Specification Testing in State-Space Models and Its Applications to Financial Data

CHIBA MASARU

INTERNATIONAL GRADUATE SCHOOL OF SOCIAL SCIENCE

YOKOHAMA NATIONAL UNIVERSITY

In the memory of Yasuhiro Terasaki

Acknowledgement

From the start of the reserch, I have been fortunate to receive a great number of critical responses and concrete suggenstions. First, from the very beginning of this thesis, I have been getting a lot of invaluable guidance and advice from my supervisor Professor Masahito Kobayashi. To him, I would like to express my deep and sincere gratitude. Professor Kobayashi has given me all kinds of instructions not only on my studies but also on my life for five years since I was a master student. Without his guidance and advice, this achievement would never have been completed. And my greatfulness must also go to my committee - supervisors, Professor Motonari Kurasawa and Professor Yoshiaki Omori for their incisive comments and encouraging during the presentations of interim report proved very stimulating. This reserch would be impossible without their help.

I am also grateful to the professors and staffs of International Gratuate School of Social Sciences. I would also like to thank all my colleagues, who made my research life a wonderful memory.

Last but not least, I would like to express my gratitude with this thesis to my parents, Akira Chiba and Toshiko Chiba, for their support and stimulation.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Stylized facts

1.1.1 State-space models

State-space models are not new in the statistics and econometric literatures. But, a growing number of published papers that employ them demonstrates their usefulness and widening application. In essence, the state-space model is one in which an observed variable is the sum of a linear function of the state variable plus an error. The state variable, in turn, evolves according to a stochastic difference equation that depends on parameters that in economic applications are generally unknown. Thus, both the path of the state variable through time and the parameters are to be inferred from the data. Harvey [1981] introduced to economists Kalman filter for obtaining maximum likelihood estimates of parameters through prediction error decomposition was introduced. It became clear from Harvey's work and others' that a surprising range of econometric models, including regression models with time-varying parameters, ARIMA models and unobserved components time series models, could be cast in state-space form and thus be rendered amenable to Kalman machinery for parameter estimation and extraction of state variables.

State-space models have a wide range of potential applications in econometrics - for example, permanent income, expectations, the ex ante real rate of interest, and the reservation wage. Engle and Watson [1981] applies it to modeling the behavior of wage rate; Garbage and Wachetel [1978] and Antonic [1986] apply it to modeling the behavior of ex ante real interests; Burmeister and Wall [1982] and Burmeister, Wall, and Hamilton [1986] apply it to modeling a time-varying monetary reaction function of the Federal Reserve. Stock and Watson's [1991] dynamic factor model of coincident economic indicators is a recent application of the state-space model. Thus, state-space models have highly productive paths for research in econometrics and finance.

However, we rarely know priori structure of an exact model. In fact, investigators estimate several models but may not undertake comprehensive testing of the adequacy of their preffered model. Thus, there are some requirements for specification tests.

1.1.2 Specification tests

Specification tests play an important role in the evaluation of econometric models. In fact, there have been rapid depelopments in the field of testing for misspecification in both applied and theoretical econometrics. The idea that a model must be tested before it can be taken to be an adequate basis for studying economic behaviour has become widely accepted. Modern empirical analysis usually include testing for a number of specification errors, and the range of tests available to applied workers has increased enormously.

Most of specification tests are based either on the Wald, Likelihood Ratio, or Lagrange Multiplier principle. These three general principles have a certain symmetry which has revolutionized the teaching of hypothesis tests and the development of new procedures. Essentially, the Lagrange Multiplier approach starts at the null and asks whether movement toward the alternative would be an improvement, while the Wald approach starts at the alternative and considers movement toward the null. The Likelihood Ratio method compares the two hypotheses directly on an equal basis. Of the three classical tests, the Lagrange Multiplier principle deserves special consideration when discussing tests for misspecification because, unlike the asymptotically equivalent Wald and Likelihood Ratio methods, it does not require the estimation of the more complex alternative in which the original model of interest has been embedded.

1.2 Purpose of this study

This paper has two purposes. One is to propose specification tests for the dynamic factor model. The weak exogeneity, linear dependency, and omitted explanatory variables tests will be presented in this paper.

Another is to apply these tests to the factor augmented APT model and the term structure model of yield curve. In the APT model, we will examine the adequacy of macroeconomic factors as systematic variables to the stock return. In the term structure model of yield curve, we will examine the nature of the linkage between factors driving the yield curve and macroeconomic factors.

Thus, we have grouped this paper into three categories: (1) specification testing in dynamic factor models, (2) an asset pricing model, with links to macroeconomy, and (3) an yield curve model, with links to macroeconomy.

1.3 Structure of this study

The rest of this paper is arranged as follows. Chapter 2 is "Specification Testing in Dynamic Factor Models." Chapter 3 is "Joint Estimation of Factor Sensitivities and Risk Premia in the Factor Augumented APT Model." Chapter 4 is "The Macroeconomy and the Yield Curve: Specification Testing based on Lagrange Multiplier Approach." The conclusion is given in the last chapter.

References

- Antonic, M. (1986), "High and volatile real interest rates: Where Does the Fed Fit In?" *Journal of Money, Credit, and Banking* 18(1), 18-27.

- Burmeister, E. and Wall, K. D. (1982), "Kalman filtering estimation of unobserved rational expectations with an application to german hyperinflation," *Journal of Econometrics* 20, 255-284.

- Burmeister, E., Wall, K. D., and Hamilton, J. D. (1986), "Estimation unobserved expected monthly inflation using kalman filtering," *Journal of Business and Economic Statistics* 4, 147-160.

- Engle, R. E. and Watson, M. W. (1981), "A one-factor multivariate time series model of metropolitan wage rates," *Journal of the American Statistical Association* 73(276), Applications Section, 774-781.

- Garbade, K. D. and Wachtel, P. (1978), "Time variation in the relationship between inflation and interest rates," *Journal of Monetary Economics* 4, 775-765.

- Harvey, A. C. (1981), *Time series model*, Oxford: Phillip Allan and Humanities Press

- Stock, J. H. and Watson, M. W. (1991), "A propability model of the coincident economic indicators," *In Leading Economic Indicators: New Approaches and Forecasting Records*, ed. K. Lahiri and G. H. Moore. Cambridge: Cambridge University Press 63-89.

Chapter 2

Specification Testing in Dynamic Factor Models

2.1 Introduction

Recently, researchers have been interested in economic and financial models in which the dynamics of large scale economic variables can be specified by a smaller number of indices or "common factors." When the dynamic factor model proposed by Stock and Watson [1989] is applied to a time series model, the result is a model based on the assumption that one latent variable causes the co-movement of four observed variables. This approach has provided some new perspectives on several economic analysis. For example, Diebold and Rudebusch [1996] and Kim and Nelson [1998] applied it in the modeling of the business cycles; Diebold et al. [2006] and Ghysels and Ng [1998] applied it in the characterization of the yield curve; and Sentana [2004] and Lehmann and Modest [1988] applied it in mimicking portfolios in an Arbitrage Pricing Theory (APT) model. However, we rarely know priori structure of an exact model. In fact, investigators estimate several models but may not undertake comprehensive testing of the adequacy of the preffered model. Thus, there are some requirements for specification tests.

Because common factors are generally unobserved, we need to extract them using statistical techniques. Estimation procedure, such as Principal component analysis and Kalman filter, are used to extract them. In the method of the former technique, a number of past studies considered the problem of verifying the adequacy of the dynamic factor model. Lewbel [1991] and Donald [1997] used the rank of a matrix to test for the number of factors. Cragg and Donald [1997], Stock and Watson [1998], and Forni et al. [2000] suggested the use of modified information criteria for the model selection. In the method of the latter technique, however, little work has been done for the hypothesis testing or the model selection on the dynamic factor model. This paper attempts to fill this gap, by proposing specification tests for weak exogeneity, linear dependency, and omitted explanatory variables based on the Lagrange Multiplier (LM) principle[1].

[1] *Note*: See Davidson and Mackinnon [1993], Breusch and Pagan [1980], Godfrey and Wickens [1981], and Engle [1984] for a detailed description of the LM test. See also Newey [1985], Tauchen [1985], and White

This paper provides an expression for the score in this model, defined as the derivative of the conditional log likelihood of the tth observation with respect to the parameter vector. This permits calculation of all of the necessary test statistic as well as an intuitive interpretation of what each test is based on. The score turns out to be a natural byproduct of the routine used to calculate the expected value with respect to smoothed density. In addition, from the same calculation we construct asymptotic standard errors for the parameter vector and specification tests. Therefore, proposed tests can be calculated together from a single pass through the data.

The LM principle deserves special consideration when discussing tests for misspecification because, unlike Likelihood Ratio and Wald methods, it does not require the estimation alternative in which the original model of interest has been embedded. In addition, the LM test proposed here employs an estimate of the information matrix based on the average outer product of the score as in Berndt et al. [1974]. Therefore, the LM test proposes a flexible and easily implemented scheme and is often regarded as the most suitable for constructing misspecification test.

The rest of the paper is organized as follows. Section 2.2 details the basic framework and notation of the model and introduces the estimation framework. Section 2.3 derives the general principles of specification tests. Section 2.4 reports Monte Carlo experiments and checks the actual size and power of the tests. Section 2.5 briefly summarizes this study.

2.2 A dynamic factor model representation

2.2.1 The notations and assumptions

The dynamic factor model approach expresses a large set of observed variables as a function of a small set of unobserved variables. Let y_{nt} be an nth observation at time t for $n = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$. Consider the following model:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{Nt} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1K} \\ \phi_{21} & \phi_{22} & \cdots & \phi_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{N1} & \phi_{N2} & \cdots & \phi_{NK} \end{bmatrix} \begin{bmatrix} c_{1t} \\ c_{2t} \\ \vdots \\ c_{Kt} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ \vdots \\ u_{Nt} \end{bmatrix}, \qquad (2.2.1)$$

where c_{kt} is a kth common factor at time t, ϕ_{nk} is an nth coefficient of factor loading associated with c_{kt} and u_{nt} is an nth idiosyncratic noise at time t for $n = 1, 2, \ldots, N,\ k = 1, 2, \ldots, K$, and $t = 1, 2, \ldots, T$. Equation (2.2.1) relates a set of N observed variables to unobserved K common factors. Therefore, the observation y_{nt} is decomposed into $\sum_{k=1}^{K} \phi_{nk} c_{kt} + u_{nt}$ of two unobservable mutually orthogonal (at all leads and lags) parts, the *common component* $\sum_{k=1}^{K} \phi_{nk} c_{kt}$, and the *idiosyncratic component* u_{nt}, respectively.

If the dynamic movement of common factors follows a stationary vector autoregressive process of first order, the model immediately forms a *state-space system*[2]. The dynamics of

[1987] for the general approach to specification testing .

[2] *Note*: As is well-known, ARMA state vector dynamics of any order may be readily accommodated in state-space form. We maintain the VAR(1) assumption only for transparency and parsimony.

common factors employed here is defined as follows:

$$\begin{bmatrix} c_{1t} \\ c_{2t} \\ \vdots \\ c_{Kt} \end{bmatrix} = \begin{bmatrix} \gamma_1 & 0 & \cdots & 0 \\ 0 & \gamma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \gamma_K \end{bmatrix} \begin{bmatrix} c_{1,t-1} \\ c_{2,t-1} \\ \vdots \\ c_{K,t-1} \end{bmatrix} + \begin{bmatrix} v_{1t} \\ v_{2t} \\ \vdots \\ v_{Kt} \end{bmatrix}, \tag{2.2.2}$$

where γ_k is a kth autoregressive coefficient and v_{kt} is a kth noise for $k = 1, 2, \ldots, K$ and $t = 1, 2, \ldots, T$. Note that our model restricts common factors to be orthogonal to each other (at all leads and lags) but does not restrict the factor loadings at all.

According to the above specification, we write a state-space system that equation (2.2.1) is the *observation equation* and equation (2.2.2) is the *transition equation*. In a vector/matrix notation, the state-space system of our model is as follows:

$$\underset{(N\times 1)}{\boldsymbol{y}_t} = \underset{(N\times K)}{\boldsymbol{\Phi}} \underset{(K\times 1)}{\boldsymbol{c}_t} + \underset{(N\times 1)}{\boldsymbol{u}_t}, \tag{2.2.3}$$

$$\underset{(K\times 1)}{\boldsymbol{c}_t} = \underset{(K\times K)}{\boldsymbol{\Gamma}} \underset{(K\times 1)}{\boldsymbol{c}_{t-1}} + \underset{(K\times 1)}{\boldsymbol{v}_t}, \tag{2.2.4}$$

where $\boldsymbol{y}_t$ is a vector of observation, $\boldsymbol{c}_t$ is a vector of common factors, $\boldsymbol{\Phi}$ is a matrix of factor loadings associated with $\boldsymbol{c}_t$, $\boldsymbol{u}_t$ is a vector of idiosyncratic noise, $\boldsymbol{\Gamma} = \text{diag}\left(\gamma_1, \gamma_2, \ldots, \gamma_K\right)$ is a matrix of autoregressive coefficients, and $\boldsymbol{v}_t$ is a vector of noise.

Further, we require that Gaussian white noises $\boldsymbol{u}_t$ and $\boldsymbol{v}_t$ be orthogonal to each other (at all leads and lags) and to the initial state:

$$\begin{bmatrix} \boldsymbol{u}_t \\ \boldsymbol{v}_t \end{bmatrix} \sim \text{i.i.d.} N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega} \end{bmatrix} \right), \tag{2.2.5}$$

$$E\left(\boldsymbol{c}_1 \boldsymbol{u}_t'\right) = \mathbf{0}, E\left(\boldsymbol{c}_1 \boldsymbol{v}_t'\right) = \mathbf{0}. \tag{2.2.6}$$

In much of our analysis, we assume that the matrix $\boldsymbol{\Sigma}$ is non-diagonal and the matrix $\boldsymbol{\Omega}$ is diagonal for simplicity. Therefore, the structure of the matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ employed here are defined as follows:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{21} & \cdots & \sigma_{N1} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{N1} & \sigma_{N2} & \cdots & \sigma_{NN} \end{bmatrix}, \boldsymbol{\Omega} = \begin{bmatrix} \omega_{11} & 0 & \cdots & 0 \\ 0 & \omega_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \omega_{KK} \end{bmatrix}, \tag{2.2.7}$$

where σ_{ij} is an ijth element of $\boldsymbol{\Sigma}$ and ω_{ij} is an ijth element of $\boldsymbol{\Omega}$. This is the basic structure we adopt for the rest of this paper[3].

In general, the state-space representation provides a powerful framework for estimation and testing of dynamic models. The recognition that the dynamic factor model is put in

[3] *Note*: This framework is also called "Approximate Dynamic Factor Model" or "Generalized Dynamic Factor Model". On the other hand, the model with orthogonal idiosyncratic noise is called "Exact Dynamic Factor Model".

state-space form is particularly useful because application of Kalman filter delivers maximum likelihood estimates (MLE) and optimal filtered and smoothed estimates of the underlying filter and smoother. In addition, we construct the test statistic from smoothed density and do not require estimation of additional parameters. Finally, the state-space representation paves the way for possible further extensions, such as allowance for heteroskedasticity and markov-switching framework, although we do not pursue those extensions in our paper[4].

2.2.2 The estimation framework

The previous subsection detailed the basic framework and notation of the model. This subsection summarizes general estimation framework of the model. Suppose we want to estimate an $(M \times 1)$ parameter vector $\boldsymbol{\theta}$ based on a time series of T observations on a vector $\boldsymbol{y}_t$. Consider the distribution $\boldsymbol{y}_t$ conditional on the information up to $t-1$,

$$f(\boldsymbol{y}_t|\boldsymbol{\psi}_{t-1};\boldsymbol{\theta}), \text{ where } \boldsymbol{\psi}_{t-1} = (\boldsymbol{y}_1', \boldsymbol{y}_2', \ldots, \boldsymbol{y}_{t-1}')'. \tag{2.2.8}$$

The task is to choose a parameter vector $\boldsymbol{\theta}$ so as to maximize the log likelihood function

$$\mathcal{L}(\boldsymbol{\theta}) \equiv \sum_{t=1}^{T} \log f(\boldsymbol{y}_t|\boldsymbol{\psi}_{t-1};\boldsymbol{\theta}). \tag{2.2.9}$$

To derive the log likelifood function (2.2.9), we specify conditional densities of the observed variable $\boldsymbol{y}_t$ and the common factor $\boldsymbol{c}_t$. From equation (2.2.3), (2.2.4), and (2.2.5), these conditional densities are defined as follows:

$$f(\boldsymbol{y}_t|\boldsymbol{c}_t;\boldsymbol{\lambda}) = (2\pi)^{-N/2} \left|\boldsymbol{\Sigma}^{-1}\right|^{1/2} \exp\left[-\frac{1}{2}(\boldsymbol{y}_t - \boldsymbol{\Phi}\boldsymbol{c}_t)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_t - \boldsymbol{\Phi}\boldsymbol{c}_t)\right], \tag{2.2.10}$$

$$f(\boldsymbol{c}_t|\boldsymbol{c}_{t-1};\boldsymbol{\eta}) = (2\pi)^{-K/2} \left|\boldsymbol{\Omega}^{-1}\right|^{1/2} \exp\left[-\frac{1}{2}(\boldsymbol{c}_t - \boldsymbol{\Gamma}\boldsymbol{c}_{t-1})'\boldsymbol{\Omega}^{-1}(\boldsymbol{c}_t - \boldsymbol{\Gamma}\boldsymbol{c}_{t-1})\right], \tag{2.2.11}$$

$$\text{where } \boldsymbol{\lambda} = ([\text{vec}(\boldsymbol{\Phi})]', [\text{vech}(\boldsymbol{\Sigma})]')', \boldsymbol{\eta} = (\gamma_1, \gamma_2, \ldots, \gamma_K, \omega_{11}, \omega_{22}, \ldots, \omega_{KK})',$$

for $t = 1, 2, \ldots, T$. Note that equation (2.2.10) and (2.2.11) are unobserved in actually. We here assume that

$$\boldsymbol{c}_1 \sim N\left(\boldsymbol{0}, [\boldsymbol{I}_{K^2} - (\boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma})]^{-1} \text{vec}(\boldsymbol{\Omega})\right),$$

where the symbol $\otimes$ denotes the kronecker product and the vec $(\cdot)$ denotes the vec-operator.

Note that the observed likelihood (2.2.8) is parameterized by $\boldsymbol{\theta}$, which includes both the parameter $\boldsymbol{\lambda}$ appearing in equation (2.2.10) and the parameter $\boldsymbol{\eta}$ in equation (2.2.11),

$$\boldsymbol{\theta} = (\boldsymbol{\lambda}', \boldsymbol{\eta}')'. \tag{2.2.12}$$

According to the above specification, Kalman filter and a smoother for the dynamic factor model are as follows.

[4] *Note*: See Kim and Nelson [1999] and Durbin and Koopman [2001] for a detailed further extensions.

First, given the density of $\boldsymbol{c}_{t-1}$ conditional on the information up to $t-1$, namely $f(\boldsymbol{c}_{t-1}|\boldsymbol{\psi}_{t-1};\boldsymbol{\theta})$, the density of the predicted state variable $\boldsymbol{c}_t$ is expressed as the integral

$$f(\boldsymbol{c}_t|\boldsymbol{\psi}_{t-1};\boldsymbol{\theta}) = \int_{-\infty}^{\infty} f(\boldsymbol{c}_t|\boldsymbol{c}_{t-1};\boldsymbol{\eta})f(\boldsymbol{c}_{t-1}|\boldsymbol{\psi}_{t-1};\boldsymbol{\theta})d\boldsymbol{c}_{t-1}, \tag{2.2.13}$$

since it is implicitly assumed that the transition equation (2.2.11) given $\boldsymbol{c}_{t-1}$ is independent of $\boldsymbol{\psi}_{t-1}$, namely $f(\boldsymbol{c}_t|\boldsymbol{c}_{t-1};\boldsymbol{\eta}) = f(\boldsymbol{c}_t|\boldsymbol{c}_{t-1},\boldsymbol{\psi}_{t-1};\boldsymbol{\eta})$. Next, we can update the conditional density of $\boldsymbol{c}_t$ by obtaining the new observation $\boldsymbol{y}_t$ as

$$f(\boldsymbol{c}_t|\boldsymbol{\psi}_t;\boldsymbol{\theta}) = \frac{f(\boldsymbol{y}_t|\boldsymbol{c}_t;\boldsymbol{\lambda})f(\boldsymbol{c}_t|\boldsymbol{\psi}_{t-1};\boldsymbol{\theta})}{f(\boldsymbol{y}_t|\boldsymbol{\psi}_{t-1};\boldsymbol{\theta})}, \tag{2.2.14}$$

where

$$f(\boldsymbol{y}_t|\boldsymbol{\psi}_{t-1};\boldsymbol{\theta}) = \int_{-\infty}^{\infty} f(\boldsymbol{y}_t|\boldsymbol{c}_t;\boldsymbol{\lambda})f(\boldsymbol{c}_t|\boldsymbol{\psi}_{t-1};\boldsymbol{\theta})d\boldsymbol{c}_t, \tag{2.2.15}$$

since it is assumed implicitly that the measurment equation (2.2.10) is independent of the past information $\boldsymbol{\psi}_{t-1}$, namely $f(\boldsymbol{y}_t|\boldsymbol{c}_t;\boldsymbol{\lambda}) = f(\boldsymbol{y}_t|\boldsymbol{c}_t,\boldsymbol{\psi}_{t-1};\boldsymbol{\lambda})$. The process of obtaining the conditional density of the state variable $\boldsymbol{c}_t$ given $\boldsymbol{\psi}_t$ is called "filtering" in the state-space framework, and we have the conditional likelihood of $\boldsymbol{y}_t$ in the denominator of equation (2.2.14) as a byproduct of filtering. We can obtain the conditional likelihood $f(\boldsymbol{y}_1;\boldsymbol{\theta}), f(\boldsymbol{y}_2|\boldsymbol{\psi}_1;\boldsymbol{\theta}),\ldots,f(\boldsymbol{y}_T|\boldsymbol{\psi}_{T-1};\boldsymbol{\theta})$ using equation (2.2.13) and (2.2.14) recursively. Thus, we obtain the unconditional density $f(\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots,\boldsymbol{y}_T)$ and the maximum likelihood estimator. This filtering algorithm was proposed by Kitagawa [1987].

In obtaining the test statistic, the conditional density of the state variable $\boldsymbol{c}_t$ given $\boldsymbol{\psi}_T$, namely $f(\boldsymbol{c}_t|\boldsymbol{\psi}_T;\boldsymbol{\theta})$, is required, as will be shown in the next section. This process is referred to by "smoothing." We see that

$$\begin{aligned} f(\boldsymbol{c}_t,\boldsymbol{c}_{t+1}|\boldsymbol{\psi}_T;\boldsymbol{\theta}) &= f(\boldsymbol{c}_{t+1}|\boldsymbol{\psi}_T;\boldsymbol{\theta})f(\boldsymbol{c}_t|\boldsymbol{c}_{t+1},\boldsymbol{\psi}_T;\boldsymbol{\theta}) \\ &= \frac{f(\boldsymbol{c}_{t+1}|\boldsymbol{\psi}_T;\boldsymbol{\theta})f(\boldsymbol{c}_{t+1}|\boldsymbol{c}_t;\boldsymbol{\eta})f(\boldsymbol{c}_t|\boldsymbol{\psi}_t;\boldsymbol{\theta})}{f(\boldsymbol{c}_{t+1}|\boldsymbol{\psi}_t;\boldsymbol{\theta})}, \end{aligned} \tag{2.2.16}$$

from the Bayes theorem and

$$f(\boldsymbol{c}_t|\boldsymbol{c}_{t+1},\boldsymbol{\psi}_t;\boldsymbol{\theta}) = f(\boldsymbol{c}_t|\boldsymbol{c}_{t+1},\boldsymbol{\psi}_T;\boldsymbol{\theta}). \tag{2.2.17}$$

The equation (2.2.17) is intuitive, since it is evident from equation (2.2.10) and (2.2.11) that, given $\boldsymbol{c}_{t+1}$, the future observations $\boldsymbol{y}_{t+1},\boldsymbol{y}_{t+2},\ldots,\boldsymbol{y}_T$ have no additional information with respect to $\boldsymbol{c}_t$. Therefore the smoothed density

$$f(\boldsymbol{c}_t|\boldsymbol{\psi}_T;\boldsymbol{\theta}) = \int_{-\infty}^{\infty} f(\boldsymbol{c}_t,\boldsymbol{c}_{t+1}|\boldsymbol{\psi}_T;\boldsymbol{\theta})d\boldsymbol{c}_{t+1}, \tag{2.2.18}$$

is derived by integrating out $\boldsymbol{c}_{t+1}$ in equation (2.2.16). Then we obtain the smoothed density at t using the smoothed density at $t+1$, the transition density $f(\boldsymbol{c}_{t+1}|\boldsymbol{c}_t;\boldsymbol{\eta})$, the filtered density $f(\boldsymbol{c}_t|\boldsymbol{\psi}_t;\boldsymbol{\theta})$ and the predicted density $f(\boldsymbol{c}_{t+1}|\boldsymbol{\psi}_t;\boldsymbol{\theta})$. This is a general estimation framework.

Note that our model is linear and assumed to be normally distributed. Therefore, conditional mean and covariance matrix can be computed through linear gaussian Kalman filter and a smoother. Hence, there is *no need for numerical integration.* For details, see Hamilton [1994] or Kim and Nelson [1999].

The use of state-space representation of the dynamic factor model is essential in this paper, since we use the formula of Hamilton [1996], who showed that the LM statistic of the Markov-switching model, a special case of the state-space model, can be obtained as expected value with respect to smoothed density. In the next section, we will show this result.

2.3 Form of the specification tests for the dynamic factor model

This section provides an expression for the score in the dynamic factor model and derives the general specification framework based on the LM principle. Subsection 2.3.1 provides an expression for the score in terms of the derivative of unobserved densities. Subsection 2.3.2 derives the general LM test statistic. Subection 2.3.3 introduces the framework of several specification tests. Subsection 2.3.4 summarizes LM tests.

Before detailing discussions, it is worth calling to three standard "regularity conditions" behind the formulas presented in this section. First, it is assumed that observations are strict stationary. Second, it is assumed that the likelihood function allows two term Taylor series expansions and the interchange of integral and derivative. Third, it is assumed that the information matrix is non-singular, so that parameters are locally identified.

2.3.1 The score for the dynamic factor model

This subsection provides an expression for the score in the dynamic factor model. The score of the tth observation is defined as

$$\boldsymbol{g}_t(\boldsymbol{\theta}) \equiv \frac{\partial \log f(\boldsymbol{y}_t|\boldsymbol{\psi}_{t-1};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}. \tag{2.3.1}$$

This score refers to the derivative of the observed conditional density (2.2.8). It turns out that the element of the score corresponding to $\boldsymbol{\lambda}$ and $\boldsymbol{\eta}$ have a simple relation to derivatives of unobserved conditional densities (2.2.10) and (2.2.11). We will show this relation below.

