeatedly estimating each set of factor loadings using all other observations except for the first p observations as instruments.

For simplicity let us focus on a model with only one factor:

$$
\begin{aligned}
R_{1,t} &= f_t + u_{1,t} \\
R_{2,t} &= \lambda_2 f_t + u_{2,t} \\
&\ldots \\
R_{N,t} &= \lambda_N f_t + u_{N,t},
\end{aligned} \tag{3.79}
$$

where $\lambda_2, \lambda_3, ..., \lambda_N$ are the factor loadings that we are interested in estimating and f is the unobserved latent factor. Let us assume that $u_{i,t}$ are distributed jointly Normal with $E(u_{i,t}) = 0$ and $E(u_{i,t} u_{j,t}) = 0$ and $E(u_{i,t}^2) = \sigma_i^2$. These assumptions allow us to investigate the case where the maximum number of possible instruments, $N - 2$, is allowed. More general correlations restrict the number of instruments that can be employed but will not affect the intuition behind the next result. Identification in such settings follows the usual rules for the identification of systems of equations.

Consider the estimation of λ_k for $k = 2 \ldots N$. Using $R_{1,t}$ as a proxy for the latent factor we obtain:

$$
R_{k,t} = \lambda_k R_{1,t} + \epsilon_{k,t} \tag{3.80}
$$

where $\epsilon_{k,t} = u_{k,t} - \lambda_2 u_{1,t}$. Furthermore, notice that for each equation $j > 1, j \neq k$ we have $f_t = (1/\lambda_j)(R_{j,t} - u_{j,t})$. Hence we can write the reduced form equation for $R_{1,t}$ as

$$
R_{1,t} = z_t \pi + v_t, \tag{3.81}
$$

where

$$
z_t = (R_{2,t}, ..., R_{k-1,t}, R_{k+1,t}, ... R_{N,t}) \tag{3.82}
$$

$$
v_t = u_{1,t} - U_t \pi \tag{3.83}
$$

$$
U_t = (u_{2,t}, ..., u_{k-1,t}, u_{k+1,t}, ..., u_{N,t}) . \tag{3.84}
$$

Notice that the dimension of z_t is $N - 2$ which is potentially large relative to the sample size T. Let us first assume that the number of potential instruments is large relative to the sample size but less than the sample size, i.e. $c = N/T < 1$. Let $E(v_t^2) = \sigma_v^2$ and let $\chi^2 = \pi' z' z \pi / \sigma_v^2$ be the concentration parameter.

We wish to investigate the bias in the estimation of λ_k using the most common IV estimation procedure, Two-Stage Least Squares (2SLS):

$$\hat{\lambda}_k^{\text{2SLS}} = \frac{R_1 P_z R_k}{R_1 P_z R_1}, \tag{3.85}$$

where $P_z = z(z'z)^{-1}z'$, and $R_j = (R_{1,1}, ..., R_{j,T})'$.

The next proposition gives the second-order bias of the 2SLS estimate.

Proposition 3.10 (Second Order Bias of 2SLS in Factor Models): *We expect the bias in estimating the factor loadings by 2SLS to be approximately:*

$$E(\hat{\lambda}_k^{\text{2SLS}}) - \lambda_k \cong \frac{E(\epsilon v)}{\sigma_v^2} \frac{K - 2}{\chi^2} = -\lambda_k \frac{\sigma_1^2}{\sigma_v^2} (N - 4) \frac{(1 - R^2)}{(T - N - 2) R^2}, \tag{3.86}$$

where R^2 is the theoretical R^2 of the first stage regression.

This expression reveals that the 2SLS estimate of the factor loadings is biased downward. Moreover, the amount of bias is proportional to the degree of over-identification and monotonically increasing in N. This explains why estimation by 2SLS suffers relative to estimation by PCA even for strong factors as the 2SLS bias increases with N.

Note that the expression given above is similar to the expressions found by Hahn and Hausman (2002 a,b) in their investigation of estimation bias in simultaneous equations. The relationship is due to the asymptotic framework used in our analysis of high-dimensional factor models. We allow the number of cross-sectional observations to grow with the sample size at the same rate. Estimation of the factor loadings by 2SLS uses some of these observations as instruments, thereby generating a model with many instruments very similar to that of Bekker (1994). Note however, that there are also differences due to the very specific form which the correlation between the structural equation and the reduced form takes in the factor model. In particular, the bias expression depends on the variances in each equation.

In Table 3.6 we explore the estimation by IV methods of a factor model with one weak factor constructed using the design outlined in the Appendix for different choices of N and T such that $c \in \{0.3, 0.5, 0.7, 0.9\}$. We report the mean bias, the median bias and the means squared error (MSE) of the factor loadings estimates. In the first column we report the results for the standard 2SLS procedure. As predicted by the bias expression in equation 3.15 above we observe a negative bias which is increasing in $c = N/T$.

In order to correct for the bias due to the large number of instruments em-

ployed in the estimation of the factor loadings, we evaluate the performance of additional estimators familiar from the literature on many instruments as possible solutions to the many instruments problem. We consider the Fuller (1977) estimator, which is a modification of the Limited Information Maximum Likelihood (LIML) estimator with parameter $a = 1$. Additionally, we compute a Bias Corrected 2SLS derived by solving for λ_k in equation 3.15 of Proposition 3.10. The exact expression for these estimators are given in the Appendix. Furthermore, we also estimate the Continuously Updating Estimator (CUE) (Newey and Windmeijer, 2005), a particular choice of a generalized empirical likelihood estimator. In order to overcome the computational issues associated with the CUE estimator we employ a bounded Nelder-Mead algorithm.

Table 3.6 shows that all the estimators considered above perform well and can be used to obtain second order unbiased estimates of the factor loadings. Fuller seems to have comparable MSE to CUE, but the mean bias and MSE performance of CUE suffers for $c = 0.9$. This may indicate that CUE has a moment problem when the number of instruments is close to the sample size. Additionally, the estimators considered appear to be median unbiased. CUE appears to dominate the performance of the other estimators in terms of MSE except for the case with $c = 0.9$, while Fuller dominates in terms of mean bias.

So far we have only considered cases where $c < 1$, i.e. where the number of possible instruments is less than the sample size. In many applications, however, the number of individuals may be larger than the number of time periods over which the sample is observed. In this case the standard IV estimation methods fail since they employ the $P_z = z(z'z)^{-1}z'$ projection and $z'z$ becomes singular when $N > T$. At first glance this appears to be a limitation of IV procedures since they restrict the number of instruments to be less than the sample size.

In Table 3.6 and Table 3.7 we investigate a proposal of Theil (1973) which advocates the use of an incomplete projection which avoids the singularity of $z'z$ by using $P_{D,z} = z(D)^{-1}z'$ for some positive definite matrix D. The resulting 2SLS estimator which we call T-2SLS is given by $\hat{\lambda}_k^{2\text{SLS}} = (R_1 P_{D,z} R_k)/(R_1 P_{Dz} R_1)$. A similar modification can be applied to the CUE estimator by choosing a moment weighting matrix which depends on D rather than on $z'z$. The exact expression is given in the Appendix and we label the resulting estimator T-CUE. We use Monte-Carlo to investigate the behavior of the estimate of the factor loadings for a simple choice of $D = I$ for both $c < 1$ and $c > 1$.

For $c < 1$ the bias of T-2SLS is only slightly higher than that of BC2SLS while delivering a lower MSE than Fuller, BC2SLS and CUE. For $c > 1$ the bias of T-2SLS becomes substantial. By contrast T-CUE performs extremely well for both $c < 1$ and $c > 1$ in terms of both bias and MSE. For $c < 1$ T-CUE seems to avoid the moment problem of CUE for $c = 0.9$ and seems to outperform Fuller and BC2SLS. We plan to explore the performance of estimators employing Theiler's modification in future work to establish their distributional
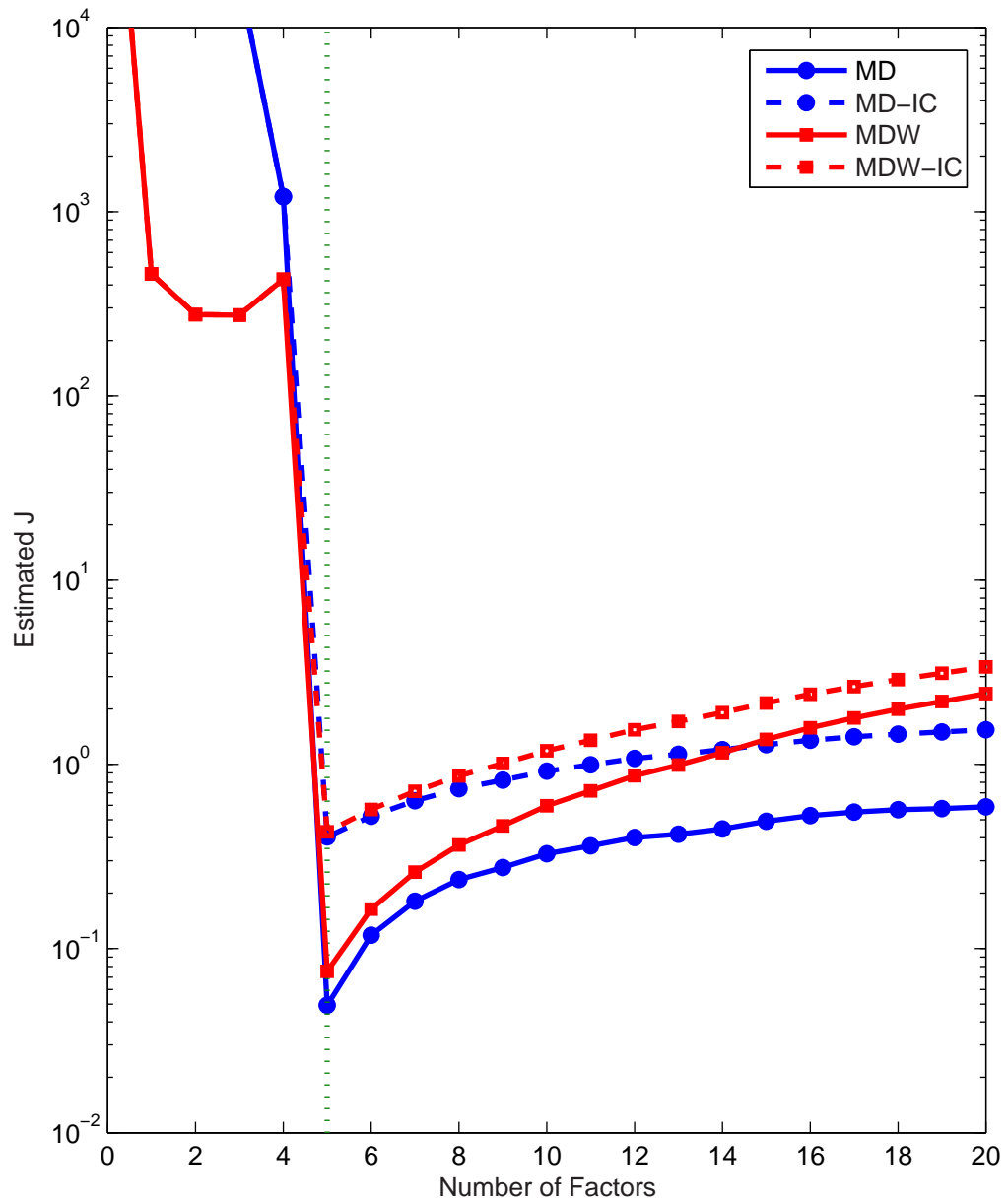
Figure 3-1. Estimation of the Number of Factors

Table 3.1. Estimating the Number of Factors: Design with 5 Strong Factors

# Strong Factors = 5			Number of Factors							
			Mean				MSE			
N	T	c	MD	MDW	MD-IC	MDW-IC	MD	MDW	MD-IC	MDW-IC
30	100	0.3	5.198	5.074	5.020	5.028	0.586	0.082	0.020	0.028
50	100	0.5	5.344	5.246	5.136	5.042	0.660	0.394	0.156	0.046
70	100	0.7	5.488	5.532	5.312	5.036	0.716	0.940	0.416	0.044
90	100	0.9	5.712	6.054	5.484	5.040	1.288	2.666	0.704	0.048
90	300	0.3	5.118	5.100	5.006	5.002	0.122	0.104	0.006	0.002
150	300	0.5	5.248	5.190	5.096	5.012	0.308	0.234	0.112	0.012
210	300	0.7	5.428	5.434	5.244	5.004	0.612	0.578	0.300	0.004
270	300	0.9	5.586	5.662	5.446	5.006	0.890	1.050	0.626	0.006
150	500	0.3	5.102	5.072	5.002	5.002	0.106	0.076	0.002	0.002
250	500	0.5	5.256	5.212	5.090	5.002	0.308	0.248	0.102	0.002
350	500	0.7	5.446	5.412	5.252	5.000	0.586	0.496	0.284	0.000
450	500	0.9	5.566	5.624	5.418	5.006	0.846	0.948	0.570	0.006

			Idiosyncratic Variance							
			Mean Bias				MSE			
N	T	c	MD	MDW	MD-IC	MDW-IC	MD	MDW	MD-IC	MDW-IC
30	100	0.3	-0.081	-0.060	-0.072	-0.058	0.008	0.004	0.006	0.004
50	100	0.5	-0.078	-0.066	-0.069	-0.058	0.007	0.005	0.005	0.004
70	100	0.7	-0.074	-0.071	-0.067	-0.056	0.006	0.006	0.005	0.003
90	100	0.9	-0.077	-0.085	-0.069	-0.056	0.007	0.009	0.005	0.003
90	300	0.3	-0.025	-0.020	-0.022	-0.018	0.001	0.000	0.001	0.000
150	300	0.5	-0.024	-0.020	-0.022	-0.018	0.001	0.000	0.001	0.000
210	300	0.7	-0.024	-0.023	-0.022	-0.018	0.001	0.001	0.001	0.000
270	300	0.9	-0.024	-0.024	-0.022	-0.018	0.001	0.001	0.001	0.000
150	500	0.3	-0.015	-0.011	-0.014	-0.011	0.000	0.000	0.000	0.000
250	500	0.5	-0.015	-0.012	-0.013	-0.011	0.000	0.000	0.000	0.000
350	500	0.7	-0.015	-0.014	-0.013	-0.011	0.000	0.000	0.000	0.000
450	500	0.9	-0.014	-0.015	-0.013	-0.011	0.000	0.000	0.000	0.000

Estimators used
     a) MD   Minimum Distance Parameter Estimation and J-test Objective Function
     b) MDW   Two-Step Minimum Distance Parameter Estimation with Optimal Covariance Matrix and J-test
     c) MD-IC   Augmentation of J-test in Estimator a) with Panel Information Criterion
     d) MDW-IC   Augmentation of J-test in Estimator b) with Panel Information Criterion

Table 3.2. Estimating the Number of Factors: Design with 5 Weak Factors

# Weak Factors = 5			Number of Factors							
			Mean				MSE			
N	T	c	MD	MDW	MD-IC	MDW-IC	MD	MDW	MD-IC	MDW-IC
30	100	0.3	5.176	5.064	5.028	5.016	0.444	0.076	0.040	0.024
50	100	0.5	5.256	5.178	5.088	4.992	0.528	0.306	0.180	0.072
70	100	0.7	5.304	5.314	5.110	4.812	0.696	0.754	0.354	0.260
90	100	0.9	5.444	5.738	5.294	4.774	0.816	2.010	0.586	0.274
90	300	0.3	5.082	5.084	5.008	5.008	0.086	0.088	0.008	0.008
150	300	0.5	5.196	5.146	5.062	4.998	0.264	0.162	0.110	0.014
210	300	0.7	5.302	5.296	5.152	4.890	0.494	0.424	0.272	0.114
270	300	0.9	5.378	5.452	5.238	4.676	0.706	0.768	0.514	0.328
150	500	0.3	5.090	5.078	5.008	5.008	0.102	0.086	0.008	0.008
250	500	0.5	5.190	5.126	5.034	4.992	0.234	0.146	0.058	0.008
350	500	0.7	5.274	5.232	5.120	4.890	0.426	0.316	0.272	0.114
450	500	0.9	5.308	5.414	5.182	4.644	0.572	0.646	0.410	0.356

			Idiosyncratic Variance							
			Mean Bias				MSE			
N	T	c	MD	MDW	MD-IC	MDW-IC	MD	MDW	MD-IC	MDW-IC
30	100	0.3	-0.088	-0.067	-0.080	-0.065	0.009	0.005	0.008	0.005
50	100	0.5	-0.081	-0.070	-0.073	-0.063	0.008	0.006	0.006	0.005
70	100	0.7	-0.076	-0.073	-0.068	-0.057	0.007	0.006	0.005	0.004
90	100	0.9	-0.077	-0.083	-0.071	-0.056	0.007	0.008	0.006	0.004
90	300	0.3	-0.028	-0.022	-0.026	-0.021	0.001	0.001	0.001	0.001
150	300	0.5	-0.026	-0.022	-0.024	-0.020	0.001	0.001	0.001	0.000
210	300	0.7	-0.026	-0.025	-0.024	-0.020	0.001	0.001	0.001	0.000
270	300	0.9	-0.025	-0.026	-0.023	-0.018	0.001	0.001	0.001	0.000
150	500	0.3	-0.017	-0.013	-0.016	-0.013	0.000	0.000	0.000	0.000
250	500	0.5	-0.016	-0.013	-0.014	-0.012	0.000	0.000	0.000	0.000
350	500	0.7	-0.015	-0.014	-0.014	-0.012	0.000	0.000	0.000	0.000
450	500	0.9	-0.015	-0.016	-0.014	-0.011	0.000	0.000	0.000	0.000

Estimators used
a) MD   Minimum Distance Parameter Estimation and J-test Objective Function
b) MDW   Two-Step Minimum Distance Parameter Estimation with Optimal Covariance Matrix and J-test
c) MD-IC Augmentation of J-test in Estimator a) with Panel Information Criterion
d) MDW-IC Augmentation of J-test in Estimator b) with Panel Information Criterion

Table 3.3. Estimating the Number of Factors: Misspecified Model

# Strong Factors = 5			Number of Factors							
			Mean				MSE			
N	T	c	MD	MDW	MD-IC	MDW-IC	MD	MDW	MD-IC	MDW-IC
30	100	0.3	5.226	5.120	5.062	5.034	0.394	0.144	0.066	0.034
50	100	0.5	5.460	5.384	5.212	5.044	0.664	0.572	0.244	0.048
70	100	0.7	5.792	5.878	5.480	5.092	1.500	1.846	0.696	0.108
90	100	0.9	6.272	6.792	5.914	5.152	2.964	5.328	1.762	0.172
90	300	0.3	5.344	5.292	5.040	5.028	0.448	0.368	0.044	0.028
150	300	0.5	5.886	5.804	5.476	5.052	1.358	1.228	0.592	0.056
210	300	0.7	6.538	6.670	6.146	5.066	3.262	3.802	2.034	0.070
270	300	0.9	7.338	7.884	6.978	5.146	6.606	9.864	4.902	0.150
150	500	0.3	5.596	5.536	5.088	5.028	0.796	0.680	0.092	0.028
250	500	0.5	6.500	6.450	5.916	5.112	2.956	2.866	1.344	0.116
350	500	0.7	7.574	7.784	7.064	5.268	7.598	8.944	5.044	0.320
450	500	0.9	8.486	9.028	8.126	5.366	13.118	17.192	10.758	0.426

			Idiosyncratic Variance							
			Mean Bias				MSE			
N	T	c	MD	MDW	MD-IC	MDW-IC	MD	MDW	MD-IC	MDW-IC
30	100	0.3	-0.084	-0.065	-0.075	-0.061	0.009	0.005	0.007	0.005
50	100	0.5	-0.079	-0.070	-0.067	-0.058	0.007	0.006	0.005	0.004
70	100	0.7	-0.087	-0.087	-0.074	-0.062	0.009	0.009	0.006	0.004
90	100	0.9	-0.093	-0.109	-0.081	-0.063	0.010	0.014	0.008	0.004
90	300	0.3	-0.028	-0.024	-0.021	-0.020	0.001	0.001	0.001	0.001
150	300	0.5	-0.031	-0.029	-0.023	-0.020	0.001	0.001	0.001	0.000
210	300	0.7	-0.035	-0.037	-0.029	-0.020	0.001	0.002	0.001	0.000
270	300	0.9	-0.040	-0.047	-0.035	-0.021	0.002	0.002	0.001	0.000
150	500	0.3	-0.017	-0.016	-0.010	-0.011	0.000	0.000	0.000	0.000
250	500	0.5	-0.022	-0.022	-0.016	-0.012	0.001	0.001	0.000	0.000
350	500	0.7	-0.027	-0.030	-0.023	-0.013	0.001	0.001	0.001	0.000
450	500	0.9	-0.030	-0.036	-0.027	-0.014	0.001	0.001	0.001	0.000

Misspecified with AR1(0.1) Idiosyncratic Errors.

Estimators used
a) MD — Minimum Distance Parameter Estimation and J-test Objective Function
b) MDW — Two-Step Minimum Distance Parameter Estimation with Optimal Covariance Matrix and J-test
c) MD-IC — Augmentation of J-test in Estimator a) with Panel Information Criterion
d) MDW-IC — Augmentation of J-test in Estimator b) with Panel Information Criterion

properties and the optimal choice of the matrix D. The simulations presented in this section, however, indicate that Fuller, BC2SLS, CUE and T-CUE can be used as alternatives for the accurate estimation of the factor loadings. Moreover, T-CUE can be used even when $N > T$, and the number of instruments exceeds the sample size. Given the difficulties inherent in estimating weak factors using PCA or imposing structural restrictions on the factor loadings, the methods explored in this section provide much needed accuracy and flexibility for the estimation of factor models.

■ 3.3 Conclusion

In this chapter we introduce new econometric theory for the estimation of large panel data models with unobserved latent variables. We show that it is possible to estimate the number of factors consistently for both the exact and approximate factor model without having to estimate first the factor loadings or factor scores. Our procedure allows for arbitrary models of heteroskedasticity and autocorrelation.

The new approach to the estimation of high-dimensional factor models introduced in this chapter allows for new extensions such as confidence intervals for the number of factors. Ongoing research to complement this chapter focuses on developing confidence intervals for the number of factors using the distribution of the minimized objective function. Another immediate extension of this approach, currently pursued, is to extend it to the estimation of dynamic factor structures by conducting the analysis in the frequency domain.

Additionally, we have shown that in factor models with weak factors, the estimation of factor loadings by PCA is inconsistent. To solve this problem we develop alternative IV based procedures with excellent finite sample properties. We relate the IV estimation of the factor model to current research on many and weak instruments.

■ 3.4 Appendix

Simulation design

We simulate a models $R_t = \Lambda F_t + U_t$, where $U = A_N^{1/2} \epsilon B_T^{1/2}$ for $\epsilon_{j,t}$ iid$N(0,1)$ and $F_{j,t}$ iid$N(0,1)$. A_N and B_T are as discussed in the text. Factor loadings are generated as follows: let $\Lambda_{j,k} = \sqrt{m_1/\sqrt{m_2}}$ for $j = 1 \ldots p_0$ and $k = 1 \ldots p_0$, where p_0 is the number of factors. For $j = p_0 + 1 \ldots N$ and $k = 1 \ldots p_0$ we have $\Lambda_{j,k} = a\sqrt{m_1/(N-k)}$ where $a = -1$ if $j = rk$ and $a = +1$ if $j \neq rk$ for $r = 1, 2, 3\ldots$. We can generate weak factors by setting $m_1 = 3$ and $m_2 = N$ and strong factors by setting $m_1 = 10$ and $m_2 = 1$.

Proof of Equation 3.37: If a and b are freely independent then so is the product $abab$. We can apply Definition 3.2 since all adjacent terms of the product

Table 3.4. Estimating the Number of Factors: Design with Autocorrelated Idiosyncratic Errors

 # Strong Factors = 5

			Factors				Idiosyncratic Variance				Time Series Correlation			
			Mean		MSE		Mean Bias		MSE		Mean Bias		MSE	
N	T	c	MD	MD-IC	MD	MD-IC	MD	MD-IC	MD	MD-IC	MD	MD-IC	MD	MD-IC
30	100	0.3	5.460	5.000	0.612	0.000	-0.092	-0.068	0.010	0.006	-0.168	-0.104	0.039	0.025
50	100	0.5	6.640	5.004	4.312	0.004	-0.130	-0.067	0.020	0.005	-0.148	-0.028	0.029	0.005
70	100	0.7	7.358	5.022	9.306	0.030	-0.144	-0.068	0.025	0.005	-0.106	-0.015	0.017	0.003
90	100	0.9	6.730	5.014	6.406	0.014	-0.117	-0.064	0.017	0.004	-0.046	-0.001	0.005	0.001
90	300	0.3	6.770	5.000	6.134	0.000	-0.049	-0.021	0.003	0.001	-0.094	-0.021	0.013	0.001
150	300	0.5	6.098	5.000	3.150	0.000	-0.036	-0.022	0.002	0.001	-0.031	-0.005	0.002	0.000
210	300	0.7	5.720	5.000	1.548	0.000	-0.031	-0.023	0.001	0.001	-0.015	-0.002	0.000	0.000
270	300	0.9	5.560	5.002	1.096	0.002	-0.027	-0.022	0.001	0.001	-0.008	0.001	0.000	0.000
150	500	0.3	6.284	5.000	3.968	0.000	-0.026	-0.014	0.001	0.000	-0.042	-0.012	0.003	0.000
250	500	0.5	5.840	5.000	1.928	0.000	-0.020	-0.013	0.001	0.000	-0.016	-0.004	0.000	0.000
350	500	0.7	5.710	5.000	1.338	0.000	-0.018	-0.013	0.000	0.000	-0.008	0.000	0.000	0.000
450	500	0.9	5.580	5.002	1.008	0.002	-0.016	-0.013	0.000	0.000	-0.004	0.001	0.000	0.000

Estimators used
a) MD   Minimum Distance Parameter Estimation and J-test Objective Function
b) MD-IC Augmentation of J-test in Estimator a) with Panel Information Criterion

Table 3.5. Inconsistency of PCA for Weak Factors

Mean Bias

N	T	c	PC1	PC2	PC3	PC4	PC5	IV1	IV2	IV3	IV4	IV5
70	100	0.7	0.202	0.067	-0.186	-0.016	-0.140	-0.094	0.073	-0.090	-0.084	0.094
90	100	0.9	-1.189	0.409	-1.678	-0.425	3.726	-0.128	0.096	0.122	-0.072	0.126
210	300	0.7	39.957	32.094	6.431	-52.497	-8.939	-0.014	0.023	0.006	-0.054	0.003
350	500	0.7	0.091	0.059	-0.022	-0.069	-0.058	-0.001	-0.011	0.009	-0.009	0.018
450	500	0.9	0.006	0.002	0.005	-0.014	-0.001	-0.002	0.045	0.019	-0.024	0.044

MSE

N	T	c	PC1	PC2	PC3	PC4	PC5	IV1	IV2	IV3	IV4	IV5
70	100	0.7	19.102	4.802	12.596	5.541	13.308	0.373	0.329	0.297	0.301	0.351
90	100	0.9	949.050	173.110	2414.200	606.960	12735.000	0.490	0.472	0.393	0.462	0.455
210	300	0.7	1337400.000	895190.000	50012.000	2451500.000	75091.000	0.423	0.293	0.268	0.348	0.297
350	500	0.7	4.771	2.633	0.421	3.836	2.689	0.352	0.289	0.192	0.346	0.239
450	500	0.9	0.627	0.168	0.076	0.480	0.165	0.848	0.689	0.548	0.738	0.764

Estimators used
a) PC   Principal Components
b) IV   Fuller IV Estimator

Table 3.6. Estimation of Factor Loadings $(c < 1)$

N	T	c	2SLS	Fuller	BC2SLS	CUE	T-2SLS	T-CUE
					Mean Bias			
30	100	0.3	-0.230	-0.003	0.028	0.010	-0.014	0.017
50	100	0.5	-0.338	0.009	0.047	0.024	-0.040	0.010
70	100	0.7	-0.413	0.022	0.105	0.036	-0.054	0.014
90	100	0.9	-0.470	0.020	0.157	-0.004	-0.072	0.010
90	300	0.3	-0.237	0.000	0.012	0.004	-0.026	0.003
150	300	0.5	-0.341	-0.001	0.021	0.005	-0.041	0.005
210	300	0.7	-0.416	0.005	0.022	0.013	-0.055	0.008
270	300	0.9	-0.474	0.038	0.036	0.111	-0.074	0.005
150	500	0.3	-0.240	-0.001	0.005	0.002	-0.028	0.001
250	500	0.5	-0.342	-0.001	0.011	0.004	-0.047	-0.002
350	500	0.7	-0.417	0.000	0.016	0.006	-0.059	0.004
450	500	0.9	-0.473	0.016	0.022	0.102	-0.076	0.001

N	T	c	2SLS	Fuller	BC2SLS	CUE	T-2SLS	T-CUE
					Median Bias			
30	100	0.3	-0.236	-0.018	-0.011	-0.005	-0.021	0.012
50	100	0.5	-0.339	-0.006	-0.021	0.006	-0.048	0.004
70	100	0.7	-0.414	-0.019	-0.006	-0.005	-0.063	-0.001
90	100	0.9	-0.469	-0.023	-0.030	-0.052	-0.079	-0.006
90	300	0.3	-0.237	-0.005	-0.002	-0.001	-0.031	-0.003
150	300	0.5	-0.342	-0.008	0.000	-0.001	-0.041	0.004
210	300	0.7	-0.419	-0.002	-0.012	0.008	-0.059	0.004
270	300	0.9	-0.474	0.002	-0.004	0.025	-0.075	0.000
150	500	0.3	-0.240	-0.005	-0.001	-0.003	-0.030	-0.002
250	500	0.5	-0.344	-0.005	-0.002	0.000	-0.049	-0.004
350	500	0.7	-0.415	-0.004	0.003	0.001	-0.061	0.002
450	500	0.9	-0.473	-0.007	-0.001	0.011	-0.076	-0.002

N	T	c	2SLS	Fuller	BC2SLS	CUE	T-2SLS	T-CUE
					MSE			
30	100	0.3	0.066	0.028	0.065	0.030	0.022	0.025
50	100	0.5	0.125	0.039	0.127	0.045	0.022	0.025
70	100	0.7	0.180	0.078	0.334	0.094	0.021	0.024
90	100	0.9	0.229	0.394	0.459	0.289	0.022	0.023
90	300	0.3	0.061	0.009	0.019	0.009	0.008	0.008
150	300	0.5	0.120	0.013	0.033	0.013	0.008	0.007
210	300	0.7	0.176	0.018	0.047	0.019	0.009	0.007
270	300	0.9	0.228	0.117	0.065	0.205	0.012	0.009
150	500	0.3	0.060	0.005	0.010	0.005	0.005	0.004
250	500	0.5	0.119	0.007	0.017	0.007	0.006	0.005
350	500	0.7	0.175	0.011	0.025	0.012	0.007	0.005
450	500	0.9	0.225	0.036	0.033	0.143	0.009	0.005

Table 3.7. Estimation of Factor Loadings ($c > 1$)

N	T	c	Mean Bias		Median Bias		MSE	
			T-2SLS	T-CUE	T-2SLS	T-CUE	T-2SLS	T-CUE
150	100	1.5	-0.116	0.009	-0.120	-0.002	0.029	0.025
300	100	3	-0.194	0.005	-0.195	-0.005	0.050	0.025
500	100	5	-0.251	0.023	-0.256	0.004	0.074	0.027
450	300	1.5	-0.115	0.004	-0.117	-0.001	0.019	0.008
900	300	3	-0.191	-0.001	-0.192	-0.007	0.040	0.008
1500	300	5	-0.250	0.005	-0.252	0.003	0.066	0.009
750	500	1.5	-0.116	0.002	-0.117	0.001	0.017	0.005
1500	500	3	-0.189	0.000	0.189	0.000	0.038	0.005
2500	500	5	-0.250	0.004	-0.251	0.000	0.063	0.005

are free and hence:

$$\phi((a - \phi(a)1)(b - \phi(b)1)(a - \phi(a)1)(b - \phi(b)1)) = 0. \tag{3.87}$$

We can expand this expression to obtain:

$$\begin{aligned}
\phi(abab) - \phi(a)\phi(bab) - \phi(b)\phi(aab) + \phi(a)\phi(b)\phi(ab) - \phi(a)\phi(abb) + \\
\phi^2(a)\phi(bb) + \phi(a)\phi(b)\phi(ab) - \phi^2(a)\phi^2(b) - \phi(b)\phi(aba) + \\
\phi(a)\phi(b)\phi(ba) + \phi^2(b)\phi(aa) - \phi^2(a)\phi^2(b) + \phi(a)\phi(b)\phi(ab) - \\
\phi^2(a)\phi(b)\phi(b) - \phi(a)\phi^2(b)\phi(a) + \phi^2(a)\phi^2(b) = 0
\end{aligned} \tag{3.88}$$

Using the fact that $\phi(ab) = \phi(a)\phi(b)$ and that $\phi(aba) = \phi(aa)\phi(b)$ we can simplify this expression as:

$$\begin{aligned}
\phi(abab) - \phi^2(a)\phi(bb) - \phi^2(b)\phi(aa) + \phi^2(a)\phi^2(b) - \phi^2(a)\phi(bb) + \\
\phi^2(a)\phi(bb) + \phi^2(a)\phi^2(b) - \phi^2(a)\phi^2(b) - \\
\phi^2(b)\phi(aa) + \phi^2(a)\phi^2(b) + \phi^2(b)\phi(aa) - \phi^2(a)\phi^2(b) + \\
\phi^2(a)\phi^2(b) - \phi^2(a)\phi^2(b) - \phi^2(a)\phi^2(b) + \phi^2(a)\phi^2(b) = 0
\end{aligned} \tag{3.89}$$

Since most of the terms in this expression cancel, we obtain

$$\phi(abab) = \phi^2(a)\phi(bb) + \phi^2(b)\phi(aa) - \phi^2(a)\phi^2(b). \tag{3.90}$$

Proof of Proposition 3.6: Define the series $\varrho_F(w) = 1/\sum_{s=1}^{\infty} m_F^s w^s$ where m_F^s are the moments of some probability distribution F. Let $S_F(w) = \varrho_F(w)(1+w)/w$. Let X and Y be two free random variables with associated probability measures F and G. Then, $S_F(w)S_G(w) = S_{FG}(w)$ (Voiculescu, 1998).

Now consider the Cauchy Transform G_Ω of Ω_N as $N \to \infty$ and let m_Ω^s be the moments of the asymptotic eigenvalue distribution of Ω. Then $G_\Omega(w) = \sum_{s=0}^{\infty} m_\Omega^s / w^{s+1}$. Following Burda et. al. (2006) we can let $G_\Omega(w) = M_\Omega(w)/w + 1$ such that $M_\Omega(w) = \sum_{s=1}^{\infty} m_\Omega^s / w^s$. Note also that $M_\Omega(\varrho_\Omega(w)) = w$. If we now apply equation 2.6 to the heteroskedasticity matrix A we have $N^{-1} \sum_{p=1}^{N} \frac{1}{1 - \lambda_p/w} = 1 + w$.

Furthermore, $N^{-1} \sum_{p=1}^{N} \frac{1}{1 - \lambda_p/\varrho_A(w)} = 1 + w$. If we now multiply both the numerator and denominator by $1/\varrho_\Psi(w)$ we have $N^{-1} \sum_{p=1}^{N} \frac{1/\varrho_\Psi(w)}{1/\varrho_\Psi(w) - \lambda_p/[\varrho_A(w)\varrho_\Psi(w)]} = 1 + w$.

Since $S_A(w)S_\Psi(w) = S_{A\Psi}(w)$ we have, $N^{-1} \sum_{p=1}^{N} \frac{1/\varrho_\Psi(w)}{1/\varrho_\Psi(w) - z\lambda_p/((1+z)\varrho_{A\Psi}(w))} = 1 + w$. Furthermore, we have $N^{-1} \sum_{p=1}^{N} \frac{1}{1 - (\lambda_p \varrho_\Psi(w))/((\frac{1+z}{z})\varrho_{A\Psi}(w))} = 1 + w$. If we now substitute $M_\Omega(w)$ for w we have $N^{-1} \sum_{p=1}^{N} \frac{1}{1 - \frac{\lambda_p \varrho_\Psi(M(w))}{\frac{1+M_\Omega(w)}{M_\Omega(w)} w}} = 1 + M_\Omega(w)$. Re-writing this expression we obtain $M_\Omega(w) = M_A(\frac{w(1+M_\Omega(w))}{M_\Omega(w)\varrho_\Psi(M(w))})$. Moreover, it ca be shown that $\varrho_\Psi(w) = (1+w)(c+w)/w$. Substituting in the previous expression (and after some further cancelations) we obtain $M_\Omega(w) = M_A(\frac{w}{1+cM_\Omega(w)})$. Re-writing this expression as a series we obtain equation 2.54: $\sum_{s=1}^{\infty} \frac{m_\Omega^s}{w^s} = \sum_{s=1}^{\infty} \frac{m_A^s}{w^s}(1 + c\sum_{r=1}^{\infty} \frac{m_\Omega^r}{w^r})^s$.

Proof of Proposition 3.7: We need to show that $\lim_{N,T\to\infty} \Pr(\hat{J}(\hat{\theta}, \mathrm{Sp}_p(\Sigma_N)) > \hat{J}(\hat{\theta}, \mathrm{Sp}_{p_0}(\Sigma_N))$ for all $p \neq p_0$ and $p \leq p_{\max}$. Let $\tilde{\Pi} = [Nm_\Omega^1, Nm_\Omega^2, ..., Nm_\Omega^s]$ and

$$\tilde{\Pi}_p = [\sum_{\lambda_j \in \mathrm{Sp}_p(\Sigma_N)} \lambda_j^1, \sum_{\lambda_j \in \mathrm{Sp}_p(\Sigma_N)} \lambda_j^2, ..., \sum_{\lambda_j \in \mathrm{Sp}_p(\Sigma_N)} \lambda_j^s].$$

For simplicity we consider the case of the minimum distance estimator with equal weighting. First consider the case where $p < p_0$. Then,

$$\hat{J}(\hat{\theta}, \mathrm{Sp}_p(\Sigma_N)) = \sum_{r=1}^{s} \left(Nm_\Omega^r - \sum_{\lambda_j \in \mathrm{Sp}_p(\Sigma_N)} \lambda_j^r \right)^2$$

$$= \sum_{r=1}^{s} \left(Nm_\Omega^r - \sum_{\lambda_j \in \mathrm{Sp}_{p_0}(\Sigma_N)} \lambda_j^r \right)^2 + \mathcal{J}(\lambda_{p_0+1}, ..., \lambda_p),$$

$$\hat{J}(\hat{\theta}, \mathrm{Sp}_p(\Sigma_N)) = \hat{J}(\hat{\theta}, \mathrm{Sp}_{p_0}(\Sigma_N)) + \mathcal{J}(\lambda_{p_0+1}, ..., \lambda_p). \tag{3.92}$$

(3.91)

The term $\mathcal{J}(\lambda_{p_0+1}, ..., \lambda_p)$ consists of polynomial terms which depend on eigenvalues resulting from the latent factors. Hence by the pervasiveness of the factors (Assumption 3.4) $\mathcal{J}(\lambda_{p_0+1}, ..., \lambda_p) \to \infty$ as $N \to \infty$ and $T \to \infty$. A fortiori, $\hat{J}(\hat{\theta}, \mathrm{Sp}_p(\Sigma_N)) > \hat{J}(\hat{\theta}, \mathrm{Sp}_{p_0}(\Sigma_N))$ with probability 1.

Now consider the case with $p > p_0$. Repeating the steps above we have $\hat{J}(\hat{\theta}, \mathrm{Sp}_p(\Sigma_N)) > \hat{J}(\hat{\theta}, \mathrm{Sp}_{p_0}(\Sigma_N))$ with probability 1. In this case however, it is no longer true that $\mathcal{J}(\lambda_{p_0+1}, ..., \lambda_p) \to \infty$. Rather, we have $\mathcal{J}(\lambda_{p_0+1}, ..., \lambda_p) > 0$. This is a consequence of the fact that for $p \neq p_0$ the moment conditions on the eigenvalue distribution are going to be misspecified. The objective function is however asymmetric since it diverges for $p < p_0$ but is increasing for $p > p_0$. Adding an information criterion does not change affect the case where $p < p_0$. For the case where $p > p_0$ however adding an information criterion penalized the

objective function proportionally to $(p - p_0)$ which is advantageous for values of p close to p_0 such that $\mathcal{J}(\lambda_{p_0+1}, ..., \lambda_p)$ is small.

Proof of Proposition 3.8: The proof of Equation 2.66 is very similar to that of given in Proposition 3.6 with (Ψ, B) substituted for (A, Ψ) and using the fact that $M_{\Psi B}(w) = cM_{\frac{1}{T}\epsilon B\epsilon'}(w)$. In order to derive equations 2.67, 2.68, 2.69 and 2.70 we can expand equation 2.66 as follows:

$$c(\mathcal{M}_\Omega - 1) = c\frac{m_B^1}{w}\mathcal{M} + c^2\frac{m_B^2}{w^2}\mathcal{M}^2 + c^3\frac{m_B^3}{w^3}\mathcal{M}^3 + c^4\frac{m_B^4}{w^4}\mathcal{M}^4 + O(w^{-5}), \quad (3.93)$$

where

$$\mathcal{M} = 1 + \frac{m_\Omega^1}{w} + \frac{m_\Omega^2}{w^2} + \frac{m_\Omega^3}{w^3} + \frac{m_\Omega^4}{w^4} + O(w^{-5}). \quad (3.94)$$

Expanding the RHS of equation 5.7 in terms of $1/w$ we obtain:

$$c(\mathcal{M} - 1) = \frac{1}{w}cm_B^1 + \frac{1}{w^2}\left(cm_B^1 m_\Omega^1 + c^2 m_B^2\right) + \frac{1}{w^3}\left(cm_B^1 m_\Omega^2 + c^3 m_B^3 + 2c^2 m_B^2 m_\Omega^1\right) + \quad (3.95)$$

$$+\frac{1}{w^4}\left(3c^3 m_B^3 m_\Omega^1 + cm_B^1 m_\Omega^3 + c^2 m_B^2(2m_\Omega^2 + (m_\Omega^1)^2) + c^3 m_B^4\right) + O(w^{-5}).$$

Dividing both sides by c, equating terms in powers of $1/w$ and substituting recursively for earlier terms we obtain:

$$m_\Omega^1 = m_B^1 \quad (3.96)$$

$$m_\Omega^2 = cm_B^2 + m_B^1 m_\Omega^1 = cm_B^2 + (m_B^1)^2 \quad (3.97)$$

$$m_\Omega^3 = c^2 m_B^3 + 2cm_B^2 m_\Omega^1 + m_\Omega^2 m_B^1 = c^2 m_B^3 + 2cm_B^2 m_B^1 + \left(cm_B^2 + (m_B^1)^2\right)m_B^1 \quad (3.98)$$

which can be further simplified as

$$m_\Omega^3 = c^2 m_B^3 + 2cm_B^2 m_\Omega^1 + m_\Omega^2 m_B^1 = c^2 m_B^3 + 2cm_B^2 m_B^1 + \left(cm_B^2 + (m_B^1)^2\right)m_B^1 \quad (3.99)$$

$$m_\Omega^3 = c^2 m_B^3 + 3cm_B^2 m_B^1 + (m_B^1)^3 \quad (3.100)$$

Additionally, we have:

$$m_\Omega^4 = m_B^1 m_\Omega^3 + cm_B^2(2m_\Omega^2 + (m_\Omega^1)^2) + 3c^2 m_B^3 m_\Omega^1 + c^3 m_B^4 \quad (3.101)$$

which after substituting the values of m_Ω^3, m_Ω^2 and m_Ω^1 derived above and

simplifying leads to:

$$m_\Omega^4 = c^3 m_B^4 + 4c^2(m_B^3 m_B^1 + \frac{1}{2}(m_B^2)^2) + 6cm_B^2(m_B^1)^2 + (m_B^1)^4. \tag{3.102}$$

Proof of Proposition 3.9: A detailed discussion of the inconsistency of PCA in large N, T models can be found in Paul (2006) and Onatski (2006). Here we show that the maximum degree of inconsistency depends on our measure $\tilde{\mu}$ of the spectral gap. Let $d_j = \lim_{N\to\infty} \hat{\Lambda}' \hat{\Lambda}$. Since the non-zero eigenvalues of $\hat{\Lambda}' \hat{\Lambda}$ are the same as the eigenvalues of $\hat{\Lambda} \hat{\Lambda}'$ we have that $\mathrm{Sp}(\lim_{N\to\infty}(\hat{\Lambda}\hat{\Lambda}')) = \{d_1, d_2, ..., d_{p_0}\}$, where p_0 is the number of factors in our factor model. Let $\tilde{d} = \min \mathrm{Sp}(\lim_{N\to\infty}\hat{\Lambda}\hat{\Lambda}')$. If we restrict our attention to the exact factor model we have $\sigma^2 = \max(\mathrm{Sp}(\lim_{N\to\infty}(\Omega_N)))$. By Theorem 1 of Onatski (2006) we have that the degree of inconsistency for the $j - \mathrm{th}$ Principal Component is given by

$$\varnothing(\hat{\Lambda}_j, \Lambda_j) = \sqrt{\frac{d_j^2 - \sigma^4 c}{d_j(d_j + \sigma^2 c)}} \tag{3.103}$$

Note however that $\varnothing(\hat{\Lambda}_j, \Lambda_j)$ is monotonically increasing in the eigenvalues d_j due to the factors:

$$0 < \frac{\partial \varnothing(\hat{\Lambda}_j, \Lambda_j)}{\partial d_j} = \frac{c\sigma^2(d_j^2 + 2\sigma^2 d_j + \sigma^4 c)}{2\sqrt{\frac{d_j^2 - \sigma^4 c}{d_j(d_j + \sigma^2)}}(d_j + c\sigma^2)^2 d_j^2}, \tag{3.104}$$

since $d_j > 0$. Hence,

$$\sqrt{\frac{\tilde{d}^2 - \sigma^4 c}{\tilde{d}^2 + \sigma^2 \tilde{d} c}} < \varnothing(\hat{\Lambda}_j, \Lambda_j). \tag{3.105}$$

We can re-write the expression on the left as:

$$\sqrt{\frac{1 - \left(\frac{\sigma^2}{\tilde{d}}\right)^2 c}{1 + \frac{\sigma^2}{\tilde{d}} c}} < \varnothing(\hat{\Lambda}_j, \Lambda_j). \tag{3.106}$$

If we define $\tilde{\mu} = \tilde{d}/\sigma^2$ we obtain the expression for the inconsistency of PCA as given in Equation 3.3.