The density function of $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_t$ is expressed as follows:

$$f(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_t;\boldsymbol{\theta}) = \int^{(t)} f(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_t|\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_t;\boldsymbol{\lambda}) f(\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_t;\boldsymbol{\eta}) d\boldsymbol{c}^{(t)}, \tag{2.3.2}$$

where

$$f(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_t|\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_t;\boldsymbol{\lambda}) = \prod_{\tau=1}^{t} f(\boldsymbol{y}_\tau|\boldsymbol{c}_\tau;\boldsymbol{\lambda}), \tag{2.3.3}$$

$$f(\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_t;\boldsymbol{\lambda}) = \prod_{\tau=2}^{t} f(\boldsymbol{y}_\tau|\boldsymbol{c}_{\tau-1};\boldsymbol{\lambda}) \cdot f(\boldsymbol{c}_1;\boldsymbol{\lambda}). \tag{2.3.4}$$

Note that $\int^{(t)} d\boldsymbol{c}^{(t)}$ denotes multiple integration:

$$\int^{(t)} d\boldsymbol{c}^{(t)} \equiv \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} d\boldsymbol{c}_1 d\boldsymbol{c}_2 \cdots d\boldsymbol{c}_t.$$

Then, the derivative of the first t observation is

$$\begin{aligned}\frac{\partial \log f(\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots,\boldsymbol{y}_t;\boldsymbol{\theta})}{\partial \boldsymbol{\lambda}} &= \frac{1}{f(\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots,\boldsymbol{y}_t;\boldsymbol{\theta})} \\ &\times \int^{(t)} \frac{\partial f(\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots,\boldsymbol{y}_t|\boldsymbol{c}_1,\boldsymbol{c}_2,\ldots,\boldsymbol{c}_t;\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} f(\boldsymbol{c}_1,\boldsymbol{c}_2,\ldots,\boldsymbol{c}_t;\boldsymbol{\eta}) d\boldsymbol{c}^{(t)} \\ &= \int^{(t)} \frac{\partial \log f(\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots,\boldsymbol{y}_t|\boldsymbol{c}_1,\boldsymbol{c}_2,\ldots,\boldsymbol{c}_t;\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} \\ &\times \frac{f(\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots,\boldsymbol{y}_t|\boldsymbol{c}_1,\boldsymbol{c}_2,\ldots,\boldsymbol{c}_t;\boldsymbol{\lambda}) f(\boldsymbol{c}_1,\boldsymbol{c}_2,\ldots,\boldsymbol{c}_t;\boldsymbol{\eta})}{f(\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots,\boldsymbol{y}_t;\boldsymbol{\theta})} d\boldsymbol{c}^{(t)}. \quad (2.3.5)\end{aligned}$$

Using equation (2.3.3), equation (2.3.5) becomes

$$\begin{aligned}&\frac{\partial \log f(\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots,\boldsymbol{y}_t;\boldsymbol{\theta})}{\partial \boldsymbol{\lambda}} \\ &= \sum_{\tau=1}^{t} \left[\int^{(t)} \frac{\partial \log f(\boldsymbol{y}_\tau|\boldsymbol{c}_\tau;\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} \cdot f(\boldsymbol{c}_1,\boldsymbol{c}_2,\ldots,\boldsymbol{c}_t|\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots,\boldsymbol{y}_t;\boldsymbol{\theta}) d\boldsymbol{c}^{(t)} \right], \quad (2.3.6)\end{aligned}$$

and the conditional density of $\boldsymbol{c}_1,\boldsymbol{c}_2,\ldots,\boldsymbol{c}_t$ given $\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots,\boldsymbol{y}_t$ is defined by

$$f(\boldsymbol{c}_1,\boldsymbol{c}_2,\ldots,\boldsymbol{c}_t|\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots,\boldsymbol{y}_t;\boldsymbol{\theta}) = \frac{f(\boldsymbol{y}_t|\boldsymbol{c}_t;\boldsymbol{\lambda}) f(\boldsymbol{c}_t|\boldsymbol{c}_{t-1};\boldsymbol{\eta}) \cdots f(\boldsymbol{y}_1|\boldsymbol{c}_1;\boldsymbol{\lambda}) f(\boldsymbol{c}_1;\boldsymbol{\eta})}{f(\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots,\boldsymbol{y}_t;\boldsymbol{\theta})}. \quad (2.3.7)$$

The multiple integral in (2.3.6) is simplified to a one-dimensional integral, since $\partial \log f(\boldsymbol{y}_\tau|\boldsymbol{c}_\tau;\boldsymbol{\lambda})/\partial \boldsymbol{\lambda}$ depends only upon $\boldsymbol{c}_\tau$. Therefore, equation (2.3.6) becomes

$$\begin{aligned}\frac{\partial \log f(\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots,\boldsymbol{y}_t;\boldsymbol{\theta})}{\partial \boldsymbol{\lambda}} &= \sum_{\tau=1}^{t} \left[\int_{-\infty}^{\infty} \frac{\partial \log f(\boldsymbol{y}_\tau|\boldsymbol{c}_\tau;\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} \cdot f(\boldsymbol{c}_\tau|\boldsymbol{\psi}_t;\boldsymbol{\theta}) d\boldsymbol{c}_\tau \right] \\ &= \sum_{\tau=1}^{t} E[\boldsymbol{\xi}_\tau(\boldsymbol{\theta})|\boldsymbol{\psi}_t], \text{ for } t = 1,2,\ldots,T, \quad (2.3.8)\end{aligned}$$

where

$$\boldsymbol{\xi}_\tau(\boldsymbol{\theta}) = \frac{\partial \log f(\boldsymbol{y}_\tau|\boldsymbol{c}_\tau;\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}}.$$

Note that the conditional density function $f(\boldsymbol{c}_\tau|\boldsymbol{\psi}_t;\boldsymbol{\theta})$ in equation (2.3.8) can be obtained as the smoothed density of $\boldsymbol{c}_\tau$ in equation (2.2.18).

Further, the score of the tth observation with respect to $\boldsymbol{\lambda}$ can be evaluated by differencing equation (2.3.6) as

$$\frac{\partial \log f(\boldsymbol{y}_t|\boldsymbol{\psi}_{t-1};\boldsymbol{\theta})}{\partial \boldsymbol{\lambda}} = \frac{\partial \log f(\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots,\boldsymbol{y}_t;\boldsymbol{\theta})}{\partial \boldsymbol{\lambda}} - \frac{\partial \log f(\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots,\boldsymbol{y}_{t-1};\boldsymbol{\theta})}{\partial \boldsymbol{\lambda}}, \quad (2.3.9)$$

where the derivative of the log likelihood for $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_t$ for $t < T$ on the right hand side can be evaluated iteratively by applying the same routine used in obtaining equation (2.3.8) for $t = 1, 2, \ldots, T$.

Similarly, using equation (2.3.4), the score with respect to $\boldsymbol{\eta}$ becomes

$$\begin{aligned} \frac{\partial \log f\left(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_t; \boldsymbol{\theta}\right)}{\partial \boldsymbol{\eta}} &= \sum_{\tau=1}^{t} \left[\int_{-\infty}^{\infty} \frac{\partial \log f\left(\boldsymbol{c}_\tau | \, \boldsymbol{c}_{\tau-1}; \boldsymbol{\eta}\right)}{\partial \boldsymbol{\eta}} \cdot f\left(\boldsymbol{c}_\tau | \, \boldsymbol{\psi}_t; \boldsymbol{\theta}\right) d\boldsymbol{c}_\tau \right] \\ &= \sum_{\tau=1}^{t} E[\boldsymbol{\xi}_\tau(\boldsymbol{\theta}) | \boldsymbol{\psi}_t], \text{ for } t = 1, 2, \ldots, T, \end{aligned} \tag{2.3.10}$$

where

$$\boldsymbol{\xi}_\tau(\boldsymbol{\theta}) = \frac{\partial \log f\left(\boldsymbol{c}_\tau | \, \boldsymbol{c}_{\tau-1}; \boldsymbol{\eta}\right)}{\partial \boldsymbol{\eta}}.$$

Then, the score of the tth observation with respect to $\boldsymbol{\eta}$ can be evaluated as

$$\frac{\partial \log f(\boldsymbol{y}_t | \boldsymbol{\psi}_{t-1}; \boldsymbol{\theta})}{\partial \boldsymbol{\eta}} = \frac{\partial \log f(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_t; \boldsymbol{\theta})}{\partial \boldsymbol{\eta}} - \frac{\partial \log f(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_{t-1}; \boldsymbol{\theta})}{\partial \boldsymbol{\eta}}. \tag{2.3.11}$$

Therefore, all of the elements of the score take the form of simple function of smoothed density. Note that the integration is necessary for derivation of equation (2.3.8) and (2.3.10). But the whole score vector for Gaussian linear state-space model can be computed exactly in a single pass of the Kalman filter and a smoother, so there is *no need for numerical integration*. For details, see appendix A.

Because the score derived here can be used to construct an estimate of the information matrix, this step is the most important part of the calculation of test statistic.

2.3.2 The general principle of specification tests

This subsection provides general approaches that can be used to test specification based on the LM principle. Suppose that the conditional density of the tth observation is given by

$$\int_{-\infty}^{\infty} f(\boldsymbol{y}_t | \boldsymbol{\psi}_{t-1}; \boldsymbol{\theta}) d\boldsymbol{y}_t = 1, \tag{2.3.12}$$

Since equation (2.3.12) holds for all admissible values of $\boldsymbol{\theta}$, we can differentiate both sides with respect to $\boldsymbol{\theta}$ to conclude that

$$\int_{-\infty}^{\infty} \frac{\partial f(\boldsymbol{y}_t | \boldsymbol{\psi}_{t-1}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} d\boldsymbol{y}_t = \mathbf{0}. \tag{2.3.13}$$

In addition, we multiply and divide the integrand in equation (2.3.13) by the conditional density of $\boldsymbol{y}_t$:

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{\partial f(\boldsymbol{y}_t | \boldsymbol{\psi}_{t-1}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \cdot \frac{1}{f(\boldsymbol{y}_t | \boldsymbol{\psi}_{t-1}; \boldsymbol{\theta})} \cdot f(\boldsymbol{y}_t | \boldsymbol{\psi}_{t-1}; \boldsymbol{\theta}) d\boldsymbol{y}_t = \mathbf{0} \\ \int_{-\infty}^{\infty} \frac{\partial \log f(\boldsymbol{y}_t | \boldsymbol{\psi}_{t-1}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \cdot f(\boldsymbol{y}_t | \boldsymbol{\psi}_{t-1}; \boldsymbol{\theta}) d\boldsymbol{y}_t = \mathbf{0}. \end{aligned} \tag{2.3.14}$$

Substitution of equation (2.3.1) into equation (2.3.14) reveals that

$$\int_{-\infty}^{\infty} \boldsymbol{g}_t(\boldsymbol{\theta}) \cdot f(\boldsymbol{y}_t|\boldsymbol{\psi}_{t-1};\boldsymbol{\theta})d\boldsymbol{y}_t = \boldsymbol{0}. \tag{2.3.15}$$

Equation (2.3.15) indicates that if the data were really generated by the density (2.2.11), then the expected value of the score conditional on information observed through date $t-1$ should be zero:

$$E\left[\boldsymbol{g}_t(\boldsymbol{\theta}_0)|\, \boldsymbol{\psi}_{t-1}\right] = \boldsymbol{0}, \tag{2.3.16}$$

where $\boldsymbol{\theta}_0$ denotes true parameter value. In other words, the score function $\{\boldsymbol{g}_t(\boldsymbol{\theta})\}_{t=1}^{T}$ should form a martingale difference sequence. Thus if the model is correctly specified, the score $\boldsymbol{g}_t(\boldsymbol{\theta}_0)$ should be impossible to forecast on the basis of any information available at date $t-1$, such as elements of the lagged score $\boldsymbol{g}_{t-1}(\boldsymbol{\theta}_0)$.

The vast majority of all testing problem is composite so that only a subset of parameters is fixed under the null. Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$. The first $(M_1 \times 1)$ parameters are specified under the null hypothesis to be $\boldsymbol{\theta}_{10}$, whereas the remaining $(M_2 \times 1)$ parameters are unrestricted under both the null and the alternative. Thus, the maximum likelihood estimate of $\boldsymbol{\theta}$ under the null is denoted $\tilde{\boldsymbol{\theta}} \equiv (\boldsymbol{\theta}_{10}', \tilde{\boldsymbol{\theta}}_2')'$. Note that $M_1 + M_2 = M$.

Then at the constrained MLE $\tilde{\boldsymbol{\theta}}$, the first M_1 elements of the average score $T^{-1}\sum_{t=1}^{T}\boldsymbol{g}_t(\tilde{\boldsymbol{\theta}})$ are nonzero, whereas the last M_2 are zero. The magnitude of these first M_1 elements reflects how much the likelihood function might increase if constraints were relaxed, and can be used to assess the validity of constraints. Then, if we use the BHHH (outer product of gradients) estimator to estimate the information matrix, the Lagrange Multiplier test of the null hypothesis that restrictions are true is given by the following statistic:

$$LM = \left[T^{-1/2}\sum_{t=1}^{T}\boldsymbol{g}_t(\tilde{\boldsymbol{\theta}})\right]' \cdot \left[T^{-1}\sum_{t=1}^{T}\boldsymbol{g}_t(\tilde{\boldsymbol{\theta}})\boldsymbol{g}_t'(\tilde{\boldsymbol{\theta}})\right]^{-1} \cdot \left[T^{-1/2}\sum_{t=1}^{T}\boldsymbol{g}_t(\tilde{\boldsymbol{\theta}})\right] \xrightarrow{d} \chi^2_{M_1}. \tag{2.3.17}$$

Namely, under the null hypothesis, the LM test statistic has a limiting χ^2 distribution with degrees of freedom equal to the number of restrictions.

In addition, let $\boldsymbol{g}_t(\tilde{\boldsymbol{\theta}})$ denote the tth term in the gradient of the log-likelihood function at the constrained MLE $\tilde{\boldsymbol{\theta}}$. Then

$$\boldsymbol{g}(\tilde{\boldsymbol{\theta}}) = \sum_{t=1}^{T}\boldsymbol{g}_t(\tilde{\boldsymbol{\theta}}) = \boldsymbol{G}'(\tilde{\boldsymbol{\theta}})\boldsymbol{i}_T. \tag{2.3.18}$$

where $\boldsymbol{G}(\tilde{\boldsymbol{\theta}})$ is the $(T \times M)$ matrix with tth row equal to $\boldsymbol{g}_t'(\tilde{\boldsymbol{\theta}})$ and $\boldsymbol{i}_T$ is a column of 1s of order T. Then, equation (2.3.17) becomes

$$LM = \boldsymbol{i}_T'\boldsymbol{G}(\tilde{\boldsymbol{\theta}})\left[\boldsymbol{G}'(\tilde{\boldsymbol{\theta}})\boldsymbol{G}(\tilde{\boldsymbol{\theta}})\right]^{-1}\boldsymbol{G}'(\tilde{\boldsymbol{\theta}})\boldsymbol{i}_T \xrightarrow{d} \chi^2_{M_1}. \tag{2.3.19}$$

Furthermore, let the matrix $\boldsymbol{G}(\tilde{\boldsymbol{\theta}})$ be partitioned as follows.

$$\boldsymbol{G}(\tilde{\boldsymbol{\theta}}) = \left[\begin{array}{cc} \underset{(T\times M_1)}{\boldsymbol{G}_1(\tilde{\boldsymbol{\theta}})} & \underset{(T\times M_2)}{\boldsymbol{G}_2(\tilde{\boldsymbol{\theta}})} \end{array}\right]. \tag{2.3.20}$$

Then, equation (2.3.19) can be rewritten as follows[5]:

$$LM = \boldsymbol{s}_1'(\tilde{\boldsymbol{\theta}})\mathscr{I}_{OP}^{11}(\tilde{\boldsymbol{\theta}})\boldsymbol{s}_1(\tilde{\boldsymbol{\theta}}) \xrightarrow{d} \chi^2_{M_1}, \tag{2.3.21}$$

where

$$\boldsymbol{s}_1(\tilde{\boldsymbol{\theta}}) \equiv \left(\begin{bmatrix} \boldsymbol{I}_{M_1} & \boldsymbol{O}_{M_1,M_2} \end{bmatrix} \begin{bmatrix} \boldsymbol{G}_1'(\tilde{\boldsymbol{\theta}}) \\ \boldsymbol{G}_2'(\tilde{\boldsymbol{\theta}}) \end{bmatrix} \right) \boldsymbol{i}_T, \tag{2.3.22}$$

$$\mathscr{I}_{OP}^{11}(\tilde{\boldsymbol{\theta}}) \equiv \begin{bmatrix} \boldsymbol{I}_{M_1} & \boldsymbol{O}_{M_1,M_2} \end{bmatrix} \left[\boldsymbol{G}'(\tilde{\boldsymbol{\theta}})\boldsymbol{G}(\tilde{\boldsymbol{\theta}})\right]^{-1} \begin{bmatrix} \boldsymbol{I}_{M_1} \\ \boldsymbol{O}_{M_1,M_2}' \end{bmatrix}. \tag{2.3.23}$$

Note that $\boldsymbol{I}_{M_1}$ is the identity matrix of order M_1 and $\boldsymbol{O}_{M_1,M_2}$ is the $(M_1 \times M_2)$ null matrix. Equation (2.3.19) and (2.3.21) provide an extremely useful class of diagnostic tests, enabling one to estimate a restricted model and test it against alternative without having to estimate the more general model. Further, the test proposed here employs an estimate of the information matrix based on the average outer product of the score. This form has the advantage that it is very simple to compute, but has the disadvantage that it may have substantially poorer finite sample properties than an estimate of the information matrix based on the average of the matrix of second derivative.

2.3.3 Specification tests for the dynamic factor model

The partitioned framework and the notation of the model

As we might be interested in a restricted system, we partition the set of observed variables into two groups of N_1 and N_2 variables and the set of common factors into two groups of K_1 and K_2 variables. Then, the system in equation (2.2.3) and (2.2.4) are partitioned as

$$\begin{bmatrix} \underset{(N_1\times 1)}{\boldsymbol{y}_{1t}} \\ \underset{(N_2\times 1)}{\boldsymbol{y}_{2t}} \end{bmatrix} = \begin{bmatrix} \underset{(N_1\times K_1)}{\boldsymbol{\Phi}_{11}} & \underset{(N_1\times K_2)}{\boldsymbol{\Phi}_{12}} \\ \underset{(N_2\times K_1)}{\boldsymbol{\Phi}_{21}} & \underset{(N_2\times K_2)}{\boldsymbol{\Phi}_{22}} \end{bmatrix} \begin{bmatrix} \underset{(K_1\times 1)}{\boldsymbol{c}_{1t}} \\ \underset{(K_2\times 1)}{\boldsymbol{c}_{2t}} \end{bmatrix} + \begin{bmatrix} \underset{(N_1\times 1)}{\boldsymbol{u}_{1t}} \\ \underset{(N_2\times 1)}{\boldsymbol{u}_{2t}} \end{bmatrix}, \tag{2.3.24}$$

$$\begin{bmatrix} \underset{(K_1\times 1)}{\boldsymbol{c}_{1t}} \\ \underset{(K_2\times 1)}{\boldsymbol{c}_{2t}} \end{bmatrix} = \begin{bmatrix} \underset{(K_1\times K_1)}{\boldsymbol{\Gamma}_1} & \underset{(K_1\times K_2)}{\boldsymbol{0}} \\ \underset{(K_2\times K_1)}{\boldsymbol{0}} & \underset{(K_2\times K_2)}{\boldsymbol{\Gamma}_2} \end{bmatrix} \begin{bmatrix} \underset{(K_1\times 1)}{\boldsymbol{c}_{1,t-1}} \\ \underset{(K_2\times 1)}{\boldsymbol{c}_{2,t-1}} \end{bmatrix} + \begin{bmatrix} \underset{(K_1\times 1)}{\boldsymbol{v}_{1t}} \\ \underset{(K_2\times 1)}{\boldsymbol{v}_{2t}} \end{bmatrix}. \tag{2.3.25}$$

Further, the matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ are partitioned as

$$\boldsymbol{\Sigma} = \begin{bmatrix} \underset{(N_1\times N_1)}{\boldsymbol{\Sigma}_{11}} & \underset{(N_1\times N_2)}{\boldsymbol{\Sigma}_{12}} \\ \underset{(N_2\times N_1)}{\boldsymbol{\Sigma}_{12}'} & \underset{(N_2\times N_2)}{\boldsymbol{\Sigma}_{22}} \end{bmatrix}, \boldsymbol{\Omega} = \begin{bmatrix} \underset{(K_1\times K_1)}{\boldsymbol{\Omega}_1} & \underset{(K_1\times K_2)}{\boldsymbol{0}} \\ \underset{(K_1\times K_2)}{\boldsymbol{0}} & \underset{(K_2\times K_2)}{\boldsymbol{\Omega}_2} \end{bmatrix}. \tag{2.3.26}$$

[5] *Note*: The small-sample properties of specification tests are often better approximated by an F distribution than by an asymptotically equivalent χ^2 distribution-see for example Mackinnon and White [1985], Kiviet [1986], and Ericsson [1991]. This small-sample adjustment involves two steps. First, there is a degree-of-freedom adjustment to the estimate of the variance-covariance matrix that appear in a denominator of the test statistics. Second, an F distribution rather than χ^2 distribution is used to interpret the resulting test statistics. Consequently, asymptotic Lagrange Multiplier tests with better small-sample performance are obtained by multiplying (2.3.21) by $(T - M_2)/(TM_1)$ and comparing the results with an $F(M_1, T - M_2)$ distribution.

Note that $N_1 + N_2 = N, K_1 + K_2 = K$, and the matrices $\mathbf{\Gamma}_1, \mathbf{\Gamma}_2, \mathbf{\Omega}_1$, and $\mathbf{\Omega}_2$ are diagonal.

According to the above specification, the conditional joint log likelihood of $\boldsymbol{y}_{1t}$ and $\boldsymbol{y}_{2t}$ is

$$\begin{aligned}\log f(\boldsymbol{y}_{1t}, \boldsymbol{y}_{2t}|\boldsymbol{c}_{1t}, \boldsymbol{c}_{2t}) = &-\frac{N}{2}\log 2\pi - \frac{1}{2}\log\left|\begin{array}{cc}\mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12}\\ \mathbf{\Sigma}_{12}' & \mathbf{\Sigma}_{22}\end{array}\right|\\ &-\frac{1}{2}\left[\begin{array}{c}\boldsymbol{y}_{1t} - (\mathbf{\Phi}_{11}\boldsymbol{c}_{1t} + \mathbf{\Phi}_{12}\boldsymbol{c}_{2t})\\ \boldsymbol{y}_{2t} - (\mathbf{\Phi}_{21}\boldsymbol{c}_{1t} + \mathbf{\Phi}_{22}\boldsymbol{c}_{2t})\end{array}\right]'\left[\begin{array}{cc}\mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12}\\ \mathbf{\Sigma}_{12}' & \mathbf{\Sigma}_{22}\end{array}\right]^{-1}\\ &\times\left[\begin{array}{c}\boldsymbol{y}_{1t} - (\mathbf{\Phi}_{11}\boldsymbol{c}_{1t} + \mathbf{\Phi}_{12}\boldsymbol{c}_{2t})\\ \boldsymbol{y}_{2t} - (\mathbf{\Phi}_{21}\boldsymbol{c}_{1t} + \mathbf{\Phi}_{22}\boldsymbol{c}_{2t})\end{array}\right].\end{aligned} \tag{2.3.27}$$

Then, the determinant of partitioned variance covariance matrix of (2.3.27) is

$$\left|\begin{array}{cc}\mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12}\\ \mathbf{\Sigma}_{12}' & \mathbf{\Sigma}_{22}\end{array}\right| = |\mathbf{\Sigma}_{22}| \cdot \left|\mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{12}'\right|. \tag{2.3.28}$$

And, the inverse of partitioned variance covariance matrix of (2.3.27) is

$$\begin{aligned}\left[\begin{array}{cc}\mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12}\\ \mathbf{\Sigma}_{12}' & \mathbf{\Sigma}_{22}\end{array}\right]^{-1} &\equiv \left[\begin{array}{cc}\mathbf{\Sigma}^{11} & \mathbf{\Sigma}^{12}\\ (\mathbf{\Sigma}^{12})' & \mathbf{\Sigma}^{22}\end{array}\right]\\ &= \left[\begin{array}{cc}\mathbf{\Sigma}^{11} & -\mathbf{\Sigma}^{11}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\\ (-\mathbf{\Sigma}^{11}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1})' & \mathbf{\Sigma}_{22}^{-1} + \mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{12}'\mathbf{\Sigma}^{11}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\end{array}\right],\end{aligned} \tag{2.3.29}$$

where $\mathbf{\Sigma}^{11} = (\mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{12}')^{-1}$ and superscripts denote partitioned inverses.

Thus, substituting (2.3.28) and (2.3.29) into (2.3.27), the conditional joint log likelihood function of $\boldsymbol{y}_{1t}$ and $\boldsymbol{y}_{2t}$ can be written as

$$\begin{aligned}\log f(\boldsymbol{y}_{1t}, \boldsymbol{y}_{2t}|\boldsymbol{c}_{1t}, \boldsymbol{c}_{2t}) = &-\frac{N}{2}\log 2\pi - \frac{1}{2}\log|\mathbf{\Sigma}_{22}| - \frac{1}{2}\log\left|\mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{12}'\right|\\ &-\frac{1}{2}\boldsymbol{u}_{1t}'\mathbf{\Sigma}^{11}\boldsymbol{u}_{1t} - \frac{1}{2}\boldsymbol{u}_{2t}'\mathbf{\Sigma}^{22}\boldsymbol{u}_{2t} - \boldsymbol{u}_{1t}'\mathbf{\Sigma}^{12}\boldsymbol{u}_{2t},\end{aligned} \tag{2.3.30}$$

where

$$\boldsymbol{u}_{1t} \equiv \boldsymbol{y}_{1t} - (\mathbf{\Phi}_{11}\boldsymbol{c}_{1t} + \mathbf{\Phi}_{12}\boldsymbol{c}_{2t}), \tag{2.3.31}$$

$$\boldsymbol{u}_{2t} \equiv \boldsymbol{y}_{2t} - (\mathbf{\Phi}_{21}\boldsymbol{c}_{1t} + \mathbf{\Phi}_{22}\boldsymbol{c}_{2t}). \tag{2.3.32}$$

Similarly, the conditional joint log likelihood function of $\boldsymbol{c}_{1t}$ and $\boldsymbol{c}_{2t}$ is

$$\begin{aligned}\log f(\boldsymbol{c}_{1t}, \boldsymbol{c}_{2t}|\,\boldsymbol{c}_{1,t-1}, \boldsymbol{c}_{2,t-1}) = &-\frac{K}{2}\log 2\pi - \frac{1}{2}\log\left|\begin{array}{cc}\mathbf{\Omega}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{\Omega}_2\end{array}\right|\\ &-\frac{1}{2}\left[\begin{array}{c}\boldsymbol{c}_{1t} - \mathbf{\Gamma}_1\boldsymbol{c}_{1,t-1}\\ \boldsymbol{c}_{2t} - \mathbf{\Gamma}_2\boldsymbol{c}_{2,t-1}\end{array}\right]'\left[\begin{array}{cc}\mathbf{\Omega}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{\Omega}_2\end{array}\right]^{-1}\left[\begin{array}{c}\boldsymbol{c}_{1t} - \mathbf{\Gamma}_1\boldsymbol{c}_{1,t-1}\\ \boldsymbol{c}_{2t} - \mathbf{\Gamma}_2\boldsymbol{c}_{2,t-1}\end{array}\right].\end{aligned} \tag{2.3.33}$$

Then, the determinant of partitioned variance covariance matrix of (2.3.33) is

$$\begin{vmatrix} \boldsymbol{\Omega}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_2 \end{vmatrix} = |\boldsymbol{\Omega}_1| \cdot |\boldsymbol{\Omega}_2| . \tag{2.3.34}$$

And, the inverse of partitioned variance covariance matrix of (2.3.33) is

$$\begin{bmatrix} \boldsymbol{\Omega}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_2 \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{\Omega}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_2^{-1} \end{bmatrix} . \tag{2.3.35}$$

Thus, substituting (2.3.34) and (2.3.35) into (2.3.33), the conditional joint log likelihood function of $\boldsymbol{c}_{1t}$ and $\boldsymbol{c}_{2t}$ can be written as

$$\begin{aligned} \log f(\boldsymbol{c}_{1t}, \boldsymbol{c}_{2t} | \boldsymbol{c}_{1,t-1}, \boldsymbol{c}_{2,t-1}) = & -\frac{K}{2} \log 2\pi - \frac{1}{2} \log |\boldsymbol{\Omega}_1| - \frac{1}{2} \log |\boldsymbol{\Omega}_2| \\ & - \frac{1}{2} \boldsymbol{v}_{1t}' \boldsymbol{\Omega}_1^{-1} \boldsymbol{v}_{1t} - \frac{1}{2} \boldsymbol{v}_{2t}' \boldsymbol{\Omega}_2^{-1} \boldsymbol{v}_{2t}, \end{aligned} \tag{2.3.36}$$

where

$$\boldsymbol{v}_{1t} \equiv \boldsymbol{c}_{1t} - \boldsymbol{\Gamma}_1 \boldsymbol{c}_{1,t-1}, \tag{2.3.37}$$

$$\boldsymbol{v}_{2t} \equiv \boldsymbol{c}_{2t} - \boldsymbol{\Gamma}_2 \boldsymbol{c}_{2,t-1}. \tag{2.3.38}$$

According to above specifications, we will propose specification tests for weak exogeneity and linear dependency.

The extended framework and the notation of the model

As we might be interested in omitted variable problem, we extend *exogenous variable* to the observation equation. Then the system in equation (2.2.3) and (2.2.4) are augmented as follows:

$$\underset{(N\times 1)}{\boldsymbol{y}_t} = \underset{(N\times K)}{\boldsymbol{\Phi}} \underset{(K\times 1)}{\boldsymbol{c}_t} + \underset{(N\times J)}{\boldsymbol{\Pi}} \underset{(J\times 1)}{\boldsymbol{x}_t} + \underset{(N\times 1)}{\boldsymbol{u}_t}, \tag{2.3.39}$$

$$\underset{(K\times 1)}{\boldsymbol{c}_t} = \underset{(K\times K)}{\boldsymbol{\Gamma}} \underset{(K\times 1)}{\boldsymbol{c}_{t-1}} + \underset{(K\times 1)}{\boldsymbol{v}_t}. \tag{2.3.40}$$

According to the above specification, the conditional likelihood of $\boldsymbol{y}_t$ and $\boldsymbol{c}_t$ are

$$\log f(\boldsymbol{y}_t | \boldsymbol{c}_t, \boldsymbol{x}_t) = -\frac{N}{2} \log 2\pi - \frac{1}{2} \log |\boldsymbol{\Sigma}| - \frac{1}{2} \boldsymbol{u}_t' \boldsymbol{\Sigma}^{-1} \boldsymbol{u}_t, \tag{2.3.41}$$

$$\log f(\boldsymbol{c}_t | \boldsymbol{c}_{t-1}) = -\frac{K}{2} \log 2\pi - \frac{1}{2} \log |\boldsymbol{\Omega}| - \frac{1}{2} \boldsymbol{v}_t' \boldsymbol{\Omega}^{-1} \boldsymbol{v}_t, \tag{2.3.42}$$

where

$$\boldsymbol{u}_t \equiv \boldsymbol{y}_t - \boldsymbol{\Phi} \boldsymbol{c}_t - \boldsymbol{\Pi} \boldsymbol{x}_t, \tag{2.3.43}$$

$$\boldsymbol{v}_t \equiv \boldsymbol{c}_t - \boldsymbol{\Gamma} \boldsymbol{c}_{t-1}. \tag{2.3.44}$$

According to above specification, we will propose a specification test for omitted explanatory variable.

Several specification tests in the model

The purpose of this subsection is to bring together a number of results on specification tests for weak exogeneity, linear dependency, and omitted explanatory variables.

Before detailing discussions, it is worth to notice that weak exogeneity test is source of controversy partly because of the variety of definitions of exogeneity implicit in the formulation of the hypothesis. In this paper, the notion of weak exogeneity as formulated by Engle et al. [1983] is used in the context of the dynamic factor model. In this case, weak exogeneity is essentially that the equations defining weak exogeneity implies, in addition, that the variables in equation cannot be forecast by past values of endogenous variables which is the definitions implicit in Granger [1968] "non-causality."

In addition, if the data are generated by equation (2.2.3) and (2.2.4) with $\boldsymbol{\Pi} \neq \mathbf{0}$, then the omission of variable of $\boldsymbol{\Pi}$ will render the parameters biased and inconsistent and cause autocorrelated disturbances and heteroskedasticity. Therefore, testing for omitted variables is particularly important.

Several specification tests are as follows.

Example 1: Specification test for weak exogeneity of $\boldsymbol{y}_{1t}$

The group of the variable represented by $\boldsymbol{y}_1$ is said to *block-exogeneous in the time series sense* with respect to the variable in $\boldsymbol{c}_2$ if the element in $\boldsymbol{c}_2$ is of no help improving a forecast of any variable contained in $\boldsymbol{y}_1$ alone. In the system of (2.3.24), (2.3.25), and (2.3.26), $\boldsymbol{y}_1$ is block-exogenous when $\boldsymbol{\Phi}_{12} = \boldsymbol{\Sigma}_{12} = \mathbf{0}$. Therefore, we conduct a LM test of the null hypothesis

$$H_0 : \boldsymbol{\Phi}_{12} = \boldsymbol{\Sigma}_{12} = \mathbf{0}. \tag{2.3.45}$$

Thus, the differential with respect to $\mathrm{vec}(\boldsymbol{\Phi}_{12})$ under the null is

$$\begin{aligned}
d \log f\left(\boldsymbol{y}_{1t}, \boldsymbol{y}_{2t} \middle| \boldsymbol{c}_{1t}, \boldsymbol{c}_{2t}\right) &= -\frac{1}{2} d\left(\boldsymbol{u}_{1t}' \boldsymbol{\Sigma}^{11} \boldsymbol{u}_{1t}\right) = -\frac{1}{2} \mathrm{tr}\left[d\left(\boldsymbol{u}_{1t}' \boldsymbol{\Sigma}^{11} \boldsymbol{u}_{1t}\right)\right] \\
&= -\frac{1}{2} \mathrm{tr}\left[\left(d\boldsymbol{u}_{1t}\right)' \boldsymbol{\Sigma}_{11}^{-1}\left(\boldsymbol{y}_{1t} - \boldsymbol{\Phi}_{11}\boldsymbol{c}_{1t}\right) + \left(\boldsymbol{y}_{1t} - \boldsymbol{\Phi}_{11}\boldsymbol{c}_{1t}\right)' \boldsymbol{\Sigma}_{11}^{-1}\left(d\boldsymbol{u}_{1t}\right)\right] \\
&= -\mathrm{tr}\left[\left(\boldsymbol{y}_{1t} - \boldsymbol{\Phi}_{11}\boldsymbol{c}_{1t}\right)' \boldsymbol{\Sigma}_{11}^{-1}\left(d\boldsymbol{u}_{1t}\right)\right] = -\mathrm{tr}\left[\left(\boldsymbol{y}_{1t} - \boldsymbol{\Phi}_{11}\boldsymbol{c}_{1t}\right)' \boldsymbol{\Sigma}_{11}^{-1}\left(-d\boldsymbol{\Phi}_{12}\right)\boldsymbol{c}_{2t}\right] \\
&= \mathrm{tr}\left[\boldsymbol{c}_{2t}\left(\boldsymbol{y}_{1t} - \boldsymbol{\Phi}_{11}\boldsymbol{c}_{1t}\right)' \boldsymbol{\Sigma}_{11}^{-1}\left(d\boldsymbol{\Phi}_{12}\right)\right] = \mathrm{tr}\left[\left\{\boldsymbol{\Sigma}_{11}^{-1}\left(\boldsymbol{y}_{1t} - \boldsymbol{\Phi}_{11}\boldsymbol{c}_{1t}\right)\boldsymbol{c}_{2t}'\right\}'\left(d\boldsymbol{\Phi}_{12}\right)\right] \\
&= \left(\mathrm{vec}\left[\boldsymbol{\Sigma}_{11}^{-1}\left(\boldsymbol{y}_{1t} - \boldsymbol{\Phi}_{11}\boldsymbol{c}_{1t}\right)\boldsymbol{c}_{2t}'\right]\right)' d\mathrm{vec}(\boldsymbol{\Phi}_{12}),
\end{aligned} \tag{2.3.46}$$

using identity $\boldsymbol{\Sigma}^{11} = \boldsymbol{\Sigma}_{11}^{-1}$ and $\mathrm{tr}(\boldsymbol{AB}) = [\mathrm{vec}(\boldsymbol{A}')]' \,\mathrm{vec}(\boldsymbol{B})$. Note that $\mathrm{tr}(\cdot)$ denotes the trace-operator. Then, the derivative with respect to $\mathrm{vec}(\boldsymbol{\Phi}_{12})$ is

$$\left.\frac{\partial \log f\left(\boldsymbol{y}_{1t}, \boldsymbol{y}_{2t} \middle| \boldsymbol{c}_{1t}, \boldsymbol{c}_{2t}\right)}{\partial \mathrm{vec}(\boldsymbol{\Phi}_{12})}\right|_{\boldsymbol{\Phi}_{12}=\boldsymbol{\Sigma}_{12}=\mathbf{0}} = \mathrm{vec}\left[\boldsymbol{\Sigma}_{11}^{-1}\left(\boldsymbol{y}_{1t} - \boldsymbol{\Phi}_{11}\boldsymbol{c}_{1t}\right)\boldsymbol{c}_{2t}'\right]. \tag{2.3.47}$$

The differential with respect to vec($\boldsymbol{\Sigma}_{12}$) under the null is

$$\begin{aligned} d\log f\left(\left.\boldsymbol{y}_{1t},\boldsymbol{y}_{2t}\right|\boldsymbol{c}_{1t},\boldsymbol{c}_{2t}\right) &= -d\left(\boldsymbol{u}_{1t}'\boldsymbol{\Sigma}^{12}\boldsymbol{u}_{2t}\right) = -\mathrm{tr}\left[\left(\boldsymbol{y}_{1t}-\boldsymbol{\Phi}_{11}\boldsymbol{c}_{1t}\right)'\left(d\boldsymbol{\Sigma}^{12}\right)\boldsymbol{u}_{2t}\right] \\ &= -\mathrm{tr}\left(\left(\boldsymbol{y}_{1t}-\boldsymbol{\Phi}_{11}\boldsymbol{c}_{1t}\right)'\left[-\boldsymbol{\Sigma}_{11}^{-1}\left(d\boldsymbol{\Sigma}_{12}\right)\boldsymbol{\Sigma}_{22}^{-1}\right]\boldsymbol{u}_{2t}\right) \\ &= \mathrm{tr}\left[\left(\boldsymbol{y}_{1t}-\boldsymbol{\Phi}_{11}\boldsymbol{c}_{1t}\right)'\boldsymbol{\Sigma}_{11}^{-1}\left(d\boldsymbol{\Sigma}_{12}\right)\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{u}_{2t}\right] \\ &= \mathrm{tr}\left[\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{u}_{2t}\left(\boldsymbol{y}_{1t}-\boldsymbol{\Phi}_{11}\boldsymbol{c}_{1t}\right)'\boldsymbol{\Sigma}_{11}^{-1}\left(d\boldsymbol{\Sigma}_{12}\right)\right] \\ &= \mathrm{tr}\left(\left[\boldsymbol{\Sigma}_{11}^{-1}\left(\boldsymbol{y}_{1t}-\boldsymbol{\Phi}_{11}\boldsymbol{c}_{1t}\right)\boldsymbol{u}_{2t}'\boldsymbol{\Sigma}_{22}^{-1}\right]'\left(d\boldsymbol{\Sigma}_{12}\right)\right) \\ &= \left(\mathrm{vec}\left[\boldsymbol{\Sigma}_{11}^{-1}\left(\boldsymbol{y}_{1t}-\boldsymbol{\Phi}_{11}\boldsymbol{c}_{1t}\right)\boldsymbol{u}_{2t}'\boldsymbol{\Sigma}_{22}^{-1}\right]\right)'d\mathrm{vec}(\boldsymbol{\Sigma}_{12}), \end{aligned} \tag{2.3.48}$$

using identity $\boldsymbol{\Sigma}^{11}=\boldsymbol{\Sigma}_{11}^{-1}$ and $\boldsymbol{\Sigma}^{12}=-\boldsymbol{\Sigma}^{11}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}$. Then, the derivative with respect to vec($\boldsymbol{\Sigma}_{12}$) is

$$\left.\frac{\partial\log f\left(\left.\boldsymbol{y}_{1t},\boldsymbol{y}_{2t}\right|\boldsymbol{c}_{1t},\boldsymbol{c}_{2t}\right)}{\partial\mathrm{vec}(\boldsymbol{\Sigma}_{12})}\right|_{\boldsymbol{\Phi}_{12}=\boldsymbol{\Sigma}_{12}=\mathbf{0}} = \mathrm{vec}\left[\boldsymbol{\Sigma}_{11}^{-1}\left(\boldsymbol{y}_{1t}-\boldsymbol{\Phi}_{11}\boldsymbol{c}_{1t}\right)\boldsymbol{u}_{2t}'\boldsymbol{\Sigma}_{22}^{-1}\right]. \tag{2.3.49}$$

And, the number of restricted parameters under the null hypothesis is $N_1(N_2+K_2)$.

Example 2: Specification test for weak exogeneity of $\boldsymbol{y}_2$

As well as example 1, the group of the variable represented by $\boldsymbol{y}_2$ is said to *block-exogeneous in the time series sense* with respect to the variable in $\boldsymbol{c}_1$ if the element in $\boldsymbol{c}_1$ is of no help improving a forecast of any variable contained in $\boldsymbol{y}_2$ alone. In the system of (2.3.24), (2.3.25), and (2.3.26), $\boldsymbol{y}_2$ is block-exogenous when $\boldsymbol{\Phi}_{21}=\boldsymbol{\Sigma}_{12}=\mathbf{0}$. Therefore, we conduct a LM test of the null hypothesis

$$H_0:\ \boldsymbol{\Phi}_{21}=\boldsymbol{\Sigma}_{12}=\mathbf{0}. \tag{2.3.50}$$

Then, derivatives with respect to vec($\boldsymbol{\Phi}_{21}$) and vec($\boldsymbol{\Sigma}_{12}$) are

$$\left.\frac{\partial\log f\left(\left.\boldsymbol{y}_{1t},\boldsymbol{y}_{2t}\right|\boldsymbol{c}_{1t},\boldsymbol{c}_{2t}\right)}{\partial\mathrm{vec}\left(\boldsymbol{\Phi}_{21}\right)}\right|_{\boldsymbol{\Phi}_{21}=\boldsymbol{\Sigma}_{12}=\mathbf{0}} = \mathrm{vec}\left[\boldsymbol{\Sigma}_{22}^{-1}\left(\boldsymbol{y}_{2t}-\boldsymbol{\Phi}_{22}\boldsymbol{c}_{2t}\right)\boldsymbol{c}_{1t}'\right], \tag{2.3.51}$$

$$\left.\frac{\partial\log f\left(\left.\boldsymbol{y}_{1t},\boldsymbol{y}_{2t}\right|\boldsymbol{c}_{1t},\boldsymbol{c}_{2t}\right)}{\partial\mathrm{vec}\left(\boldsymbol{\Sigma}_{12}\right)}\right|_{\boldsymbol{\Phi}_{21}=\boldsymbol{\Sigma}_{12}=\mathbf{0}} = \mathrm{vec}\left[\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{u}_{1t}\left(\boldsymbol{y}_{2t}-\boldsymbol{\Phi}_{22}\boldsymbol{c}_{2t}\right)\boldsymbol{\Sigma}_{22}^{-1}\right]. \tag{2.3.52}$$

And, the number of restricted parameters under the null hypothesis is $N_2(N_1+K_1)$.

Example 3: Specification test for linear dependency between $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$

In the system of (2.3.24), (2.3.25), and (2.3.26), there is no relation between $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ when $\boldsymbol{\Phi}_{12}=\boldsymbol{\Phi}_{21}=\boldsymbol{\Sigma}_{12}=\mathbf{0}$. Therefore, we conduct a LM test of the null hypothesis

$$H_0:\ \boldsymbol{\Phi}_{12}=\boldsymbol{\Phi}_{21}=\boldsymbol{\Sigma}_{12}=\mathbf{0}. \tag{2.3.53}$$

Then, derivatives with respect to vec($\boldsymbol{\Phi}_{12}$), vec($\boldsymbol{\Phi}_{21}$), and vec($\boldsymbol{\Sigma}_{12}$) are

$$\left.\frac{\partial \log f\left(\boldsymbol{y}_{1t}, \boldsymbol{y}_{2t} \middle| \boldsymbol{c}_{1t}, \boldsymbol{c}_{2t}\right)}{\partial \text{vec}\left(\boldsymbol{\Phi}_{12}\right)}\right|_{\boldsymbol{\Phi}_{12}=\boldsymbol{\Phi}_{21}=\boldsymbol{\Sigma}_{12}=\mathbf{0}} = \text{vec}\left[\boldsymbol{\Sigma}_{11}^{-1}\left(\boldsymbol{y}_{1t} - \boldsymbol{\Phi}_{11}\boldsymbol{c}_{1t}\right)\boldsymbol{c}_{2t}'\right], \tag{2.3.54}$$

$$\left.\frac{\partial \log f\left(\boldsymbol{y}_{1t}, \boldsymbol{y}_{2t} \middle| \boldsymbol{c}_{1t}, \boldsymbol{c}_{2t}\right)}{\partial \text{vec}\left(\boldsymbol{\Phi}_{21}\right)}\right|_{\boldsymbol{\Phi}_{12}=\boldsymbol{\Phi}_{21}=\boldsymbol{\Sigma}_{12}=\mathbf{0}} = \text{vec}\left[\boldsymbol{\Sigma}_{22}^{-1}\left(\boldsymbol{y}_{2t} - \boldsymbol{\Phi}_{22}\boldsymbol{c}_{2t}\right)\boldsymbol{c}_{1t}'\right], \tag{2.3.55}$$

$$\left.\frac{\partial \log f\left(\boldsymbol{y}_{1t}, \boldsymbol{y}_{2t} \middle| \boldsymbol{c}_{1t}, \boldsymbol{c}_{2t}\right)}{\partial \text{vec}\left(\boldsymbol{\Sigma}_{12}\right)}\right|_{\boldsymbol{\Phi}_{12}=\boldsymbol{\Phi}_{21}=\boldsymbol{\Sigma}_{12}=\mathbf{0}} = \text{vec}\left[\boldsymbol{\Sigma}_{11}^{-1}\left(\boldsymbol{y}_{1t} - \boldsymbol{\Phi}_{11}\boldsymbol{c}_{1t}\right)\left(\boldsymbol{y}_{2t} - \boldsymbol{\Phi}_{22}\boldsymbol{c}_{2t}\right)\boldsymbol{\Sigma}_{22}^{-1}\right]. \tag{2.3.56}$$

And the number of restricted parameters under the null hypothesis is $N_1N_2 + N_1K_2 + N_2K_1$.

Example 4: Specification test for omitted explanatory variable

In the system of (2.3.39) and (2.3.40), observed variables $\boldsymbol{y}_t$ are specified by common factor $\boldsymbol{c}_t$ only, that is restriction that $\boldsymbol{\Pi} = \mathbf{0}$. Therefore, we conduct a LM test of the null hypothesis

$$H_0 : \boldsymbol{\Pi} = \mathbf{0}. \tag{2.3.57}$$

Then, the differential with respect to vec($\boldsymbol{\Pi}$) under the null is

$$\begin{aligned}
d\log f(\boldsymbol{y}_t|\,\boldsymbol{c}_t, \boldsymbol{x}_t) &= -\frac{1}{2}d(\boldsymbol{u}_t'\boldsymbol{\Sigma}^{-1}\boldsymbol{u}_t) = -\frac{1}{2}\text{tr}[d(\boldsymbol{u}_t'\boldsymbol{\Sigma}^{-1}\boldsymbol{u}_t)] \\
&= -\frac{1}{2}\text{tr}\left[(d\boldsymbol{u}_t)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_t - \boldsymbol{\Phi}\boldsymbol{c}_t) + (\boldsymbol{y}_t - \boldsymbol{\Phi}\boldsymbol{c}_t)'\boldsymbol{\Sigma}^{-1}(d\boldsymbol{u}_t)\right] \\
&= -\text{tr}\left[(\boldsymbol{y}_t - \boldsymbol{\Phi}\boldsymbol{c}_t)'\boldsymbol{\Sigma}^{-1}(d\boldsymbol{u}_t)\right] = -\text{tr}\left[(\boldsymbol{y}_t - \boldsymbol{\Phi}\boldsymbol{c}_t)'\boldsymbol{\Sigma}^{-1}(-d\boldsymbol{\Pi})\boldsymbol{x}_t\right] \\
&= \text{tr}\left[\boldsymbol{x}_t(\boldsymbol{y}_t - \boldsymbol{\Phi}\boldsymbol{c}_t)'\boldsymbol{\Sigma}^{-1}(d\boldsymbol{\Pi})\right] = \text{tr}\left[\left\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_t - \boldsymbol{\Phi}\boldsymbol{c}_t)\boldsymbol{x}_t'\right\}'(d\boldsymbol{\Pi})\right] \\
&= \left(\text{vec}\left[\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_t - \boldsymbol{\Phi}\boldsymbol{c}_t)\boldsymbol{x}_t'\right]\right)' d\text{vec}(\boldsymbol{\Pi}).
\end{aligned} \tag{2.3.58}$$

Therefore, the derivative with respect to vec($\boldsymbol{\Pi}$) under the null is

$$\left.\frac{\partial \log f(\boldsymbol{y}_t|\,\boldsymbol{c}_t, \boldsymbol{x}_t)}{\partial \text{vec}(\boldsymbol{\Pi})}\right|_{\boldsymbol{\Pi}=\mathbf{0}} = \text{vec}\left[\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_t - \boldsymbol{\Phi}\boldsymbol{c}_t)\boldsymbol{x}_t'\right]. \tag{2.3.59}$$

And, the number of restricted parameters under the null is NJ.

2.3.4 Lagrange Multiplier tests

In this subsection, we test the dynamic factor model (2.2.3) and (2.2.4) against the alternative such as weak exogeneity, linear dependency, and omitted explanatory variables, according to the specification of subsection 2.3.3.

Suppose that parameter vector is partitioned as in subsection 2.3.2. Then, when $\boldsymbol{\theta}_1 \subseteq \boldsymbol{\lambda}$, the score of the tth observation with respect to $\boldsymbol{\theta}_1$ under the null hypothesis is given by

$$\boldsymbol{g}_t(\tilde{\boldsymbol{\theta}}) = \sum_{\tau=1}^{t} E[\boldsymbol{\xi}_\tau(\tilde{\boldsymbol{\theta}})|\boldsymbol{\psi}_t] - \sum_{\tau=1}^{t-1} E[\boldsymbol{\xi}_\tau(\tilde{\boldsymbol{\theta}})|\boldsymbol{\psi}_t], \text{ for } t = 2, 3, \ldots, T \tag{2.3.60}$$

where

$$\boldsymbol{\xi}_\tau(\tilde{\boldsymbol{\theta}}) = \left. \frac{\partial \log f\left(\boldsymbol{y}_\tau | \boldsymbol{c}_\tau; \boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}_1} \right|_{\boldsymbol{\theta}_1 = \mathbf{0}}.$$

Similarly, when $\boldsymbol{\theta}_1 \subseteq \boldsymbol{\eta}$, the score of the tth observation with respect to $\boldsymbol{\theta}_1$ under the null hypothesis is given by

$$\boldsymbol{g}_t(\tilde{\boldsymbol{\theta}}) = \sum_{\tau=1}^{t} E[\boldsymbol{\xi}_\tau(\tilde{\boldsymbol{\theta}})|\boldsymbol{\psi}_t] - \sum_{\tau=1}^{t-1} E[\boldsymbol{\xi}_\tau(\tilde{\boldsymbol{\theta}})|\boldsymbol{\psi}_t], \text{ for } t = 2, 3, \ldots, T \tag{2.3.61}$$

where

$$\boldsymbol{\xi}_\tau(\tilde{\boldsymbol{\theta}}) = \left. \frac{\partial \log f\left(\boldsymbol{c}_\tau | \boldsymbol{c}_{\tau-1}; \boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}_1} \right|_{\boldsymbol{\theta}_1 = \mathbf{0}}.$$

Then, LM test statistic is derived by stacking scores (2.3.60) and (2.3.61) as in equation (2.3.19) and (2.3.21) for $t = 1, 2, \cdots, T$. Table 2.1 presents formulas necessary to implement LM tests against the alternative. To implement any of the tests, one only needs to estimate the model under the null and calculate $\boldsymbol{g}_t(\tilde{\boldsymbol{\theta}})$ in the manner described in subsection 2.3.1 using the restricted estimates $\tilde{\boldsymbol{\theta}}$ to form the smoothed density.

Table 2.1: Values to use for $\boldsymbol{\xi}_t$ in Lagrange Multiplier tests

1. *To test for weak exogeneity of* $\boldsymbol{y}_1$
$H_0 : \boldsymbol{\Phi}_{12} = \boldsymbol{\Sigma}_{12} = \mathbf{0}$ $\boldsymbol{\xi}_t(\tilde{\boldsymbol{\theta}}) = \text{vec}\left[\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{y}_{1t} - \boldsymbol{\Phi}_{11}\boldsymbol{c}_{1t})\boldsymbol{c}_{2t}'\right], \text{vec}\left[\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{y}_{1t} - \boldsymbol{\Phi}_{11}\boldsymbol{c}_{1t})\boldsymbol{u}_{2t}'\boldsymbol{\Sigma}_{22}^{-1}\right]$ $M_1 = N_1(N_2 + K_2)$
2. *To test for weak exogeneity of* $\boldsymbol{y}_2$
$H_0 : \boldsymbol{\Phi}_{21} = \boldsymbol{\Sigma}_{12} = \mathbf{0}$ $\boldsymbol{\xi}_t(\tilde{\boldsymbol{\theta}}) = \text{vec}\left[\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{y}_{2t} - \boldsymbol{\Phi}_{22}\boldsymbol{c}_{2t})\boldsymbol{c}_{1t}'\right], \text{vec}\left[\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{u}_{1t}(\boldsymbol{y}_{2t} - \boldsymbol{\Phi}_{22}\boldsymbol{c}_{2t})'\boldsymbol{\Sigma}_{22}^{-1}\right]$ $M_1 = N_2(N_1 + K_1)$
3. *To test for linear dependency between* $\boldsymbol{y}_1$ *and* $\boldsymbol{y}_2$
$H_0 : \boldsymbol{\Phi}_{12} = \boldsymbol{\Phi}_{21} = \boldsymbol{\Sigma}_{12} = \mathbf{0}$ $\boldsymbol{\xi}_t(\tilde{\boldsymbol{\theta}}) = \text{vec}\left[\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{y}_{1t} - \boldsymbol{\Phi}_{11}\boldsymbol{c}_{1t})\boldsymbol{c}_{2t}'\right], \text{vec}\left[\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{y}_{2t} - \boldsymbol{\Phi}_{22}\boldsymbol{c}_{2t})\boldsymbol{c}_{1t}'\right], \text{vec}\left[\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{u}_{1t}(\boldsymbol{y}_{2t} - \boldsymbol{\Phi}_{22}\boldsymbol{c}_{2t})'\boldsymbol{\Sigma}_{22}^{-1}\right]$ $M_1 = N_1N_2 + N_1K_2 + N_2K_1$
4. *To test for omitted explanatory variable* $\boldsymbol{x}$
$H_0 : \boldsymbol{\Pi} = \mathbf{0}$ $\boldsymbol{\xi}_t(\tilde{\boldsymbol{\theta}}) = \text{vec}[\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_t - \boldsymbol{\Phi}\boldsymbol{c}_t)\boldsymbol{x}_t']$ $M_1 = NJ$

2.4 Monte Carlo analysis of the tests

This section examines finite sample properties of LM tests for the weak exogeneity, linear dependency, and omitted explanatory variable through simulation experiments. All the work described subsequently was conducted using GAUSS programming language. The true model in the experiment is as follows.

- *Observation equation*

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 1.5 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} c_{1t} \\ c_{2t} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \cdot x_t + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix},$$

- *Transition equation*

$$\begin{bmatrix} c_{1t} \\ c_{2t} \end{bmatrix} = \begin{bmatrix} 0.8 & 0 \\ 0 & 0.8 \end{bmatrix} \begin{bmatrix} c_{1,t-1} \\ c_{2,t-1} \end{bmatrix} + \begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix},$$

- *Structure of covariance matrices*

$$\begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix} \sim \text{i.i.d.}N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.8 & 0 \\ 0.8 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right),$$

$$\begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix} \sim \text{i.i.d.}N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right),$$

for $t = 1, 2, \ldots, T$. The expression x_t is generated from a unifrom distribution, and u_t and v_t are generated from a normal distribution. The sample size is 500, 1000, and 1500, and the number of iteration is 1000 in our experiments. One iteration under the null hypothesis takes approximately 40 seconds using GAUSS 8.0 on a PC with Pentium 4, but it takes longer under the alternative hypothesis. In our experiment, all parameter values are set near to the estimates of the empirical analysis reported later.

First, we consider the size of specification tests. Table 2.2 through 2.4 present simulation results for the sizes of the specification tests. The null hypothesis of Table 2.2 specifies $(\phi_{12}, \phi_{22}, \phi_{31}, \sigma_{31}, \sigma_{32}) = (0, 0, 0, 0, 0)$, the null hypothesis of Table 2.3 specifies $(\phi_{31}, \sigma_{31}, \sigma_{32}) = (0, 0, 0)$, and the null hypothesis of Table 2.4 specifies $(\beta_1, \beta_2, \beta_3) = (0, 0, 0)$.