Proof of Proposition 3.10: Consider the estimation of the factor loadings in a strict factor model with 1 latent factor given by 3.8. We can sequentially estimate the factor loadings by estimating λ_k for $k = 2..N$ using $R_{k,t} = \lambda_k R_{1,t} + \epsilon_{k,t}$ for $k = 2..N$ and $\epsilon_{k,t} = u_{k,t} - \lambda_k u_{1,t}$. Focusing on the estimation of equation

k we can re-write the estimating equation as a system of equations with two endogenous variables y_1 and y_2 in a form familiar to the IV literature:

$$y_1 = \beta y_2 + \epsilon = \beta z \pi + w \tag{3.107}$$

$$y_2 = z\pi + v \tag{3.108}$$

where $y_1 = R_k$, $y_2 = R_1$, $z = (R_2, R_{3,...}, R_{k-1}, R_{k+1}, ..., R_N)$ and $v = u_{1,t} - U_t \pi$. Let $K = \dim(\pi)$. For the strict factor model $K = N - \#\text{factors} - 1$. In our case $K = N - 2$.

Furthermore, assume that the reduced form errors (w, v) are i.i.d Normal distributed as

$$\begin{pmatrix} w \\ v \end{pmatrix} \sim N(0, \Omega) = N\left(0, \begin{pmatrix} \sigma_w^2 & \sigma_{wv} \\ \sigma_{wv} & \sigma_v^2 \end{pmatrix}\right), \tag{3.109}$$

and let the covariance between the structural equation (1) and the reduced form equation (2) for the endogenous variable y_2, be $\sigma_{\epsilon v}$. Note that for the exact factor model with 1 factor $\sigma_{\epsilon v} = -\lambda_k \sigma_1^2$ where $\sigma_1^2 = \text{Var}(u_1)$. Let the concentration parameter be given by $\chi^2 = \pi' z' z \pi / \sigma_v^2$. We are interested in the bias of the 2SLS estimator as a function of the concentration parameter. The finite sample bias of the 2SLS estimator $\hat{\beta}_{2\text{SLS}}$ was derived by Richardson (1968) and is given by the following expression:

$$E(\hat{\beta}_{2\text{SLS}}) - \beta = \frac{\sigma_{\epsilon v}}{\sigma_{vv}} \exp(-\chi^2/2) \; {}_1F_1(K/2 - 1; K/2; \chi^2/2), \tag{3.110}$$

where ${}_1F_1(a; b; c)$ denotes the confluent hypergeometric function ${}_1F_1(a; b; c)$ given by the following expansion

$${}_1F_1(a; b; c) = \sum_{j=1}^{\infty} \frac{(a)_j}{(b)_j} \frac{c^j}{j!}, \tag{3.111}$$

Note that the confluent hypergeometric function is defined in terms of Pochammer's symbol $(a)_j$ which corresponds to the ascending factorial:

$$(a)_j = \prod_{k=0}^{j-1} (a + k) = a\,(a+1)\,(a+2)\ldots(a+j-1) \quad \text{for} (a)_0 = 1. \tag{3.112}$$

Let $A = \exp(-\chi^2/2) \; {}_1F_1(K/2 - 1; K/2; \chi^2/2)$ and consider an expansion of

A for large values of χ^2:

$$A = \exp(-\chi^2/2) \ \frac{\Gamma(K/2)}{\Gamma(K/2 - 1)} \exp(\chi^2/2) \left(\frac{\chi^2}{2}\right)^{(K/2-1-K/2)} + O[\left(\frac{\chi^2}{2}\right)^{-2}] \tag{3.113}$$

$$A = 2 \ \frac{\Gamma(K/2)}{\Gamma(K/2 - 1)} \left(\frac{1}{\chi^2}\right) + O[\left(\frac{\chi^2}{2}\right)^{-2}] \tag{3.114}$$

But since $\Gamma(K/2) = (K/2 - 1)\Gamma(K/2 - 1)$ we have

$$A = \frac{K - 2}{\chi^2} + O[\left(\frac{\chi^2}{2}\right)^{-2}] \tag{3.115}$$

If we now substitute the first term of A in our bias expression we obtain

$$E(\hat{\beta}_{2\text{SLS}}) - \beta \cong \frac{\sigma_{\epsilon v}}{\sigma_v^2} \frac{K - 2}{\chi^2} = \frac{\sigma_{\epsilon v}}{\sigma_v^2} (K - 2) \frac{(1 - R^2)}{(N - K) R^2}, \tag{3.116}$$

where R^2 corresponds to the R^2 of the first stage regression. This expression corresponds to the approximate bias expression given in Hahn and Hausman (2002 a, b). Recall that $K = N - 2$ and $\sigma_{\epsilon v} = -\lambda_k \sigma_1^2$. Hence we obtain the expression in Proposition 3.10:

$$E(\hat{\lambda}_k^{2\text{SLS}}) - \lambda_k \cong -\lambda_k \frac{\sigma_1^2}{\sigma_v^2} (N - 4) \frac{(1 - R^2)}{(T - N - 2) R^2}. \tag{3.117}$$

IV Estimators used in Section 3.2:

The estimation of the loadings in a factor model using IV methods requires applying IV estimators recursively to determine the set of loadings λ_k for each cross-sectional unit. The complete set of loadings is obtained by adding the normalizations on the first p loadings. For simplicity we give the IV estimators under the setup of equations 5.21 and 5.22. At each step however a different set of observations are substituted for y_1 and z. Let $P_z = z(z'z)^{-1}z'$ and $Q_z = I - P_z$.

2SLS: $\hat{\beta} = (y_2' P_z y_1)/(y_2' P_z y_2)$

Fuller: $\hat{\beta} = (y_2' P_z y_1 - \kappa y_2' Q_z y_1)/(y_2' P_z y_2 - \kappa y_2' Q_z y_1)$ for $\kappa = \phi - 1/(T - N - 2)$, where $\phi = \min \text{Sp}\{W' P_z W (W' Q_z W)^{-1}\}$ and $W = (y_1, y_2)$.

BC2SLS: Uses the idea in Hahn and Hausman (2002b) to solve for the population coefficient from the second-order bias expression.

$$\hat{\beta}_{\text{BC}} = \hat{\beta}_{2\text{SLS}}/ \left(1 - \frac{\sigma_1^2}{\sigma_v^2} (N - 4) \frac{(1 - R^2)}{(T - N - 2) R^2}\right) \tag{3.118}$$

where $\hat{\beta}_{2\text{SLS}}$ is the 2SLS estimator. We can make this estimator feasible by

substituting R^2 with the estimated first stage R^2, σ_v^2 with the estimated first stage variance and estimate $\sigma_1^2 = \text{Var}(y_1) - 1$, if we normalize the variance of the factors to 1.

CUE: $\hat{\beta} = \text{argmin}_{\beta \in B} \hat{g}(\beta)' \hat{\Omega}(\beta) \hat{g}(\beta)/2$ for $\hat{g}(\beta) = z'(y_1 - \beta y_2)$ and $\hat{\Omega}(\beta) = E(\hat{\sigma}_t^2 z_t' z_t)$, where $\hat{\sigma}_t^2 = E(\epsilon_t^2)$. In order to optimize this objective function we applied a bounded version of the Nelder-Mead algorithm which allows us to impose restrictions on the parameter space B in order to avoid the multiple minima occasionally found in the simulations.

T-2SLS: The idea behind Theil's modification is to replace the term $(z'z)^{-1}$ by a matrix D. Here we take $D = I$ and therefore we have $\hat{\beta} = (y_2' z z' y_1)/(y_2' z z' y_2)$. Note that this estimator is consistent since $\hat{\beta} = \beta_0 + (y_2' z z' \epsilon)/(y_2' z z' y_2)$ and plim $T^{-1} z' \epsilon = 0$.

T-CUE: The estimator is the same as CUE but uses a different weighting matrix $\hat{\Omega}(\beta) = E(\hat{\sigma}_t^2)$.

Bibliography

AMENGUAL, D., AND M. W. WATSON (2006): "Consistent Estimation of the Number of Dynamic Factors in a Large N and T Panel," mimeo.

ANDERSON, T. W. (2003): *An Introduction to Multivariate Statistical Analysis*. Wiley, New York.

ANDERSON, T. W., AND H. RUBIN (1956): "Statistical Inference in Factor Analysis," Cowles Foundation Working Paper.

ANDREWS, D. W. K. (1999): "Consistent Moment Selection Procedures for Generalized Method of Moments Estimation," *Econometrica*, 67(3), 543–564.

ARNOLD, S. F. (1981): *The Theory of Linear Models and Multivariate Analysis*. Wiley, New York.

BAI, J. (2003): "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71, 135–171.

———— (2005): "Panel Data Models with Interactive Fixed Effects," mimeo.

BAI, J., AND S. NG (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70(1), 191–221.

BAI, Z. D., AND J. W. SILVERSTEIN (1998): "No Eigenvalues Outside the Support of the Limiting Spectral Distribution of Large Dimensional Sample Covariance Matrices," *Annals of Probability*, 26(1), 316–345.

———— (2004): "CLT for Linear Spectral Statistics of Large-Dimensional Sample Covariance Matrices," *Annals of Probability*, 32(1A), 553–605.

———— (2005): "Exact Separation of Eigenvalues of Large Dimensional Covariance Matrices," mimeo.

BAIK, J., AND J. W. SILVERSTEIN (2006): "Eigenvalues of Large Sample Covariance Matrices of Spiked Population Models," *ournal of Multivariate Analysis*, J97, 1382–1408.

BILLINGSLEY, P. (1995): *Probability and Measure*. Wiley, New York.

BROWN, S. J. (1989): "The Number of Factors in Security Returns," *Journal of Finance*, 54, 1247–1262.

BURDA, Z., A. JAROSZ, AND J. JURKIEWICZ (2006): "Applying Free Random Variables to Random Matrix Analysis of Financial Data," mimeo.

CHAMBERLAIN, G., AND M. ROTHSCHILD (1983): "Arbitrage, Factor Structure and Mean-Variance Analysis in Large Asset Markets," *Econometrica*, 51(5), 1305–324.

CONNOR, G., AND R. A. KORAJCZYK (1986): "Performance Measurement with the Arbitrage Pricing Theory: A New Framework for Analysis," *Journal of Financial Economics*, 15, 373–394.

DOZIER, R. B., AND J. W. SILVERSTEIN (2004): "On the Empirical Distribution of Eigenvalues of Large Dimensional Information-Plus-Noise Type Matrices," mimeo.

EDELMAN, A., AND N. R. RAO (2005): "Random Matrix Theory," *Acta Numerica*, pp. 233–297.

FULLER, W. A. (1977): "Some Properties of a Modification of the Limited Information Estimator," *Econometrica*, 45(4), 939–953.

GOLDBERGER, A. S. (1972): "Maximum-Likelihood Estimation of Regressions Containing Unobservable Independent Variables," *International Economic Review*, 13(1), 1–15.

GORMAN, W. M. (1980): "A Possible Procedure for Analysing Quality Differentials in the Egg Market," *Review of Economic Studies*, 47, 843–856.

GRENANDER, U., AND G. SZEGO (1958): *Toeplitz Forms and their Applications*. University of California Press, Berkeley.

HAHN, J., AND J. HAUSMAN (2002a): "A New Specification Test for the Validity of Instrumental Variables," *Econometrica*, 70(1), 163–189.

――――― (2002b): "Notes on Bias in Estimators for Simultaneous Equation Models," *Economics Letters*, 75, 237–241.

HAHN, J., AND G. KUERSTEINER (2002): "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects when both n and T are Large," *Econometrica*, 70, 1639–1657.

HAHN, J., G. KUERSTEINER, AND J. HAUSMAN (2004): "Estimation with Weak Instruments: Accuracy of Higher Order Bias and MSE Approximations," *Econometrics Journal*, 7(1), 272–306.

HANSEN, C., J. HAUSMAN, AND W. NEWEY (2006): "Estimation with Many Instrumental Variables," mimeo.

HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

HARDING, M. C. (2006): "Agreement Beyond Polarization: Spectral Analysis of Congressional Roll Call Votes," mimeo.

HECKMAN, J. J., AND S. NAVARRO (forthcoming): "Dynamic Discrete Choice and Dynamic Treatment Effects," *Journal of Econometrics*.

HOYLE, D., AND M. RATTRAY (2004): "Principal Component Analysis Eigenvalue Spectra From Data with Symmetry Breaking Structure," *Physical Review E*, 69.

JOLLIFFE, I. T. (2002): *Principal Component Analysis*. Springer, New York.

JONSSON, D. (1982): "Some Limit Theorems for the Eigenvalues of a Sample Covariance Matrix," *Journal of Multivariate Analysis*, 12, 1–38.

MADANSKY, A. (1964): "Instrumental Variables in Factor Analysis," *Psychometrika*, 29(2), 105–113.

ONATSKI, A. (2006): "Determining the Number of Factors from Empirical Distribution of Eigenvalues," mimeo.

PAUL, D. (2005): "Asymptotics of the Leading Sample Eigenvalues of a Spiked Covariance Model," mimeo.

RICHARDSON, D. H. (1968): "The Exact Distribution of a Structural Coefficient Estimator," *American Statistical Association Journal*, December, 1214–1226.

ROBINSON, P. M. (1974): "Identification, Estimation and Large-Sample Theory for Regressions Containing Unobservable Variables," *International Economic Review*, 15(3), 680–692.

SILVERSTEIN, J. (1995): "Strong Convergence of the Empirical Distribution of Large Dimensional Random Matrices," *Journal of Multivariate Analysis*, 55, 331–339.

SILVERSTEIN, J. W., AND Z. D. BAI (54): "On the Empirical Distribution of Eigenvalues of a Class of Large Dimensional Random Matrices," *Journal of Multivariate Analysis*, pp. 175–192.

SPEICHER, R. (2005): "Free Probability Theory," Notes on Operator Algebras, Ottawa, mimeo.

STOCK, J., AND M. WATSON (2003): "Forecasting Output and Inflation: The Role of Asset Prices," *Journal of Economic Literature*, 41, 788–829.

——— (2005): "Implications of Dynamic Factor Models for VAR Analysis," mimeo.

——— (2006): "Forecasting with Many Predictors," in *Handbook of Economic Forecasting*, ed. by G. Elliott, C. J. Granger, and A. Timmermann. Elsevier, Amsterdam.

THEIL, H. (1973): "A Simple Modification of the Two-Stage Least Squares Procedure for Undersized Samples," in *Structural Equation Models in the Social Sciences*, ed. by A. Goldberger, and O. Duncan. Seminar Press, New York.

VOICULESCU, D. V. (1985): "Symmetries of Some Reduced Free Product C*-Algebras," in *Operator Algebras and their Connections, with Topology and Ergodic Theory, Lecture Notes in Mathematics*, vol. 1132, pp. 556–588. Springer, New York.

——— (1998): "Lectures on Free Probability Theory," in *Lectures on Probability Theory and Statistics*, vol. 1738, pp. 279–349. Springer, Saint Flour 28.

WACHTER, K. W. (1978): "The Strong Central Limits of Random Matrix Spectra for Sample Matrices of Independent Elements," *Annals of Probability*, 6(1), 1–18.

ZELLNER, A. (1970): "Estimation of Regression Relationships Containing Unobservable Independent Variables," *International Economic Review*, 11(3), 441–454.

CHAPTER 4

Uncovering the Effect of Global Factors on the US Economy

A quintessential issue of modern macroeconomics is the elusive relationship between inflation and unemployment. Over the past few decades this relationship seems to have been subject to many changes of perspective both in our models and in the empirical reality underlying it. Globalization is undoubtedly one of the most remarkable processes in recent history and we cannot overlook the potential impact of global factors in shaping the economic experience of the US. Thus, the importance of understanding the role of global factors as drivers of US inflation is at the core of recent macroeconomic research (Bernanke, 2006; Rogoff, 2006).

In this chapter we apply our econometric methodology to identify, estimate and characterize global factors and to understand how the inflation-unemployment relationship is affected by the presence of these factors. In doing so we develop an approach, which is less centered on US markets, as recently advocated by Bernanke (2006), and which allows us to understand the effect of global factors by combining structural macroeconomic modeling with information extracted from financial markets.

The forward looking nature of financial markets may contain information with important implications for output and inflation in the medium term. Extracting relevant economic information from financial markets presents a powerful opportunity for the improved identification of shocks to the economy. It may also allow us to improve our understanding of the relationship between financial system risk and the macroeconomy. However, relating financial markets to the economy also presents limitations due to the extremely noisy nature of these markets (Stock and Watson, 2003).

In this chapter we show how the econometric methods derived in this paper can be employed to overcome this limitation and augment a structural New

Keynesian model of the economy with global factors identified from international stock markets. The econometric approach developed in this paper proves crucial in separating the noise in international financial markets from pervasive global factors and in estimating their structural effect.

■ 4.1 A Structural New Keynesian Macro-Finance Model

The concept of fusing a small structural macroeconomic model with financial information has been successfully employed by a number of recent authors in macro-finance (Ang and Piazzesi, 2003; Bekaert, Cho and Moreno, 2006; Rudebusch and Wu, 2004). In spite of this similarity, our model presents a number of important innovations which we will briefly mention here and the technical aspects of which will be discussed in detail below, as we develop the model and the econometric estimation strategy.

The focus of our investigation is on the information revealed by international stock markets. We believe that the increasing global and financial integration of the last few decades has important implications for the description of shocks to the US economy. This adds increased importance to our attempt to understand economic forces potentially outside the control of US monetary policy and in plotting the future course inflation and employment. The current macro-finance literature seems to have focused exclusively on extracting information from yield curves. We believe that by focusing our attention on the stock market we can understand the nature of pervasive global factors, since much of modern asset pricing theory emphasizes that a few pervasive factors are the dominant source of correlations between asset returns. In future work we hope to explore joint stock and bond markets.

While stock prices incorporate a substantial amount of rational information about the future of the global economy, they also contain a non-negligible amount of noise which could potentially make them unreliable. Thus our ability to correctly identify the number of global factors and estimate them accurately is of paramount importance. The international dimension of our model is also crucial in that it employs the correlation between different international stock markets in order to identify the global factors.

Another important characteristic of our approach consists in the attempt to improve our understanding of the nature of shocks affecting the macro-economy. By specifying a multi-factor error structure we can estimate the structural effect of pervasive latent factors on the demand and supply equations for the economy. This allows us to investigate the possibility of correlated movements in output and inflation. Understanding the nature of global shocks and having the ability to distinguish between supply and demand shocks is crucial for the conduct of optimal monetary policy (Clarida, Gali and Gertler, 1999).

Following the standard macro-finance approach we shall balance the need for a

structural specification with an empirically successful set of estimating equations. Our approach will focus on a structural model of the US economy and a reduced form set of equations describing asset pricing in international financial markets using a set of observed and latent factors. The structural model for the US economy follows in the tradition of recent small scale New Keynesian models (Bekaert, Cho and Moreno, 2006; Carlin and Soskice, 2006). The model for the international financial markets corresponds to an International Arbitrage Pricing Theory approach (Ikeda, 1991; Solnik, 1981).

The supply side of our economy is characterized by the following Phillips curve:

$$(\text{PC}) \qquad \pi_t = \alpha E_t \pi_{t+1} + (1 - \alpha) \pi_{t-1} + \beta y_t + \epsilon_t, \qquad (4.1)$$

where π_t corresponds to the current period inflation and y_t is the current period unemployment or output gap. The parameter α measures the trade-off between lagged inflation π_{t-1} and the agents' rational expectation $E_t \pi_{t+1}$ of the next period inflation with respect to the current period information set. The term ϵ_t corresponds to an aggregate supply shock which will be discussed further below. Notice that current inflation depends on the weighted average between past inflation and rational expectations of future inflation with weights α and $1 - \alpha$. This reflects the trade-off between a theoretically justified New Keynesian specification based on rational expectations (Clarida, Gali and Gertler, 1999) and an empirically successful adaptive specification. While the exact origins of inflation persistence are still a matter of debate, existing models seek to explain inertia through overlapping contracts (Fuhrer and Moore, 1995) or informational limitations (Mankiw and Reis, 2002).

The demand side of the economy is characterized by the following aggregate demand curve:

$$(\text{AD}) \qquad y_t = \gamma E_t y_{t+1} + (1 - \gamma) y_{t-1} + \delta(i_t - E_t \pi_{t+1}) + \omega_t, \qquad (4.2)$$

where the current period unemployment or output gap depends on a weighted average between the previous period unemployment and the agents' rational expectations of future unemployment $E_t y_{t+1}$ with weights $1 - \gamma$ and γ respectively. This captures the trade-off between the standard Keynesian multiplier effect of external shocks and the dampening effect of rational consumption smoothing on output which results from the rational expectations term $E_t y_{t+1}$. The amplifying effect of inertial adjustment may be due to credit constraints or habit persistence (Fuhrer, 2000). The last term in the equation corresponds to the real interest rate $i_t - E_t \pi_{t+1}$, which balances the demand for consumption with that for savings.

The monetary policy curve captures the objective of the Fed to stabilize interest rates while taking into account inflation and unemployment (Taylor, 1993).

Hence the short-term interest rate is given by:

$$\text{(MP)} \qquad i_t = \rho i_{t-1} + (1 - \rho) \left[\theta \pi_t + \phi y_t\right] + v_t, \qquad (4.3)$$

where the parameter ρ controls the extent to which the Fed engages in interest rate smoothing. Additionally, the Fed adjusts interest rates with a factor $1 - \rho$ towards the desired level of interest rates given by $\theta \pi_t + \phi y_t$. This target level for the short rate may be derived as the optimal level of nominal interest rates in a model where a Central Bank minimizes a quadratic loss function over unemployment and inflation with relative weights given by ϕ and θ respectively. The version of the monetary policy curve used here corresponds to the "benchmark Taylor rule" of Ang, Dong and Piazzesi (2005).

So far we have not imposed any assumptions on the error terms in the three equations $(\epsilon_t, \omega_t, v_t)$. We will pay particular attention to the errors given by ϵ_t and ω_t and wish to interpret them as structural aggregate supply and demand shocks respectively, subject to the caveat that ω_t may confound true demand shocks with shocks to the natural rate (Woodford, 2001). We improve, however, on the current macro-finance literature which models these shocks as autoregressive processes by imposing a more economically insightful multifactor error structure (Pesaran, 2006). In addition to avoiding the need to impose long lag structures from the more familiar VAR approach, these additional set of assumptions allow us to incorporate the effect of globally determined factors in the specification of the shocks to the US economy. Thus we assume that the supply and demand shocks are given in terms of a set of unobserved common global factors F_t^G and an idiosyncratic component (e_t, w_t):

$$\epsilon_t = A' F_t^G + e_t \qquad (4.4)$$

$$\omega_t = B' F_t^G + w_t. \qquad (4.5)$$

For the purposes of this application we assume that (e_t, w_t, v_t) are distributed independently of the the global factors F_t^G and the other right hand side variables in the structural model. This assumption underlies the applicability of a factor analysis based estimation procedure for the number of factors. As we previously remarked however, we do not need to assume a strict factor structure and in principle our procedure is robust to more complex patterns of time dependence, structural breaks with unknown change points or volatility regime switching processes which are currently investigated in the macro-finance literature (Bibkov, 2005).

The multifactor error specification allows the possibility of correlated common effects due to the global factors to shock both the demand and supply equations. The extent to which they affect supply and demand is determined by the magnitude of the respective factor loadings coefficients A and B.

Our approach of extracting useful economic information from the global fac-

tors is enhanced by the ability to treat these factors as exogenous shocks to the macroeconomic system (Bibkov and Chernov, 2004). This justifies our assumption of independence of the global factors with respect to the idiosyncratic shocks. We may wish however to consider correlations between the unobserved latent factors and past values of the macroeconomic variables for inflation, output and interest rates. Given the central position of the US economy in global markets this allows for potentially long term effects of the US economy on the global market.