We can see in Table 2.2 and 2.3 that LM_{s1} and LM_{s2} show similar results. Both tests have correct size except when ϕ_{11} or ϕ_{32} takes small value. For $\phi_{11} = 1.4$ or $\phi_{32} = 1.4$, there are considerable size distortions (too many rejections) even for $T = 1000$. Conversely, for Table 2.4, there are considerable size distortions (too few rejections) for any parameter values.

Next, we consider the power of the tests. Table 2.4 through 2.6 present simulation results giving the power of the test. We can see in Table 2.5 and 2.6 that powers of LM_{p1} and LM_{p2} have appreciable power except when $\phi_{31} = 0.1$ and $\sigma_{31} = 0.1$. For $\phi_{31} = 0.1$ and $\sigma_{31} = 0.1$, power increases as T increases. In Table 2.7, we can also see that LM_{p3} shows good peformance for any parameter values.

Thus, specification tests proposed in this paper are reliable in the presence of i.i.d. errors.

2.5 Summary

In this peper, we provided an analytic representation for the score statistic for the dynamic factor model, and suggested a variety of specification tests based on these scores. These scores turned out to be a natural byproduct of the routine used to calculate the expected value with respect to smoothed density. In addition, from the same calculation we constructed asymptotic standard errors for the parameter vector and specification tests. Therefore, proposed tests could be calculated together from a single pass through the data, and they do not require the estimation of additional parameters by maximum likelihood. We also investigated finite sample properties of the tests. Monte Carlo results showed that the tests are reliable in terms of both size and power performance.

It should be noted that it is open to question whether the tests work under the autocorrelation or heteroskedasticity of the disturbances. This question awaits further research.

Appendix A: Score vector for gaussian linear state-space model

This appendix shows that the score vector for Gaussian linear state-space model takes on a simple form which can be computed in a single pass of the Kalman filter and a smoother. The *state-space representation* is given by

$$\boldsymbol{y}_t = \boldsymbol{\Phi}\boldsymbol{c}_t + \boldsymbol{\Pi}\boldsymbol{x}_t + \boldsymbol{u}_t, \tag{2.A.1}$$

$$\boldsymbol{c}_t = \boldsymbol{\Gamma}\boldsymbol{c}_{t-1} + \boldsymbol{v}_t \tag{2.A.2}$$

for $t = 1, 2, \ldots, T$. Here, $\boldsymbol{\Phi}$, $\boldsymbol{\Pi}$, and $\boldsymbol{\Gamma}$ are matrices of parameters, $\boldsymbol{y}_t$ is a vector of observed variables, $\boldsymbol{c}_t$ is a vector of unobserved state variables, and $\boldsymbol{x}_t$ is a vector of exogenous or predetermined variables. Equation (2.A.1) is an *observation equation*, and (2.A.2) is a *state equation*. The vector $\boldsymbol{u}_t$ and $\boldsymbol{v}_t$ are vectors of Gaussian white noise, respectively. We require that $\boldsymbol{u}_t$ and $\boldsymbol{v}_t$ be orthogonal to each other (at all leads and lags) and to the initial state:

$$\begin{bmatrix} \boldsymbol{u}_t \\ \boldsymbol{v}_t \end{bmatrix} \sim \text{i.i.d.}N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega} \end{bmatrix}\right), \tag{2.A.3}$$

$$E\left(\boldsymbol{c}_1\boldsymbol{u}_t'\right) = \mathbf{0}, E\left(\boldsymbol{c}_1\boldsymbol{v}_t'\right) = \mathbf{0}. \tag{2.A.4}$$

where $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ are positive definite symmetric matrices, respectively.

Therefore, the conditional densities of the observed vector $\boldsymbol{y}_t$ and the unobserved state vector $\boldsymbol{c}_t$ are defined as follows:

$$f(\boldsymbol{y}_t|\boldsymbol{c}_t, \boldsymbol{x}_t) = (2\pi)^{-N/2}\left|\boldsymbol{\Sigma}^{-1}\right|^{1/2} \exp\left[-\frac{1}{2}(\boldsymbol{y}_t - \boldsymbol{\Phi}\boldsymbol{c}_t - \boldsymbol{\Pi}\boldsymbol{x}_t)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_t - \boldsymbol{\Phi}\boldsymbol{c}_t - \boldsymbol{\Pi}\boldsymbol{x}_t)\right], \tag{2.A.6}$$

$$f(\boldsymbol{c}_t|\boldsymbol{c}_{t-1}; \boldsymbol{\eta}) = (2\pi)^{-K/2}\left|\boldsymbol{\Omega}^{-1}\right|^{1/2} \exp\left[-\frac{1}{2}(\boldsymbol{c}_t - \boldsymbol{\Gamma}\boldsymbol{c}_{t-1})'\boldsymbol{\Omega}^{-1}(\boldsymbol{c}_t - \boldsymbol{\Gamma}\boldsymbol{c}_{t-1})\right], \tag{2.A.7}$$

where $\boldsymbol{\lambda} = ([\text{vec}(\boldsymbol{\Phi})]', [\text{vec}(\boldsymbol{\Pi})]', [\text{vech}(\boldsymbol{\Sigma})]')', \boldsymbol{\eta} = ([\text{vec}(\boldsymbol{\Gamma})]', [\text{vech}(\boldsymbol{\Omega})]')'$.

Note that parameter vector $\boldsymbol{\theta}$ includes both the parameter $\boldsymbol{\lambda}$ appearing in equation (2.A.6) and the parameter $\boldsymbol{\eta}$ in equation (2.A.7),

$$\boldsymbol{\theta} = (\boldsymbol{\lambda}', \boldsymbol{\eta}')'. \tag{2.A.8}$$

According to the above specification, we will show the whole score vector can be computed exactly in a single pass of the Kalman filter and a smoother below. From the Bayes theorem, $f(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T; \boldsymbol{\theta})$ is given by

$$\begin{aligned} &f(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T; \boldsymbol{\theta}) \\ &= \frac{f(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T | \boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_T, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_T; \boldsymbol{\lambda}) f(\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_T, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_T; \boldsymbol{\eta})}{f(\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_T, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_T | \boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T; \boldsymbol{\theta})} \end{aligned} \tag{2.A.9}$$

Then, $\log f(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T; \boldsymbol{\theta})$ is as follows

$$\begin{aligned} \log f(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T; \boldsymbol{\theta}) &= \sum_{t=1}^{T} \log f(\boldsymbol{y}_t | \boldsymbol{c}_t, \boldsymbol{x}_t; \boldsymbol{\lambda}) + \sum_{t=1}^{T} \log f(\boldsymbol{c}_t, \boldsymbol{x}_t | \boldsymbol{c}_{t-1}, \boldsymbol{x}_{t-1}; \boldsymbol{\lambda}) \\ &\quad - \log f(\, \boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_T, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_T |\, \boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T; \boldsymbol{\theta}) \\ &= \sum_{t=1}^{T} \log f(\boldsymbol{y}_t | \boldsymbol{c}_t, \boldsymbol{x}_t; \boldsymbol{\lambda}) + \sum_{t=1}^{T} \log f(\boldsymbol{c}_t | \boldsymbol{c}_{t-1}; \boldsymbol{\lambda}) \\ &\quad - \log f(\, \boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_T |\, \boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T; \boldsymbol{\theta}) \end{aligned} \tag{2.A.10}$$

Further, substituting equation (2.A.6) and (2.A.7), equation (2.A.10) becomes

$$\begin{aligned} \log f(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T; \boldsymbol{\theta}) &= \sum_{t=1}^{T} \left(-\frac{N}{2} \log 2\pi - \frac{1}{2} \log |\boldsymbol{\Sigma}| - \frac{1}{2} \mathrm{tr} \left[(\boldsymbol{y}_t - \boldsymbol{\Phi} \boldsymbol{c}_t - \boldsymbol{\Pi} \boldsymbol{x}_t)' \, \boldsymbol{\Sigma}^{-1} \, (\boldsymbol{y}_t - \boldsymbol{\Phi} \boldsymbol{c}_t - \boldsymbol{\Pi} \boldsymbol{x}_t) \right] \right) \\ &+ \sum_{t=1}^{T} \left(-\frac{K}{2} \log 2\pi - \frac{1}{2} \log |\boldsymbol{\Omega}| - \frac{1}{2} \mathrm{tr} \left[(\boldsymbol{c}_t - \boldsymbol{\Gamma} \boldsymbol{c}_{t-1})' \, \boldsymbol{\Omega}^{-1} \, (\boldsymbol{c}_t - \boldsymbol{\Gamma} \boldsymbol{c}_{t-1}) \right] \right) \\ &- \log f\left(\, \boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_T |\, \boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T; \boldsymbol{\theta}\right) \\ &= - \left[\frac{T\,(N+K)}{2} \right] \log 2\pi - \left(\frac{T}{2} \right) (\log |\boldsymbol{\Sigma}| + \log |\boldsymbol{\Omega}|) \\ &- \frac{1}{2} \sum_{t=1}^{T} \mathrm{tr} \left[\boldsymbol{\Sigma}^{-1} \, (\boldsymbol{y}_t - \boldsymbol{\Phi} \boldsymbol{c}_t - \boldsymbol{\Pi} \boldsymbol{x}_t) \, (\boldsymbol{y}_t - \boldsymbol{\Phi} \boldsymbol{c}_t - \boldsymbol{\Pi} \boldsymbol{x}_t)' \right] \\ &- \frac{1}{2} \sum_{t=1}^{T} \mathrm{tr} \left[\boldsymbol{\Omega}^{-1} \, (\boldsymbol{c}_t - \boldsymbol{\Gamma} \boldsymbol{c}_{t-1}) \, (\boldsymbol{c}_t - \boldsymbol{\Gamma} \boldsymbol{c}_{t-1})' \right] \\ &- \log f\left(\, \boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_T |\, \boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T; \boldsymbol{\theta}\right). \end{aligned} \tag{2.A.11}$$

To be able to derive the score at a point $\hat{\boldsymbol{\theta}}$, we will first integrate both sides of equation (A.11) with respect to the joint density of the smoother $f(\, \boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_T |\, \boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T; \hat{\boldsymbol{\theta}})$, to get

$$\log f(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T; \boldsymbol{\theta}) = Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) - R(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}), \tag{2.A.12}$$

where

$$Q(\boldsymbol{\theta},\hat{\boldsymbol{\theta}}) = -\frac{T}{2}(\log|\boldsymbol{\Sigma}| + \log|\boldsymbol{\Omega}|) - \frac{1}{2}\sum_{t=1}^{T}\operatorname{tr}(\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{W}}_{t|T}) - \frac{1}{2}\sum_{t=1}^{T}\operatorname{tr}(\boldsymbol{\Omega}^{-1}\hat{\boldsymbol{Z}}_{t|T}),$$

$$R(\boldsymbol{\theta},\hat{\boldsymbol{\theta}}) = \int^{(T)} \log f(\boldsymbol{c}_1,\boldsymbol{c}_2,\ldots,\boldsymbol{c}_T|\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots,\boldsymbol{y}_T;\boldsymbol{\theta}) f(\boldsymbol{c}_1,\boldsymbol{c}_2,\ldots,\boldsymbol{c}_T|\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots,\boldsymbol{y}_T;\hat{\boldsymbol{\theta}}) d\boldsymbol{c}^{(T)},$$

$$\hat{\boldsymbol{W}}_{t|T} \equiv \hat{\boldsymbol{u}}_{t|T}\hat{\boldsymbol{u}}'_{t|T} + \boldsymbol{\Phi}\boldsymbol{P}_{t|T}\boldsymbol{\Phi}',$$

$$\hat{\boldsymbol{Z}}_{t|T} \equiv \hat{\boldsymbol{v}}_{t|T}\hat{\boldsymbol{v}}_{t|T} + \boldsymbol{P}_{t|T} - \boldsymbol{\Gamma}\boldsymbol{P}_{t-1,t|T} - \boldsymbol{P}_{t-1,t|T}\boldsymbol{\Gamma}' + \boldsymbol{\Gamma}\boldsymbol{P}_{t-1|T}\boldsymbol{\Gamma}'.$$

Here, $\hat{\boldsymbol{u}}_{t|T}$ and $\hat{\boldsymbol{v}}_{t|T}$ are the smoothed estimates of $\boldsymbol{u}_t$ and $\boldsymbol{v}_t$ respectively, $\boldsymbol{P}_{t|T}$ is the mean squared error (mse) of $\hat{\boldsymbol{c}}_{t|T}$, and $\boldsymbol{P}_{t,t-1|T}$ is the covariance between the estimates of $\hat{\boldsymbol{c}}_{t|T}$ and $\hat{\boldsymbol{c}}_{t-1|T}$. All these smoothed quantities are computed with $\boldsymbol{\theta}$ taken to be $\hat{\boldsymbol{\theta}}$ by first running a Kalman filter which is given by

$$\begin{gathered}\hat{\boldsymbol{c}}_{t+1|t} = \boldsymbol{\Gamma}\hat{\boldsymbol{c}}_{t|t-1} + \boldsymbol{K}_t\boldsymbol{\zeta}_{t|t-1}, \boldsymbol{P}_{t+1|t} = \boldsymbol{\Gamma}\boldsymbol{P}_{t|t-1}\boldsymbol{L}'_t + \boldsymbol{\Omega},\\ \boldsymbol{\zeta}_{t|t-1} = \boldsymbol{y}_t - \boldsymbol{\Phi}\hat{\boldsymbol{c}}_{t|t-1} - \boldsymbol{\Pi}\boldsymbol{x}_t, \boldsymbol{F}_{t|t-1} = \boldsymbol{\Phi}\boldsymbol{P}_{t|t-1}\boldsymbol{\Phi}' + \boldsymbol{\Sigma},\\ \boldsymbol{K}_t = \boldsymbol{\Gamma}\boldsymbol{P}_{t|t-1}\boldsymbol{\Phi}'\boldsymbol{F}^{-1}_{t|t-1}, \boldsymbol{L}_t = \boldsymbol{\Gamma} - \boldsymbol{K}_t\boldsymbol{\Phi},\end{gathered} \tag{2.A.13}$$

for $t = 1, 2, \ldots, T$. Then, the De Jong [1989] and De Jong and MacKinnon [1988] smoothing algorithms deliver

$$\begin{gathered}\hat{\boldsymbol{u}}_{t|T} = \boldsymbol{y}_t - \boldsymbol{\Phi}\hat{\boldsymbol{c}}_{t|T} - \boldsymbol{\Pi}\boldsymbol{x}_t, \hat{\boldsymbol{v}}_{t|T} = \hat{\boldsymbol{c}}_{t|T} - \boldsymbol{\Gamma}\hat{\boldsymbol{c}}_{t-1|T},\\ \hat{\boldsymbol{c}}_{t|T} = \hat{\boldsymbol{c}}_{t|t-1} + \boldsymbol{P}_{t|t-1}\boldsymbol{r}_{t-1}, \boldsymbol{P}_{t|T} = \boldsymbol{P}_{t|t-1} - \boldsymbol{P}_{t|t-1}\boldsymbol{N}_{t-1}\boldsymbol{P}_{t|t-1},\\ \boldsymbol{r}_{t-1} = \boldsymbol{\Phi}'\boldsymbol{F}^{-1}_{t|t-1}\boldsymbol{\zeta}_{t|t-1} + \boldsymbol{L}'_t\boldsymbol{r}_t, \boldsymbol{N}_{t-1} = \boldsymbol{\Phi}'\boldsymbol{F}^{-1}_{t|t-1}\boldsymbol{\Phi} + \boldsymbol{L}'_t\boldsymbol{N}_t\boldsymbol{L}_t, \boldsymbol{P}_{t-1,t|T} = \boldsymbol{P}_{t-1|t-2}\boldsymbol{L}'_{t-1}(\boldsymbol{I} - \boldsymbol{N}_{t-1}\boldsymbol{P}_{t|t-1})\end{gathered} \tag{2.A.14}$$

for $t = T, T-1, \ldots, 1$. If equation (2.A.12) is differentiated with respect to $\boldsymbol{\theta}$, then

$$\frac{\partial \log f(\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots,\boldsymbol{y}_T;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \frac{\partial Q(\boldsymbol{\theta},\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}} - \frac{\partial R(\boldsymbol{\theta},\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}},$$

but it can be shown that

$$\begin{aligned}&\left.\frac{\partial R(\boldsymbol{\theta},\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\\ &= \frac{\partial}{\partial\boldsymbol{\theta}}\left[\int^{(T)} \log f(\boldsymbol{c}_1,\boldsymbol{c}_2,\ldots,\boldsymbol{c}_T|\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots,\boldsymbol{y}_T;\boldsymbol{\theta}) f(\boldsymbol{c}_1,\boldsymbol{c}_2,\ldots,\boldsymbol{c}_T|\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots,\boldsymbol{y}_T;\boldsymbol{\theta}) d\boldsymbol{c}^{(T)}\right]\Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\\ &= \int^{(T)} \left.\frac{\partial \log f(\boldsymbol{c}_1,\boldsymbol{c}_2,\ldots,\boldsymbol{c}_T|\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots,\boldsymbol{y}_T;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} f(\boldsymbol{c}_1,\boldsymbol{c}_2,\ldots,\boldsymbol{c}_T|\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots,\boldsymbol{y}_T;\boldsymbol{\theta}) d\boldsymbol{c}^{(T)}\\ &= \int^{(T)} \left.\frac{\partial f(\boldsymbol{c}_1,\boldsymbol{c}_2,\ldots,\boldsymbol{c}_T|\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots,\boldsymbol{y}_T;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} d\boldsymbol{c}^{(T)}\\ &= \frac{\partial}{\partial\boldsymbol{\theta}}\left[\int^{(T)} f(\boldsymbol{c}_1,\boldsymbol{c}_2,\ldots,\boldsymbol{c}_T|\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots,\boldsymbol{y}_T;\boldsymbol{\theta}) d\boldsymbol{c}^{(T)}\right]\Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \left.\frac{\partial 1}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \boldsymbol{0}\end{aligned}$$

and so

$$\left.\frac{\partial \log f\left(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T; \boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \left.\frac{\partial Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} - \left.\frac{\partial R(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \left.\frac{\partial Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}.$$

Then, the differential with respect to vec($\boldsymbol{\Phi}$) is

$$\begin{aligned} d\log f(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T, \boldsymbol{\theta}) &= -\frac{1}{2}\sum_{t=1}^{T} d\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{W}}_{t|T}) = -\frac{1}{2}\sum_{t=1}^{T} \mathrm{tr}\left[\boldsymbol{\Sigma}^{-1} d(\hat{\boldsymbol{u}}_{t|T}\hat{\boldsymbol{u}}'_{t|T} + \boldsymbol{\Phi}\boldsymbol{P}_{t|T}\boldsymbol{\Phi}')\right] \\ &= -\frac{1}{2}\sum_{t=1}^{T} \mathrm{tr}\left[-2\left\{\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{u}}_{t|T}\hat{\boldsymbol{c}}'_{t|T} - \boldsymbol{\Phi}\boldsymbol{P}_{t|T})\right\}'(d\boldsymbol{\Phi})\right] \\ &= \sum_{t=1}^{T}\left(\mathrm{vec}\left[\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{u}}_{t|T}\hat{\boldsymbol{c}}'_{t|T} - \boldsymbol{\Phi}\boldsymbol{P}_{t|T})\right]\right)' d\mathrm{vec}(\boldsymbol{\Phi}). \end{aligned} \tag{2.A.15}$$

Therefore, the score vector with respect to vec ($\boldsymbol{\Phi}$) evaluated at $\hat{\boldsymbol{\theta}}$ is

$$\left.\frac{\partial \log f(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T; \boldsymbol{\theta})}{\partial \mathrm{vec}(\boldsymbol{\Phi})}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \sum_{t=1}^{T} \mathrm{vec}\left[\hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{u}}_{t|T}\hat{\boldsymbol{c}}'_{t|T} - \hat{\boldsymbol{\Phi}}\boldsymbol{P}_{t|T})\right]. \tag{2.A.16}$$

Next, the differential with respect to vec($\boldsymbol{\Pi}$) is similarly found by setting

$$\begin{aligned} d\log f(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T, \boldsymbol{\theta}) &= -\frac{1}{2}\sum_{t=1}^{T} d\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{W}}_{t|T}) = -\frac{1}{2}\sum_{t=1}^{T} \mathrm{tr}\left[\boldsymbol{\Sigma}^{-1} d(\hat{\boldsymbol{u}}_{t|T}\hat{\boldsymbol{u}}'_{t|T})\right] \\ &= -\frac{1}{2}\sum_{t=1}^{T} \mathrm{tr}\left[-2\left(\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{u}}_{t|T}\boldsymbol{x}'_t\right)'(d\boldsymbol{\Pi})\right] \\ &= \sum_{t=1}^{T}\left[\mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{u}}_{t|T}\boldsymbol{x}'_t\right)\right]' d\mathrm{vec}(\boldsymbol{\Pi}). \end{aligned} \tag{2.A.17}$$

Then, the score vector with respect to vec ($\boldsymbol{\Pi}$) evaluated at $\hat{\boldsymbol{\theta}}$ is

$$\left.\frac{\partial \log f(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T, \boldsymbol{\theta})}{\partial \mathrm{vec}(\boldsymbol{\Pi})}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \sum_{t=1}^{T} \mathrm{vec}\left(\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{u}}_{t|T}\boldsymbol{x}'_t\right). \tag{2.A.18}$$

The differeintial with respect to vec $(\boldsymbol{\Sigma})$ is

$$\begin{aligned} d\log f(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T; \boldsymbol{\theta}) &= -\frac{T}{2} d(\log|\boldsymbol{\Sigma}|) - \frac{1}{2}\sum_{t=1}^{T} \mathrm{tr}\left[(d\boldsymbol{\Sigma}^{-1})\hat{\boldsymbol{W}}_{t|T}\right] \\ &= -\frac{T}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}(d\boldsymbol{\Sigma})\right] + \frac{1}{2}\sum_{t=1}^{T} \mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}(d\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{W}}_{t|T}\right] \\ &= -\frac{1}{2}\sum_{t=1}^{T} \mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}(\boldsymbol{I}_N - \hat{\boldsymbol{W}}_{t|T}\boldsymbol{\Sigma}^{-1})(d\boldsymbol{\Sigma})\right] \\ &= -\frac{1}{2}\sum_{t=1}^{T}\left(\mathrm{vec}\left[\boldsymbol{\Sigma}^{-1}(\boldsymbol{I}_N - \hat{\boldsymbol{W}}_{t|T}\boldsymbol{\Sigma}^{-1})\right]\right)' d\mathrm{vec}(\boldsymbol{\Sigma}). \end{aligned} \tag{2.A.19}$$

using identity $d\log|\boldsymbol{A}| = \mathrm{tr}[\boldsymbol{A}^{-1}(d\boldsymbol{A})]$, $d\boldsymbol{A}^{-1} = -\boldsymbol{A}^{-1}(d\boldsymbol{A})\boldsymbol{A}^{-1}$. Then, the score vector with respect to vec $(\boldsymbol{\Sigma})$ evaluated at $\hat{\boldsymbol{\theta}}$ is

$$\left.\frac{\partial \log f(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T; \boldsymbol{\theta})}{\partial \mathrm{vec}(\boldsymbol{\Sigma})}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = -\frac{1}{2}\sum_{t=1}^{T} \mathrm{vec}\left[\hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{I}_N - \hat{\boldsymbol{W}}_{t|T}\hat{\boldsymbol{\Sigma}}^{-1})\right]. \tag{2.A.20}$$

The differential with respect to vec $(\boldsymbol{\Gamma})$ is similarly found by setting

$$\begin{aligned} & d\log f(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T; \boldsymbol{\theta}) \\ =& -\frac{1}{2}\sum_{t=1}^{T} d\mathrm{tr}(\boldsymbol{\Omega}^{-1}\hat{\boldsymbol{Z}}_{t|T}) \\ =& -\frac{1}{2}\sum_{t=1}^{T} \mathrm{tr}\left[\boldsymbol{\Omega}^{-1} d\left(\hat{\boldsymbol{v}}_{t|T}\hat{\boldsymbol{v}}'_{t|T} - \boldsymbol{\Gamma}\boldsymbol{P}_{t,t-1|T} - \boldsymbol{P}_{t,t-1|T}\boldsymbol{\Gamma}' + \boldsymbol{\Gamma}\boldsymbol{P}_{t-1|T}\boldsymbol{\Gamma}'\right)\right] \\ =& -\frac{1}{2}\sum_{t=1}^{T} \mathrm{tr}\left[-2\left\{\boldsymbol{\Omega}^{-1}\left(\hat{\boldsymbol{v}}_{t|T}\hat{\boldsymbol{c}}'_{t-1|T} + \boldsymbol{P}_{t,t-1|T} - \boldsymbol{\Gamma}\boldsymbol{P}_{t-1|T}\right)\right\}'(d\boldsymbol{\Gamma})\right] \\ =& \sum_{t=1}^{T}\left(\mathrm{vec}\left[\boldsymbol{\Omega}^{-1}\left(\hat{\boldsymbol{v}}_{t|T}\hat{\boldsymbol{c}}'_{t-1|T} + \boldsymbol{P}_{t,t-1|T} - \boldsymbol{\Gamma}\boldsymbol{P}_{t-1|T}\right)\right]\right)' d\mathrm{vec}(\boldsymbol{\Gamma}). \end{aligned} \tag{2.A.21}$$

Then, the score vector with respect to vec $(\boldsymbol{\Gamma})$ is

$$\left.\frac{\partial \log f\left(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T; \boldsymbol{\theta}\right)}{\partial \mathrm{vec}\left(\boldsymbol{\Gamma}\right)}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \sum_{t=1}^{T} \mathrm{vec}\left[\hat{\boldsymbol{\Omega}}^{-1}\left(\hat{\boldsymbol{v}}_{t|T}\hat{\boldsymbol{c}}'_{t-1|T} + \boldsymbol{P}_{t-1,t|T} - \hat{\boldsymbol{\Gamma}}\boldsymbol{P}_{t-1|T}\right)\right]. \tag{2.A.22}$$

Finally, the differential with respect to vec $(\boldsymbol{\Omega})$ is

$$
\begin{aligned}
d \log f(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T; \boldsymbol{\theta}) &= -\frac{T}{2} d(\log |\boldsymbol{\Omega}|) - \frac{1}{2} \sum_{t=1}^{T} \operatorname{tr}\left[(d\boldsymbol{\Omega}^{-1}) \hat{\boldsymbol{Z}}_{t|T} \right] \\
&= -\frac{T}{2} \operatorname{tr}\left[\boldsymbol{\Omega}^{-1} (d\boldsymbol{\Omega}) \right] + \frac{1}{2} \sum_{t=1}^{T} \operatorname{tr}\left[\boldsymbol{\Omega}^{-1} (d\boldsymbol{\Omega}) \boldsymbol{\Omega}^{-1} \hat{\boldsymbol{Z}}_{t|T} \right] \\
&= -\frac{1}{2} \sum_{t=1}^{T} \operatorname{tr}\left[\boldsymbol{\Omega}^{-1} (\boldsymbol{I}_N - \hat{\boldsymbol{Z}}_{t|T} \boldsymbol{\Omega}^{-1}) (d\boldsymbol{\Omega}) \right] \\
&= -\frac{1}{2} \sum_{t=1}^{T} \left(\operatorname{vec}\left[\boldsymbol{\Omega}^{-1} (\boldsymbol{I}_K - \hat{\boldsymbol{Z}}_{t|T} \boldsymbol{\Omega}^{-1}) \right] \right)' d\operatorname{vec}(\boldsymbol{\Omega}). \qquad (2.A.23)
\end{aligned}
$$

Then, the score vector with respect to vec $(\boldsymbol{\Omega})$ evaluated at $\hat{\boldsymbol{\theta}}$ is

$$
\left. \frac{\partial \log f(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T; \boldsymbol{\theta})}{\partial \operatorname{vec}(\boldsymbol{\Omega})} \right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}} = -\frac{1}{2} \sum_{t=1}^{T} \operatorname{vec}\left[\hat{\boldsymbol{\Omega}}^{-1} (\boldsymbol{I}_K - \hat{\boldsymbol{Z}}_{t|T} \hat{\boldsymbol{\Omega}}^{-1}) \right]. \qquad (2.A.24)
$$

Therefore, we conclude that the whole score vector can be computed exactly in a single pass of the Kalman filter and a smoother.

Appendix B: Some useful results of matrix calculus

This appendix introduces some useful results of matrix calculus. We use these results for calculation of log likelihood functions. For a complete exposition, see Magnus and Neudecker [1988].

B.1. Relationship of vec-operator and trace

For any two matrices of the same order, the following relation holds:

$$
[\operatorname{vec}(\boldsymbol{A})]' \operatorname{vec}(\boldsymbol{B}) = \operatorname{tr}(\boldsymbol{A}'\boldsymbol{B}). \qquad (2.B.1)
$$

Proof: Let $\boldsymbol{A} = (a_{ij}), \boldsymbol{B} = (b_{ij})$ be $(I \times J)$ matrices. Then,

$$
[\operatorname{vec}(\boldsymbol{A})]' \operatorname{vec}(\boldsymbol{B}) = \sum_{i=1}^{I} \sum_{j=1}^{J} a_{ij} b_{ij} = \sum_{j=1}^{J} (\boldsymbol{A}'\boldsymbol{B})_{jj} = \operatorname{tr}(\boldsymbol{A}'\boldsymbol{B}).
$$

B.2. Differential of the determinant

For any nonsingular matrix $\boldsymbol{A}$, the following relation holds:

$$
d\,|\boldsymbol{A}| = |\boldsymbol{A}| \operatorname{tr}\left[\boldsymbol{A}^{-1} (d\boldsymbol{A}) \right]. \qquad (2.B.2)
$$

Proof: Let $\boldsymbol{A} = (a_{ij})$ be an $(I \times I)$ matrix. Recall that the minor of a_{ij} is the determinant of the $[(I-1)\times(I-1)]$ submatrix of $\boldsymbol{A}$ obtained by deleting the i-th row and the j-th column, and the cofactor c_{ij} of a_{ij} is $(-1)^{i+j}$ times the minor. The cofactors are can be put into an $(I \times I)$ matrix $\boldsymbol{C} = (c_{ij})$. Then,

$$|\boldsymbol{A}| = \sum_{i=1}^{I} c_{ij} a_{ij} \text{ for } j = 1, 2, \ldots, I.$$

The crucial step is to realize that, for given j, $c_{1j}, c_{2j}, \cdots, c_{Ij}$ do not depend on a_{ij}. This gives

$$\frac{\partial |\boldsymbol{A}|}{\partial a_{ij}} = \frac{\partial(c_{1j}a_{1j} + c_{2j}a_{2j} + \cdots + c_{ij}a_{ij} + \cdots + c_{Ij}a_{Ij})}{\partial a_{ij}} = c_{ij},$$

and hence

$$\begin{aligned} d\,|\boldsymbol{A}| &= \sum_{i=1}^{I}\sum_{j=1}^{I} c_{ij} da_{ij} = \mathrm{tr}[\boldsymbol{C}'(d\boldsymbol{A})] = \mathrm{tr}[(\,|\boldsymbol{A}|\,\boldsymbol{A}^{-1})(d\boldsymbol{A})] \\ &= |\boldsymbol{A}|\,\mathrm{tr}[\boldsymbol{A}^{-1}(d\boldsymbol{A})].\ (\because \boldsymbol{C}' = |\boldsymbol{A}|\,\boldsymbol{A}^{-1}) \end{aligned}$$

B.3. Differential of $\log |\boldsymbol{A}|$

For any nonsingular matrix $\boldsymbol{A}$ and $|\boldsymbol{A}| > 0$, the following relation holds:

$$d \log |\boldsymbol{A}| = \mathrm{tr}\left[\boldsymbol{A}^{-1}(d\boldsymbol{A})\right]. \tag{2.B.3}$$

Proof: Using the results of equation (B.2), we have

$$d \log |\boldsymbol{A}| = \frac{d\,|\boldsymbol{A}|}{|\boldsymbol{A}|} = \frac{|\boldsymbol{A}|\,\mathrm{tr}[\boldsymbol{A}^{-1}(d\boldsymbol{A})]}{|\boldsymbol{A}|} = \mathrm{tr}[\boldsymbol{A}^{-1}(d\boldsymbol{A})].$$

B.4. Differential of the inverse

For any nonsingular matrix $\boldsymbol{A}$, the following relation holds:

$$d\boldsymbol{A}^{-1} = -\boldsymbol{A}^{-1}(d\boldsymbol{A})\boldsymbol{A}^{-1}. \tag{2.B.4}$$

Proof: Since $\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$, we have

$$\begin{aligned} d(\boldsymbol{A}^{-1}\boldsymbol{A}) &= d\boldsymbol{I} \\ (d\boldsymbol{A}^{-1})\boldsymbol{A} + \boldsymbol{A}^{-1}(d\boldsymbol{A}) &= \boldsymbol{O} \\ d\boldsymbol{A}^{-1} &= -\boldsymbol{A}^{-1}(d\boldsymbol{A})\boldsymbol{A}^{-1}. \end{aligned}$$

References

- Bai, J. (2003), "Inference on factor models of large dimensions," *Econometrica* 71(1), 135-172.
- Bai, J. and Ng, S. (2002), "Determining the number of factors inapproximate factor models," *Econometrica* 70, 191-221.
- Berndt, E. K., B. H. Hall., R. E. Hall., and J. A. Hausman. (1974), "Estimation and inference in nonlinear structural models," *Annals of Economic and Social Management* 3, 653-665.
- Breusch, T. S. and Pagan, A. R. (1980), "The Lagrange multiplier test and its applications to model sepcification in econometrics," *Review of Economic Studies* 47, 239-253.
- Connor, G. and Korajczyk, R. (1986), "Performance measurement with the arbitrage pricing theory: a new framework for analysis," *Journal of Financial Economics* 15, 373-394.
- Cragg, J. and Donald, S. (1997), "Inferring the Rank of a Matrix," *Journal of Econometrics* 76, 223-250.
- Davidson, R. and MacKinnon, J. G. (1983), "Small sample properties of alternative forms of the Lagrange multiplier test," *Economics Letters* 12, 269-275.
- Davidson, R. and MacKinnon, J. G. (1993), *Estimation and inference in econometrics*, Oxford University Press, New York.
- De Jong, P. (1989), "Smoothing and Interpolation with the State-Space Model," *Journal of American Statistical Association* 84, 1085-1088.
- De Jong, P. and Mackinnon, M. (1988), "Covariances for Smoothed Estimates in State-Space Models," *Biometrica* 75, 601-602.
- Diebold, F. X. and Rudebusch, G. D. (1996), "Measuring Business Cycles: A Modern Perspective," *The Review of Economics and Statistics* 78(1) 67-77.
- Diebold, F. X., Rudebusch, G. D., and Aruoba., S. B. (2006), "The macroeconomy and the yield curve: a dynamic latent factor approach," *Journal of Econometrics* 131(1), 309-338.
- Donald, S. (1997), "Inferring Concerning the Number of Factors in a Multivariate Nonparametric Relationship," *Econometrica* 65, 103-132.
- Durbin, J. and Koopman, S. J. (2001), *Time Series Analysis by State-Space Models*, Oxford University Press, New York.
- Engle, R. F. (1984), "Wald, likelihood ratio, and Lagrange multiplier tests in econometrics," in: Z. Griliches and M.D. Intriligator, eds., *Handbook of econometrics*, Vol. II (North-Holland, Amsterdam).

- Engle, R. F., Hendry., D. F., and Richard, J. F. (1983), "Exogeneity," *Econometrica* 51(2), 277-304.
- Engle, R. F. and Watson, M. F. (1981), "A one-factor multivariate time series model of metropolitan wage rates," *Journal of the American Statistical Association* 76, 774-781.
- Ericsson, N. R. (1991), "Monte Carlo methodology and the finite sample properties of instrumental statistics for testing nested and non-nested hypothesis," *Econometrica* 59, 1249-1277.
- Forni, M. and Lippi, M. (2001), "The generalized factor model: representation theory," *Econometric Theory* 17, 1113-1141.
- Forni, M. and Reichlin, L. (1998), "Let's get real: a factor analytical approach to disaggregated business cycle dynamics," *Review of Economic Studies* 65, 453-473.
- Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000), "The generalized factor model: identification and estimation," *The Review of Economics and Statistics* 80, 540-554.
- Forni, M., Hallin, M., Lippi, M., and Reichlin, L., (2001), "Coincident and leading indicators for the Euro area," *The economic Journal* 111, 62-85.
- Gallant, A. R. and White, H. (1988), "A unified theory of estimation and inference for nonlinear dyanmic models," Basis Blackwell, Oxford.
- Geweke, J. (1977), "The dynamic factor analysis of economic time series," In: Aigner, D. J., Goldberger, A. S. (Eds.), *Latent Variables in Socio-Economic Models.* North-Holland, Amsterdam.
- Ghysels, E. and Ng, S. (1998), "A Semi-parametric Factor Model for Interest Rates and Spreads," *Review of Economics and Statistics* 80, 489-502.
- Granger, C. W. J. (1968), "Investigating Causal Relations by Econometric Models and Cross-spectral Methods," *Econometrica* 37(3) 424-438.
- Granger, C. W. J. (2001), "Macroeconometrics: past and futures," *Journal of Econometrics* 100, 17-19.
- Greene, W. (2006), *Econometric Analysis*, Sixth ed., Prentice Hall.
- Godfrey, L. G. and Wickens, M. R. (1981), "Testing linear and log-linear regressions for functional form," *Review of Economic Studies* 48, 487-496.
- Godfrey, L. G. and McAleer, M., and McKenzie, C. R. (1988), "Variable addition and and Lagrange multiplier tests for linear and logarithmic regression models," *Review of Economics and Statistics* 70, 492-503.
- Hallin, M. and Liška, R. (2003), "Dynamic factor models: the number of factors and related issues," *Unpublished manuscript.*

- Hamilton, J. D. (1989), “A new approach to the economic analysis of nonstationary time series and the business cycle,” *Econometrica* 57, 357-384.

- Hamilton, J. D. (1990), “ Analysis of time series subject to changes in regime,” *Journal of Econometrics* 45, 39-70.

- Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press, New Jersey.

- Hamilton, J. D. (1996), “Specification testing in Markov-switching time-series models,” *Journal of Econometrics* 70(1), 127-157.

- Kim, C. J. and Nelson, C. R. (1998), “Business Cycle Turning Points, A New Coincident Index, and Tests of Duration Dependence Based on A Dynamic Factor Model with Regime-Switching,” *Review of Business and Economic Statistics* 80, 188-201.

- Kim, C. J. and Nelson, C. R. (1999), *State-Space Models with Regime Switching*, Cambridge: M.I.T. Press.

- Kitagawa, G. (1987), “Non Gaussian State-Space Modelling of Nonstationary Time Series,” *Journal of American Statistical Association* 82, 1032-1063.

- Kiviet, J. (1986), “On the rigour of some misspecification tests for modelling dynamics relationships,” *Review of Economic Studies* 53, 241-261.

- Kobayashi, M. and Shi, X. (2005), “Testing for EGARCH against stochastic volatility models,” *Journal of Time Series Analysis* 26, 135-150.

- Koopman., S. J. and Shephard, N. (1992), “Exact Score for Time Series Models in state-space Form,” *Biometrica* 79(4), 823-826.

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

- Lewbel A. (1991), “The Rank of Demand Systems: Theory and Nonparametric Estimation,” *Econometrica* 59, 711-730.

- Mackinnon, J. G. and White, H. (1985), “Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties,” *Journal of Econometrics* 29, 305-325.

- Magnus, J. and Neudecker, H. (1988), *Matrix Calculus with Applications in Statistics and Econometrics* Wiley, Chichester.

- Newey, W. K. (1985), “Maximum likelihood specification testing and conditional moment tests,” *Econometrica* 53, 1047-1070.

- Ruud, P. A. (2000), *An Introduction To Classical Econometric Theory*, Oxford University Press, New York.

- Sentana, E. (2004), “Factor representing portfolios in large asset market,” *Journal of Econometrics* 119, 257-289.
- Stock, J. H. and Watson, M. W. (1989), “New Indexes of Coincident and Laeding Economic Indications,” in *NBER Macroeoconomics Annals 1989*, ed. by O. J. Blanchard and S. Fischer. Cambridge: M.I.T. Press.
- Stock, J. H. and Watson, M. W. (1998), “Diffusion Indexes,” *NBER Working Paper* 6702.
- Stock, J. H. and Watson, M. W. (2002a), “Macroeconomic Forecasting using Diffusion Indexes. *Journal of Business and Economic Statistics* 20, 147-162.
- Stock, J. H. and Watson, M. W. (2002b), “Forecasting using principal components from a large number of predictors,” *Journal of the American Statistical Association* 97, 1167-1179.
- Tauchen, G. (1985), “Diagnostic testing and evaluation of maximum likelihood models,” *Journal of Econometrics* 30, 415-443.
- White, H. (1982), “Maximum likelihood estimation of misspecified models,” *Econometrica* 50, 1-26.
- White, H. (1987), “Specification testing in dynamic models,” in: Truman F. Bewley, ed., *Advances in econometrics*, Fifth world congress, Vol II, (Cambridge University Press, Cambridge).
- White, H. (1984), *Asymptotic Theory for Econometricians*, Academic Press, Orlando, Fla.
- Wooldridge, J. M. (1990), “A unified approach to robust, regression based specification tests,” *Econometric Theory* 6, 17-43.
- Wooldridge, J. M. (1991), “On the application of robust, regression-based on diagnostic to models of conditional means and conditional variances,” *Journal of Econometrics* 47, 5-46.

Table 2.2: Size of the test for linear dependency between $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$: LM_{s1}

T	ϕ_{11}	ϕ_{32}	γ_1	σ_{21}	σ_{33}	10 % level	5 % level	1 % level
500	2	2	0.8	0.8	1	15.1	9.7	2.7
1000	2	2	0.8	0.8	1	12.9	7.8	3.1
1500	2	2	0.8	0.8	1	11.9	6	2.3
1000	**1.8**	2	0.8	0.8	1	13.1	8.2	3.5
1000	**1.6**	2	0.8	0.8	1	14.7	8.6	2.4
1000	**1.4**	2	0.8	0.8	1	15.2	8.6	2.5
1000	2	**1.8**	0.8	0.8	1	12.7	6.9	2.5
1000	2	**1.6**	0.8	0.8	1	14.2	9.4	3.4
1000	2	**1.4**	0.8	0.8	1	13.2	7.5	2.7
1000	2	2	**0.83**	0.8	1	12.4	7	2.4
1000	2	2	**0.85**	0.8	1	13.2	8.9	3
1000	2	2	**0.87**	0.8	1	11.3	6.2	2
1000	2	2	0.8	**0.6**	1	15.3	9.9	2.3
1000	2	2	0.8	**0.4**	1	10.3	5.7	1.9
1000	2	2	0.8	**0.2**	1	12.4	7.1	2.2
1000	2	2	0.8	0.8	**0.8**	13	7.6	2.5
1000	2	2	0.8	0.8	**0.6**	12.5	7.5	2.8
1000	2	2	0.8	0.8	**0.4**	13.2	7.4	2

Table 2.3: Size of the test for weak exogeneity of $\boldsymbol{y}_2$: LM_{s2}

T	ϕ_{11}	ϕ_{32}	γ_1	σ_{21}	σ_{33}	10 % level	5 % level	1 % level
500	2	2	0.8	0.8	1	20.1	13.3	5.1
1000	2	2	0.8	0.8	1	20.3	13	3.3
1500	2	2	0.8	0.8	1	18.9	12.1	5
1000	**1.8**	2	0.8	0.8	1	20.4	12.6	4.6
1000	**1.6**	2	0.8	0.8	1	23.2	13.9	5.4
1000	**1.4**	2	0.8	0.8	1	19.7	12.6	4.7
1000	2	**1.8**	0.8	0.8	1	20	12.2	4.3
1000	2	**1.6**	0.8	0.8	1	18.5	11.6	4.1
1000	2	**1.4**	0.8	0.8	1	19.3	11.1	3.8
1000	2	2	**0.83**	0.8	1	20.4	13	4.4
1000	2	2	**0.85**	0.8	1	19.2	12.7	3.4
1000	2	2	**0.87**	0.8	1	18.6	11.9	2.1
1000	2	2	0.8	**0.6**	1	19.8	12.3	3.7
1000	2	2	0.8	**0.4**	1	18.9	11	2.9
1000	2	2	0.8	**0.2**	1	19.8	10.5	3.8
1000	2	2	0.8	0.8	**0.8**	20.5	12.3	4.7
1000	2	2	0.8	0.8	**0.6**	19.8	12.4	5.4
1000	2	2	0.8	0.8	**0.4**	19.4	11.9	4.1

Table 2.4: Size of the test for omitted explanatory variable $\boldsymbol{x}_t$: LM_{s3}

T	ϕ_{11}	ϕ_{32}	γ_1	σ_{21}	σ_{33}	10 % level	5 % level	1 % level
500	2	2	0.8	0.8	1	4	1.6	0.7
1000	2	2	0.8	0.8	1	2.6	1	0.3
1500	2	2	0.8	0.8	1	2	0.6	0.1
1000	**1.8**	2	0.8	0.8	1	2.8	0.6	0.1
1000	**1.6**	2	0.8	0.8	1	2.5	1.1	0.1
1000	**1.4**	2	0.8	0.8	1	2	0.7	0.3
1000	2	**1.8**	0.8	0.8	1	2.8	0.6	0.1
1000	2	**1.6**	0.8	0.8	1	2.8	0.9	0.1
1000	2	**1.4**	0.8	0.8	1	2.2	1.5	0.1
1000	2	2	**0.83**	0.8	1	2.4	1.1	0.3
1000	2	2	**0.85**	0.8	1	2	0.6	0.1
1000	2	2	**0.87**	0.8	1	2.6	1.5	0.4
1000	2	2	0.8	**0.6**	1	2.3	0.9	0.1
1000	2	2	0.8	**0.4**	1	2.1	1.1	0.1
1000	2	2	0.8	**0.2**	1	1.8	0.6	0.1
1000	2	2	0.8	0.8	**0.8**	2.3	0.9	0.2
1000	2	2	0.8	0.8	**0.6**	1.7	0.5	0.1
1000	2	2	0.8	0.8	**0.4**	2.2	1.1	0.3

Table 2.5: Power of the test for linear dependency between $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$: LM_{p1}

T	ϕ_{12}	ϕ_{31}	σ_{31}	10 % level	5 % level	1 % level
500	**0.1**	0	0	98.1	96.4	89.5
500	**0.2**	0	0	100	100	100
500	**0.3**	0	0	100	100	100
500	0	**0.1**	0	26.3	13.8	6.5
500	0	**0.2**	0	38.2	27.1	14.3
500	0	**0.3**	0	64.6	54.1	33.3
500	0	0	**0.1**	25.4	17.2	7.3
500	0	0	**0.2**	40.5	33.2	19.6
500	0	0	**0.3**	84.9	77	59
1000	**0.1**	0	0	100	100	99.6
1000	**0.2**	0	0	100	100	100
1000	**0.3**	0	0	100	100	100
1000	0	**0.1**	0	21.8	14.5	5.5
1000	0	**0.2**	0	56.3	45.4	26
1000	0	**0.3**	0	86.3	79.9	63.7
1000	0	0	**0.1**	40.9	34.5	22.5
1000	0	0	**0.2**	79.5	69.9	47
1000	0	0	**0.3**	98.7	97.7	92.6
1500	**0.1**	0	0	100	100	100
1500	**0.2**	0	0	100	100	100
1500	**0.3**	0	0	100	100	100
1500	0	**0.1**	0	28.2	20	8.9
1500	0	**0.2**	0	72.8	62.6	41.9
1500	0	**0.3**	0	96.5	93.8	83.8
1500	0	0	**0.1**	41.1	29.5	13.7
1500	0	0	**0.2**	94	88.5	75.3
1500	0	0	**0.3**	100	100	100

Table 2.6: Power of the test for weak exogeneity of $\boldsymbol{y}_2$: LM_{p2}

T	ϕ_{31}	σ_{31}	10 % level	5 % level	1 % level
500	**0.1**	0	30.2	24.2	18.6
500	**0.2**	0	54.1	46.4	29.1
500	**0.3**	0	79.9	72.1	53.8
500	0	**0.1**	33.7	24.6	10.5
500	0	**0.2**	57.8	46.5	28.7
500	0	**0.3**	70.6	61.3	54.9
1000	**0.1**	0	37.8	29.1	14.9
1000	**0.2**	0	75.4	67.6	47.8
1000	**0.3**	0	97	94.5	85.3
1000	0	**0.1**	42.8	33	16.8
1000	0	**0.2**	80.6	73.4	57.1
1000	0	**0.3**	95.6	93.6	86.1
1500	**0.1**	0	42.5	33.3	16.8
1500	**0.2**	0	86.8	81	65.7
1500	**0.3**	0	98.8	97.9	95.5
1500	0	**0.1**	54.8	44.6	24.3
1500	0	**0.2**	95.3	91.3	79.6
1500	0	**0.3**	99.7	99.4	98.1

Table 2.7: Power of the test for omitted explanatory variable $\boldsymbol{x}_t$: LM_{p3}

T	β_2	10 % level	5 % level	1 % level
500	0.1	43.9	33.2	16
500	0.2	80.6	74.7	69.9
500	0.3	99.8	99.8	99.6
1000	0.1	75.5	64.2	41.1
1000	0.2	100	100	99.8
1000	0.3	100	100	100
1500	0.1	90.6	85.5	66.1
1500	0.2	100	100	100
1500	0.3	100	100	100

Chapter 3

Joint Estimation of Factor Sensitivities and Risk Premia in the Factor Augmented APT Model

3.1 Introduction

There is a long tradition of factor or multi-index models in finance where they were originally introduced to simplify the computation of the covariance of returns in a mean-variance portfolio allocation framework. In this context, two major theories provide a rigorous foundation for computing the trade-off between risk and return: the capital asset pricing model (CAPM) and the arbitrage pricing theory (APT).

The CAPM, for which William F. Sharp shared the 1990 Nobel Memorial Prize in Economic Sciences, predicts that only one type of nondiversifiable risk influences expected returns, and that single type of risk is "market risk." In 1976, a little more than a decade after the CAPM was proposed, Stephen A. Ross invented the APT. The APT is more general than the CAPM in accepting a variety of different sources. A number of paper indicate that the APT performs better than the CAPM in terms of describing the expected returns of risky assets. For example, Chen [1983] estimates a version of the APT and finds this model outperforms the CAPM in the U.S. stock market.

The APT takes the view that there need not be any single way to measure systematic risk. There are two alternative approaches to estimate them. The first approach relies in statistical techniques such as factor analysis or principal component to estimate risk exposure profiles and associated risk premiums. The second approach estimates them from available macroeconomic and financial data.

Each of these approaches has its merits and is appropriate for certain types of analysis. In particular, the first approach is useful for determining an appropriate number of factors that are significant in pricing and statistically characterizing them. The estimates extracted using statistical techniques have undesirable property, however, that render them difficult to interpret; This problem arises because, by the nature of the technique, the estimated risk exposure profiles and the associated risk premiums are nonunique linear combinations of more fundamental underlying economic forces. Therefore, the extracted factors lack intuitive economic meaning.

The advantage of the second approach is that it provides an intuitively appealing set of factors that admit economic interpretation of the risk exposure profiles and the associated risk premiums. From a purely statistical view, this approach has the advantage of using economic information in addition to stock returns. This additional information will, in general, lead to statistical estimates of better properties, but of cource, insofar as economic variables are measured with errors, these advantages are diminished.

The objective of this paper is to unify these two approaches and overcome the weakness of them. We propose a framework composed of observed macroeconomic factors and unobserved factors based on state-space system. Employing our framework, we estimate risk exposure profiles and associated risk premiums and extract unobserved factors simultaneously. In addition, we examine the adequacy of macroeconomic factors as the systematic variables based on Chiba [2007]'s framework.

The rest of the paper is organized as follows. Section 3.2 reviews the literature on the APT model. Section 3.3 details the APT basics and derives the equations of the APT. Section 3.4 describes empirical hypothesis and derives observed macroeconomic factors that are the underlying sources of risk. In section 3.5, data are described and in section 3.6 we present main results. Section 3.7 briefly summarizes this study.

3.2 Literature review

A number of reserchers have investigated empirically the relationship between unanticipated innovations in observed economic variables (or unobserved fundamentals) and stock returns based on the APT model. For example, Roll and Ross [1980] conducts a factor analysis to extract systematic factors influencing stock returns. Brown and Weinstein [1983] and Chen [1983] also utilize factor analysis. These results are confirmed by Connor and Korajczyk [1989], who use principal components to estimate the APT. Given that statistical techniques such as factor analysis and principal components offer little in the way of economic intuition when attempting to interpret the estimated risk premia, attention has focused on prespecifying observed macroeconomic and financial factors as candidates for systematic risk factors. For example, Chen et al. [1986] find five statistically significant observed risk factors in the U.S. stock market. There are, shocks to real industrial production, twists in the yield curve, a measure of default risk, unexpected inflation and the change in expected inflation. Following this pioneering study, a plethora of related work has confirmed the result that stock returns are related to observed macroeconomic and financial factors.

Beenstock and Chan [1988] conducted similar analysis for the London stock market and found that significant risk factors were interest rate, input cost, money supply, and inflation. Similarly, Antoniou et al. [1998] analyzed two samples of assets and found that three factors relating money supply, inflation, and excess return on the stock market were priced in the APT model and these factors carried the same price of risk in both sample.

The papers listed above are based on the two-stage cross-sectional estimation technique introduced by Fama and Macbeth [1973]. However, this technique has a number of economic flaws. First, the estimation of asset exposures to the factors in one period and the resulting estimates of the prices of risk in another period lead to an errors-in-variables problem. Second, forming portfolios does not necessarily remove errors-in-variables problem and consequently failure to correct fully for errors in variables may well to invalid inferences regarding which factors are statistically significant. Thus, there is some requirement for estimation framework of the APT, which is robust to the above problem. In next section, we attempt to overcome these problems, by proposing the framework based on the state-space system.

3.3 APT model representation with observed and unobserved factors

This section details the APT basics and derives the equations of the APT. While some theoretical formulation of the APT can be demanding, appealing basics behind the APT are easy to understand. Moreover, the APT provides a portfolio manager with a variety of new and easily implemented tools to control risks and to enhance portfolio performance.

The APT model expresses a set of observed asset returns as a function of risk factors. Let $\boldsymbol{r}_t$ be a vector of actual asset return for $t = 1, 2, \ldots, T$. Consider the following model:

$$\underset{(N\times 1)}{\boldsymbol{r}_t} = \underset{(N\times 1)}{\bar{\boldsymbol{r}}} + \underset{(N\times K)}{\boldsymbol{\Phi}} \underset{(K\times 1)}{\boldsymbol{f}_t} + \underset{(N\times 1)}{\boldsymbol{u}_t}, \tag{3.3.1}$$

where $\boldsymbol{f}_t$ is a vector of systematic factors, $\boldsymbol{\Phi}$ is a matrix of factor loadings (risk exposure profiles) associated with $\boldsymbol{f}_t$, and $\boldsymbol{u}_t$ is a vector of idiosyncratic risks. We assume that $\boldsymbol{f}_t$ and $\boldsymbol{u}_t$ have zero means so that $\bar{\boldsymbol{r}}_t$ denotes a vector of expected return. Equation (3.3.1) merely says that the actual return equals to the sum, the expected return, plus over all risk factors, of the risk exposure multiplied by the realization for that risk factor, plus idiosyncratic error term. The model addresses how expected return behave in a market with no arbitrage opportunities[1] and predicts that an asset's expected return is linearly related to the factor loading:

$$\underset{(N\times 1)}{\bar{\boldsymbol{r}}} = \underset{(N\times 1)}{\boldsymbol{r}_f} + \underset{(N\times K)}{\boldsymbol{\Phi}} \underset{(K\times 1)}{\boldsymbol{p}}, \tag{3.3.2}$$

where $\boldsymbol{r}_f$ is a vector of risk-free return and $\boldsymbol{p}$ is a vector of risk premiums. Equation (3.3.2) is based on the rationale that unsystematic risk is diversifiable and therefore should have a zero price in the market with no arbitrage opportunities.