It is not possible to identify the unknown factors from the three structural equations above. Thus, we will need to augment the structural New Keynesian model above with a model of international financial markets. The existence of global economic factors manifested in international financial markets has been suggested in a number of previous studies, yet little is known about their precise macroeconomic effect. We aim to use the econometric methods developed in this paper in order to address these questions. In particular we wish to identify the number and intensity of such factors. The corresponding factor loadings allow us to measure the strength of correlated shocks to the US economy as well as distinguish between cost-push shocks and demand shocks.

Recently, Diebold, Li and Yue (2006) argue for economically important global factors describing country yield curve dynamics, while Bekaert, Hodrick and Zhang (2005) find that co-movements between international stock returns are best explained using parsimonious risk-based factor models. The existence of global economic forces driving asset prices across countries is often thought to be a consequence of the increasing capital market integration of recent decades (Brooks and Catao, 2000; Grinold, Rudd and Stafek, 1989).

In order to fully account for the effects of exchange rates in the estimation of global factors from an international portfolio of stock returns, we shall employ a model of international arbitrage pricing (Clyman, 1997; Ikeda, 1991; Solnik, 1982). We assume that the global market is divided between K countries, each with its own currency. In each country there is a large number of risky assets denominated in the currency of the country where they are issued. In particular, we assume that each country has a riskless national bond and a large number of traded stocks issued by companies operating in different industries. Let one of the currencies, $j = 1$, be the nominal numeraire. In total the global economy consists of N assets labelled $j = 1 \ldots N$.

Consider an arbitrary asset j issued in country k. We assume that the process generating this return, when expressed in local currency, is given by the following linear relationship:

$$R_j^k = r_j^k + b_{j0}f_0 + b_{j1}f_1 + \ldots + b_{jS}f_S + u_j, \qquad (4.6)$$

where R_j^k is the stochastic return to asset j in country k with expected value r_j^k and f_s for $s = 0 \ldots S$ are a set of common mean-zero factors. The coefficients

$b_{i0}, b_{i1}, ..., b_{iS}$,correspond to factor loadings on the common risk factors f_s, $s = 0 \ldots S$. The idiosyncratic risk component u_j is assumed to be mean zero and uncorrelated with the common factors f_s.

We cannot, however, assume that all the factors f_s in this specification are in fact global factors shared across countries. We note that the CAPM is nested within the APT structure of the model for asset returns (Burmeister and McElroy, 1988) and in the presence of transportation costs, tariffs or other restrictions to free trade, we expect to find differences between representative consumption configurations across countries (Levine, 1989). Hence we let f_0 be equal to M_k, the return to the market portfolio for country k where asset j is located.

The above model of returns applies to assets in local currency. If we are to consider the same model from the perspective of an investor in the numeraire country we have to add considerations of exchange rates. Let Υ_t be the (log) exchange rate at time t for country j. Consider the Uncovered Interest Parity (UIP) condition:

$$E(\Upsilon_{j,t+1}) - \Upsilon_{j,t} = r_{1,t} - r_{j,t}. \tag{4.7}$$

We assume that the linear return process above also applies to unhedged returns, that is the return to a stock j in country $k > 1$ in terms of the numeraire is given by

$$\left(R^k_{j,t+1}\right)_1 = R^k_{j,t+1} + E(\Upsilon_{j,t+1}) - \Upsilon_{j,t} \tag{4.8}$$

A common explanation behind the empirical failure of UIP suggests the existence of additional risk premia $\mathrm{RP}_{j,t}$, such that

$$E(\Upsilon_{j,t+1}) - \Upsilon_{j,t} = r_{1,t} - r_{j,t} + \mathrm{RP}_{j,t}. \tag{4.9}$$

As is common in the literature on international economics, we assume that the additional risk premia $\mathrm{RP}_{j,t}$ are also subject to common factors (Driessen, Melenberg, Nijman, 2003). Some of the underlying factors are likely to be the global factors, while some may be pure exchange rate factors. Noticing that both the unhedged returns in terms of the numeraire and the returns in local currency share a common factor structure with the addition of exchange rate factors and after taking into account for the different local risk-free rates:

$$\left(R^k_{j,t}\right)_1 - r_{1,t} = R^k_{j,t} - r_{k,t} + \mathrm{RP}_{j,t}. \tag{4.10}$$

This implies an estimating equation for stock return j in country k which is given by:

$$R^k_{j,t} = r_{k,t} + c_j M_{k,t} + \Lambda^k_j F^G_t + u_{j,t}, \tag{4.11}$$

where $r_{k,t}$ is the local risk free rate, M_{kt} is the market index for country k and F^G_t are global factors. We let the estimation procedure detect if some of the factors labeled as global may in fact be pure exchange rate factors. The coefficients

c_j and Λ_j^k correspond to asset specific factor loadings on the observable country factor M_k and the global factors F^G. Furthermore, we assume that

$$E(u_{j,t}|F_t^G, M_{kt}) = E(F_t^G) = E(M_{kt}) = E(u_{jt}) = 0. \qquad (4.12)$$

Similar to the Chamberlain and Rothschild (1983) model it is possible to consider both a strict factor model where $u_{j,t}$ are both serially uncorrelated and homoscedastic or alternatively allow for weak correlations in $u_{j,t}$ in order to accommodate an approximate factor structure.

■ 4.2 Rational Expectations Solution

Let us now re-state the main estimating equations of our structural macro-finance model for ease of reference:

$$(\text{PC}) \qquad \pi_t = \alpha E_t \pi_{t+1} + (1 - \alpha) \pi_{t-1} + \beta y_t + \epsilon_t, \qquad (4.13)$$

$$(\text{AD}) \qquad y_t = \gamma E_t y_{t+1} + (1 - \gamma) y_{t-1} + \delta(i_t - E_t \pi_{t+1}) + \omega_t, \qquad (4.14)$$

$$(\text{MP}) \qquad i_t = \rho i_{t-1} + (1 - \rho)[\theta \pi_t + \phi y_t] + v_t, \qquad (4.15)$$

$$\epsilon_t = A' F_t^G + e_t, \qquad (4.16)$$

$$\omega_t = B' F_t^G + w_t, \qquad (4.17)$$

$$R_{j,t}^k = r_{k,t} + c_j M_{k,t} + \Lambda_j^k F_t^G + u_{j,t}, \qquad (4.18)$$

$$(\text{UIP}) \qquad E(\Upsilon_{k,t+1}) - \Upsilon_{k,t} = \psi(r_{1,t} - r_{k,t}) + \Lambda_k^e F_t^G + \xi_{k,t}. \qquad (4.19)$$

Since our model combines observed factors such as the lagged values of macroeconomic variables like inflation with latent factors corresponding to the global and exchange rate factors, we need to develop a strategy for estimating both the number of these factors and their corresponding time series. We shall first derive the rational expectations solution to the set of structural macro-equations above. This will play the role of a reduced form specification which can be combined with the equations for stock returns. In the next section we will then describe the econometric methods employed in order to "partial out" the effect of observed factors and apply the factor analytic methods developed in this paper.

First let us write the system corresponding to the first three structural equa-

tions for the US in matrix form:

$$
\begin{pmatrix} 1 & -\beta & 0 \\ 0 & 1 & \delta \\ (\rho-1)\theta & (\rho-1)\phi & 1 \end{pmatrix} \begin{pmatrix} \pi_t \\ y_t \\ i_t \end{pmatrix} =
$$

$$
\begin{pmatrix} \alpha & 0 & 0 \\ \delta & \gamma & 0 \\ 0 & 0 & 0 \end{pmatrix} E_t \begin{pmatrix} \pi_{t+1} \\ y_{t+1} \\ i_{t+1} \end{pmatrix} + \begin{pmatrix} 1-\alpha & 0 & 0 \\ 0 & 1-\gamma & \delta \\ 0 & 0 & \rho \end{pmatrix} \begin{pmatrix} \pi_{t-1} \\ y_{t-1} \\ i_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ \omega_t \\ v_t \end{pmatrix}. \quad (4.20)
$$

This is equivalent to the following equation in matrix notation:

$$
\Gamma_1 X_t = \Gamma_2 E_t X_{t+1} + \Gamma_3 X_{t-1} + \xi_t, \quad (4.21)
$$

where the coefficient matrices $\Gamma_1, \Gamma_2, \Gamma_3$ encode the structural coefficients. We need to show that the solution of this system is of the form:

$$
X_t = \Omega X_{t-1} + \Sigma \xi_t. \quad (4.22)
$$

Substituting equation 4.22 in 4.21 above we obtain under the assumption of rational expectations:

$$
\Gamma_1 X_t = \Gamma_2 E_t \left\{ \Omega X_t + \Sigma \xi_{t+1} \right\} + \Gamma_3 X_{t-1} + \xi_t, \quad (4.23)
$$

$$
\left(\Gamma_1 - \Gamma_2 \Omega \right) X_t = \Gamma_3 X_{t-1} + \xi_t, \quad (4.24)
$$

$$
X_t = \left(\Gamma_1 - \Gamma_2 \Omega \right)^{-1} \Gamma_3 X_{t-1} + \left(\Gamma_1 - \Gamma_2 \Omega \right)^{-1} \xi_t. \quad (4.25)
$$

Matching coefficients between 4.22 and 4.25 we derive a solution to the system given by the following matrix equations:

$$
\Omega = \left(\Gamma_1 - \Gamma_2 \Omega \right)^{-1} \Gamma_3, \quad (4.26)
$$

$$
\Sigma = \left(\Gamma_1 - \Gamma_2 \Omega \right)^{-1}. \quad (4.27)
$$

Note that the first equation can be re-written as a quadratic matrix equation in Ω:

$$
\Gamma_2 \Omega^2 - \Gamma_1 \Omega + \Gamma_3 = 0. \quad (4.28)
$$

We will now investigate the effect of writing the rational expectations solution to the structural model as a VAR(1) process on the multi-factor error structure

of the error terms. Consider the error term in equation 108 above:

$$\Sigma\xi_t = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix} \begin{pmatrix} \epsilon_t \\ \omega_t \\ v_t \end{pmatrix} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix} \begin{pmatrix} A'F_t^G + e_t \\ B'F_t^G + w_t \\ v_t \end{pmatrix} \tag{4.29}$$

$$\Sigma\xi_t = \begin{pmatrix} (\sigma_{11}A + \sigma_{12}B)' F_t^G + \sigma_{11}e_t + \sigma_{12}w_t + \sigma_{13}v_t \\ (\sigma_{21}A + \sigma_{22}B)' F_t^G + \sigma_{21}e_t + \sigma_{22}w_t + \sigma_{23}v_t \\ (\sigma_{31}A + \sigma_{32}B)' F_t^G + \sigma_{31}e_t + \sigma_{32}w_t + \sigma_{33}v_t \end{pmatrix}. \tag{4.30}$$

Let us now make the following notational definitions:

$$\Lambda_\pi = \sigma_{11}A + \sigma_{12}B, \tag{4.31}$$

$$\Lambda_y = \sigma_{21}A + \sigma_{22}B, \tag{4.32}$$

$$\Lambda_i = \sigma_{31}A + \sigma_{32}B, \tag{4.33}$$

$$u_t^\pi = \sigma_{11}e_t + \sigma_{12}w_t + \sigma_{13}v_t, \tag{4.34}$$

$$u_t^y = \sigma_{21}e_t + \sigma_{22}w_t + \sigma_{23}v_t, \tag{4.35}$$

$$u_t^i = \sigma_{31}e_t + \sigma_{32}w_t + \sigma_{33}v_t. \tag{4.36}$$

Then we can write the error term as:

$$\Sigma\xi_t = \begin{pmatrix} \Lambda'_\pi F_t^G + u_t^\pi \\ \Lambda'_y F_t^G + u_t^y \\ \Lambda'_i F_t^G + u_t^i \end{pmatrix}. \tag{4.37}$$

This shows that the multi-factor error structure of the macro-equations for the US also implies a linear factor structure in the unobserved global factors for the reduced form VAR equations. Note, however, that the error terms in the VAR form, (u_t^π, u_t^y, u_t^i) will be correlated with each other.

In order to identify and estimate the unobserved latent factors we need to augment the reduced form VAR specification derived above using rational expectations with the set of linear APT equations for stock returns from our model of international financial markets and the corresponding UIP conditions:

$$X_t = \Omega X_{t-1} + \Lambda'_X F_t^G + u_t, \tag{4.38}$$

$$S_{j,t}^k = c_j M_{k,t} + \Lambda_j^k F_t^G + u_{j,t}, \tag{4.39}$$

$$E(\Upsilon_{k,t+1}) - \Upsilon_{k,t} = \psi(r_{1,t} - r_{k,t}) + \Lambda_k^e F_t^G + \xi_{k,t}, \tag{4.40}$$

for $\Lambda_X = [\Lambda_\pi \ \Lambda_y \ \Lambda_i]$ and $S_{jt} = R_{j,t}^k - r_{k,t}$.

■ 4.3 Econometric Estimation Strategy

Since the complete model can be characterized as a hybrid factor model with both observed and unobserved factors, the first stage of our procedure involves the consistent estimation of the parameters on the observed factors. Using the corresponding residuals after partialling out the observed factors, we can then estimate a factor model employing the methods introduced in this paper. This will identify the number of factors and provide an estimate of their time series process. These first two steps of our procedure only involve the reduced form specification. We can then use the estimated global factors and compute the structural parameters for the US macroeconomic model.

However we cannot simply assume that the unobserved latent factors are uncorrelated with the other observed factors in our model, and hence we cannot employ a direct least squares procedure for the estimation of the coefficients $(\Omega, c_1, ..., c_K, \psi)$ in equations 4.38-4.40 above. In order to account for these possible correlations we will employ the recently developed *common correlated effects* (CCE) estimator of Pesaran (2006). The idea of this estimator is to augment each estimating equation with cross-sectional averages that proxy for the unobserved factors in order to obtain consistent residuals on which further factor analytic procedures may be applied.

Consider the stock returns equations for a set of firms $j \in J_k = \{j_1, j_2, ..., j_k\}$ based in some country k in our model. The returns minus risk-free rate specification is given by:

$$S_{j,t}^k = c_j M_{k,t} + \Lambda_j^k F_t^G + u_{j,t}. \tag{4.41}$$

Now consider the cross-section averages of this equation:

$$\overline{S}_{w,t} = \overline{c}_w M_{k,t} + \overline{\Lambda}_w^{\ j} F_t^G + \overline{u}_{w,t}, \tag{4.42}$$

where,

$$\overline{S}_{w,t} = \sum_{j \in J_k} w_j S_{j,t}^k, \overline{c}_w = \sum_{j \in J_k} w_j c_j, \overline{\Lambda}_w = \sum_{j \in J_k} w_j \Lambda_j, \overline{u}_{w,t} = \sum_{j \in J_k} w_j u_{j,t}, \tag{4.43}$$

for some set of weights w_j. Pesaran (2006) shows that as $||J_k|| \to \infty$, that is as the number of stocks for each country k included in our model increases, we can obtain consistent estimates of the residuals from regressing the LHS variables of the model on the observed factors if we additionally condition on the cross-

sectional averages defined above. The cross-sectional averages provide an effective way of proxying for the unobserved factors in the reduced form equations. Using this method we can compute the following residuals of interest:

$$\hat{u}_{j,t} = S_{j,t} - \hat{c}_j M_{k,t}, \tag{4.44}$$

where the estimated coefficients $(\hat{c}_0, \hat{c}_1, ..., \hat{c}_K)$ are obtained by least squares regressions that include the cross-sectionally averaged terms above. Similarly, we can augment the macro-equations for the US with the appropriate cross-sectional averages based on US stocks and the UIP conditions with stock market cross-sections in order to estimate the relevant coefficients on the observed variables.

The second step in our procedure involves estimating the number of unobserved factors and the time series of the factors. We will apply the econometric procedure described in the theoretical part of this paper to the set of residuals estimated using the CCE estimator for US macroeconomic equations, international stock returns and UIP conditions. Using the eigenvalue minimum distance procedure introduced in this paper and the IV procedures for the estimation of the factor loadings we can identify and estimate the global factors.

Once we have identified the global factors we can estimate the structural parameters of the aggregate supply and demand equations in our US macroeconomic model. Under rational expectations future forecasts of macro-variables are uncorrelated with lagged information. Following Clarida and Gertler (1999), we can define the following moment conditions:

$$E_t \left\{ \left[\begin{pmatrix} \pi_t - \alpha\pi_{t+1} - (1-\alpha)\,\pi_{t-1} - \beta y_t - A'\hat{F}_t^G \\ y_t - \gamma y_{t+1} - (1-\gamma)\,y_{t-1} - \delta(i_t - \pi_{t+1}) - B'\hat{F}_t^G \end{pmatrix} \right] Z_{t-2}, Z_{t-3}, ... \right\} = 0, \tag{4.45}$$

for an appropriately chosen set of instruments Z_{t-2}, $Z_{t-3}...$ dated $t-2$ and earlier.

■ 4.4 Data

We construct a dataset consisting of macroeconomic variables for the US and financial variables for 8 countries: Australia, Canada, France, Germany, Hong Kong, Japan, UK and US at monthly frequency over the period 02/1973 to 09/2006. The financial variables consist of monthly closing stock prices in local currencies for a sample of firms in each country downloaded from Datastream. We restrict our attention to firms for which complete time series are available in order to avoid the biases resulting from missing observations in factor models, which tend to be very severe. In total we use data on 1680 firms divided as follows: Australia (52), Canada (29), France (35), Germany (82), Hong Kong (37), Japan (626), UK (209), US (610). We also use the exchange rates for the local currency with respect to the US dollar as provided by the Federal Reserve. The data on

the risk free rates is constructed as the return on short-term (1 or 3 months) government bills in the respective country and are provided by Global Financial. Additionally we use stock market indexes provided by Global Financial for each country in our sample.

For the US economy we use data on CPI-U, all items, US city average for urban consumers provided by the BLS, monthly unemployment as measured by the BLS and the end of month federal funds rate (overnight rate in the market for commercial reserves) as reported by the Federal Reserve Board. The first two variables are also seasonally adjusted.

■ 4.5 Empirical Results

We employ the methods outlined above to estimate the number of factors from a set of residuals obtained after partialling out the observed variables for each of the 3 sets of equations: US macroeconomic, international stocks, UIP/exchange rates. We estimated the number of factors using different assumptions on idiosyncratic error term such as independence (strict factor models) and weak autocorrelations (approximate factor model). The number of identified factors varied between 23 and 27. For the rest of this section we take the estimated number of factors to be 23.

This number of factors is much larger than the number of factors previously estimated in similar models. We perceive this to be a strength of our improved econometric methodology for the identification and estimation of factors in large panel data. The global economy is a complex phenomenon and the number of factors is thus appropriately large.

Recall that the estimated factors are properly speaking innovations. Thus, in Figure 4.1 we plot the cumulative (integrated) factors over the period 1973-2006. Notice how the cumulative effect of some factors is stronger than that of other factors. The factors with the strongest cumulative effect also tend to be the first few commonly identified factors, which we label as "strong" factors. Ultimately this is only a nominal issue however and we need to investigate the relationship of all the factors to the US economy.

An often misunderstood issue in using factor analysis concerns our ability to label the estimated factors. In Figure 4.2 we display the individual time series for the cumulative effect of the first 6 factors. In general it is not possible for us to relate each factor to a specific time series since the estimated factors will be functions of many different variables. Indeed, the estimation method itself is dependent on the identifying assumptions which fix the rotation of the factors. Therefore, a slightly different identification scheme would produce alternative time series profiles. Nevertheless, we hope to address this issue in future research by developing methods of relating the estimated factors to sets of international macroeconomic variables, thereby providing a more in-depth description of the
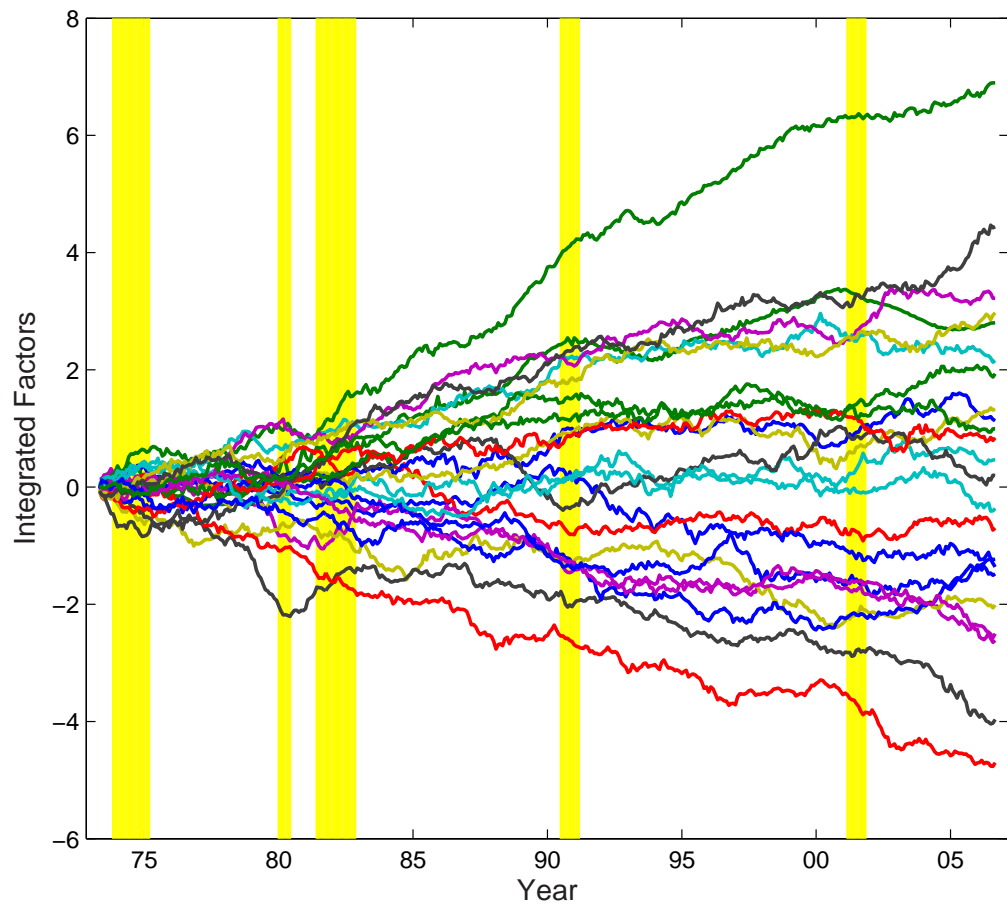
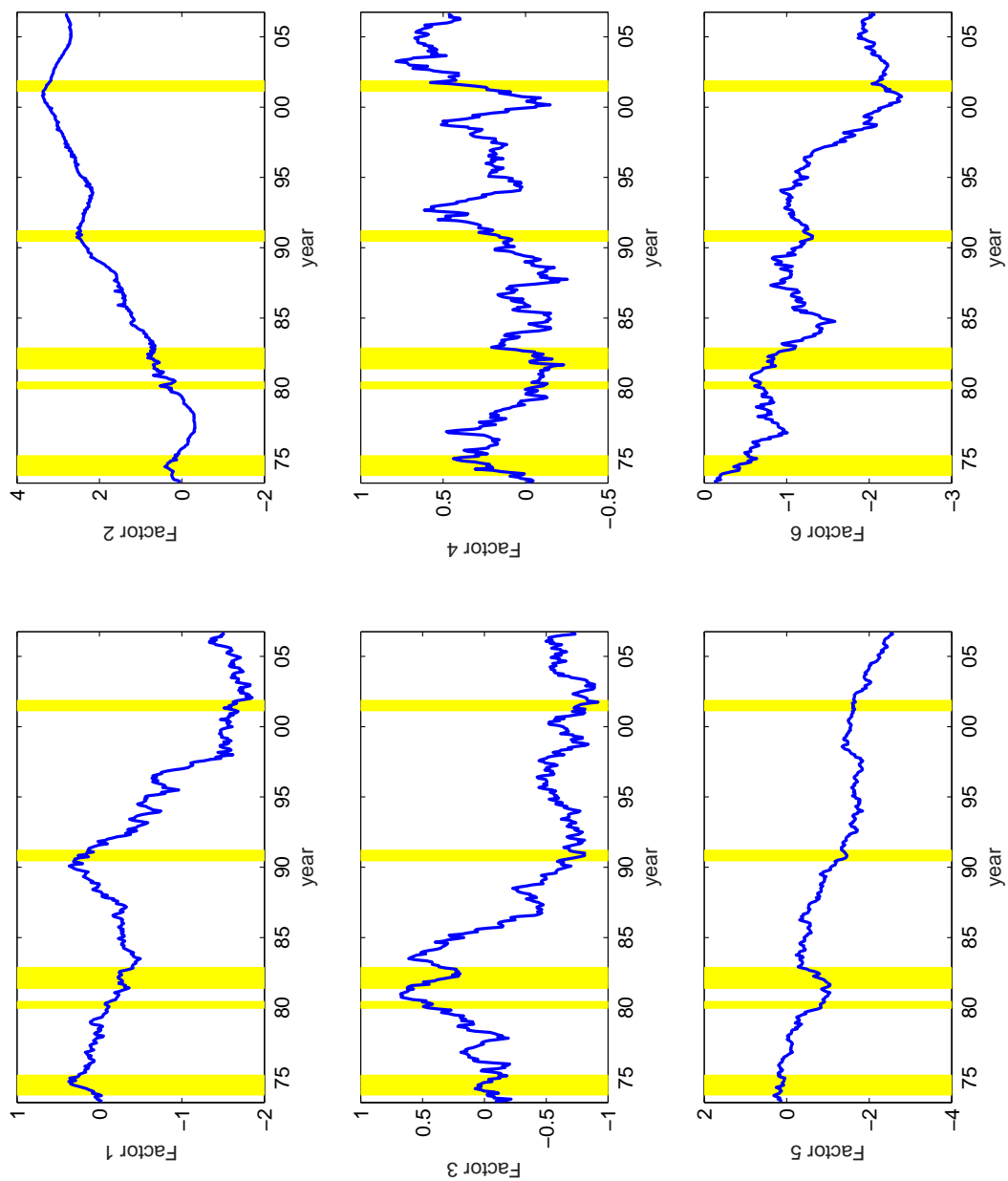**Figure 4-1.** Estimated Global Factors