In addition, in order to distinguish between observed and unobserved factor, we partition the set of K factors into two groups of K_1 unobserved factors and K_2 observed factors. Then

[1] *Note*: No arbitrage opportunities mean that because of competition in financial markets, it is impossible for an investor to earn a positive expected return on any combination of assets without undertaking some risk and without making some net investment of funds.

equation (3.3.1) is partitoned as

$$\begin{aligned}
\boldsymbol{r}_t &= \bar{\boldsymbol{r}} + \begin{bmatrix} \underset{(N\times K_1)}{\boldsymbol{\Phi}_1} & \underset{(N\times K_2)}{\boldsymbol{\Phi}_2} \end{bmatrix} \begin{bmatrix} \underset{(K_1\times 1)}{\boldsymbol{f}_{1t}} \\ \underset{(K_2\times 1)}{\boldsymbol{f}_{2t}} \end{bmatrix} + \boldsymbol{u}_t \\
&= \bar{\boldsymbol{r}} + \boldsymbol{\Phi}_1 \boldsymbol{f}_{1t} + \boldsymbol{\Phi}_2 \boldsymbol{f}_{2t} + \boldsymbol{u}_t. \qquad (3.3.3)
\end{aligned}$$

Note that $K = K_1 + K_2$, $\boldsymbol{f}_{1t}$ is a vector of unobserved factor, and $\boldsymbol{f}_{2t}$ is a vector of observed factor. If the unoberved factor follows a stationary vector autoregressive process of first order, the model forms *state-space system.* Thus, substituting (3.3.3) into (3.3.2), the state-space system of our model is as follows:

$$\boldsymbol{R}_t = \boldsymbol{\mu} + \boldsymbol{\Phi}_1 \boldsymbol{f}_{1t} + \boldsymbol{u}_t, \qquad (3.3.4)$$

$$\boldsymbol{f}_{1t} = \boldsymbol{\Gamma} \boldsymbol{f}_{1,t-1} + \boldsymbol{v}_t, \qquad (3.3.5)$$

where

$$\boldsymbol{\mu} \equiv \boldsymbol{\Phi p} + \boldsymbol{\Phi}_2 \boldsymbol{f}_{2t}, \boldsymbol{R}_t \equiv \boldsymbol{r}_t - \boldsymbol{r}_f.$$

Here, $\boldsymbol{R}_t$ is a vector of asset returns in excess of the risk-free rate, $\boldsymbol{\mu}$ is a vector of observed components, $\boldsymbol{\Gamma} = \text{diag}(\gamma_1, \gamma_2, \ldots, \gamma_{K_1})$ is a matrix of autoregressive coefficients, and $\boldsymbol{v}_t$ is a vector of noises.

Further, we require that Gaussian white noise $\boldsymbol{u}_t$ and $\boldsymbol{v}_t$ be orthogonal to each other and to the initial state:

$$\begin{bmatrix} \boldsymbol{u}_t \\ \boldsymbol{v}_t \end{bmatrix} \sim \text{i.i.d.} N \left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Omega} \end{bmatrix} \right), \qquad (3.3.6)$$

$$E(\boldsymbol{f}_{11} \boldsymbol{u}_t') = E(\boldsymbol{f}_{11} \boldsymbol{v}_t') = \boldsymbol{0}. \qquad (3.3.7)$$

In our analysis, we assume that $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ are diagonal matrices for simplicity. Therefore, the structures of the matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ employed here are defined as follows:

$$\boldsymbol{\Sigma} = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_N),\ \boldsymbol{\Omega} = \text{diag}(\omega_1, \omega_1, \ldots, \omega_K).$$

This is the basic structure we adopt for the rest of this paper. Consequently, the APT is completely general and does not specify exactly what the systematic risks are, or even how many such risks exist, academic and commercial research suggests that there are several primary sources of risk which consistently impact returns.

In general, the state-space representation provides a powerful framework for estimation and testing of the models. The recognition that the APT model is easily put in state-space form is particularly useful, because the application of Kalman filter then delivers maximum likelihood estimates (MLE) and optimal filtered and smoothed estimates of parameters and unobserved risk factors. In addition, the one-step Kalman filter approach of this paper is preferable to the Fama and Macbeth approach, because the simultaneous estimation of all parameters produces correct inference via the standard theory. Whereas, the two-step procedure suffers from the errors-in-variable as discussed in Section 3.2. In addition, state-space representation facilitates specification testing regarding the adequacy of macroeconomic factors on the systematic variables based on the LM principle. Therefore, our framework has a number of advantages over the Fama and MacBeth methodology.

3.4 Examines factors

3.4.1 Theory

We have defined that asset returns are function of unobserved and observed factors in section 3.2. But, what are observed factors? If we knew them, we could attribute a particular fraction of the observed price movement in a given stock (or portfolio) to movements in observed factors.

Unfortunately, there is no formal theoretical guidance in choosing the appropriate group of observed factors to be included in the APT model. However, stock prices are usually considered as responding to external forces (even though they may have a feedback on other variables). Therefore, our goal is to "model stock returns as functions of observed economic factors and unobserved factors."

Our choice of economic factors was based on the general hypothesis that stock returns are influenced by macroeconomic factors. Changes in any of these factors could be expected to change investor's perceptions of cash flows and therefore affect asset prices. In other words, any systematic factors that affect the economy's pricing operator or that influence dividends would also influence stock market returns. Additionally, any factors that are necessary to complete the description of the state of nature will also be part of the description of systematic risk factors. Therefore, the choice of factors is typically justified by reference to the traditional discounted cash flows (DCF) valuation formula. The DCF valuation formula is given by:

$$p_t = \sum_{\tau=0}^{\infty} \left(\frac{1}{1+r} \right)^{\tau+1} E_t(d_{t+\tau}), \tag{3.4.1}$$

where $E_t(\cdot)$ denotes an expectation of unknown future quantity based on information available to investors at date t, r denotes the appropriate discount rate, and $d_{t+\tau}$ is the dividend paid at the end of period $t+\tau$. Any economic announcements will affect stock price movements if the new information revealed by announcement affects either expectations of future dividends or discount rate or both.

Expected dividends are affected by anything which influences cash flows. Changes in industrial production influences profits and hence dividends. Unexpected inflation would influence nominal expected cash flows as well as the nominal rate of interest. Perhaps a more important reason to expect a relationship between stock returns and unexpected inflation is that unexpected inflation contains new information about future levels of expected information. If unexpected inflation is bad news for the stock market, and if the announcement of the Consumer Price Index (CPI) contains new information about inflation, then unexpected inflation (deflation) should be associated with a decrease (increase) in stock prices at the time of the announcement. The relationship between inflation and common stock returns has been studied extensively.

The discount rate in equation (3.4.1) is constructed from the prevailing risk-free rate and a risk premium, and is an average of rates over time, and it changes with both the level of rate and the term structure spread across different maturities. Unanticipated changes in the risk-less interest rate influences returns. The discount rate also depends on the risk premium; hence, unanticipated changes in the premium influences returns.

3.4.2 Constructing macroeconomic factors

This subsection details the derivation of macroeconomic factors we use in this paper. Several primary sources of risk from unanticipated changes in the fundamental economic factors are as follows.

Business cycle risk (monthly and annual growth of industrial production)

Business cycle risk represents unanticipated changes in the level of business activity. Their basic series are computed as the growth rate in U.S. industrial activity. If IP_t denotes the rate of industrial production in month t, then the monthly growth rate is

$$MP_t = \log IP_t - \log IP_{t-1}, \tag{3.4.2}$$

and the annual growth rate is

$$YP_t = \log IP_t - \log IP_{t-12}. \tag{3.4.3}$$

Hence, a positive realization of Business cycle risk indicates that the expected rate of the economy has increased. Note that MP_t measures the change in industiral production lagged by at least a partial month. To make this variable contemporaneous with other series, subsequent statistical work will lead it by 1 month. Similarly, YP_t measures the change in industiral production in the long run. Therefore, subsequent statistical work will lead it by 1 year.

Inflation risk (expected and unexpected inflation)

Inflation risk is a combination of the unexpected components of short- and long-run inflation rate. If UI_t denotes the unexpected inflation in month t, UI_t is derived as

$$UI_t = I_t - E_{t-1}(I_t), \tag{3.4.4}$$

where I_t is the realized monthly first difference in the logarithm of the Consumer Price Index for period t. The series of expected inflation $E_{t-1}(I_t)$ is obtained from Fama and Gibbons [1984]. If RHO_t denotes the ex post real rate of interest applicable in period t and TB_{t-1} denotes the return on 1-month Treasury-bill known at the end of period $t-1$ and applying to period t, the Fisher's equation asserts that

$$TB_{t-1} = E_{t-1}(RHO_t) + E_{t-1}(I_t). \tag{3.4.5}$$

Hence, $TB_{t-1} - I_t$ measures the ex post return on Treasury-bill in the period. Another inflation variable that is unanticipated and that might have an influence separable from UI is

$$DEI_t = E_t(I_{t+1}) - E_{t-1}(I_t), \tag{3.4.6}$$

the change in expected inflation. We subscript this variable with t since it is unknown at the end of month $t-1$.

Since most stocks have negative exposures to Inflation risk, a positive inflation surprise causes a negative contribution to return, whereas a negative inflation surprise (deflation surprise) contirbutes positively toward them.

Note that chnages in industrial production is a proxy for the discounted future cash flow. And, the inflation related variables enter because assets are not risk neutral; their nominal cash flow growth rates do not always match expected inflation rates.

Time horizon risk (the term structure)

Time horizon risk is the unanticipated change in investor's desired time to payout. It is measured as follows:

$$UTS_t = LGB_t - TB_{t-1}. \tag{3.4.7}$$

Note that LGB denotes the return on twenty-year government bonds. A positive realization of Time horizon risk means that the price of long-term bonds has risen relative to the 1-month Treasury-bill price. This is a signal that investors require a lower compensation for holding investments with relatively longer time to payout.

The term structure would seem intuitively to be more related to the denominator of in the DCF formula-i.e., to the risk adjusted discount rate. The term structure of interest rate enters because most assets have multiple year cash flows, for reasons relating to time preference, the discount rate that applies to cash flows in future.

These variables make intuitive sense, and it also makes sense that they are indeed "systematic." It is possible, of course, to think of many other potential systematic factors. However, many of them influence returns only through their impact on the above five factors. The money supply, for example, is an important variable, but it is not good yardstick to measure exposures, because most of the influence of money supply changes is captured by other variables.

We summarize the result of the derivation of macroeconomic factors in Table 3.1.

3.5 The data

3.5.1 Description of the data

The basic data[2] for this study consists of rates of return on all securities trading on the all NYSE, AMEX, and NASDAQ from April 1941 through June 2000. This period is long enough, so we divide the entire time period into three a priori to discuss different magnitude of the influence of risk factors on asset pricing.

The dependent variables are six portfolios in excess of risk-free rate, which are constructed at the end of each June, are the intersections of 2 portfolios formed on size[3] (market

[2] *Note*: The data was kindly made available on Kenneth R. French's website.

[3] *Note*: The size is computed as price of stock multiplied by outstanding volume. The size is known to be strongly related to average return, and we hoped that it would provide the desired dispersion without biasing the macroeconomic factors.

Table 3.1: Hypothesized macroeconomic factors

Symbol	Macroeconomic factors	Definition or Source
Basic Series		
I	Inflation	Log relative of U.S. Consumer Price Index
TB	Treasury-bill rate	End-of-period term on 1-month bills
LGB	Long-term goverment bonds	Return on long-term goverment bonds
IP	Industrial Production	Industrial Production during month
Derived series		
MP	Monthly growth, industrial production	$\log IP_t - \log IP_{t-1}$
YP	Annual growth, industrial production	$\log IP_t - \log IP_{t-12}$
EI	Expected Inflation	Fama and Gibbons [1984]
UI	Unexpected Inflation	$I_t - E_{t-1}(I_t)$
RHO	Ex post real interest	$TB_{t-1} - I_t$
DEI	Change in expected inflation	$E_t(I_{t+1}) - E_{t-1}(I_t)$
UTS	Term structure	$LGB_t - TB_{t-1}$

equity, ME) and 3 portfolios formed on the ratio of book equity to market equity (BE/ME). The size breakpoint for year t is the median NYSE market equity at the end of June of year t. BE/ME for June of year t is the book equity for the last fiscal year end in $t-1$ divided by ME for December of $t-1$. The BE/ME breakpoints are the 30th and 70th NYSE percentiles. The risk-free rate is taken from the 1-month Treasury-bill rate that is known to at the beginning of each month. The six portfolio returns are fully adjusted for dividends. Monthly returns are used in all subsequent analysis because most macroeconomic data are only available on monthly basis. Macroeconomic factors we use are monthly and annual growth of industrial production, unexpected inflation, the change in expected inflation, and the spread between long and short interest. Data for those macroeconomic factors were obtained from the *Federal Reserve System*.

3.5.2 Statistical characteristics of the data

This subsection describes statistical characteristics of the data. Descriptive statistics for six portfolios and five macroeconomic factors are given in Table 3.2[4].

The eighth column of Table 3.2 shows that the kurtosis of all stock return portfolios is much larger than the kurtosis of the normal distribution (is equal to 3). This reflects the fact that the tails of distribution of these portfolios are fatter than the tail of the normal distribution. Put differently, large observations occur much more often than one might expect

[4] *Note*: Glossary: SV: small value portfolio, SN: small neutral portfolio, SG: small growth portfolio, BV: big value portfolio, BN: big neutral portfolio, BG: big growth portfolio. JB test statistic is Chi-squared with 2 degree of freedom and LB test statistic is Chi-squared with 12 degree of freedom. Bold entries denote test statistic are significant at the 5 percent level.

for a normally distributed variable.

This is illustrated further in Figure 3.1, which show histograms of the monthly returns on six portfolios. Clearly, all histograms are more peaked and have fatter tails than the normal distribution. Thus, both very small and very large observations occur more often compared to a normally distributed variable with the same first and second moment.

Further, all six portfolios except for portfolio BG have negative skewness, which implies that the left tail of the distribution is fatter than the right tail, or that large negative returns tend to occur more often than large positive one. This is visible in the histograms in Figure 3.1 as well as more extreme values are presented in the left tail than in the right tail. In addition, JB test rejects the null hypothesis in all portfolios, which suggests the normality hypothesis on these six portfolios are not satisfied exactly.

Table 3.2 displays correlation matrices for macroeconomic factors. The correlation matrices of Table 3.2 are computed for several different periods; part A covers the entire 711-month sample period from April 1941 through June 2000, and the remaining part cover three subperiods, with breaks at December 1960 and September 1980.

We find that the strongest correlation is between DEI and UI. This is expected since they both contain the EI series. A number of other correlations are not too strong and the variables are quite far from perfectly correlated.

Table 3.4 and Figure 3.3 illustrate that all macroeconomic factors display mild autocorrelations, in particular YP and UTS are highly autocorrelated and have seasonals: The YP has seasonals at 12-month lag. Whereas, the UTS has seasonals at 30-month lag. These results imply the existence of an errors-in-variable problem that bias estimates of the loadings of the stock returns on these variables and will bias downward estimates of statistical significance.

Table 3.2: Descriptive statistics of six portfolios and macroeconomic factors

Factor	Mean	Median	Min	Max	SD	Skew	Kurt	JB	$LB(12)$
SV	0.739	0.910	-32.820	28.490	6.199	-0.363	5.649	**173.720**	0.371
SN	1.003	1.350	-28.680	26.240	5.094	-0.439	6.073	**193.602**	4.804
SG	1.242	1.430	-28.460	29.810	5.480	-0.056	6.171	**278.943**	14.834
BV	0.709	1.030	-23.810	20.760	4.407	-0.332	4.868	**76.861**	3.235
BN	0.760	0.860	-20.900	16.520	4.028	-0.298	4.892	**80.448**	1.726
BG	1.006	1.180	-19.990	20.630	4.602	0.006	4.230	**84.557**	7.636
MP	0.313	0.360	-10.962	9.937	1.349	-0.814	18.731	**5416.768**	4.250
YP	3.730	4.177	-40.380	25.530	7.813	-1.092	8.615	**167.153**	14.632
UTS	1.248	1.280	-4.700	4.590	1.231	-0.594	4.944	**184.589**	**413.921**
UI	-0.006	-0.011	-5.282	6.130	0.760	0.578	16.245	**5252.970**	**66.600**
DEI	0.005	0.003	-5.035	5.468	0.587	0.275	23.019	**11968.739**	**107.476**

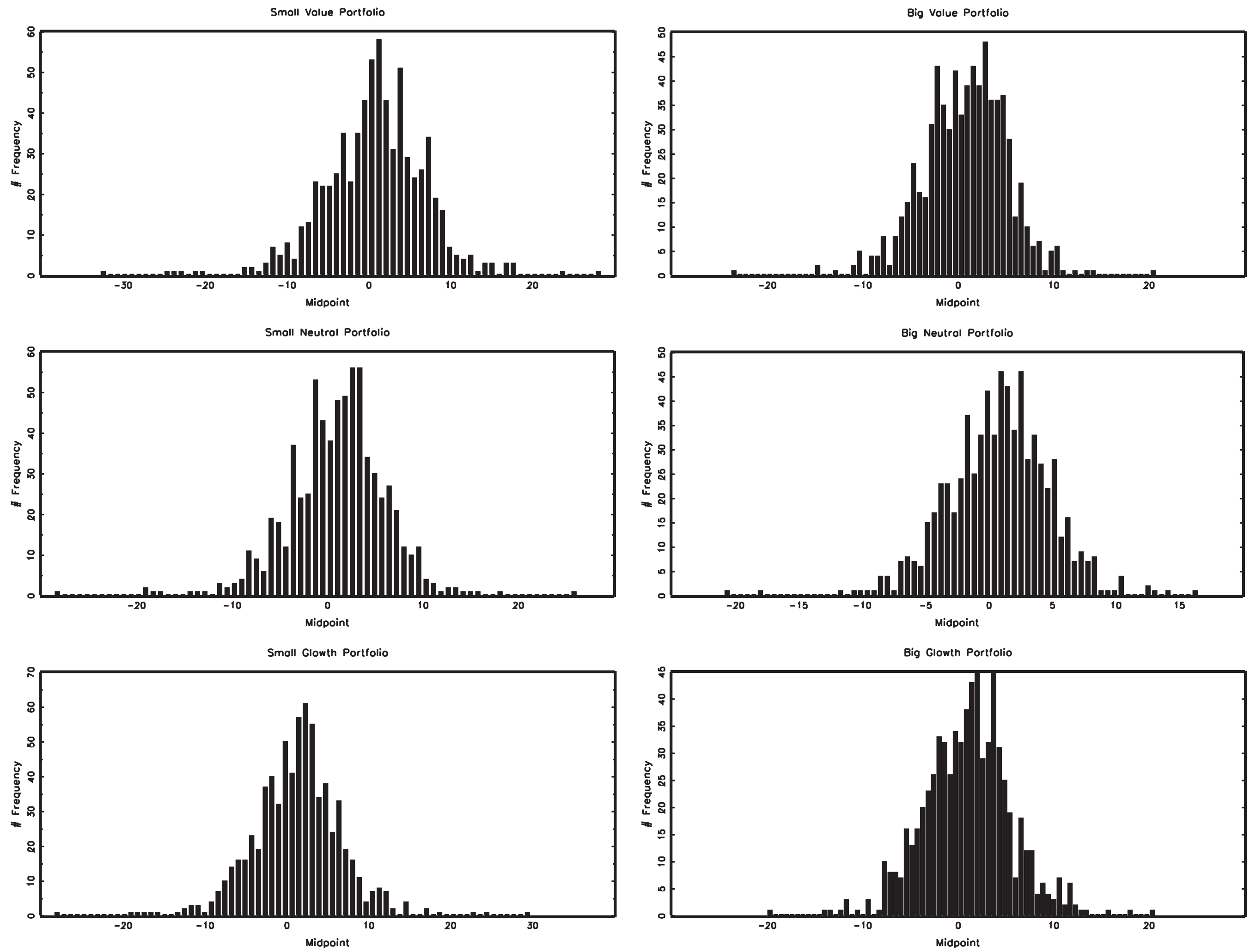

Figure 3.1: Histogram of six portfolios

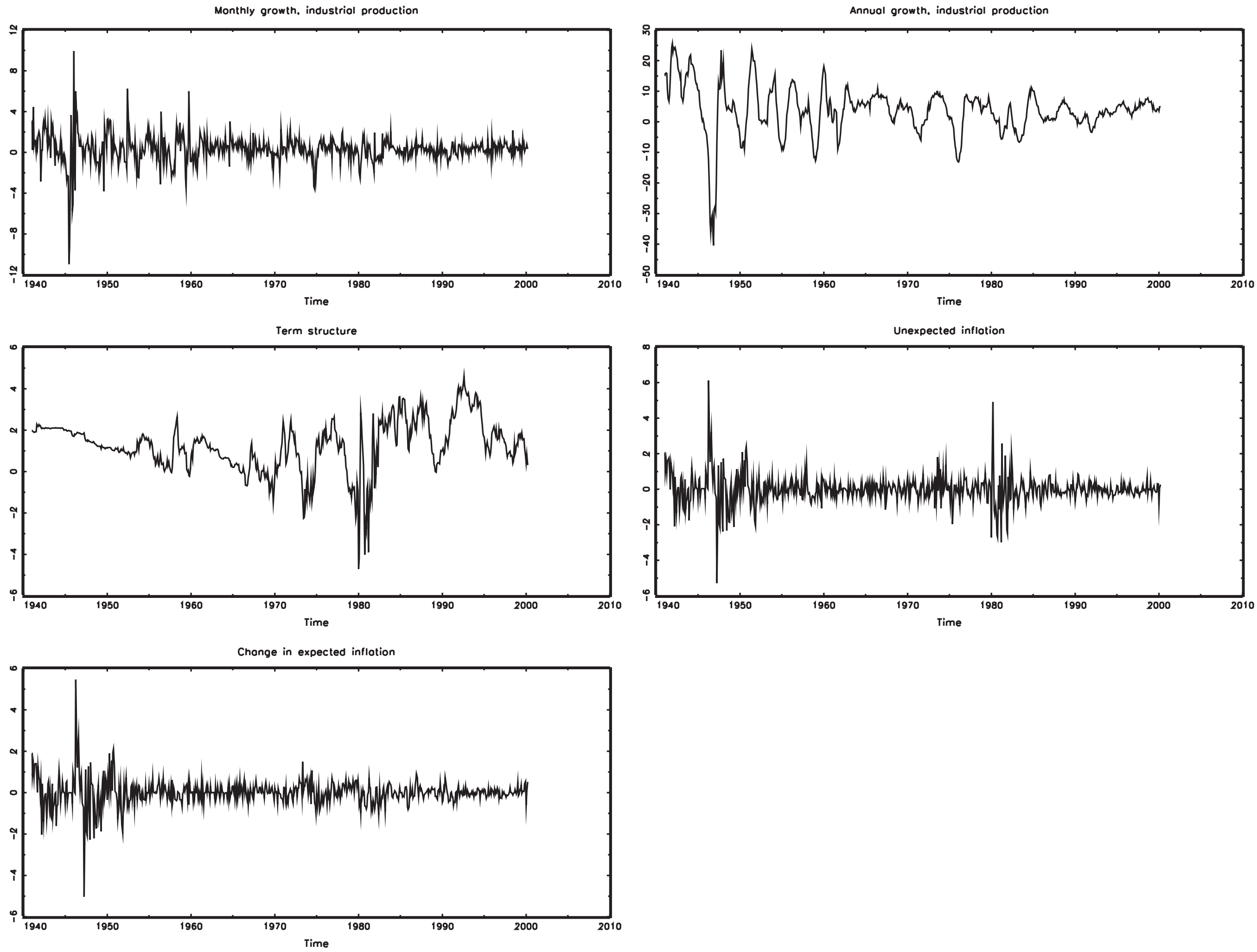

Figure 3.2: Macroeconomic factors, April 1941-June 2000

Table 3.3: Correlation matrices for macroeconomic factors

Factor	MP	YP	UTS	UI	DEI
A: 1941.04-2000.06					
MP	1.000				
YP	-0.083	1.000			
UTS	0.075	-0.079	1.000		
UI	0.025	-0.102	0.068	1.000	
DEI	0.129	-0.119	-0.050	0.772	1.000
B: 1941.04-1960.12					
MP	1.000				
YP	-0.090	1.000			
UTS	0.039	0.069	1.000		
UI	0.095	-0.197	0.060	1.000	
DEI	0.120	-0.211	-0.002	0.979	1.000
C: 1961.01-1980.09					
MP	1.000				
YP	-0.067	1.000			
UTS	0.279	-0.370	1.000		
UI	-0.138	0.103	-0.006	1.000	
DEI	0.120	0.327	-0.282	0.293	1.000
D: 1980.10-2000.06					
MP	1.000				
YP	-0.079	1.000			
UTS	0.086	0.005	1.000		
UI	-0.155	0.070	0.265	1.000	
DEI	0.233	0.042	0.144	0.365	1.000

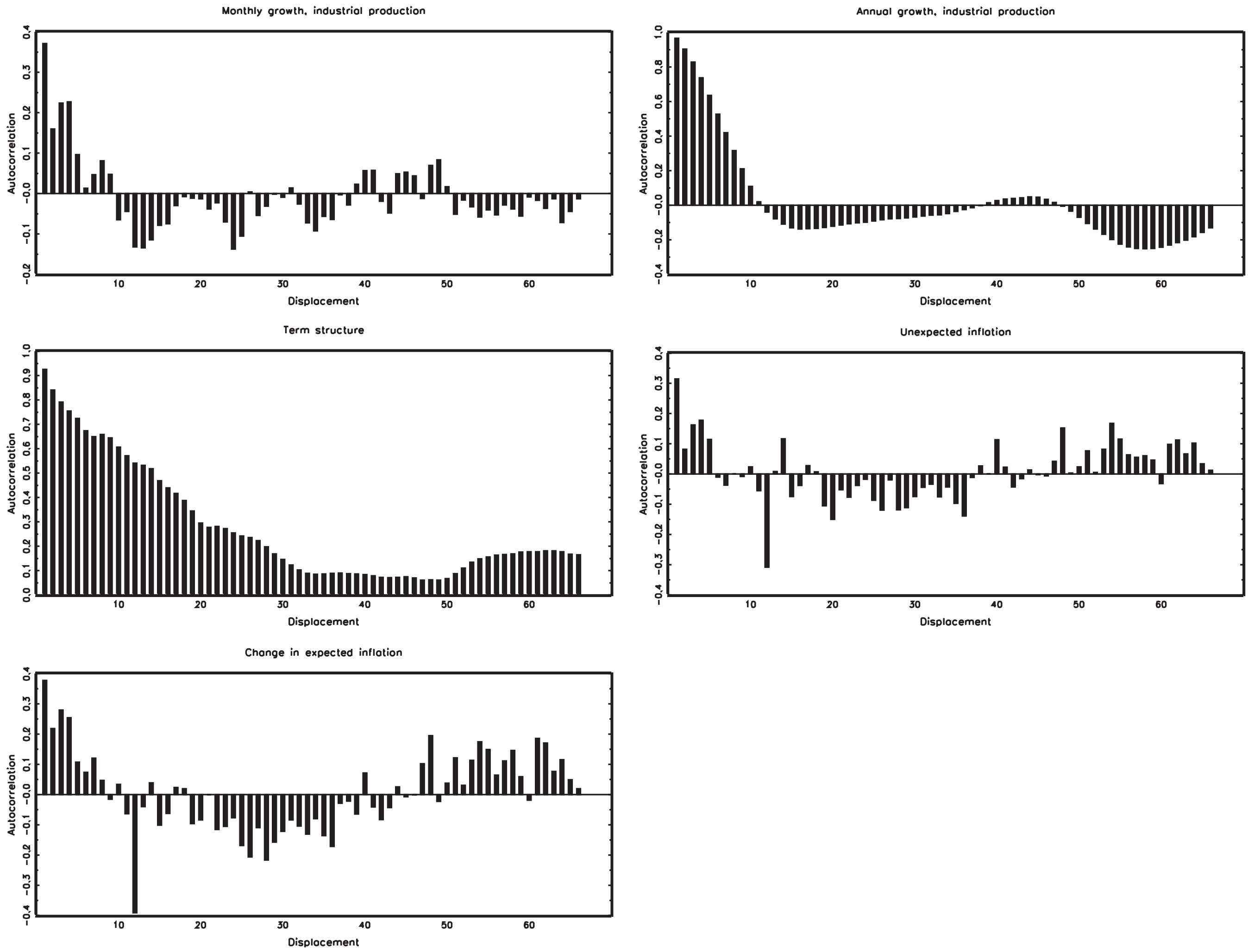

Figure 3.3: Autocorrelation of macroeconomic factors

Table 3.4: Autocorrelation of macroeconomic factors, April 1941-June 2000

Factor	$\hat{\rho}(1)$	$\hat{\rho}(6)$	$\hat{\rho}(12)$	$\hat{\rho}(18)$	$\hat{\rho}(24)$	$\hat{\rho}(30)$
MP	0.371	0.013	-0.132	-0.007	-0.137	-0.009
YP	0.966	0.527	-0.039	-0.134	-0.099	-0.068
UTS	0.925	0.673	0.541	0.387	0.254	0.146
UI	0.314	-0.010	-0.308	0.008	-0.018	-0.075
DEI	0.378	0.073	-0.391	0.020	-0.077	-0.121
Factor	$\hat{\rho}(36)$	$\hat{\rho}(42)$	$\hat{\rho}(48)$	$\hat{\rho}(54)$	$\hat{\rho}(60)$	$\hat{\rho}(66)$
MP	-0.064	-0.019	0.070	-0.058	-0.008	-0.013
YP	-0.025	0.039	-0.006	-0.199	-0.242	-0.131
UTS	0.089	0.072	0.063	0.149	0.178	0.164
UI	-0.138	-0.043	0.152	0.167	-0.031	0.012
DEI	-0.171	-0.083	0.194	0.175	-0.019	0.020

3.6 Main results

3.6.1 Estimation results

This subsection describes the relationship of stock returns and risk factors. We estimate the linear factor model with the APT restrictions imposed for six portfolio excess returns by Kalman filter and a smoother with all of the factor included.

The risk exposure profiles and the corresponding prices of risk are shown in Table 3.5 through Table 3.8. For each risk factor, the contribution to the expected return is the product of the risk exposure profiles (Column 2) and the corresponding risk premiums (Column 3), and the sum of these products is equal to the expected return is excess of 1-month Treasury-bill rate (Column 4). The risk premium of factor determines how much expected return will change because of an increase or decrease in the portfolio's exposure to that type of risk. Hence, the risk premium is particular important in the APT model.

One of the features which stand out most prominently is the sign of risk factors. Over the entire period and in any subperiod, the effects on MP and YP have positive signs. Whereas, UTS, UI, and DEI have negative signs. The positive risk premiums of MP and YP reflect the value of real systematic production risks. In addition, the negative risk premium of UTS indicates that stocks whose returns are inversely related to increases in long rates over short rate are more valuable. One interpretation of this result is that UTS measures a change in the long-term real rate of interest. After long-term real rates decrease, there is subsequently a lower real return on any form of capital. Investors who want protection against this possibility will place a relatively higher value on assets whose price increases when long-term real rates decline, and such assets carry a negative risk premium. Further, the negative signs of UI and DEI probably mean that stock market assets are generally perceived to be hedges against the adverse influence on other assets.

Next, we consider the magnitude of risk premium. From Table 3.6 through Table 3.8, we find that the absolute magnitude of risk premiums have changed in each period: for instance, the magnitude of the risk premium for YP and UTS have changed drastically, whereas, the magnitude of the risk premium for MP, UI, and DEI have not changed very much. However, the risk premium for UC have greately changed in magnitude and in sign.

Finally, we consider the contribution of each factor to the expected excess return. we find that some of the absolute contribution of UC are larger than the total one of observed factors in each period. This reflects that the large proportion of the expected excess return depends on UC in any period.

3.6.2 Specification tests

The subsection 3.6.1 discussed the contribution of the expected excess return for each risk factor. Then, which macroeconomic factor is adequate to explain the expected excess return? This subsection constructs the specification test for the adequacy of macroeconomic factors on the systematic variables based on Chiba [2007]'s framework and derive their results.

In the system of (3.3.4) and (3.3.5), when the excess asset return $\boldsymbol{R}_t$ is specified without the macroeconomic factor f_{kt}, that is restriction that $p_k = \phi_{k1} = \phi_{k2} = \cdots = \phi_{kN} = 0$. To ascertain whether macroeconomic factors are related to the stock return, we conduct a LM test of the null hypothesis as follows:

$$H_0 : p_k = \phi_{1k} = \phi_{2k} = \cdots = \phi_{Nk} = 0 \text{ for } k = K_1 + 1, K_1 + 2, \ldots, K. \tag{3.6.1}$$

If the null hypothesis is rejected, the macroeconomic factor f_{kt} has a significant effect on the stock return and their pricing influence does exist. Whereas, if the null hypothesis cannot be rejected, the macroeconomic factor does not have a significant effect on the stock return.

Part A of Table 3.9[5] displays specification results over the entire period. Part B through part D of Table 3.9 display specification results in each period. We find that MP, UI, and DEI are significant, while YP and UTS are not significant over the entire period. We also find that all macroeconomic factors are significant in period B. However, YP, UTS, and DEI are not significant in period C, further MP is not significant in Period D.

Therefore, we expect that the number of macroeconomic factor which has a significant effect on the stock return are decreasing as time goes by. One interpretation of this result is that the degree of the innovation of macroeconomic factors are decreasing. As we noted earlier, the APT relates the risk of assets to the covariance of unanticipated innovations in observed economic variables (or unobserved fundamentals). Actually, the amplitude of all macroeconomic factors except for UTS are weakening as time goes by. (see Figure 3.2.) Therefore, we conclude that the effect of macroeconomic factors on stock returns is decreasing in recent years.

[5] *Note*: LM statistic for ALL is Chi-squared with 35 degree of freedom. LM statistic for MP, YP, UTS, UI, and DEI are Chi-squared with 7 degree of freedom. Bold entries denote each test statistic are significant at the 5 percent level.

3.7 Summary

In this paper, we proposed new APT framework composed of unobserved factors and observed macroeconomic factors based on state-space system. Employing our framework, we estimate risk exposure profiles and associated risk premiums, and extract unobserved factors simultaneously. In addition, we used simple arguments to choose a set of macroeconomic factors as sources of systematic asset risk. Macroeconomic factors we use are monthly and annual growth of industrial production, unexpected inflation, the change in expected inflation, and the spread between long and short interest. From the estimation result, we found that the large proportion of the expected excess return depends on the unobserved factor. Therefore we found that the contribution by the total one of macroeconomic factors are relatively decreasing.

Further, we examined the adequacy of macroeconomic factors as systematic variables based on Chiba [2007]'s framework. From the result of specification test, we found that three macroeconomic factors have significant influence on the stock return. They are monthly growth of industrial production, unexpected inflation, and the change in expected inflation. We also found that the number of macroeconomic factor which has a significant effect on the stock return are decreasing as time goes by. Therefore, we conclude that the effect of macroeconomic factors to stock returns is decreasing in recent years.

References

- Antoniou, A., Garrette, I., and Priestly, R. (1998), "Macroeconomic variables as common pervasive risk factors and the empirical content of the arbitrage pricing content of the arbitrage pricing theory," *Journal of Empirical Finance* 5, 221-240.

- Beenstock, M. and Chan, K. (1988), "Economic forces and London stock market," *Oxford Bulletin of Economics and Studies* 50, 27-29.

- Berry, M. A., Burmeister, E., and McElroy, M. B. (1988a), "Sorting Out Risks Using Known APT Factors," *Financial Analysis Journal* 29-43.

- Berry, M. A., Burmeister, E., and McElroy, M. B. (1988b), "A Practical Perspective on Evaluating Mutual Fund Risk," *Investment Management Review* 2(2), 78-86.

- Brown, S. J. and Weinstein, M. I. (1988), "New Approach to Testing Asset Pricing Theories: The Billinear Paradigm," *Journal of Finance* 63(3), 721-733.

- Burmeister, E. and McElroy, M. B. (1988), "Joint Estimation of Factor Sensitivities and and Risk Premia for the Arbitrage Pricing Theory," *Journal of Finance* 63(3), 721-733.

- Burmeister, E., Roll, R., and Ross, S. A., (1994), "A Practitioner's Guide to Arbitrage Pricing Theory," in *A Practitioner's Guide to Factor Model*, Charlottesville, VA: Research Foundation of the Institute of Chartered Financial Analysts.

- Burmeister, E. and Wall, K. D. (1986), "The Arbitrage Pricing Theory and Macroeconomic Factor Measures," *The Financial Review* 21, 1-20.
- Burmeister, E., Wall, K. D., and Hamilton, J. D. (1986), "Estimation of Unobserved Expected Monthly Inflation Using Kalman Filtering," *Journal of Business and Economic Statistics* 4, 147-160.
- Chen, N. (1983), "Some Empirical Tests of the Theory of Arbitrage Placing." *Journal of Finance* 38, 1392-1414.
- Chen, N. and Ingersoll, J. (1983), "Exact Pricing in Linear factor Models with Finitely Many Assets: A Note," *Journal of Finance* 38, 985-988.
- Chen, N., Roll, R., and Ross, S. A. (1986), "Economic Forces and the Stock Market," *Journal of Business* 59, 383-403.
- Connor, G. (1984), "Unified Beta Pricing Theory," *Journal of Economic Theory* 34, 13-31.
- Connor, G. and Korajczyk, R. A. (1989), "An Intertemporal Equilibrium Beta Pricing Model," *The Review of Financial Studies* 2(3), 373-392.
- Chiba, M. (2007), "Specification testing in dynamic factor models," *Manuscript.*
- Cox, J. C., Ingersoll, J. E., and Ross, S. A. (1985), "An Intertemporal General Equilibrium Model of Asset Prices," *Econometrica* 53, 363-384.
- Dybvig, P. H. (1983), "An Explicit Bound on Deviations from APT Pricing in a Finite Economy," *Journal of Financial Economics* 12, 483-496.
- Dybvic, P. H. and Ross, S. A. (1985), "Yes, the APT is testable," *Journal of finance* 40, 1173-1188.
- Elton, E., J., Gruber, M., J., *Modern Portfolio Theory and Investment Analysis*, fifth ed. Wiley.
- Fama, E. F. (1981) "Stock returns, real activity, inflation, and money," *American Economic Review* 71, 545-565.
- Fama, E. F. and French, K. R. (1992), "The Cross-Section of Expected Stock Returns," *Journal of Finance* 47(2), 427-465.
- Fama, E. F. and Gibbons, M. (1984), "A comparison of inflation forecasts," *Journal of Monetary Economics* 13, 327-348.
- Fama, E. F. and MacBeth, J. (1973), "Risk and Return: some empirical tests," *Journal of Political economy* 81, 607-636.
- Gibbons, M. R. (1982), "Multivariate tests of financial models: a new approach," *Journal of Econometrics* 3, 35-50.

- Huberman, G. (1982), "A simple Approach to Arbitrage Pricing Theory," *Journal of Economic Theory* 78, 183-191.
- Ingersoll, J. (1984), "Some Results in the Theory of Arbitrage Pricing," *Journal of Finance* 39, 1021-1039.
- Li, Y. (1998), "Expected stock returns, risk premiums and volatilities of economic factors," *Journal of Empirical Finance* 5, 69-97.
- McElroy, M. B. and Burmeister, E. (1988), "Arbitrage Pricing Theory as a restricted nonlinear multivariate regression model: interated nonlinear seemingly unrelated regressions estimates," *Journal of Business and Economic Statistics* 6, 29-42.
- McElroy, M. B., Burmeister, E., and Wall, K. D. (1985), "Two Estimators for the APT model when Factors are Measured," *Economic Letters* 19, 271-275.
- Priestly, R. (1996), "The Arbitrage Pricing Theory, macroeconomic and financial factors and the expectation generating process," *Journal of Banking and Finance* 20, 869-890.
- Roll, R. and Ross, S. A. (1980), "An Empirical Investigation of the Arbitrage Pricing Theory," *Journal of Finance* 35, 1073-1103.
- Ross, S. A. (1976), "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory* 13, 341-360.
- Ross, S. A. (1977), "Return, Risk, and Arbitrage," Risk and Return in Finance, Cambridge, MA: Baillinger, 189-219.
- Ross, S. A. (1978), "Mutual Fund Separation in Financial Theory-The Separating Distributions," *Journal of Economic Theory* 17, 254-286.
- Ross, S. A., and Roll, R. (1984), "The arbitrage Pricing Theory Approach to Strategic Portfolio Planning," *Financial Analysis Journal* 14-26.
- Shanken, J. (1992), "On the estimation of beta pricing models," *Review of Financial Studies* 5, 1-33.
- Sharpe, William F. (1977), "The Capital Asset Pricing Model: Multi-Beta Interpretation," *Financial Decision Making Under Uncertainty*, eds. Haim Levy and Marshall Sarnat, New York: Academic Press.
- Wei, K. C. John. (1988), "An Asset-Pricing Theory Unifying the CAPM and the APT," *Journal of Finance* 63(4), 881-892.
- Willem, T. (1997), "On stock market returns and monetary policy," *Journal of Finance* 2, 635-655.

Table 3.5: Calculation of expected excess return for six portfolios, April 1941-June 2000

Factor	Price of risk	Exposure	Contribution	Factor	Price of risk	Exposure	Contribution
SV				BV			
UC	1.123	5.812	6.527	UC	1.123	3.840	4.313
MP	3.521	0.472	1.663	MP	3.521	0.304	1.071
YP	0.888	0.037	0.033	YP	0.888	0.038	0.034
UTS	-8.293	0.060	-0.497	UTS	-8.293	-0.039	0.324
UI	-8.434	-0.898	7.570	UI	-8.434	-0.695	5.858
DEI	-4.478	3.233	-14.479	DEI	-4.478	2.431	-10.886
Return			0.817	Return			0.714
SN				BN			
UC	1.123	4.965	5.576	UC	1.123	3.585	4.027
MP	3.521	0.349	1.230	MP	3.521	0.322	1.134
YP	0.888	0.034	0.030	YP	0.888	0.027	0.024
UTS	-8.293	-0.022	0.181	UTS	-8.293	-0.045	0.375
UI	-8.434	-0.740	6.244	UI	-8.434	-0.376	3.167
DEI	-4.478	2.745	-12.292	DEI	-4.478	1.785	-7.995
Return			0.970	Return			0.732
SG				BG			
UC	1.123	5.236	5.880	UC	1.123	4.126	4.634
MP	3.521	0.312	1.097	MP	3.521	0.301	1.061
YP	0.888	0.032	0.029	YP	0.888	0.034	0.030
UTS	-8.293	-0.022	0.184	UTS	-8.293	-0.052	0.428
UI	-8.434	-0.655	5.527	UI	-8.434	-0.271	2.283
DEI	-4.478	2.577	-11.539	DEI	-4.478	1.671	-7.481
Return			1.178	Return			0.955

Table 3.6: Calculation of expected excess return for six portfolios, April 1941-December 1960

Factor	Price of risk	Exposure	Contribution	Factor	Price of risk	Exposure	Contribution
SV				BV			
UC	-0.627	4.420	-2.772	UC	-0.627	3.431	-2.151
MP	3.852	0.452	1.741	MP	3.852	0.355	1.369
YP	10.555	-0.033	-0.351	YP	10.555	-0.051	-0.539
UTS	-0.652	0.179	-0.116	UTS	-0.652	0.352	-0.230
UI	-1.205	-0.038	0.045	UI	-1.205	0.871	-1.049
DEI	-1.641	-1.512	2.481	DEI	-1.641	-2.365	3.882
Return			1.028	Return			1.281
SN				BN			
UC	-0.627	4.581	-2.873	UC	-0.627	3.542	-2.221
MP	3.852	0.366	1.408	MP	3.852	0.345	1.328
YP	10.555	-0.036	-0.383	YP	10.555	-0.037	-0.393
UTS	-0.652	0.036	-0.023	UTS	-0.652	0.293	-0.191
UI	-1.205	0.344	-0.415	UI	-1.205	1.161	-1.400
DEI	-1.641	-1.956	3.210	DEI	-1.641	-2.541	4.170
Return			0.924	Return			1.293
SG				BG			
UC	-0.627	5.707	-3.579	UC	-0.627	4.970	-3.117
MP	3.852	0.390	1.503	MP	3.852	0.390	1.500
YP	10.555	-0.048	-0.508	YP	10.555	-0.059	-0.622
UTS	-0.652	-0.259	0.169	UTS	-0.652	-0.377	0.245
UI	-1.205	0.470	-0.566	UI	-1.205	0.147	-0.178
DEI	-1.641	-2.272	3.728	DEI	-1.641	-1.701	2.792
Return			0.747	Return			0.621

Table 3.7: Calculation of expected excess return for six portfolios, January 1961-September 1980

Factor	Price of risk	Exposure	Contribution	Factor	Price of risk	Exposure	Contribution
SV				BV			
UC	3.026	6.967	21.085	UC	3.026	4.044	12.238
MP	3.258	-2.518	-8.203	MP	3.258	-1.379	-4.491
YP	8.790	-0.187	-1.641	YP	8.790	-0.101	-0.888
UTS	-6.594	1.505	-9.924	UTS	-6.594	0.924	-6.093
UI	-0.482	0.035	-0.017	UI	-0.482	-0.026	0.012
DEI	-0.246	2.289	-0.562	DEI	-0.246	1.904	-0.468
Return			0.738	Return			0.311
SN				BN			
UC	3.026	5.956	18.025	UC	3.026	3.793	11.480
MP	3.258	-2.145	-6.988	MP	3.258	-1.609	-5.243
YP	8.790	-0.140	-1.233	YP	8.790	-0.041	-0.364
UTS	-6.594	1.281	-8.444	UTS	-6.594	0.741	-4.884
UI	-0.482	-0.160	0.077	UI	-0.482	0.103	-0.050
DEI	-0.246	1.851	-0.454	DEI	-0.246	0.666	-0.163
Return			0.983	Return			0.776
SG				BG			
UC	3.026	6.184	18.714	UC	3.026	4.610	13.951
MP	3.258	-2.252	-7.336	MP	3.258	-1.835	-5.978
YP	8.790	-0.140	-1.231	YP	8.790	-0.061	-0.534
UTS	-6.594	1.325	-8.740	UTS	-6.594	0.967	-6.374
UI	-0.482	-0.388	0.187	UI	-0.482	-0.051	0.024
DEI	-0.246	1.321	-0.324	DEI	-0.246	0.239	-0.059
Return			1.270	Return			1.031

Table 3.8: Calculation of expected excess return for six portfolios, October 1980-June 2000

Factor	Price of risk	Exposure	Contribution	Factor	Price of risk	Exposure	Contribution
SV				BV			
UC	0.576	-6.229	-3.589	UC	0.576	-4.283	-2.468
MP	1.111	0.868	0.964	MP	1.111	1.022	1.135
YP	0.660	0.142	0.094	YP	0.660	0.160	0.105
UTS	-1.713	-0.138	0.236	UTS	-1.713	-0.259	0.444
UI	-0.821	0.507	-0.416	UI	-0.821	0.785	-0.644
DEI	-1.372	-2.379	3.263	DEI	-1.372	-1.786	2.451
Return			0.552	Return			1.024
SN				BN			
UC	0.576	-4.605	-2.653	UC	0.576	-3.681	-2.121
MP	1.111	0.615	0.684	MP	1.111	0.840	0.933
YP	0.660	0.051	0.033	YP	0.660	0.096	0.064
UTS	-1.713	-0.067	0.115	UTS	-1.713	-0.018	0.031
UI	-0.821	0.194	-0.159	UI	-0.821	0.192	-0.157
DEI	-1.372	-2.194	3.010	DEI	-1.372	-1.289	1.769
Return			1.030	Return			0.518
SG				BG			
UC	0.576	-4.202	-2.421	UC	0.576	-3.143	-1.811
MP	1.111	0.565	0.628	MP	1.111	0.521	0.579
YP	0.660	0.011	0.007	YP	0.660	0.047	0.031
UTS	-1.713	0.013	-0.022	UTS	-1.713	-0.074	0.127
UI	-0.821	-0.031	0.025	UI	-0.821	-0.184	0.151
DEI	-1.372	-2.057	2.822	DEI	-1.372	-1.318	1.808
Return			1.039	Return			0.886

Table 3.9: Specification testing for omitted macroeconomic factors

Symbol	LM statistic	P-value	Symbol	LM statistic	P-value
A: 1941.04-2000.06			B: 1941.04-1960.12		
ALL	**134.586**	0.000	ALL	**510.835**	0.000
MP	**12.592**	0.201	MP	**13.156**	0.041
YP	5.189	0.520	YP	**16.760**	0.010
UTS	6.611	0.358	UTS	**24.712**	0.000
UI	**67.799**	0.000	UI	**237.506**	0.000
DEI	**31.093**	0.000	DEI	**154.144**	0.000
Symbol	LM statistic	P-value	Symbol	LM statistic	P-value
B: 1961.01-1980.09			D: 1980.10-2000.06		
ALL	**236.904**	0.000	ALL	**72.398**	0.000
MP	**44.240**	0.000	MP	11.418	0.076
YP	5.210	0.517	YP	7.341	0.290
UTS	6.060	0.417	UTS	4.875	0.560
UI	**35.567**	0.000	UI	17.807	0.007
DEI	**19.411**	0.004	DEI	**18.012**	0.006

Chapter 4

The Macroeconomy and the Yield Curve: Specification Testing based on Lagrange Multiplier Approach

4.1 Introduction

Understanding the term structure of interest rate has been a topic on the agenda of both financial economists and macroeconomists, albeit for different reasons. Therefore, they all have attempted to build good and tractable model of the yield curve, yet unusually large gap is apparent between the yield curve model developed by financial economists and the model developed by macroeconomists. Financial economists have mainly and pricing interest rate-related securities: They have developed powerful models based on the assumption of absence of opportunities, but typically left unspecified the relationship between the term structure and other economic variables. Macroeconomists, on the other hand, have focused on understanding the relationship between interest rates, monetary policy, and macroeconomic fundamentals. Thus, combining these two lines of research seems fruitful, in that there are potential gains both ways.

Recently, a number of papers take a step toward bridging this gap by formulating and estimating an yield curve model that integrates macroeconomic and financial factors. For example, Ang and Piazzesi [2003], Hördahl et al. [2002], and Wu [2002], who explicitly incorporate macro determinants into multi-factor yield curve models. However, those papers only consider a unidirectional macro linkage, because output and inflation are assumed to be determined independently of the shape of the yield curve, but not vice versa. In contrast to this assumption of a one-way macro-to-yields link, the opposite view is taken in another large literature typified by Estrella and Hardouvelis [1991] and Estrella and Mishkin [1998], which assumes yields-to-macro link and focuses only on the unidirectional predictive power of the yield curve for the economy. These literatures, however, forcus only one-way yields-to-macro or macro-to-yields links, so the interactions between macroeconomic and term-structure dynamics have also been left unexplored.

In order to redress these shortcomings, Diebold et al. [2006] constructs a dynamic term structure model entirely based on macroeconomic factors, which allows for an explicit feedback from the yield curve to macroeconomic outcomes. The joint modelling of these key macroeconomic factors - namely, capacity utilization, the federal funds rate, and annual price inflation - should allow us to obtain a more accurate (endogenous) description of the dynamics of the yield curve. Therefore, we employ Diebold's framewok in this paper for analyzing the potential bidirectional feedback from the yield curve to the economy and back again[1].

Diebold et al. [2006] estimates many parameters by numerical optimization and examine the interactions between the macroeconomy and the yield curve based on Likelihood Ratio and Wald methods. However, this approach needs high numerical task. The Lagrange Multiplier procedure is regarded as the most suitable for that situation because, unlike Likelihood Ratio and Wald procedure, it does not require the estimation of the alternative. Therefore, this paper attempts to ease this difficulty, by proposing formal test of macro and yield curve inteactions based on Chiba [2007]'s testing framework.

The rest of the paper is organized as follows. Section 4.2 describes the main features of Diebold's framework. Section 4.3 describing the data. Section 4.4 estimates the model and examines causal direction between the macroeconomy and the yield curve. Section 4.5 briefly summarizes this study.

4.2 The model

In this section, we first review the dynamic Nelson-Siegel model in subsection 4.2.1 and introduce a detailed description of the Diebold et al. [2006]'s framework in subsection 4.2.2.

4.2.1 The dynamic Nelson-Siegel model

The original Nelson-Siegel model fits the yield curve with the simple functional form

$$y_t(\tau) = \beta_1 + \beta_2 \left[\frac{1-\exp(-\lambda\tau)}{\lambda\tau}\right] + \beta_3 \left[\frac{1-\exp(-\lambda\tau)}{\lambda\tau} - \exp(-\lambda\tau)\right] \tag{4.2.1}$$

for $t = 1, 2, \ldots, T$. Here, $y_t(\tau)$ is the zero-coupon yield with τ dates to maturity at time t, and $\beta_1, \beta_2, \beta_3$, and λ are parameters. Note that the parameter λ governs the exponential rate; small value of λ produces fast decay and can better fit the curve at *long* maturities, while large value of λ produces slow decay and can better fit the curve at *short* maturities.

This representation is commonly used by financial researchers and market practitioners to fit the yield curve at a point in time. Although for some purposes such a static representation appears useful, a dynamic version is required to understand the evolution of the bond market over time. Therefore, Diebold and Li [2006] reinterpreted the coefficients β_1, β_2, and β_3 as time-varying level, slope, and curvature factors and the terms that multiply these these

[1] *Note*: Diebold's framewok is not the usual no-arbitrage factor representation typically used in the finance literature. Such no arbitrage factor models often appear to fit the cross-section of yields at a particular point in time, but they do less well in describing the dynamics of the yield curve over time.

factors are factor loadings. Thus, we write

$$y_t(\tau) = L_t + S_t\left[\frac{1-\exp(-\lambda\tau)}{\lambda\tau}\right] + C_t\left[\frac{1-\exp(-\lambda\tau)}{\lambda\tau} - \exp(-\lambda\tau)\right], \tag{4.2.2}$$

for $t = 1, 2, \ldots, T$. Diebold and Li [2006] assumes an autoregressive stucture, which yields the DNS curve. This model is very flexible and capable of accommodating several stylized facts on the term structure and its dynamics.

4.2.2 An yield curve with macroeconomic factors

We wish to characterize the dynamic interactions between the yield curve and the macroeconomy. Therefore, we need to formulate the expanded yield model that integrates macroeconomic and financial factors. Our measures of the activity include three key variables: manufacturing capacity utilization (CU_t), the federal funds rate (FFR_t), and annual price inflation ($INFR_t$). These three variables represent the level of real economic ativity relative to potential, the monetary policy instrument, and the inflation rate, which are widely considered to be minimum set of fundamentals to capture basic macroeconomic dynamics.