Figure 4-2. First 6 Integrated Global Factors
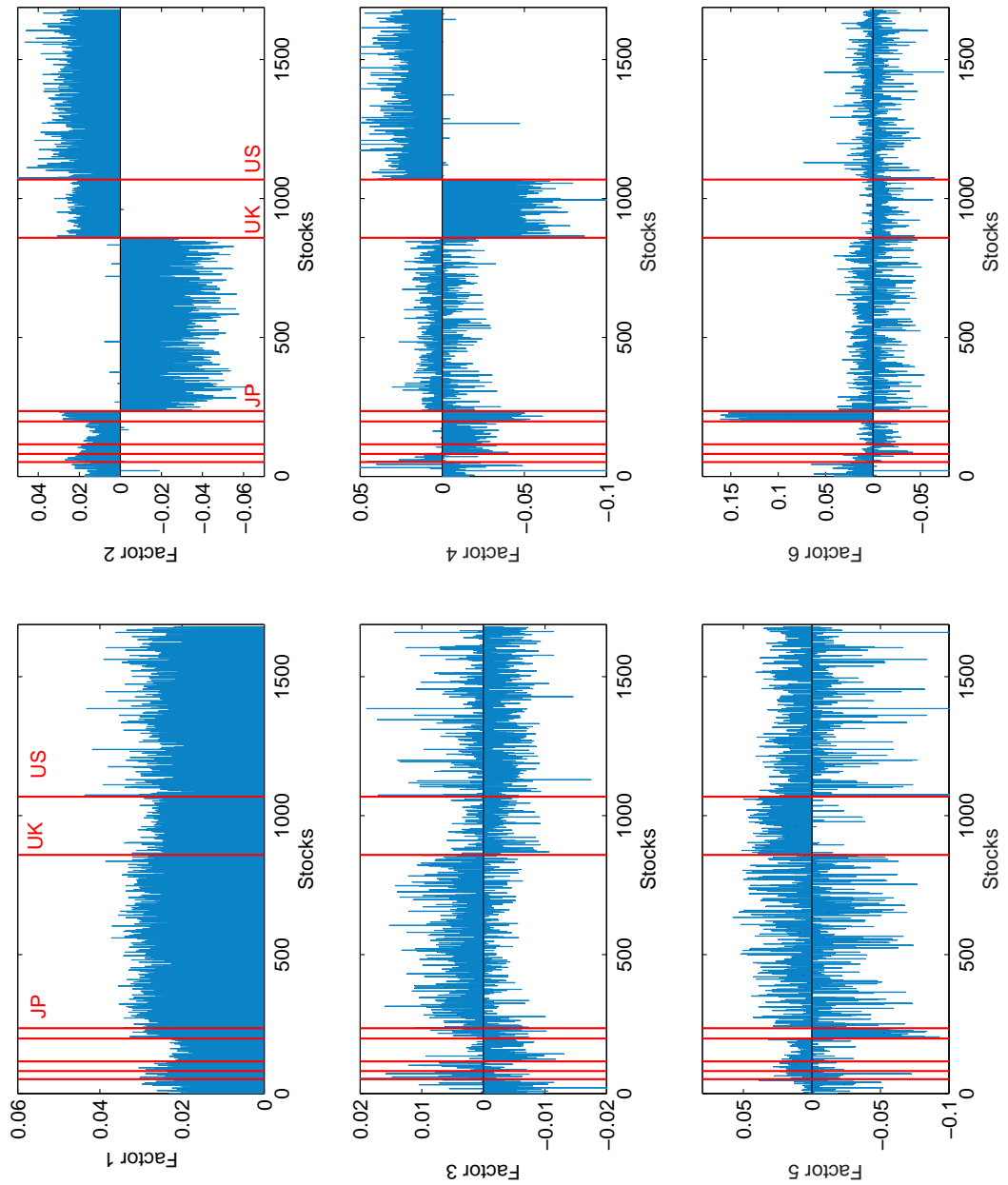
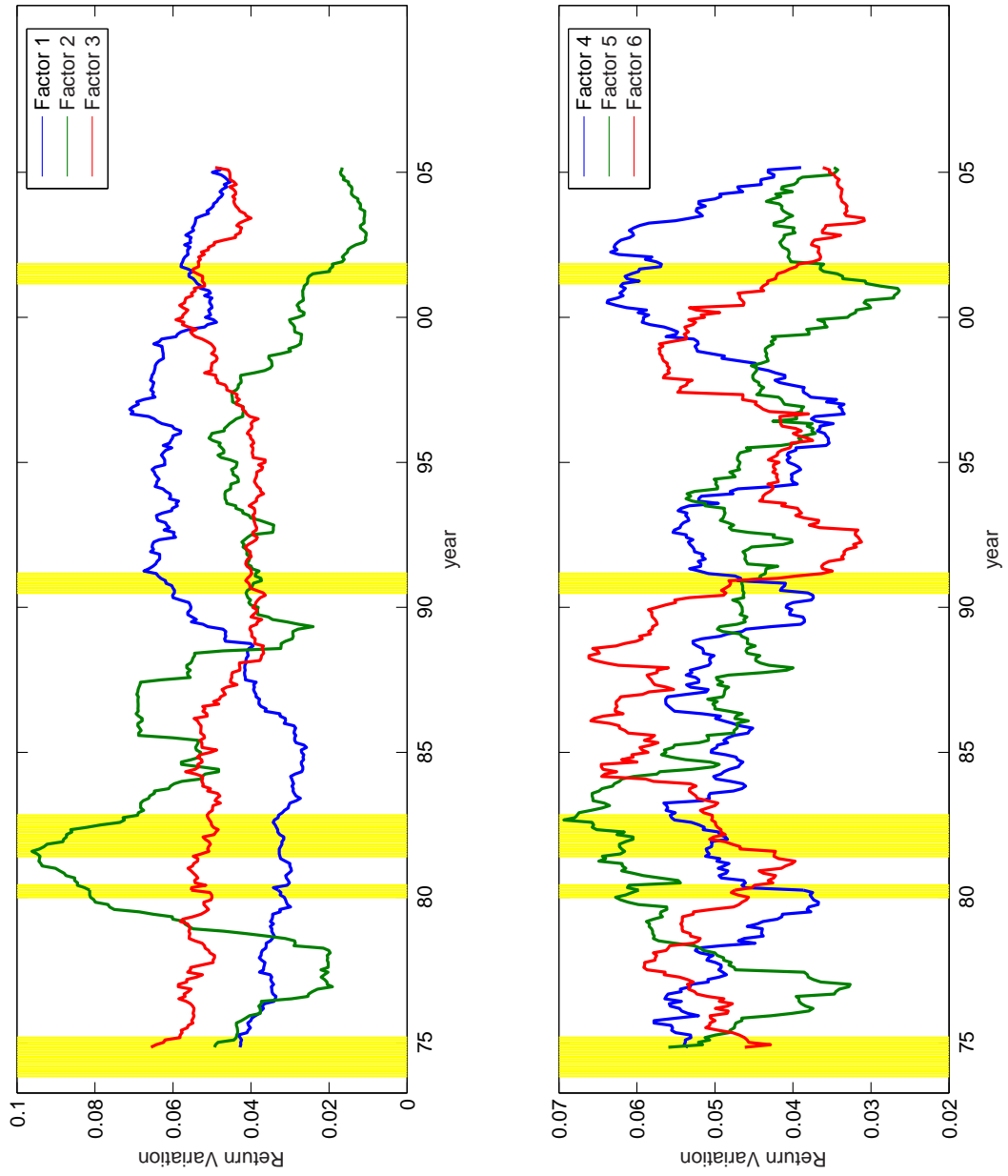**Figure 4-3.** Loadings on First 6 Global Factors

Figure 4-4. 3-year Rolling Standard Deviations

factors.

Nevertheless, we can learn more about the nature of the estimated global factors by analyzing their respective loadings. In Figure 4.3 we display the normalized loadings for each company included in the financial part of our model. The loadings are represented by vertical bars. The companies are ordered by country and each country is marked by a vertical line. We notice that the first panel indicates the presence of a global market factor which affects all companies thereby inducing an approximately uniform loading pattern. It is interesting to compare the factor loading pattern for factors 2 and 4. Factor 2 loads positively on all companies in countries other than Japan and negatively on Japanese companies. Factor 4 on the other hand loads positively on the US, negatively on the UK and has a mixed pattern for Japan. These charts illustrate that it is possible for global factors to impact companies in different countries in opposite directions. Additionally, they may have a very heterogeneous impact on companies in the same country. This indicates the complex nature of the international economy.

In Figure 4.4 we construct the 3-year rolling standard deviations for the estimated factors. This allows us to investigate whether some of the factors have increased or decreased in importance over the sample. Interestingly we find that most factors seem to exhibit a cyclical pattern rather than a time trend that would indicate their increased or decreased importance today compared to a few decades ago. Some exceptions such as factor 2 in the upper panel of Figure 4.4 are to be noted. Factor 2 for example peaked in importance during the first half of the 1980 and declined afterwards. This particular pattern may indicate that it is related to inflation.

In Figure 4.3b we have seen how the pattern of loadings varies across firms in different countries. This suggests that global shocks do not impact countries uniformly. Thus, they may increase the risk for companies in some countries while reducing it for companies in other countries. Moreover, the timing of the effect of the global factors may differ. This heterogeneity merits further investigation and in Figure 4.5 we address the issue of the synchronization of the effects of the global factors on US and international stocks.

In order to evaluate the extent to which the global shocks are synchronized we employ a statistical technique for the visualization of high-dimensional data known as an Andrews (1972) plot. We compute the mean loadings per factors for each of the 23 factors for all the stocks in the US and all the stocks in the rest of the world:

$$
\lambda_{\text{US}} = \left(\lambda_1^{\text{US}}, ..., \lambda_{23}^{\text{US}} \right) = \left\{ (1/J_{\text{US}}) \sum_{j=1}^{J_{\text{US}}} \lambda_{1,j}, (1/J_{\text{US}}) \sum_{j=1}^{J_{\text{US}}} \lambda_{2,j}, ..., (1/J_{\text{US}}) \sum_{j=1}^{J_{\text{US}}} \lambda_{23,j} \right\}
$$

(4.46)

and similarly for the stock returns of companies in countries other than the

US.

We can use these values as coefficients in the following function of parameter τ:

$$f_{\text{US}}(\tau) = \frac{\lambda_1^{\text{US}}}{\sqrt{2}} + \lambda_1^{\text{US}} \sin(2\pi\tau) + \lambda_2^{\text{US}} \cos(2\pi\tau) + \dots \tag{4.47}$$

The resulting functions $f_{\text{US}}(\tau)$ and $f_{\text{World}}(\tau)$ are then plotted for $\tau = [0, 1]$ in Figure 4.5. Andrews (1972) shows that this representation preserves a certain distance metric (proportional to the Euclidean distance) between the loadings matrices for the US and for the rest of the world. Thus, it can be used to easily compare the high-dimensional loadings matrices for the two sets of equations (US vs. World). We find a definite correlation between the two functions indicating a substantial degree of synchronization. Nevertheless, the synchronization is far from perfect indicating the some global shocks may affect the US differently from the rest of the world.

The recent literature on factor models often assumes that the latent factors are autocorrelated, implying a dynamic factor model. The factor model estimated above, however, does not make these assumptions. As Stock and Watson (2005) remark, the standard factor model is the static form of a dynamic model. That is, if a factor is dynamic we may expect to find both F_t and ρF_{t-1} identified as separate factors. This is however, subject, to the phase transition phenomenon discussed in Chapter 2 which limits the number of weak factors that can actually be identified in a given sample. Thus, even if a factor is dynamic, in a given sample we may only observe F_t and not ρF_{t-1} when the degree of autocorrelation ρ is small relative to the unobserved noise variance.

An informal visual test for the presence of dynamic factors can be constructed as follows. Consider the matrix of all correlations between factors at L lags and leads given by:

$$C = \begin{pmatrix} F_t^1 F_{t+L}^1 & & & & & & & & & \dots \\ \dots & & & & & & & & & \\ F_t^1 F_{t+1}^1 & & & & & & & & & \\ F_t^1 F_t^1 & F_t^1 F_t^2 & \dots & F_t^1 F_t^p & F_t^2 F_t^1 & \dots & F_t^2 F_t^p & \dots & F_t^p F_t^p \\ F_t^1 F_{t-1}^1 & & & & & & & & & \\ \dots & & & & & & & & & \\ F_t^1 F_{t-L}^1 & & & & & & & & & \dots \end{pmatrix}, \tag{4.48}$$

where entry $(L + 1, 2)$ corresponds to the correlation between factors 1 and 2 at the same period, while the entry $(1,1)$ corresponds to the correlation between factor 1 and factor 1 considered at L leads. In Figure 4.6 we plot this matrix of correlations for $L = 12$ lags and leads, thus providing a comprehensive visual image of correlations between factors and their potential dynamic effects. The resulting matrix C has dimension 25×23^2. We notice the central spikes corre-

sponding to the unit correlations between (F_t^k, F_t^k). While this is not a formal statistical test it nevertheless seems that some of the stronger factors also appear to be weakly autocorrelated. It was not possible for us to identify weak factors which can be thought of as lagged representations of the stronger factors. Thus, it seems that the weak factors are below the identification bound discussed in Chapter 2 and that all the 23 factors are in fact unrelated factors.

We complete the statistical characterization of the global factors by asking which sets of equations are the factors most strongly correlated to? Recall that the global factors are estimated from nearly 1700 equations. We divide the equations into 4 groups corresponding to exchange rates (UIP) equations, equations for international stock returns, equations for US stock returns and equations for the structural model of the US economy. Intuitively, this allows us to enquire as to the best use of these estimated factors, that is should we use these factors to explain exchange rates or to explain inflation? In Figure 4.7 we plot the mean factor loadings for each set of equations normalized on a [0,1] scale for each of the 4 sets. We find that the global factors carry little explanatory power for the exchange rates but may be substantially related to the US macroeconomic model. The extent to which this is the case differs by factor. Factor 3 seems to be very strongly related to the US economy, indicating that it may in fact correspond to the effect of the US economy on the world economy. All factors also seem to correspond to risk factors which are being priced by both the US market and the international financial markets.

Let us now return to the structural model of the US economy and estimate the structural coefficients in the aggregate supply and demand equations of the hybrid New Keynesian model with both forward and backward looking components. We employ two basic IV estimation procedures, GMM and Fuller and use as instruments either 4 or 6 lags of the macroeconomic variables. The recent macroeconomic literature has raised numerous questions about the estimation of the hybrid New Keynesian model, which are discussed in detail by Gali, Gertler and Lopez-Salido (2005). Most of the objections seem to concern the use of potentially weak instruments in the GMM estimation. In order to address these concerns we also report the estimates derived from the Fuller estimator.

The first two columns of Table 1 report the benchmark estimates of the structural model. Although we use unemployment instead of marginal cost or the output gap, the estimates are similar. In particular we find a dominant forward-looking behavior in both equations combined with an insignificant effect of unemployment in the Phillips curve and the real interest rate in the demand equations.