If the dynamic movement of L_t, S_t, and C_t follow a vector autoregressive process of first order, the yield curve model forms a *state-space model*. Therefore, the yield curve employed here is as follows.

$$\boldsymbol{y}_t = \begin{bmatrix} \boldsymbol{\Phi} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_t \\ \boldsymbol{x}_t \end{bmatrix} + \boldsymbol{u}_t, \tag{4.2.3}$$

$$\begin{bmatrix} \boldsymbol{\beta}_t \\ \boldsymbol{x}_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Gamma}_{11} & \boldsymbol{\Gamma}_{12} \\ \boldsymbol{\Gamma}_{21} & \boldsymbol{\Gamma}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_{t-1} \\ \boldsymbol{x}_{t-1} \end{bmatrix} + \begin{bmatrix} \boldsymbol{v}_{1t} \\ \boldsymbol{v}_{2t} \end{bmatrix}, \tag{4.2.4}$$

where

$$\boldsymbol{y}_t \equiv \begin{bmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \vdots \\ y_t(\tau_N) \end{bmatrix}, \boldsymbol{\Phi} \equiv \begin{bmatrix} 1 & \frac{1-\exp(-\lambda\tau_1)}{\lambda\tau_1} & \frac{1-\exp(-\lambda\tau_1)}{\lambda\tau_1} - \exp(-\lambda\tau_1) \\ 1 & \frac{1-\exp(-\lambda\tau_2)}{\lambda\tau_2} & \frac{1-\exp(-\lambda\tau_2)}{\lambda\tau_2} - \exp(-\lambda\tau_2) \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-\exp(-\lambda\tau_N)}{\lambda\tau_N} & \frac{1-\exp(-\lambda\tau_N)}{\lambda\tau_N} - \exp(-\lambda\tau_N) \end{bmatrix},$$

$$\boldsymbol{\beta}_t \equiv \begin{bmatrix} L_t \\ S_t \\ C_t \end{bmatrix}, \boldsymbol{x}_t \equiv \begin{bmatrix} CU_t \\ FFR_t \\ INFR_t \end{bmatrix}, \boldsymbol{u}_t \equiv \begin{bmatrix} u_{1t} \\ u_{2t} \\ \vdots \\ u_{Nt} \end{bmatrix}, \boldsymbol{\mu}_t \equiv \begin{bmatrix} \underset{(3\times 1)}{\boldsymbol{\mu}_{1t}} \\ \underset{(3\times 1)}{\boldsymbol{\mu}_{2t}} \end{bmatrix} = \begin{bmatrix} \mu_{1t} \\ \mu_{2t} \\ \mu_{3t} \\ \mu_{4t} \\ \mu_{5t} \\ \mu_{6t} \end{bmatrix},$$

$$\boldsymbol{\Gamma} \equiv \begin{bmatrix} \underset{(3\times3)}{\boldsymbol{\Gamma}_{11}} & \underset{(3\times3)}{\boldsymbol{\Gamma}_{12}} \\ \underset{(3\times3)}{\boldsymbol{\Gamma}_{21}} & \underset{(3\times3)}{\boldsymbol{\Gamma}_{22}} \end{bmatrix} = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} & \gamma_{14} & \gamma_{15} & \gamma_{16} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} & \gamma_{24} & \gamma_{25} & \gamma_{26} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} & \gamma_{34} & \gamma_{35} & \gamma_{36} \\ \gamma_{41} & \gamma_{42} & \gamma_{43} & \gamma_{44} & \gamma_{45} & \gamma_{46} \\ \gamma_{51} & \gamma_{52} & \gamma_{53} & \gamma_{54} & \gamma_{55} & \gamma_{56} \\ \gamma_{61} & \gamma_{62} & \gamma_{63} & \gamma_{64} & \gamma_{65} & \gamma_{66} \end{bmatrix}, \boldsymbol{v}_t \equiv \begin{bmatrix} \underset{(3\times1)}{\boldsymbol{v}_{1t}} \\ \underset{(3\times1)}{\boldsymbol{v}_{2t}} \end{bmatrix} = \begin{bmatrix} v_{1t} \\ v_{2t} \\ v_{3t} \\ v_{4t} \\ v_{5t} \\ v_{6t} \end{bmatrix}.$$

In this framework, equation (4.2.3) is an *observation equation*, and equation (4.2.4) is a *transition equation*. Here, $\boldsymbol{y}_t$ is the vector of zero-coupon yield, $\boldsymbol{\beta}_t$ is the vector of time-varying latent factors, and $\boldsymbol{x}_t$ is the observed macroeconomic factors. For linear least-square optimality, we require that the Gaussian white noise in observation and transition disturbances are orthogonal to each other and to the initial state:

$$\begin{bmatrix} \boldsymbol{u}_t \\ \boldsymbol{v}_t \end{bmatrix} \sim \text{i.i.d.}N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega} \end{bmatrix} \right), \tag{4.2.5}$$

$$E(\boldsymbol{\xi}_1 \boldsymbol{u}_t') = \mathbf{0}, E(\boldsymbol{\xi}_1 \boldsymbol{v}_t') = \mathbf{0}. \tag{4.2.6}$$

where $\boldsymbol{\xi}_t \equiv (\boldsymbol{\beta}_t', \boldsymbol{x}_t')'$. Furthermore, we assume that $\boldsymbol{\Sigma}$ is an identity matrix and $\boldsymbol{\Omega}$ is a potitive definite symmetric matrix for compurational tractability:

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}, \tag{4.2.6}$$

$$\boldsymbol{\Omega} \equiv \begin{bmatrix} \underset{(3\times3)}{\boldsymbol{\Omega}_{11}} & \underset{(3\times3)}{\boldsymbol{\Omega}_{12}} \\ \underset{(3\times3)}{\boldsymbol{\Omega}_{12}'} & \underset{(3\times3)}{\boldsymbol{\Omega}_{22}} \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} & \sigma_{16} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} & \sigma_{24} & \sigma_{25} & \sigma_{26} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} & \sigma_{34} & \sigma_{35} & \sigma_{36} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_{44} & \sigma_{45} & \sigma_{46} \\ \sigma_{15} & \sigma_{25} & \sigma_{35} & \sigma_{54} & \sigma_{55} & \sigma_{56} \\ \sigma_{16} & \sigma_{26} & \sigma_{36} & \sigma_{64} & \sigma_{65} & \sigma_{66} \end{bmatrix}. \tag{4.2.7}$$

According to the above parameter configuration, we proceed to evaluate the Gaussian likelihood function of the model using the prediction-error decomposition of the likelihood. However, we must estimate many parameters in this model: The measurement matrix $\boldsymbol{\Phi}$ contains one free parameter, λ. the transition matrix $\boldsymbol{\Gamma}$ contains 36 free parameters, the mean state vector $\boldsymbol{\mu}$ contains 6 free parameters. Moreover, the trasition covariance matrix $\boldsymbol{\Omega}$ contains 21 free parameters. All told, then, 64 parameters must be estimated by numerical optimization. Therefore, the parameter space is quite large and the optimization problem of maximization problem the likelihood function is non-trivial and time consuming. Then, we propose formal test of macro and yield curve inteactions based on Chiba [2007]'s framework. This framework does not require the estimation of the alternative hypothesis and constructs the test statistic from a single pass of the Kalman filter and a smoother. Therefore, our approach eases numerical task for analyzing the potential bidirectional feedback between the yield curve and the macroeconomy.

4.3 Statistical characteristics of the data

This section intorduces the data we use in this paper and discusses their statistical characteristics. We use U.S. Treasury yields with maturities of 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108, and 120 months. The yields are derived from bid-ask average price quotes, from January 1970 through December 2000, taken from Diebold's website. They are measured as of the beggining of each month, where a month is defined as 30.4375 days.

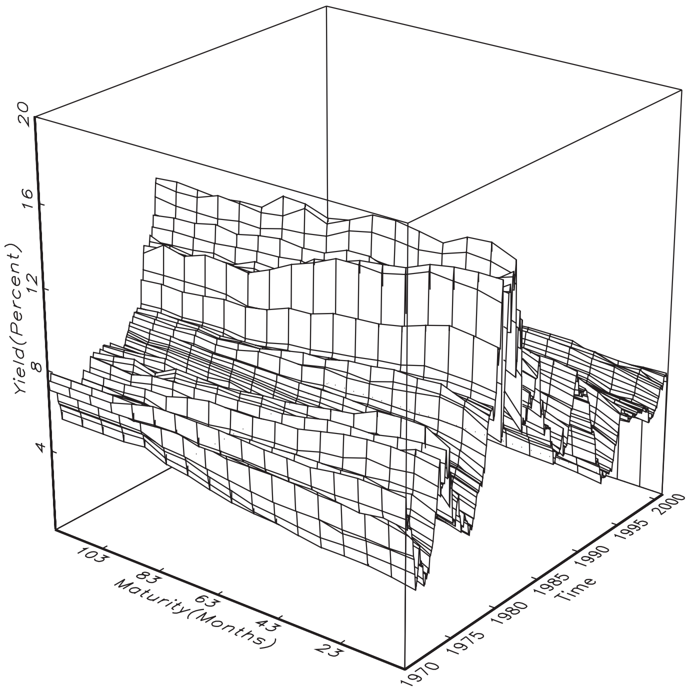

Figure 4.1: Yield Curves, January 1970-December 2000

In Figure 4.1 we provide a three-dimentional plot of our yield curve data. We find that the variation of the yield curve is quite similar. In Table 4.1[2], we present descriptive statistics for the yields. It is clear that the long rate yield curve are less volatile and more persistent than the short rate yield curve from first and second moments.

In Figure 4.2, we plot three estimated factors and macroeconomic factors as a function of time. We find that the level factor is very persistent except from 1980 through 1985 and is of cource positive - in the neighborhood of 8 percent. In contrast, the slope and curvature are volatile and assume both positive and negative values.

In Table 4.2[3], we present descriptive statistics of estimated level, slope, curvature fac-

[2] *Note*: The JB test statistic is Chi-squared with 2 degree of freedom and the LB test statistic is Chi-squared with 12 degree of freedom. Bold entries denote statistical significance at the 5 percent level.

[3] *Note*: The JB test statistic is Chi-squared with 2 degree of freedom and the LB test statistic is Chi-squared with 12 degree of freedom. Bold entries denote statistical significance at the 5 percent level.

Table 4.1: Descriptive statistics of yield curves

Yield	Mean	Med	Min	Max	SD	Skew	Kurt	JB	$LB(12)$
$M3$	6.755	5.928	2.732	16.020	2.655	1.265	4.642	**119.421**	**340.216**
$M6$	6.983	6.243	2.891	16.481	2.662	1.205	4.385	**119.312**	**343.327**
$M9$	7.105	6.404	2.984	16.394	2.640	1.165	4.254	**119.239**	**345.367**
$M12$	7.201	6.612	3.107	15.822	2.569	1.114	4.063	**119.215**	**347.351**
$M15$	7.306	6.752	3.288	16.043	2.517	1.095	3.992	**119.201**	**349.378**
$M18$	7.378	6.780	3.482	16.229	2.502	1.111	4.004	**119.172**	**350.633**
$M21$	7.441	6.807	3.638	16.177	2.490	1.123	3.996	**119.164**	**351.589**
$M24$	7.459	6.811	3.777	15.650	2.443	1.106	3.893	**119.236**	**352.159**
$M30$	7.552	6.928	4.043	15.397	2.365	1.081	3.747	**119.386**	**354.348**
$M36$	7.631	7.060	4.204	15.765	2.341	1.122	3.839	**119.356**	**355.468**
$M48$	7.769	7.220	4.308	15.821	2.284	1.117	3.750	**119.520**	**357.179**
$M60$	7.841	7.365	4.347	15.005	2.248	1.086	3.570	**119.695**	**359.000**
$M72$	7.957	7.425	4.384	14.979	2.222	1.085	3.512	**119.812**	**360.429**
$M84$	7.987	7.454	4.352	14.975	2.182	1.094	3.577	**119.875**	**359.963**
$M96$	8.046	7.506	4.433	14.936	2.171	1.054	3.412	**120.010**	**361.827**
$M108$	8.078	7.541	4.429	15.018	2.180	1.064	3.485	**119.981**	**362.131**
$M120$	8.047	7.589	4.443	14.925	2.135	1.066	3.578	**120.031**	**361.349**

Table 4.2: Descriptive statistics of estimated level, slope, curvature factors, and macroeconomic factors

Factor	Mean	Med	Min	Max	SD	Skew	Kurt	JB	$LB(12)$
L	8.196	7.848	4.542	14.385	0.623	0.926	3.268	**120.382**	**364.522**
S	-1.512	-1.643	-5.347	2.908	1.773	0.263	2.579	**112.015**	**224.391**
C	0.434	0.476	-1.280	1.778	2.106	-0.619	3.858	**67.339**	**244.249**
CU	80.405	80.872	68.494	88.271	3.728	-0.654	3.626	**123.877**	**361.486**
FFR	7.367	6.510	2.920	19.100	3.193	1.342	5.034	**120.789**	**331.183**
$INFR$	5.160	4.165	1.100	14.760	3.103	1.211	3.660	**138.705**	**321.127**

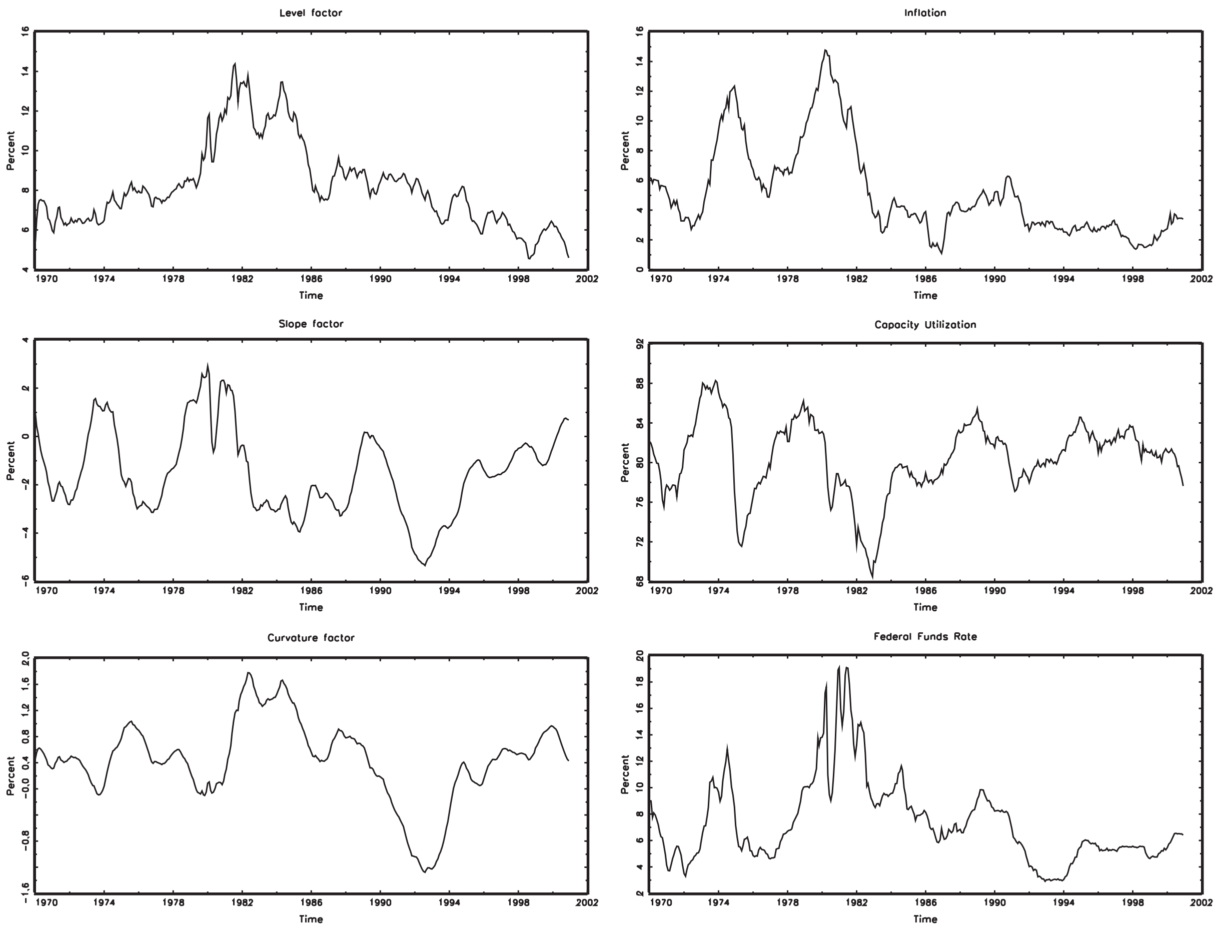

Figure 4.2: Estimates of level, slope, curvature factors, and macroeconomic factors

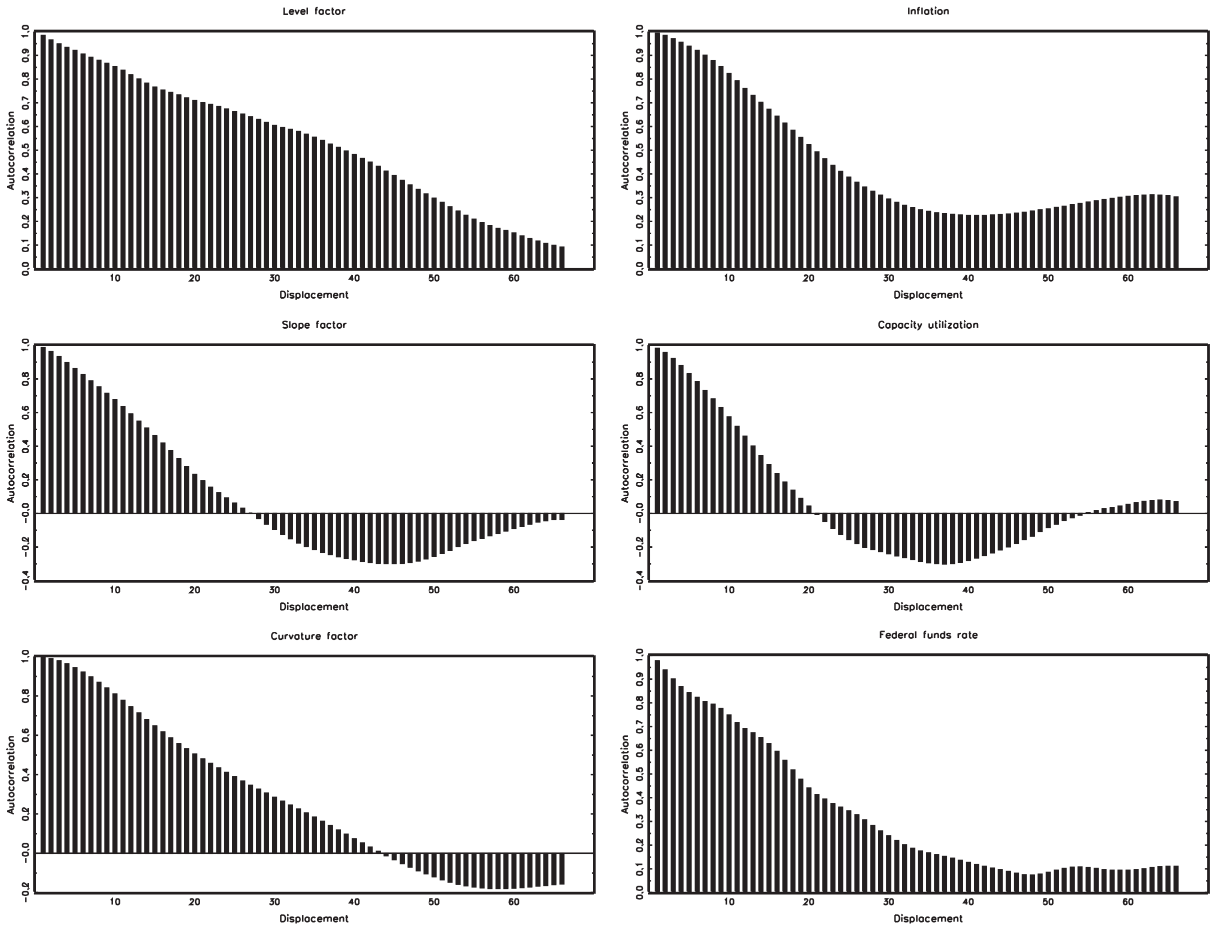

Figure 4.3: Autocorrelation of estimated level, slope, curvature factors, and macroeconomic factors

Table 4.3: Autocorrelation of estimated level, slope, curvature factors, and macroeconomic factors

Factor	$\hat{\rho}(1)$	$\hat{\rho}(6)$	$\hat{\rho}(12)$	$\hat{\rho}(18)$	$\hat{\rho}(24)$	$\hat{\rho}(30)$
L	0.984	0.905	0.818	0.733	0.674	0.604
S	0.986	0.824	0.591	0.325	0.091	-0.093
C	0.996	0.921	0.745	0.558	0.411	0.284
CU	0.983	0.782	0.458	0.137	-0.124	-0.240
FFR	0.977	0.822	0.691	0.517	0.359	0.239
$INFR$	0.993	0.920	0.759	0.584	0.410	0.293
Factor	$\hat{\rho}(36)$	$\hat{\rho}(42)$	$\hat{\rho}(48)$	$\hat{\rho}(54)$	$\hat{\rho}(60)$	$\hat{\rho}(66)$
L	0.540	0.449	0.334	0.226	0.151	0.092
S	-0.231	-0.291	-0.281	-0.178	-0.089	-0.033
C	0.162	0.030	-0.088	-0.165	-0.176	-0.155
CU	-0.298	-0.250	-0.133	-0.009	0.053	0.069
FFR	0.160	0.110	0.074	0.108	0.094	0.110
$INFR$	0.236	0.225	0.243	0.275	0.305	0.303

tors, and macroeconnomic variables. The estimates indicate that persistence decreases, and volatility increases, as we move the level factor through the curvature factor.

In Table 4.3 and Figure 4.3 we present the autocorrelations of estimated level, slope, curvature factors, and macroeconnomic variables. We can see that all variables are highly persistent, and that the level factor is the most persistent, and that the slope factor is more persistent than the curvature factor.

Table 4.4: Correlation matrix for estimated level, slope, curvature factors, and macroeconomic factors

Factor	L	S	C	CU	FFR	$INFR$
L	1.000					
S	-0.011	1.000				
C	0.414	0.147	1.000			
CU	-0.056	0.381	-0.342	1.000		
FFR	0.117	0.106	0.364	-0.112	1.000	
$INFR$	0.397	0.069	0.034	-0.037	0.677	1.000

In Table 4.4 we provide the correlation matrix for estimated level, slope, curvature factors, and macroeconnomic variables. We find that level, slope, and curvature are not highly correlated with each other; all pairwise correlations are less than 0.414. The correlation 0.397 between the level factor and actual inflation is consistent with a link between the level of the yield curve and inflationary expectations, as suggested by the Fisher equation. In addition, the correlation 0.381 between the slope factor and capacity utilization suggest that yield curve slope is intimately connected to the cyclical dynamics of the economy.

4.4 Main results

4.4.1 Estimation results

In this subsection, we propose estimation results of the model under the several null hypothesis. In Table 4.5 and 4.6[4], we present estimation results of $\mathbf{\Gamma}$ and $\mathbf{\Omega}$ matrices.

The diagonal elements of $\mathbf{\Gamma}_{11}$ matrix indicates highly persistent own dynamics of L_t, S_t, and C_t under all null hypotheses. Wheras, many of the off diagonal elements of $\mathbf{\Gamma}_{11}$ indicate relative low values. Thus, cross factor dynamics appear unimportant.

In addition, $\mathbf{\Gamma}_{12}$ and $\mathbf{\Gamma}_{21}$ matrices indicate that many of the off diagonal elements appear significant. In particular, estimates of $\mathbf{\Gamma}_{21}$ indicate relative high values. And, the significance of γ_{16} and γ_{24} suggests that links from yields to macroeconomy strongly exist as discussed in section 4.2.

The estimated $\mathbf{\Omega}$ matrix is provided in Table 4.6. The estimates of $\mathbf{\Omega}_{11}$ suggest that volatility shock increase as we move from L_t to S_t to C_t. Several of the off diagonal covariances appear significant individually.

4.4.2 Specification tests

This subsection constructs specification tests and derives their results. There are two links from yields to the macroeconomy in our setup: the contemporaneous link given by $\mathbf{\Omega}_{12}$, and the effects of lagged yields on macroeconomy embodied in $\mathbf{\Gamma}_{21}$. Conversely, links from the macroeconomy to yields are embodied in $\mathbf{\Gamma}_{12}$. Therefore, we construct the null hypotheses for analyzing a bidirectional link between the macroeconomy and the yield curve as follows.

Specification test for no links from the macroeconomy to yields

The latent factor represented by $\boldsymbol{\beta}_t$ is called "block exogenous" in the time series with respect to the lagged macro factor $\boldsymbol{x}_{t-1}$ if the elements in $\boldsymbol{x}_{t-1}$ are of no help improving a forecast of any variable contained in $\boldsymbol{\beta}_t$ alone. In our setup, $\boldsymbol{\beta}_t$ is block-exogenous when $\mathbf{\Gamma}_{12} = \mathbf{\Omega}_{12} = \mathbf{0}$, and there is no links from the macroeconomy to yields. Therefore, we conduct the LM test of the null hypothesis as follows:

$$H_0 : \mathbf{\Gamma}_{12} = \mathbf{\Omega}_{12} = \mathbf{0}. \tag{4.4.1}$$

Specification test for no links from yields to the macroeconomy

The latent factor represented by $\boldsymbol{x}_t$ is called "block exogenous" in the time series with respect to the lagged macro factor $\boldsymbol{\beta}_{t-1}$ if the elements in $\boldsymbol{\beta}_{t-1}$ are of no help improving a forecast of any variable contained in $\boldsymbol{x}_t$ alone. In our setup, $\boldsymbol{x}_t$ is block-exogenous when $\mathbf{\Gamma}_{21} = \mathbf{\Omega}_{12} = \mathbf{0}$,

[4] *Note*: Bold entries denote parameter estimates significant at the 5 percent level. Standard errors appear in parentheses.

Table 4.5: Estimated $\mathbf{\Gamma}$ matrix under the several null hypotheses

A: Diagonal	L_{t-1}	S_{t-1}	C_{t-1}	CU_{t-1}	FFR_{t-1}	$INFR_{t-1}$
L_t	**0.995**	0.023	− 0.022	− 0.005	− 0.013	**0.015**
	(0.018)	(0.021)	(0.032)	(0.005)	(0.015)	(0.008)
S_t	**0.101**	**1.090**	0.017	**0.015**	− **0.103**	0.001
	(0.013)	(0.015)	(0.022)	(0.004)	(0.010)	(0.005)
C_t	− **0.009**	− 0.006	**0.995**	− 0.001	**0.007**	0.002
	(0.003)	(0.004)	(0.006)	(0.001)	(0.003)	(0.001)
CU_t	**0.267**	**0.222**	− 0.049	**0.987**	− **0.222**	− **0.052**
	(0.037)	(0.041)	(0.060)	(0.012)	(0.029)	(0.015)
FFR_t	**0.501**	**0.550**	− **0.188**	0.009	**0.589**	− **0.030**
	(0.031)	(0.034)	(0.051)	(0.009)	(0.024)	(0.013)
$INFR_t$	0.029	0.027	− 0.016	**0.043**	− 0.004	**0.982**
	(0.020)	(0.023)	(0.037)	(0.006)	(0.016)	(0.009)
B: No intersection	L_{t-1}	S_{t-1}	C_{t-1}	CU_{t-1}	FFR_{t-1}	$INFR_{t-1}$
L_t	**0.996**	**0.021**	− 0.032			
	(0.008)	(0.009)	(0.027)			
S_t	− **0.023**	**0.990**	0.010			
	(0.006)	(0.007)	(0.022)			
C_t	0.001	**0.003**	**0.984**			
	(0.001)	(0.002)	(0.005)			
CU_t				**0.980**	− **0.033**	− **0.040**
				(0.009)	(0.014)	(0.014)
FFR_t				**0.029**	**0.968**	0.002
				(0.009)	(0.015)	(0.015)
$INFR_t$				**0.043**	**0.016**	**0.985**
				(0.004)	(0.008)	(0.008)
C: No macro to yield	L_{t-1}	S_{t-1}	C_{t-1}	CU_{t-1}	FFR_{t-1}	$INFR_{t-1}$
L_t	**0.996**	**0.021**	− 0.032			
	(0.008)	(0.009)	(0.027)			
S_t	− **0.023**	**0.990**	0.010			
	(0.006)	(0.007)	(0.022)			
C_t	0.001	**0.003**	**0.994**			
	(0.001)	(0.002)	(0.005)			
CU_t	**0.267**	**0.222**	− 0.049	**0.987**	**-0.222**	− **0.052**
	(0.037)	(0.042)	(0.062)	(0.011)	(0.030)	(0.015)
FFR_t	**0.501**	**0.550**	− **0.188**	0.009	**0.589**	− **0.030**
	(0.031)	(0.035)	(0.052)	(0.009)	(0.025)	(0.013)
$INFR_t$	0.029	0.027	− 0.016	**0.043**	− 0.004	0.982
	(0.021)	(0.023)	(0.036)	(0.006)	(0.016)	(0.009)
D: No yield to macro	L_{t-1}	S_{t-1}	C_{t-1}	CU_{t-1}	FFR_{t-1}	$INFR_{t-1}$
L_t	**0.995**	0.023	− 0.022	− 0.005	− 0.013	**0.015**
	(0.018)	(0.021)	(0.030)	(0.005)	(0.015)	(0.008)
S_t	**0.101**	**1.090**	0.017	**0.015**	− **0.103**	**0.001**
	(0.013)	(0.014)	(0.021)	(0.004)	(0.010)	(0.005)
C_t	− **0.009**	− 0.006	**0.995**	− 0.001	**0.007**	0.002
	(0.003)	(0.004)	(0.005)	(0.001)	(0.003)	(0.001)
CU_t				**0.980**	− **0.033**	− **0.040**
				(0.009)	(0.014)	(0.014)
FFR_t				**0.029**	**0.968**	**0.020**
				(0.009)	(0.014)	(0.015)
$INFR_t$				**0.043**	**0.016**	**0.985**
				(0.004)	(0.008)	(0.008)

Table 4.6: Estimated $\boldsymbol{\Omega}$