We next augment the existing specification with a factor proxy based on the Pesaran (2006) procedure and finally with all the estimated factors derived above. We notice that as we add the factors to the Phillips curve the results change substantially from the estimates without the global factors. Backward looking behavior becomes dominant and the coefficient on unemployment increases by
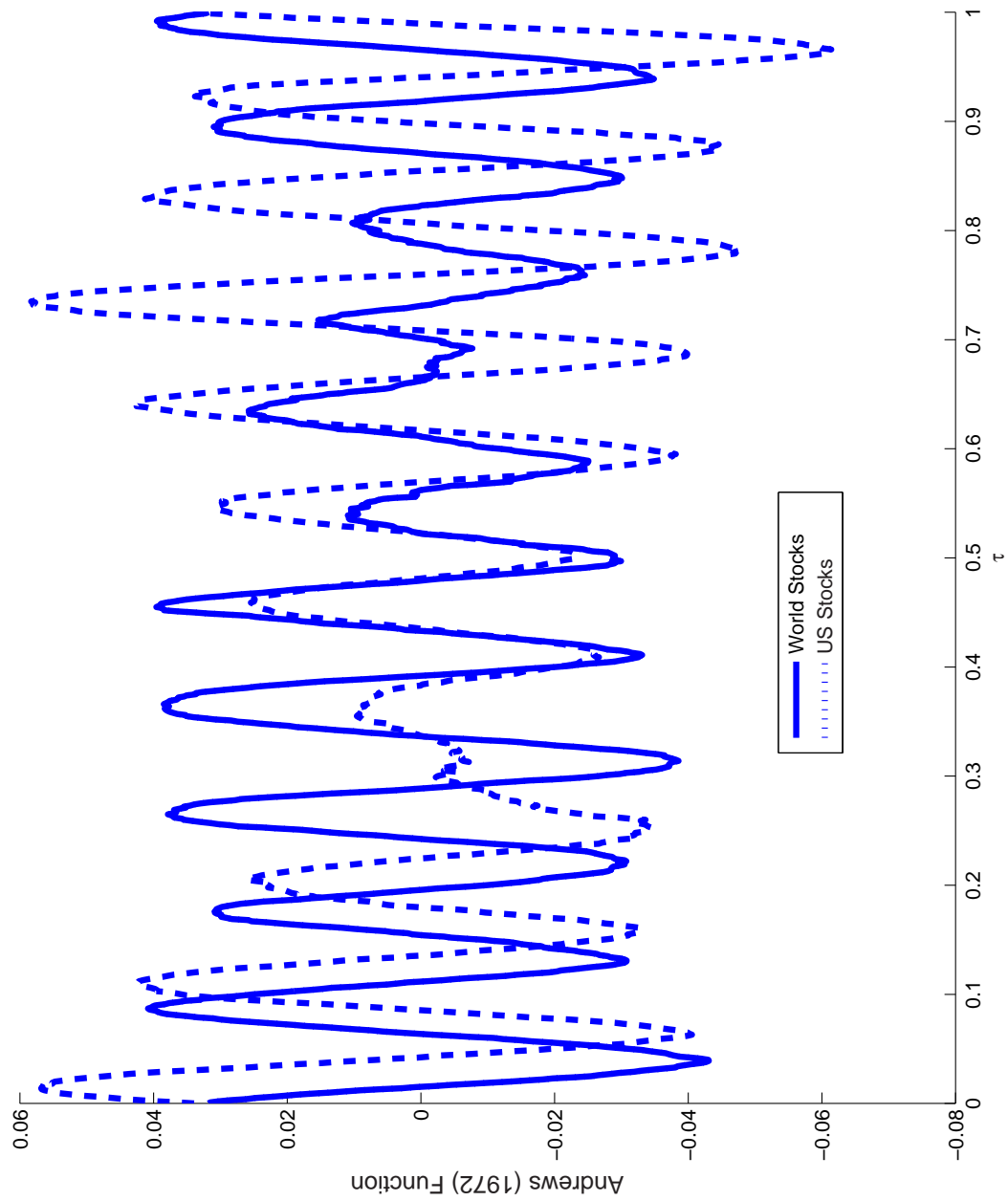
Figure 4-5. Synchronization of Global Factors
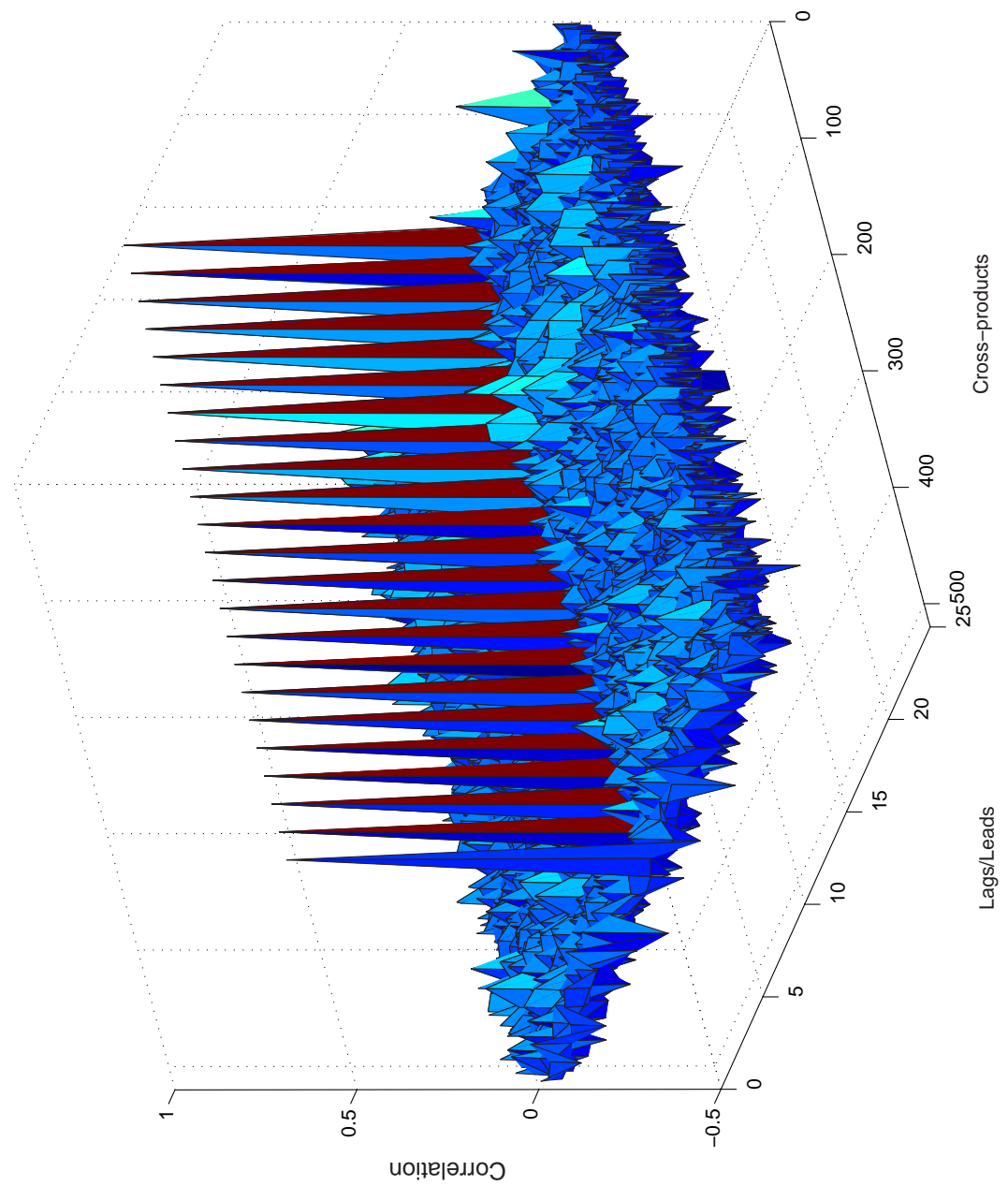
Figure 4-6. Cross-correlations of Factors Over 12 Leads/Lags
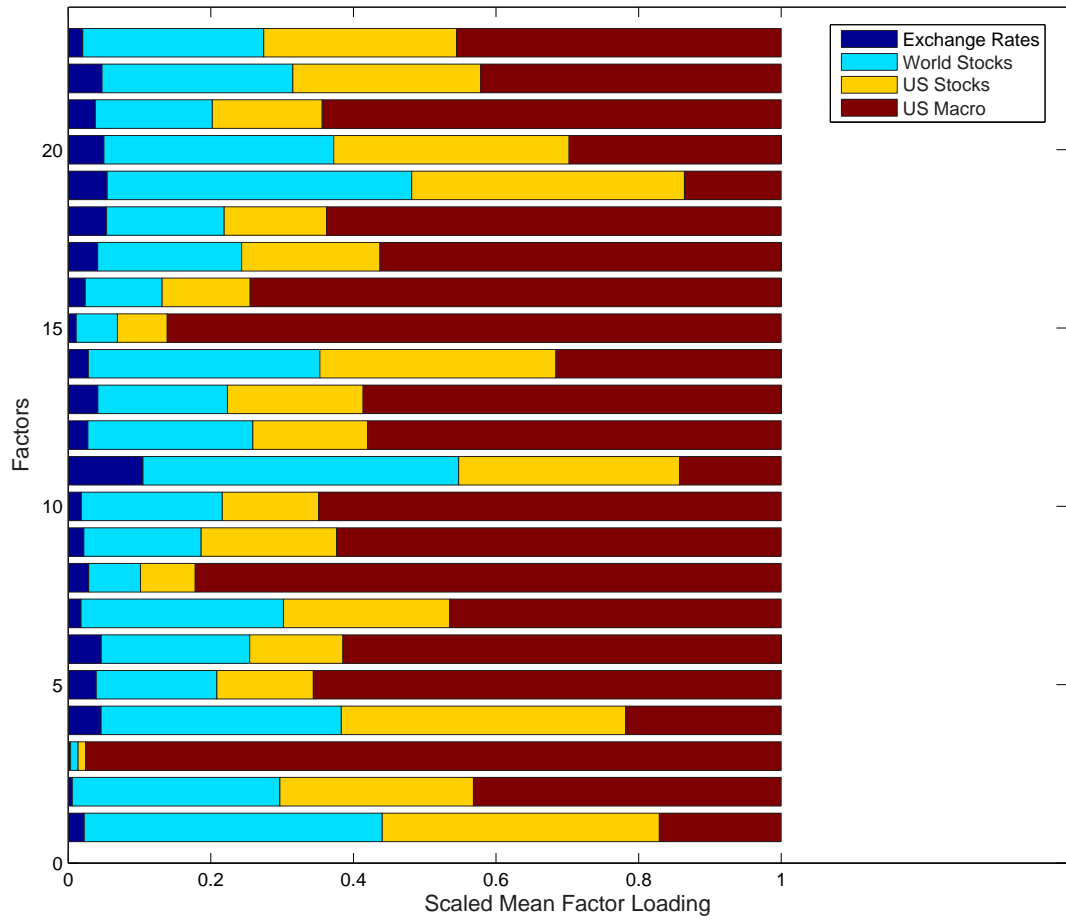
Figure 4-7. Relative Correlations of Global Factors with Model Equations
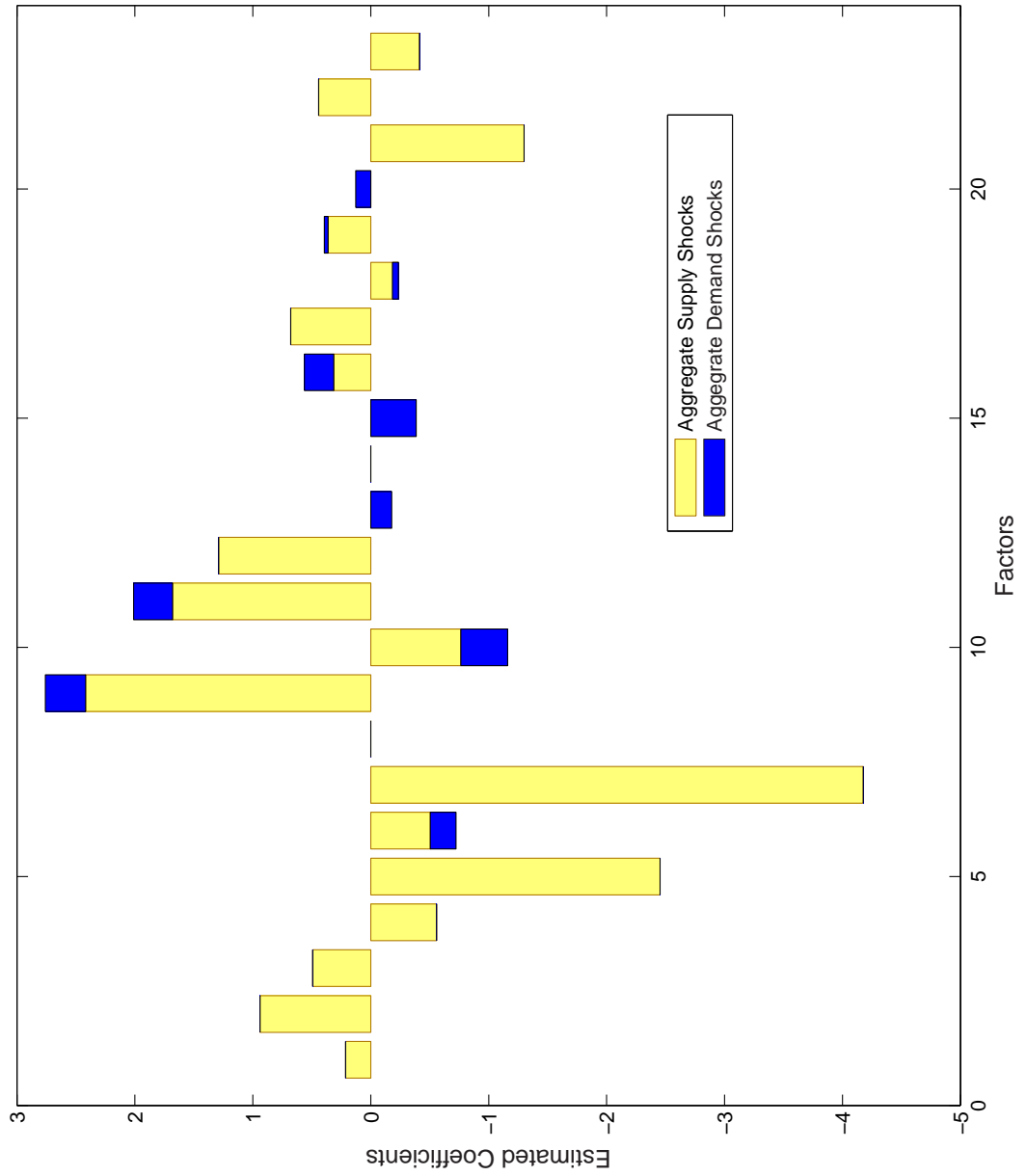
Figure 4-8. Statistically Significant Marginal Effects of Global Factors on the US Economy

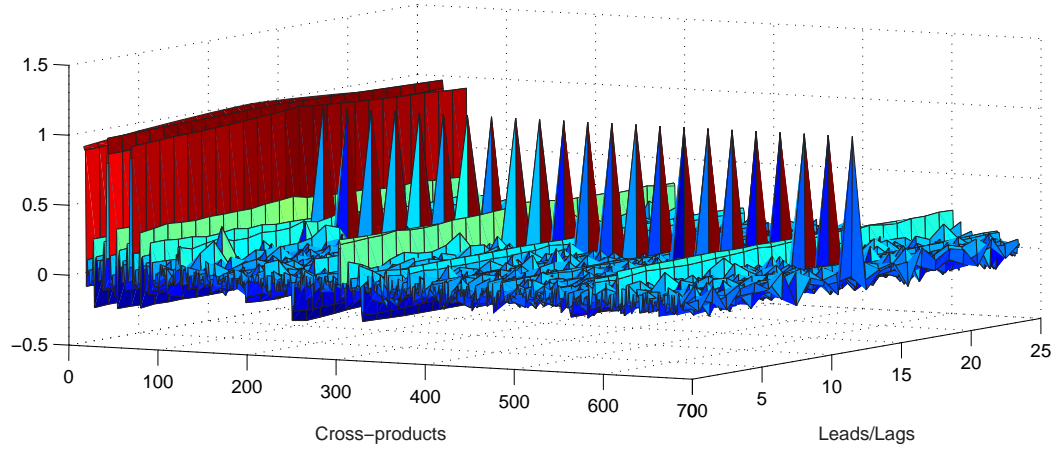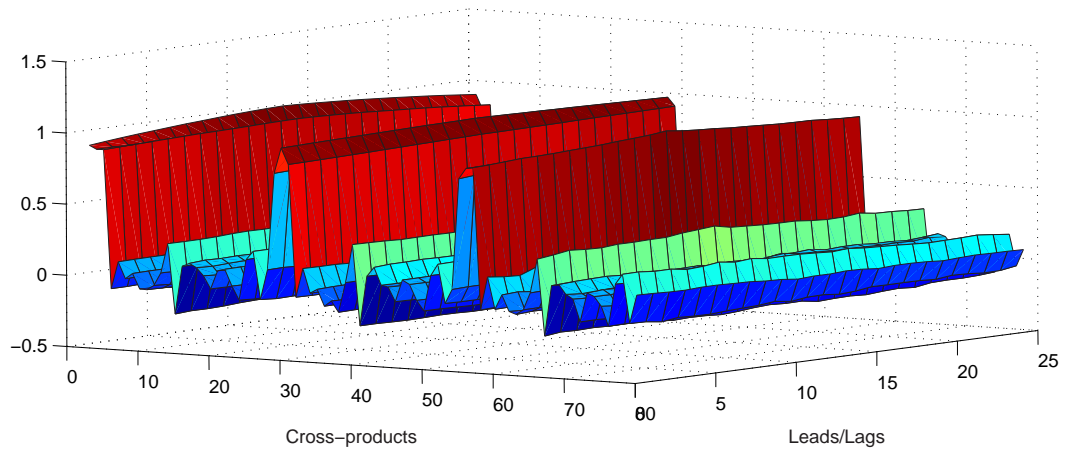Figure 4.9a: Cross–Correlations between US Macro Variables and Factors

Figure 4.9b: Detail –– Macro Variables

Figure 4-9. Relationship between Macro Variables and Factors at Different Leads/Lags

a factor of 16 to 18. We notice that the global factors carry very substantial explanatory power for inflation over the sampling period. The residual sum of squares (RSS) for the Phillips curve decreases by a factor of more than 4 as we add the global factors to the equation. The global factors also have a similar effect when added to the aggregate demand equation. In the benchmark estimates of the aggregate demand, the behavior is weakly forward looking while the effect of the real interest rate is small and insignificant. As we add the global factors the behavior switches to being weakly backward looking while the effect of the real interest rate increases by a factor of 10. The RSS also decreases by a small amount suggesting that global factors are much more important for characterizing US inflation than aggregate demand.

In Figure 4.8 we plot the estimated coefficients on the significant global factors as estimated in our structural model. The plot reveals that most of the global factors can be thought of as aggregate supply shocks, while only a minority corresponds to significant aggregate demand shocks. Moreover, their effect is not necessarily in opposite directions, thus indicating that at least some of the shocks can operate the same direction of both the supply and demand side in the US economy.

The results of Table 1 are surprising and reveal the strong effect of global factors in masking the true behavior of expectations and lags of inflation and unemployment as well the structural effect of unemployment in the Phillips curve and the real interest rate in aggregate demand. Our results show that this is not simply due to the presence of weak instruments as a number of previous authors have argued. The estimates for GMM and Fuller are very similar to each other. In fact, it seems that using lags of macroeconomic variables in the presence of global factors is inappropriate due to the global shocks leading to inconsistent estimates in the benchmark specification without global factors. In Figure 4.9 we plot the cross-correlation matrix described in equation 4.48 for both macroeconomic variables and estimated global factors. We find that both lags and leads of the macroeconomic variables are at least weakly correlated with the estimated global factors. Thus global factors appear to be driven at least in part by the previous macroeconomic performance of the US economy while at the same time affecting expectations of future performance. The benchmark specification leads to inconsistent estimates since the global factors are present in the error term but not explicitly modeled and therefore correlated with the instruments. Our specification of the Phillips curve and the aggregate demand augmented with the global factors, however, leads to consistent estimates since the global factors are explicitly taken into account in the equations.

Table 4.1. Estimation of the Structural Model with Global Factors

| Phillips Curve | IV 4 lags | | | | | | IV 6 lags | |
| | No Factors | | With Proxies | | With Factors | | With Factors | |
	Fuller	GMM	Fuller	GMM	Fuller	GMM	Fuller	GMM
Inflation (t+1)	0.530	0.539	0.519	0.528	0.139	0.131	0.182	0.143
	(0.053)*	(0.048)*	(0.054)*	(0.047)*	(0.091)	(0.062)*	(0.080)*	(0.056)*
Inflation (t-1)	0.472	0.462	0.482	0.473	0.86	0.867	0.817	0.854
	(0.053)*	(0.048)*	(0.053)*	(0.047)*	(0.090)*	(0.062)*	(0.079)*	(0.056)*
Unemployment (t)	-0.001	-0.001	-0.005	-0.004	-0.018	-0.018	-0.016	-0.017
	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)*	(0.002)*	(0.003)*	(0.002)*
RSS	16.337	16.434	16.179	16.248	3.751	3.758	3.812	3.731

| Aggregate Demand | IV 4 lags | | | | | | IV 6 lags | |
| | No Factors | | With Proxies | | With Factors | | With Factors | |
	Fuller	GMM	Fuller	GMM	Fuller	GMM	Fuller	GMM
Unemployment (t+1)	0.518	0.512	0.528	0.524	0.437	0.427	0.448	0.420
	(0.041)*	(0.043)*	(0.045)*	(0.048)*	(0.042)*	(0.043)*	(0.043)*	(0.043)*
Unemployment (t-1)	0.482	0.489	0.473	0.478	0.561	0.572	0.550	0.578
	(0.041)*	(0.043)*	(0.045)*	(0.047)*	(0.042)*	(0.043)*	(0.043)*	(0.043)*
Real Interest Rate (t)	-0.050	-0.070	0.046	0.040	-0.461	-0.582	-0.502	-0.525
	(0.258)	(0.261)	(0.297)	(0.300)	(0.276)	(0.282)*	(0.276)	(0.264)*
RSS	5.637	5.634	5.650	5.649	4.585	4.617	4.525	4.585

■ 4.6 Conclusion

Factor models have traditionally played a central role in empirical macroeconomics and finance, and recently the question of how global factors influence US economic performance has been at the forefront of policymakers' attention. The proposed new econometric procedures are used to estimate a structural macroeconomic model which provides many insights into the nature of global supply shocks and their effect on the US economy. In particular we show that ignoring the role of global factors can mask the effect of unemployment in the Phillips curve and of the real interest rate in the aggregate demand relationship.

Bibliography

ANDREWS, D. F. (1972): "Plots of High-Dimensional Data," *Biometrics*, 28, 125–136.

ANG, A., AND M. PIAZZESI (2003): "A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables," *Journal of Monetary Economics*, 50, 745–787.

BAI, J. (2003): "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71, 135–171.

——— (2005): "Panel Data Models with Interactive Fixed Effects," mimeo.

BAI, J., AND S. NG (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70(1), 191–221.

BEKAERT, G., S. CHO, AND A. MORENO (2006): "New-Keynesian Macroeconomics and the Term Structure," mimeo.

BEKAERT, G., R. J. HODRICK, AND X. ZHANG (2005): "International Stock Return Comovements," NBER WP 11906.

BERNANKE, B. S. (2006): "Reflections on the Yield Curve and Monetary Policy," Remarks before the Economic Club of New York, March.

BIBKOV, R. (2005): "Monetary Regimes and the Term Structure of Interest Rates," mimeo.

BURMEISTER, E., AND M. B. MCELROY (1988): "Joint Estimation of Factor Sensitivities and Risk Premia for the Arbitrage Pricing Theory," *Journal of Finance*, 43(3), 721–733.

CARLIN, W., AND D. SOSKICE (2006): *Macroeconomics: Imperfections, Institutions and Policies*. Oxford University Press, Oxford.

CHAMBERLAIN, G., AND M. ROTHSCHILD (1983): "Arbitrage, Factor Structure and Mean-Variance Analysis in Large Asset Markets," *Econometrica*, 51(5), 1305–324.

CLARIDA, R., J. GALI, AND M. GERTLER (1999): "The Science of Monetary Policy: A New Keynesian Perspective," *Journal of Economic Literature*, 37(4), 661–707.

CONNOR, G., AND R. A. KORAJCZYK (1986): "Performance Measurement with the Arbitrage Pricing Theory: A New Framework for Analysis," *Journal of Financial Economics*, 15, 373–394.

DRIESSEN, J., B. MELENBERG, AND T. NIJMAN (2003): "Common Factors in International Bond Returns," *Journal of International Money and Finance*, 22, 629–656.

FUHRER, J. C., AND MOORE (1995): "Inflation Persistence," *Quarterly Journal of Economics*, 440, 127–159.

GALI, J., M. GERTLER, AND J. D. LOPEZ-SALIDO (2005): "Robustness of the Estimates of the Hybrid New Keynesian Phillips Curve," *Journal of Monetary Economics*, 52, 1107–1118.

IKEDA, S. (1991): "Arbitrage Asset Pricing under Exchange Risk," *Journal of Finance*, 46(1), 447–455.

JOLLIFFE, I. T. (2002): *Principal Component Analysis*. Springer, New York.

MADANSKY, A. (1964): "Instrumental Variables in Factor Analysis," *Psychometrika*, 29(2), 105–113.

MANKIW, N. G., AND R. REIS (2002): "Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve," *Quarterly Journal of Economics*, 117(4), 1295–1328.

MILLER, G. (2006): "Factor Models and Fundamentalism," Morgan Stanley Capital International - BARRA, Publication 181.

NEWEY, W., AND F. WINDMEIJER (2005): "GMM with Many Weak Moment Conditions," mimeo.

PESARAN, M. H. (2006a): "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure," *Econometrica*, 74(4), 967–1013.

——— (2006b): "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure," *Econometrica*, 74(4), 967–1013.

ROBINSON, P. M. (1974): "Identification, Estimation and Large-Sample Theory for Regressions Containing Unobservable Variables," *International Economic Review*, 15(3), 680–692.

ROGOFF, K. (2006): "Impact of Globalization on Monetary Policy," mimeo.

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

SOLNIK, B. (1982): "International Asset Pricing Theory," *Journal of Finance*, 38(1), 449–457.

STOCK, J., AND M. WATSON (2003): "Forecasting Output and Inflation: The Role of Asset Prices," *Journal of Economic Literature*, 41, 788–829.

——— (2005): "Implications of Dynamic Factor Models for VAR Analysis," mimeo.

——— (2006): "Forecasting with Many Predictors," in *Handbook of Economic Forecasting*, ed. by G. Elliott, C. J. Granger, and A. Timmermann. Elsevier, Amsterdam.

CHAPTER 5

Estimating the Random Coefficients Logit Model

Understanding discrete economic choices is an important aspect of modern economics. McFadden (1974) introduced the *multinomial logit* model as a model of choice behavior derived from a random utility framework. An individual i faces the choice between K different goods $i = 1..K$. The utility to individual i from consuming good j is given by $U_{ij} = x'_{ij}\beta + \epsilon_{ij}$, where x'_{ij} corresponds to a set of choice relevant characteristics specific to the consumer-good pair (i, j). The error component ϵ_{ij} is assumed to be independently identically distributed with an extreme value distribution $f(\epsilon_{ij}) = \exp(-\epsilon_{ij})\exp(-\exp(-\epsilon_{ij}))$.

If individual i is constrained to choose a single good within the available set, utility maximization implies that some good j will be chosen over all other goods $l \neq j$ such that $U_{ij} > U_{il}$, for all $l \neq j$. We are interested in deriving the probability that consumer i chooses good j, which is

$$P_{ij} = \Pr[x'_{ij}\beta + \epsilon_{ij} > x'_{il}\beta + \epsilon_{il}, \text{for all } l \neq j]. \tag{5.1}$$

McFadden (1974) shows that the resulting integral can be solved in closed form implying the familiar expression:

$$P_{ij} = \frac{\exp(x'_{ij}\beta)}{\sum\limits_{k=1}^{K} \exp(x'_{ik}\beta)} \left(= s_{ij}\right). \tag{5.2}$$

In some analyses it is also useful to think of the market shares of different firms. Without loss of generality we can also consider the choice probability described above to be the share of the total market demand which goes to good j in market i

and we will denote this by s_{ij}. All the results derived in this chapter will be valid for either interpretation. For convenience we shall focus on the market shares interpretation of the above equation.

The vector of coefficients β can be thought of as a representation of the individual tastes and determines the choice, conditional on the observable consumer-good characteristics. Although an extremely useful model, the multinomial model suffers from an important limitation: it is built around the the assumption of independence of irrelevant alternatives (IIA), which implies equal cross price elasticities across all choices, as demonstrated by Hausman (1975). Additionally, it does not allow for correlations between the random components of utility, thus limiting the complexity of individual choice which can be modeled (Hausman and Wise, 1978).

While a number of more flexible specifications have been proposed, few proved to be computationally tractable. The addition of a random coefficients framework to the logit model provides an attractive alternative (Cardell and Dunbar, 1980). In many applications however it is important to think of tastes as varying in the population of consumers according to a distribution $F(\beta)$. It is particularly important not to assume the taste parameters to be independent. The estimation of correlations between the components of the vector β is of major interest. The resulting correlations describe patterns of substitution between different product characteristics.

In practice, we often assume that the distribution $F(\beta)$ is Normal with mean b and covariance Σ. The purpose of random coefficients models is to estimate the unknown parameters b and Σ from the available sample. From a computational point of view, the aim is to obtain the expected share of good j in market i from the evaluation of the following expectation:

$$E_{\beta}(s_{\text{ij}}) = \int\limits_{-\infty}^{+\infty} \frac{\exp(x'_{\text{ij}}\beta)}{\sum\limits_{k=1}^{K} \exp(x'_{\text{ik}}\beta)} dF(\beta) \tag{5.3}$$

We denote this model to be the random coefficients logit model. The above expression corresponds to a multivariate integral over the dimension of the space of the taste parameters. Since the integral does not have a known analytic solution, the use of simulation methods currently plays an important part in the implementation of these models (Lerman and Manski, 1981) with recent applications employing pseudo-random Halton sequences (Small, Winston and Yan, 2005; Train, 2003).

The random coefficients logit model is an extremely versatile tool for the analysis of discrete choices since it can be thought of as an arbitrarily close approximate representation of any random utility model consistent with choice probabilities (McFadden and Train, 2000). This has prompted researchers to think of this

model as "one of the most promising state of the art discrete choice models" (Hensher and Green, 2003). Applications of the random coefficients logit model abound, not only within economics, but also in related disciplines such as marketing or transportation research (Hess and Polak, 2005). The random coefficients model is also an important building block for more complex models. Thus, Berry, Levinsohn and Pakes (1995) employ the random coefficients logit model to analyse demand based on market-level price and quantity data. Bajari, Hong and Ryan (2005) incorporate it into an econometric model of discrete games with perfect information, where it selects the probability of different equilibria.

The implementation of the random coefficients model remains a challenging application of the method of simulated moments. In particular the estimation of a full covariance matrix of the taste parameters, which fully incorporates all the possible correlations between parameters, seems to elude most researchers and appears to be a serious limitation of the simulation approach. In Section 5.1 of this chapter we will derive an analytic approximation of the integral expression in Equation 5.3 which can be incorporated into an extremely convenient non-linear least squares framework for the estimation of all mean and variance-covariance parameters of the taste distribution. Section 5.2 shows the superior performance of the new method based on the Laplace method compared to the simulation alternative in cases where the model is specified with non-zero correlations.

■ 5.1 A Laplace Approximation of the Expected Share

Consider the expected share of product j in market i under the random coefficients logit model introduced above.

$$E_\beta(s_{ij}) = E_\beta \left\{ \frac{\exp(x'_{ij}\beta)}{\sum\limits_{k=1}^{K} \exp(x'_{ik}\beta)} \right\} = E_\beta \left\{ \left(\sum_{k=1}^{K} \exp(x'_{ijk}\beta) \right)^{-1} \right\}, \qquad (5.4)$$

where $x_{ijk} = x_{ik} - x_{ij}$ for all k. Assume that the taste parameters β are drawn from a normal multivariate distribution with mean b and covariance matrix Σ,

$$f(\beta) = (2\pi)^{-p/2} \left| \Sigma \right|^{-1/2} \exp\left\{ -\frac{1}{2}(\beta - b)' \Sigma^{-1} (\beta - b) \right\}. \qquad (5.5)$$

For simplicity we focus in our derivations on the case where all coefficients are random. More generally, we may wish to allow for mixture of fixed and random coefficients. The results in this chapter will continue to hold in this case too and we restate the main result of this chapter in terms of both random and fixed coefficients in Appendix B.

Then the expected share is given by the following multivariate integral:

$$E_\beta(s_{ij}) = (2\pi)^{-p/2} \left| \Sigma \right|^{-1/2} \int\limits_{-\infty}^{+\infty} \exp[-g(\beta)]d\beta, \text{ where} \qquad (5.6)$$

$$g(\beta) = \frac{1}{2}(\beta - b)' \Sigma^{-1}(\beta - b) + \log\left(\sum_{k=1}^{K} \exp(x'_{ijk}\beta)\right) \qquad (5.7)$$

In this section we provide an approximation to the integral expression above using the asymptotic method of Laplace. While univariate applications of this method are common to mathematics and physics, where they are routinely applied to the complex functions in order to derive "saddle-point approximations", few applications to econometrics or statistics have been attempted. The extension of the method to multivariate settings was developed by Hsu (1948) and Glynn (1980). A statement of the main theorem is given in Appendix A together with the technical conditions required for the approximation to exist. Statistical applications of the Laplace approximation were developed by Daniels (1954) and Barndorff-Nielsen and Cox (1979) who employ the Laplace approximation to derive the *indirect Edgeworth expansion*, a generalization of the Edgeworth expansion method for distributions to exponential families. The Laplace method was also applied in Bayesian statistics to derive approximations to posterior moments and distributions (Tierney and Kadane, 1986; Efstathiou, Guthierrez-Pena and Smith, 1998). More recently, Butler (2002) noticed that the Laplace approximation often produces accurate results in sub-asymptotic situations which are not covered by the traditional setting. It is this insight which we will use below.

Now perform a Taylor expansion of the function $g(\beta)$ around the point $\tilde{\beta}_{ij}$, such that $g(\tilde{\beta}_{ij}) < g(\beta)$ for all $\beta \neq \tilde{\beta}$. This expansion is given by:

$$g(\beta) \cong g(\tilde{\beta}_{ij}) + \left(\beta - \tilde{\beta}_{ij}\right)' [\frac{\partial g}{\partial \beta}|_{\beta = \tilde{\beta}_{ij}}] +$$
$$+ \frac{1}{2}\left(\beta - \tilde{\beta}_{ij}\right)' [\left(\frac{\partial^2 g(\beta)}{\partial\beta\partial\beta'}\right)_{\beta=\tilde{\beta}_{ij}}] \left(\beta - \tilde{\beta}_{ij}\right) + O(\left(\beta - \tilde{\beta}_{ij}\right)^3). \quad (5.8)$$

Substituting in the integral expression above we obtain:

$$E_\beta(s_{ij}) \cong |\Sigma|^{-1/2} \exp(-g(\tilde{\beta}_{ij})) \times$$
$$\int\limits_{-\infty}^{+\infty} (2\pi)^{-p/2} \exp\left\{-\frac{1}{2}\left(\beta - \tilde{\beta}_{ij}\right)' [\left(\frac{\partial^2 g(\beta)}{\partial\beta\partial\beta'}\right)_{\beta=\tilde{\beta}_{ij}}] \left(\beta - \tilde{\beta}_{ij}\right) + O(\left(\beta - \tilde{\beta}_{ij}\right)^3)\right\} d\beta$$

The intuition for this approach is given by the fact that if $g(\beta)$ has a minimum

at the point $\tilde{\beta}_{ij}$, then the contribution of the function $g(\beta)$ to the exponential integral will be dominated by a small region around the point $\tilde{\beta}_{ij}$. Furthermore by using a second order Taylor expansion around $\tilde{\beta}_{ij}$, we make the further assumption that the higher order terms of the expansion may be safely ignored. Let $\tilde{\Sigma}_{ij}$ be the inverse of the Hessian of $g(\beta)$ evaluated at $\tilde{\beta}_{ij}$, i.e. $\tilde{\Sigma}_{ij}^{-1} = \left(\frac{\partial^2 g(\beta)}{\partial \beta \partial \beta'}\right)_{\beta=\tilde{\beta}_{ij}}$. Note that both $\tilde{\beta}_{ij}$ and $\tilde{\Sigma}_{ij}$ are indexed by i and j to remind us that these values depend on the covariates of product j in market i explicitly and in general will not be constant across products or markets.

Then, we can re-write the integral above as:

$$E_\beta(s_{ij}) \cong |\Sigma|^{-1/2} \exp(-g(\tilde{\beta}_{ij})) \left|\tilde{\Sigma}_{ij}\right|^{1/2} (2\pi)^{-p/2} \left|\tilde{\Sigma}_{ij}\right|^{-1/2}$$
$$\int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2}\left(\beta - \tilde{\beta}_{ij}\right)' \tilde{\Sigma}_{ij}^{-1}\left(\beta - \tilde{\beta}_{ij}\right)\right\} d\beta. \quad (5.9)$$

We recognize the right hand side of this expression to be the Gaussian integral, that is the integral over the probability density of a Normal variable β with mean $\tilde{\beta}_{ij}$ and covariance $\tilde{\Sigma}_{ij}$. Since this area integrates to 1 we have,

$$(2\pi)^{-p/2} \left|\tilde{\Sigma}_{ij}\right|^{-1/2} \left[\int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2}\left(\beta - \tilde{\beta}_{ij}\right)' \tilde{\Sigma}_{ij}^{-1}\left(\beta - \tilde{\beta}_{ij}\right)\right\} d\beta\right] = 1 \quad (5.10)$$

and we can write the expected share of product i in market j as

$$E_\beta(s_{ij}) \cong \sqrt{\frac{\left|\tilde{\Sigma}_{ij}\right|}{|\Sigma|}} \exp(-g(\tilde{\beta}_{ij})) \quad (5.11)$$

The expansion point $\tilde{\beta}_{ij}$ has to be chosen optimally for each share, that is $\tilde{\beta}_{ij}$ solves the equation $g'(\beta)|_{\beta=\tilde{\beta}_{ij}} = 0, i.e.$

$$\left(\tilde{\beta}_{ij} - b\right)' \Sigma^{-1} + \sum_{k=1}^{K}\left\{x'_{ijk} \frac{\exp(x'_{ijk}\tilde{\beta}_{ij})}{\sum_{k=1}^{K} \exp(x'_{ijk}\tilde{\beta}_{ij})}\right\} = 0 \quad (5.12)$$

In Appendix B we show that -g(β) is the sum of two strictly concave functions and thus it is also concave. Hence, the function $g(\beta)$ attains a unique minimum at the point $\tilde{\beta}_{ij}$. We can also think of the optimal expansion point $\tilde{\beta}_{ij}$ as solving

a fixed-point equation, $\tilde{\beta}_{ij} = B(\tilde{\beta}_{ij})$, where

$$B(\tilde{\beta}_{ij}) = b' - \left[\sum_{k=1}^{K} \left\{ x'_{ijk} \frac{\exp(x'_{ijk}\tilde{\beta}_{ij})}{\sum\limits_{k=1}^{K} \exp(x'_{ijk}\tilde{\beta}_{ij})} \right\} \right] \Sigma. \tag{5.13}$$

Additionally, the Hessian of $g(\beta)$ is given by

$$\frac{\partial^2 g(\beta)}{\partial \beta \partial \beta'} = \Sigma^{-1} + \frac{\sum\limits_{k=1}^{K} x_{ijk} x'_{ijk} \exp(x'_{ijk}\tilde{\beta}_{ij})}{\sum\limits_{k=1}^{K} \exp(x'_{ijk}\tilde{\beta}_{ij})} - \frac{\left[\sum\limits_{k=1}^{K} x_{ijk} \exp(x'_{ijk}\tilde{\beta}_{ij}) \right] \left[\sum\limits_{k=1}^{K} x'_{ijk} \exp(x'_{ijk}\tilde{\beta}_{ij}) \right]}{\left[\sum\limits_{k=1}^{K} \exp(x'_{ijk}\tilde{\beta}_{ij}) \right]^2}$$

$$\tag{5.14}$$

The following proposition summarizes the main result of this chapter by approximating the Gaussian integral corresponding to the expected share of product i in market j using a Laplace approximation.

Proposition 5.1: *If β has a Normal distribution with mean b and covariance Σ, we can approximate $E_\beta(s_{ij}) = E_\beta\{(\sum\limits_{k=1}^{K} \exp(x'_{ijk}\beta))^{-1}\}$ by*

$$E_\beta(s_{ij}) \cong \sqrt{\frac{|\tilde{\Sigma}_{ij}|}{|\Sigma|}} \exp\left\{ -\frac{1}{2}\left(\tilde{\beta}_{ij} - b\right)' \tilde{\Sigma}_{ij}^{-1}(\tilde{\beta}_{ij} - b) \right\} \left(\sum_{k=1}^{K} \exp(x'_{ijk}\tilde{\beta}_{ij}) \right)^{-1}, \tag{5.15}$$

where
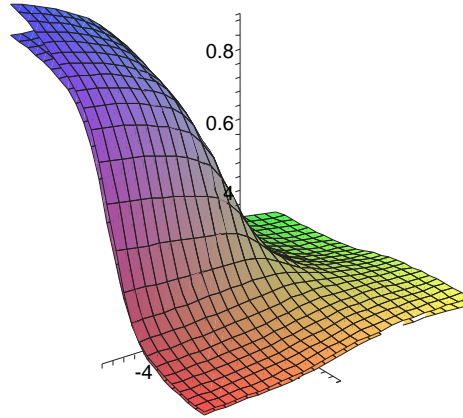
$$\tilde{\Sigma}_{ij} = \left\{ \Sigma^{-1} + \frac{\sum\limits_{k=1}^{K} x_{ijk} x'_{ijk} \exp(x'_{ijk}\tilde{\beta}_{ij})}{\sum\limits_{k=1}^{K} \exp(x'_{ijk}\tilde{\beta}_{ij})} - \frac{\left[\sum\limits_{k=1}^{K} x_{ijk} \exp(x'_{ijk}\tilde{\beta}_{ij}) \right] \left[\sum\limits_{k=1}^{K} x'_{ijk} \exp(x'_{ijk}\tilde{\beta}_{ij}) \right]}{\left[\sum\limits_{k=1}^{K} \exp(x'_{ijk}\tilde{\beta}_{ij}) \right]^2} \right\}^{-1},$$

$$\tag{5.16}$$

and $\tilde{\beta}_{ij}$ solves the fixed-point equation $\tilde{\beta}_{ij} = B(\tilde{\beta}_{ij})$ for

$$B(\tilde{\beta}_{ij}) = b' - \left[\sum_{k=1}^{K} \left\{ x'_{ijk} \frac{\exp(x'_{ijk}\tilde{\beta}_{ij})}{\sum\limits_{k=1}^{K} \exp(x'_{ijk}\tilde{\beta}_{ij})} \right\} \right] \Sigma. \tag{5.17}$$

In the next section we present detailed simulation results which show the performance of the approximation in estimating the unknown parameters b and

Figure 5-1. Comparison of expected share obtained by numerical integration and the corresponding Laplace approximation for a model with 2 covariates at fixed values of b and Σ.

Σ of the model. The figure below shows the remarkably good fit between of the Laplace approximation of the true market share at fixed values of b and Σ for two covariates.

The exact expected share obtained by numerical integration coincides with the expected share obtained by the Laplace approximation almost everywhere. The only noticeable deviation occurs for values of the expected share close to 1. Fortunately, this case is relatively infrequent in economic applications, where in multi-brand competition models we may expect to have many small shares in any given market but it is unlikely to have more than a few very large shares in the entire sample. The Laplace approximation introduced in this section has the peculiar property of being an asymmetrical approximation to a symmetrical function. This feature however proves to be extremely useful for economic applications since it provides an very close approximation to small shares which are much more likely to occur in economic data than shares close to 1, where the approximation tends to underestimate the true expected share.

The optimal expansion point $\tilde{\beta}_{ij}$ used in Proposition 5.1 can be computed by standard iterative methods which solve the fixed-point equation $\tilde{\beta}_{ij} = B(\tilde{\beta}_{ij})$. While such methods are widely available in commercial software packages and tend to be extremely fast, the optimal expansion point $\tilde{\beta}_{ij}$ needs to be computed for each firm in each market separately, which may potentially slow down numerical optimization routines if large data sets are used. To improve computational

efficiency we can further derive an approximate solution to the fixed point equation, which as we will show in the next section, performs very well.

Let $h(\beta) = \log\left(\sum_{k=1}^{K} \exp(x'_{ijk}\beta)\right)$ and perform a quadratic Taylor approximation of $g(\beta)$ around the constant parameter vector b. Then,

$$h(\beta) \cong h_{ij}(b) + (\beta - b)' J_{ij}(b) + \frac{1}{2}(\beta - b)' H_{ij}(b)(\beta - b) + O((\beta - b)^3), \quad (5.18)$$

where the Jacobian and Hessian terms are given by

$$J_{ij}(b) = \sum_{k=1}^{K} \left\{ x'_{ijk} \frac{\exp(x'_{ijk}b)}{\sum_{k=1}^{K} \exp(x'_{ijk}b)} \right\} \quad \text{and} \quad (5.19)$$

$$H_{ij}(b) = \frac{\sum_{k=1}^{K} x_{ijk} x'_{ijk} \exp(x'_{ijk}b)}{\sum_{k=1}^{K} \exp(x'_{ijk}b)} - \frac{\left[\sum_{k=1}^{K} x_{ijk} \exp(x'_{ijk}b)\right]\left[\sum_{k=1}^{K} x'_{ijk} \exp(x'_{ijk}b)\right]}{\left[\sum_{k=1}^{K} \exp(x'_{ijk}b)\right]^2}. \quad (5.20)$$

Thus, we can re-write the expression for $g(\beta)$ as

$$g(\beta) = \frac{1}{2}(\beta - b)' \Sigma^{-1}(\beta - b) + h_{ij}(b) + (\beta - b)' J_{ij}(b) + \frac{1}{2}(\beta - b)' H_{ij}(b)(\beta - b) \quad (5.21)$$

The optimal expansion point $\tilde{\beta}_{ij}$ solves the equation $\partial g(\beta)/\partial \beta = 0$. Hence,

$$\frac{\partial g(\beta)}{\partial \beta} = (\beta - b)' \Sigma^{-1} + J_{ij}(b) + (\beta - b)' H_{ij}(b) = 0. \quad (5.22)$$

Since this expression is now linear we can easily solve for the optimal expansion point $\tilde{\beta}_{ij}$,

$$\tilde{\beta}_{ij} = b + \left[\Sigma^{-1} + H_{ij}(b)\right]^{-1} J'_{ij}(b). \quad (5.23)$$

We can now re-write Proposition 5.1 to obtain an easily implementable version of the Laplace approximation of the expected share.

Proposition 5.2: *If β has a Normal distribution with mean β and covariance Σ, we can approximate $E_\beta(s_{ij}) = E_\beta\{(\sum_{k=1}^{K} \exp(x'_{ijk}\beta))^{-1}\}$ by*

$$E_\beta(s_{ij}) \cong \sqrt{\frac{|\tilde{\Sigma}_{ij}|}{|\Sigma|}} \exp\left\{-\frac{1}{2}\left(\tilde{\beta}_{ij} - b\right)' \tilde{\Sigma}_{ij}^{-1}(\tilde{\beta}_{ij} - b)\right\} \left(\sum_{k=1}^{K} \exp(x'_{ijk}\tilde{\beta}_{ij})\right)^{-1}, \quad (5.24)$$

where

$$\tilde{\beta}_{ij} = b + \left[\Sigma^{-1} + H_{ij}(b^*)_{b^*=b}\right]^{-1} J'_{ij}(b) \tag{5.25}$$

$$\tilde{\Sigma}_{ij}^{-1} = \Sigma^{-1} + H_{ij}(b^*)_{b^*=\tilde{b}_{ij}}, \tag{5.26}$$

and

$$J_{ij}(b) = \sum_{k=1}^{K} \left\{ x'_{ijk} \frac{\exp(x'_{ijk}b)}{\sum\limits_{k=1}^{K} \exp(x'_{ijk}b)} \right\} \tag{5.27}$$

$$H_{ij}(b^*) = \frac{\sum\limits_{k=1}^{K} x_{ijk}x'_{ijk}\exp(x'_{ijk}b^*)}{\sum\limits_{k=1}^{K} \exp(x'_{ijk}b^*)} - \frac{\left[\sum\limits_{k=1}^{K} x_{ijk}\exp(x'_{ijk}b^*)\right]\left[\sum\limits_{k=1}^{K} x'_{ijk}\exp(x'_{ijk}b^*)\right]}{\left[\sum\limits_{k=1}^{K} \exp(x'_{ijk}b^*)\right]^2}.$$
$$\tag{5.28}$$

Notice that the Hessian expression $H_{ij}(b^*)$ is evaluated at different points b^* in the computation of the values of $\tilde{\beta}_{ij}$ and $\tilde{\Sigma}_{ij}$. Proposition 5.2 is also insightful in that it explains why a simple Taylor expansion of the Gaussian integral around the mean b will fail. Consider the expression for $\tilde{\beta}_{ij}$, which is the optimal expansion point in the Laplace approximation. Notice that $\tilde{\beta}_{ij}=b$ only if $J_{ij}(b) = 0$. But this expression can only be zero if the vectors of covariates x_{ijk} are zero for all k. Hence a Taylor approximation of the same problem will fail since it expands each expected share around a constant value when in fact it ought to perform the expansion around an optimal value which will differ from share to share depending on the covariates. The Laplace approximation developed above performs this optimal expansion.

■ 5.2 Monte-Carlo Simulations

In this section we discuss the estimation of the random coefficients model by non-linear least squares after applying the Laplace approximation derived in the previous section to each expected market share. We will also compare its performance in Monte-Carlo simulations to that of alternative methods used for the estimation of these models in the econometric literature.

Since the model was introduced over thirty years ago, several estimation methods have been proposed which try to circumvent the problem that the integral expression for the expected shares does not have a closed form solution for most distributions of the taste parameters. While numerical integration by quadrature is implemented in numerous software packages it is also extremely time consum-

ing. In practice it is not possible to use numerical integration to solve such problems if the number of regressors is greater than two or three. We have found that even for the case of a single regressor this method is extremely slow and not always reliable.

The main attempt to estimate random coefficients models is based on the method of simulated moments (McFadden, 1989; Pakes and Pollard, 1989), where the expectation is replaced by a an average over repeated draws from the distribution of taste parameters:

$$E_\beta(s_{ij}) = \int \frac{\exp(x_{ij}'\beta)}{\sum\limits_{k=1}^{K} \exp(x_{ik}'\beta)} dF(\beta) \cong \frac{1}{R} \sum_{r=1}^{R} \frac{\exp(x_{ij}'\overline{\beta_r})}{\sum\limits_{k=1}^{K} \exp(x_{ik}'\overline{\beta_r})}, \qquad (5.29)$$

where $\overline{\beta_r}$ is drawn from the distribution $F(\beta)$. Random sampling from a distribution may nevertheless provide poor coverage of the domain of integration. There is no guarantee that in a particular set of draws the obtained sequence will uniformly cover the domain of integration and may in fact exhibit random clusters which will distort the approximation. To achieve a good approximation the number of draws R will have to be very large.

More recently the use of variance reduction techniques has been advocated in an attempt to improve the properties of simulated estimation (Train, 2003). Negatively correlated pseudo-random sequences may lead to a lower variance of the resulting estimator than traditional independent sampling methods. The method currently employed in econometrics uses *Halton sequences* (Small, Winston and Yan, 2005).

Halton sequences can be constructed as follows. For each dimension r of the vector β and some prime number k construct the sequence

$$s_{t+1} = \left\{ s_t, s_t + \frac{1}{k}, ..., s_t + \frac{(k-1)}{k^t} \right\}, \text{for } s_0 = 0. \qquad (5.30)$$

This sequence is then randomized by drawing μ uniform (0,1) and for each element s, letting $s^* = \mod(s + \mu)$.

This method provides coverage of the unit hypercube by associating each dimension with a different prime number k. In order to transform these points into draws from the relevant distribution, an inversion in then applied, e.g. if the desired distribution is Normal one would turn these points on the unit hypercube into values of β, by letting $\overline{\beta}_r = \Phi^{-1}(s_r^*)$, which corresponds to the inverse of the normal distribution.

The use of Halton sequences improves performance over the use of independent draws and yet nevertheless it suffers from the curse of dimensionality. Many thousand draws are required for each observation and the application of this

method is extremely problematic for the estimation of even a small number of parameters since it is so time consuming.

The mathematical properties of Halton sequences are not sufficiently well understood and may represent a liability in some applications. Train (2003) reports that in estimating a random coefficients logit model for households' choice of electricity supplier repeatedly, most runs provided similar estimates of the coefficients, yet some runs provided significantly different coefficients even though the algorithm was unchanged and applied to the same data set. Similarly, Chiou and Walker (2005) report that simulation based methods may falsely identify models if the number of draws is not sufficiently large. The algorithm may produce spurious results which "look" reasonable yet are not supported by the underlying data.

Additionally, to our knowledge, it was not possible so far to reliably estimate the full covariance matrix using simulation based methods. Researchers focus exclusively on the estimation of the mean and variance parameters thereby assuming a diagonal structure to the covariance matrix Σ of the taste parameters. We will show how this problem can be easily overcome by the use of the Laplace approximation method we propose in this chapter. Later on in this section we will also show how ignoring the covariances may lead to biased results and unreliable policy analysis if the taste parameters in the true data generating process are correlated.

We propose estimating the model parameters (b, Σ) by non-linear least squares. Let s_{ij} be the observed market share of firm j in market i. We can construct the approximation of the expected share using the Laplace approximation as described in Section 5.1, $\hat{s}_{ij}(b, \Sigma) = E_\beta(s_{ij})$. This will be a non-linear function in the model parameters b and Σ and can be implemented using either Proposition 5.1 or Proposition 5.2. The implementation of Proposition 5.2 is immediate and only involves the use of matrix functions. We can then proceed to estimate the model parameters by least squares or weighted least squares which can improve efficiency:

$$\left(\hat{\beta}, \hat{\Sigma}\right) = \operatorname{argmin}_{\beta, \Sigma} \sum_{i=1}^{N} \sum_{j=1}^{K} \left(s_{ij} - \hat{s}_{ij}(b, \Sigma)\right)^2. \tag{5.31}$$

The optimization can be achieved using a Newton type constrained optimization routine. Some parameters may require linear constraints (e.g. if the optimization is performed over variance parameters, then $(\sigma^2)_p > 0$ for all taste parameters β_p). The optimization needs to ensure that the estimated covariance matrix is positive definite at each step, for example by employing an appropriate re-parameterization or the Cholesky decomposition

This can be achieved by an appropriate penalization at the edges of the allowable domain. The model can also be estimated by minimum chi-square techniques or by maximum likelihood given our evaluation of the expected shares. Simula-

Table 5.1. Estimation of the one variable random coefficients model. $N = 1000, K = 6$.

Mean Bias	Quadrature	Fixed Point Laplace	Laplace	Halton
b	0.00778	0.00492	0.00555	0.00269
σ^2	0.04957	0.02348	0.00634	-0.02011

MSE	Quadrature	Fixed Point Laplace	Laplace	Halton
b	0.01003	0.00297	0.00281	0.00301
σ^2	0.09833	0.06641	0.08357	0.07381

tion results suggest no significant performance differences between these different methods of implementation.

In Table 5.1 we estimate a random coefficients model with a single taste parameter using the methods discussed above. The covariate is drawn from a mixture distribution of a normal and a uniform random variable. This particular construction is performed in order to correct for unreliable estimates that have been reported when only normal covariates are being used. Since the model only requires univariate integration we can also perform numerical integration. We use a second order Newton-Coates algorithm to perform the integration by quadrature for each expected share. Additionally we compute estimates using the two versions of the Laplace approximation of the expected share as described in Section 5.1 in Propositions 5.1 and 5.2 respectively. The results labelled as "Fixed Point Laplace" compute the optimal expansion points $\tilde{\beta}_{ij}$ using iterative fixed point techniques. The results labelled "Laplace" approximate this fixed point calculation using the analytic expression of Proposition 5.2. We also compute estimates using Halton sequences as implemented by Whinston, Small and Yan (2005). We perform 500 draws for each observation.

The results in Table 5.1 show that all four methods produce comparable results. Interestingly though, numerical integration tends to be outperformed by either of the approximation methods presented here. In particular the Laplace approximation we proposed performs very similarly to the simulated estimation based on Halton sequences both in terms of mean bias and mean squared error. This result was confirmed in additional simulations were the number of taste parameters was increased. The Laplace approximation introduced in this chapter outperforms the method of simulated moments in terms of computational time. Even in this simple one dimensional example the Laplace method runs about three times faster than the corresponding estimation using Halton sequences.

We have found no significantly different performance results between the Laplace approximation using the fixed point calculation and that using the approximation to the optimal expansion point. The Laplace approximation of Proposition 5.2 nevertheless outperformed all other methods in terms of com-

Table 5.2. Estimation of the three variable random coefficients model with covariances. $N = 2000, K = 6$.

Laplace	Mean Bias	MSE
b_1	0.01167	0.00233
b_2	0.00679	0.00201
b_3	-0.00371	0.00298
σ_1^2	-0.06889	0.09499
σ_2^2	-0.08245	0.07016
σ_3^2	0.03880	0.03180
ρ_{12}	0.04918	0.00774
ρ_{13}	0.04317	0.00350
ρ_{23}	-0.00702	0.00551

putational time, being 3 to 5 times faster than the simulation approach.

Once we allow for multiple taste parameters we can ask the question whether these taste parameters are correlated with each other. Consider a model with 3 taste parameters, drawn from a distribution with mean $(b_1, b_2, b_3)'$ and variances $(\sigma_1^2, \sigma_2^2, \sigma_3^2)$. In many cases of interest there is no a priori reason to constrain the covariance matrix of this distribution to be diagonal. We can allow for correlations between taste parameters by setting the off-diagonal elements of the covariance matrix equal to $\sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$ for $-1 < \rho_{ij} < 1$. The parameter ρ_{ij} measures the strength of the correlation between the different taste parameters. The full covariance matrix which needs to be estimated in this case is:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 \\ \rho_{13}\sigma_1\sigma_3 & \rho_{23}\sigma_2\sigma_3 & \sigma_3^2 \end{pmatrix} \tag{5.32}$$

We use the Laplace approximation method to estimate all 9 parameters and report results for mean bias and MSE in Table 5.2. We were not able to estimate the same parameters using the method of simulated moments with Halton sequences. The algorithm failed to converge for Halton sequences under different model parameters and different starting values.

Computational issues involving the use of simulated moments seem to have prevented empirical work involving the estimation of the full covariance matrix. We now wish to explore to what extent this may bias the results. To this purpose we estimate the same model as in the above example but ignore the covariances. Thus the true model has $\rho_{ij} \neq 0$ but we only estimate the restricted model where we assume $\rho_{ij} = 0$ for all $i, j, i \neq j$.

The results are presented in Table 5.3. We were able to obtain estimates of

Table 5.3. Estimation of the three variable random coefficients model without covariances. The true model contains covariances but these are not estimated. $N = 2000, K = 6$.

	Mean Bias Laplace	Mean Bias Halton	MSE Laplace	MSE Halton
b_1	0.02037	0.01003	0.00321	0.00256
b_2	0.01582	0.00778	0.00201	0.00258
b_3	0.00651	0.00212	0.00122	0.00197
σ_1^2	-0.01032	-0.21102	0.10192	0.18226
σ_2^2	-0.50883	-0.43340	0.32381	0.27094
σ_3^2	-0.12991	-0.14967	0.03900	0.09577

the restricted model using both the new Laplace approximation we propose and by using the simulation approach involving Halton sequences. Once again both methods produce comparable results. While the estimates of the mean parameters $(b_1, b_2, b_3)'$ seem to be sufficiently robust to the misspecification of the covariance matrix, the estimates of the variance parameters $(\sigma_1^2, \sigma_2^2, \sigma_3^2)$ seem to be strongly affected by the non-inclusion of the covariance terms in the optimization. The size of the bias is model dependent and we have found an absolute value of the bias between 30-60% in most simulations. Additionally, it seems that negative correlations which are falsely excluded bias the results much more than positive ones.

The failure to include the correlations between taste parameters may also lead to incorrect policy recommendations. Thus, consider the three variable described above where the true data generating process has non-zero correlation terms and a full covariance matrix. We can interpret the model as follows.

We label the first variable as "price" and consider the policy experiment whereby the government has to decide whether to impose a 10% tax on a specific good. The tax is fully passed on to the consumers in the form of a 10% price increase. There are $K = 6$ competing firms in each market producing differentiated brands of the good on which the tax was imposed. We wish to simulate the ex post effect of the tax on the market shares of each firm. In order to do so we collect a sample of observations consisting of the market shares of each firm in different markets and the product characteristics of the differentiated good produced by brand and market. We estimate the random coefficients model with a full covariance matrix which allows for correlations between taste parameters. We also estimate the same model but limit ourselves to estimating a diagonal covariance thus restricting the correlations to be zero and also derive the logit estimates of the means corresponding to the case where the taste parameters are assumed to be constant in the population. We can use these estimates to simulate the distribution of market shares of each firm across the markets and compare
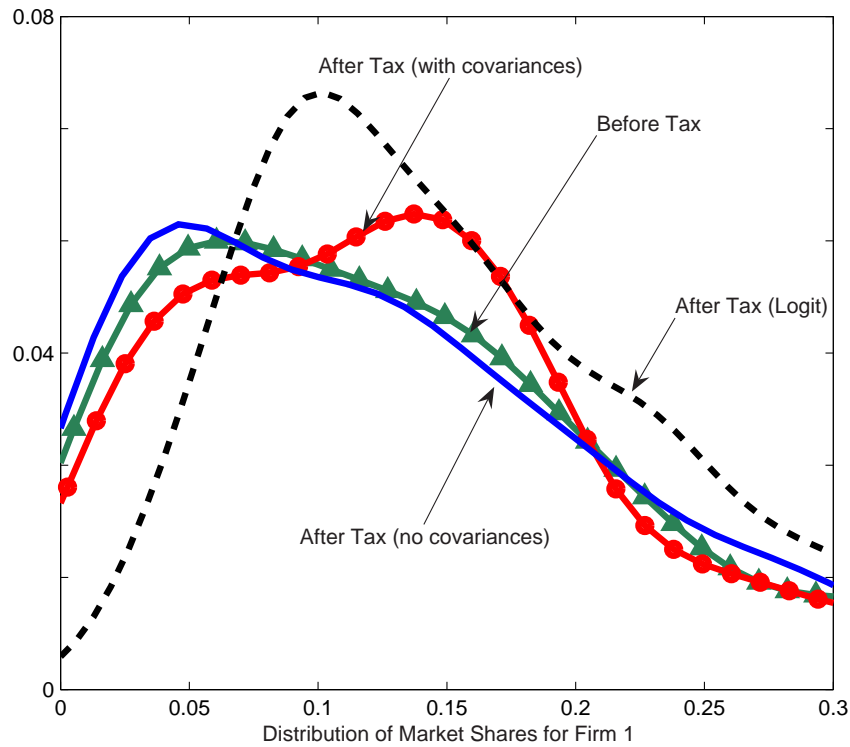
Figure 5-2. Market shares of Firm 1 before and after tax

them to the initial distribution of market shares before the tax was implemented. We present the resulting distributions in Figure 5.2.

If we estimate any of the misspecified models by using either the logit estimates of Equation 5.2 or the random coefficients logit estimates of Equation 5.3 under the assumption of no correlation we would reach very different conclusions from the case when we take into account the full covariance matrix between taste parameters. Thus we can see how ignoring the correlations may lead to incorrect policy recommendations when the random coefficients model is used to estimate the distribution of taste parameters.

■ 5.3 Conclusion

In this chapter we have introduced a new analytic approximation to the choice probability in a random coefficients logit model. The approximation was derived using a multivariate extension of the Laplace approximation for sub-asymptotic domains. The expression results in a non-linear function of the data and parameters which can be conveniently estimated using non-linear least squares.

This new method of estimating random coefficients logit models allows for the estimation of correlations between taste parameters. The estimation of a full covariance matrix seems to have eluded many previous implementations of the random coefficients logit model employing simulations of the underlying taste distributions.

Simulation results show that our new method performs extremely well, both in terms of numerical accuracy and computational time. We also provide an example of the importance of estimating correlations between taste parameters through a tax simulation where very different policy implications would be reached if the estimated model is misspecified by restricting the correlations to be zero.

■ 5.4 Appendix A: Laplace Approximation Theorem

This appendix states the multivariate Laplace approximation theorem. For additional discussions of the theorem and applications to statistics see Muirhead (2005) and Jensen (1995). A proof is given in Hsu (1948).

Laplace Approximation Theorem. Let D be a subset of R^p and let f and g be real-valued functions on D and T a real parameter. Consider the integral

$$I = \int_{\beta \in D} f(\beta) \exp(-Tg(\beta)) d\beta \tag{5.33}$$

a) g has an absolute minimum at an interior point $\tilde{\beta}$ of D;

b) there exists $T \geq 0$ such that $f(\beta) \exp(-Tg(\beta))$ is absolutely integrable over the domain D;

c) all first and second order partial derivatives of $g(x)$, $\frac{\partial g}{\partial \beta_i}, \frac{\partial^2 g}{\partial \beta_i \partial \beta_j}$, for $i = 1 \dots p$ and $j = 1 \dots p$ exist and are continuous in the neighborhood $N(\tilde{\beta})$ of $\tilde{\beta}$.

d) there is a constant $\gamma < 1$ such that $|\frac{\exp(-g(\beta))}{\exp(-g(\tilde{\beta}))}| < \gamma$ for all $x \in D \backslash N(\tilde{\beta})$

e) f is continuous in a neighborhood $N(\tilde{\beta})$ of $\tilde{\beta}$.

Then for large T, we have:

$$\tilde{I} = \left(\frac{2\pi}{T}\right)^{p/2} [\det(H(\tilde{\beta}))]^{-1/2} f(\tilde{\beta}) \exp(-Tg(\tilde{\beta})), \text{where} H(\tilde{\beta}) = \frac{\partial^2 g(\tilde{\beta})}{\partial \tilde{\beta} \partial \tilde{\beta}'} \tag{5.34}$$

and

$$I = \tilde{I}(1 + O(T^{-1})) \quad \text{as} T \to \infty. \tag{5.35}$$

In Section 5.1 we let $f(\beta) = 1$ and $g(\beta) = \frac{1}{2}(\beta - b)'\Sigma^{-1}(\beta - b) + \log\left(\sum_{k=1}^{K} \exp(x'_{\text{ijk}}\beta)\right)$. This is sometimes referred to as an *exponential form Laplace approximation.*

Moreover we use the observation of Butler (2002) that in many cases of interest this approximation performs very well even in sub-asymptotic cases where T remains small. In our case $T = 1$.

■ 5.5 Appendix B: Extensions and Proofs

In some applications we may wish to allow for a mixture of fixed and random coefficients. We can partition the $p \times 1$ dimensional vector of taste parameters into two subvectors b^0 and β^1 of lengths p_0 and p_1 respectively, where $p_0 + p_1 = p$. The vector b^0 contains the fixed (unknown) parameters corresponding to the non-random coefficients of the model, while the vector β^1 captures the random coefficients. Furthermore, we can assume that β^1 is Normally distributed with mean b^1 and variance Σ. The results derived in this chapter extend to the case of a model specification with both random and fixed coefficients by performing the integration over the random coefficients while treating the fixed coefficients as constant for the purpose of deriving the Laplace approximation.

We now re-state Proposition 5.2 for the case with both fixed and random coefficients, $\beta = (b^0, \beta^1)$. The unknown parameters to be estimated are (b^0, b^1, Σ), where b^1 is the vector of mean parameters of the random coefficients β^1 and Σ is the corresponding covariance matrix of β^1.

Proposition 5.2.1: *We can approximate* $E_\beta(s_{\text{ij}}) = E_\beta\{(\sum_{k=1}^{K} \exp(x'_{\text{ijk}}\beta))^{-1}\}$ *by*

$$E_\beta(s_{\text{ij}}) \cong \sqrt{\frac{\left|\tilde{\Sigma}_{\text{ij}}\right|}{|\Sigma|}} \exp\left\{-\frac{1}{2}\left(\tilde{\beta}^1_{\text{ij}} - b^1\right)' \tilde{\Sigma}^{-1}_{\text{ij}}(\tilde{\beta}^1_{\text{ij}} - b^1)\right\} \left(\sum_{k=1}^{K} \exp(x'_{\text{ijk}}\tilde{\beta}_{\text{ij}})\right)^{-1},$$

(5.36)

where $\tilde{\beta}_{\text{ij}} = (b^0, \tilde{\beta}^1_{\text{ij}})$ *and* $\check{\beta} = (b^0, b^1)$ *and*

$$\tilde{\beta}^1_{\text{ij}} = b^1 + \left[\Sigma^{-1} + H_{\text{ij}}(b^*)_{b^*=\check{b}}\right]^{-1} J'_{\text{ij}}(\check{b}) \tag{5.37}$$

$$\tilde{\Sigma}^{-1}_{\text{ij}} = \Sigma^{-1} + H_{\text{ij}}(b^*)_{b^*=\tilde{b}_{\text{ij}}}, \tag{5.38}$$

and

$$J_{\text{ij}}(b^*) = \sum_{k=1}^{K} \left\{x'_{\text{ijk}} \frac{\exp(x'_{\text{ijk}}b^*)}{\sum_{k=1}^{K} \exp(x'_{\text{ijk}}b^*)}\right\} \tag{5.39}$$

$$H_{ij}(b^*) = \frac{\sum_{k=1}^{K} x_{ijk} x'_{ijk} \exp(x'_{ijk}b^*)}{\sum_{k=1}^{K} \exp(x'_{ijk}b^*)} - \frac{\left[\sum_{k=1}^{K} x_{ijk} \exp(x'_{ijk}b^*)\right]\left[\sum_{k=1}^{K} x'_{ijk} \exp(x'_{ijk}b^*)\right]}{\left[\sum_{k=1}^{K} \exp(x'_{ijk}b^*)\right]^2}.$$

(5.40)

In Section 5.1 we assert one of the conditions required for the existence of a Laplace approximation with a unique expansion point, the concavity of the function $-g(\beta)$. The Lemma below proves this result.

Lemma: *The function $g(\beta)$ is convex, where*

$$g(\beta) = \frac{1}{2}(\beta - b)' \Sigma^{-1}(\beta - b) + \log\left(\sum_{k=1}^{K} \exp(x'_{ijk}\beta)\right)$$ (5.41)

Proof: $g(\beta)$ is the sum of two convex functions, a quadratic form in β and the function $g_1(\beta) = \log\left(\sum_{k=1}^{K} \exp(x'_{ijk}\beta)\right)$. The Hessian of this function is given by $H_{ij}(\beta)$ defined in equation 5.20 above. In order to see that $H_{ij}(\beta) \geq 0$ notice that,

$$\left[\sum_{k=1}^{K} \exp(x'_{ijk}b)\right]^2 H_{ij}(b) = \sum_{k=1}^{K} \exp(x'_{ijk}b) \sum_{k=1}^{K} x_{ijk} x'_{ijk} \exp(x'_{ijk}b) -$$ (5.42)

$$- \left[\sum_{k=1}^{K} x_{ijk} \exp(x'_{ijk}b)\right]\left[\sum_{k=1}^{K} x'_{ijk} \exp(x'_{ijk}b)\right].$$ (5.43)

If we expand the right hand side of equation 5.43 and cancel the terms in $x_{ijk} x'_{ijk}(\exp(x'_{ijk}b))^2$ we can re-arrange this expression as:

$$\left[\sum_{k=1}^{K} \exp(x'_{ijk}b)\right]^2 H_{ij}(b) = \sum_{r=1}^{K-1} \sum_{s=r+1}^{K} (x_{ijr} - x_{ijs})(x_{ijr} - x_{ijs})' \exp(x'_{ijr}b) \exp(x'_{ijs}b) \geq 0.$$

(5.44)

Bibliography

BAJARI, P., H. HONG, AND S. RYAN (2005): "Identification and Estimation of Discrete Games of Complete Information," Mimeo, MIT.

BARNDORFF-NIELSEN, O., AND D. R. COX (1979): "Edgeworth and Saddle-point Approximations with Statistical Applications," *Journal of the Royal Statistical Society B*, 41, 179–312.

BUTLER, R., AND A. T. A. WOOD (2001): "Laplace Approximations for Hypergeometric Functions of Matrix Argument," *The Annals of Statistics*, 30, 1155–1177.

CARDELL, S., AND F. DUNBAR (1980): "Measuring the Societal Impacts of Automobile Downsizing," *Transportation Research A*, 14, 423–434.

CHIOU, L., AND J. WALKER (2005): "Identification and Estimation of Mixed Logit Models under Simulation Methods," Mimeo, MIT.

DANIELS, H. E. (1954): "Saddlepoint Approximations in Statistics," *The Annals of Mathematical Statistics*, 25, 631–650.

DE BRUIJN, N. (1981): *Asymptotic Methods in Analysis*. Dover Publications, New York.

EFSTATIOU, M., E. GUTIERREZ-PENA, AND A. F. M. SMITH (1998): "Laplace Approximations for Natural Exponential Families with Cuts," *Scandinavian Journal of Statistics*, 25.

GLYNN, W. J. (1980): "Asymptotic Representations of the Densities of canonical Correlations and Latent Roots in Manova When the Population Parameters have Arbitrary Multiplicity," *The Annals of Statistics*, 8, 958–976.

HARDING, M. C., AND J. A. HAUSMAN (2007): "Flexible Parametric Estimation of the Taste Distribution in Random Coefficients Logit Models," Mimeo, MIT.

HAUSMAN, J. A. (1975): "Project Independence Report: A Review of U.S. Energy Needs up to 1985," *Bell Journal of Economics*.

HAUSMAN, J. A., AND D. A. WISE (1978): "A Conditional Probit Model for Qualitative Choice-Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," *Econometrica*, (46), 403–426.

HENSHER, D., AND W. GREENE (2003): "The Mixed Logit Model: The State of Practice," *Transportation*, (30), 133–176.

HESS, J., AND J. POLAK (2005): "Mixed Logit Modelling of Airport Choice in Multi-Airport Regions," *Journal of Air Transport Management*, (11), 59–68.

HSU, L. C. (1948): "A Theorem on the Asymptotic Behavior of a Multiple Integral," *Duke Mathematical Journal*, (15), 623–632.

JENSEN, J. L. (1995): *Saddlepoint Approxiations*. Clarendon Press, Oxford.

LERMAN, S., AND C. MANSKI (1981): "On the Use of Simulated Frequencies to Approximate Choice Probabilities," in *Structural Analysis of Discrete Data with Econometric*, ed. by C. Manski, and D. McFadden, pp. 305–320. MIT Press, Cambridge.

MANSKI, C., AND D. MCFADDEN (eds.) (1981): *Structural Analysis of Discrete Data with Econometric Applications*. MIT Press, Cambridge.

MCFADDEN, D. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, ed. by A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull, pp. 75–96. Academic Press, Amsterdam.

——— (1989): "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration," *Econometrica*, (57), 995–1026.

MCFADDEN, D., AND K. TRAIN (2000): "Mixed MNL Models for Discrete Response," *Journal of Applied Econometrics*, (15), 447–470.

MUIRHEAD, R. J. (ed.) (2005): *Aspects of Multivariate Statistical Theory*. Wiley, New York.

MURRAY, J. (ed.) (1984): *Asymptotic Analysis*. Springer-Verlag, New York.

PAKES, A., AND D. POLLARD (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, (57), 1027–1057.

REVELT, D., AND K. TRAIN (1998): "Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level," *Review of Economics and Statistics*, pp. 647–657.

SMALL, K. A., C. WINSTON, AND J. YAN (2005): "Uncovering the Distribution of Motorists' Preferences for Travel Time and Reliability," *Econometrica*, (73), 1367–1382.

TIERNEY, L., AND J. KADANE (1986): "Accurate Approximations for Posterior Moments and Marginal Densities," *Journal of the Americal Statistical Association*, (81), 82–86.

TRAIN, E. K. (2003): *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge.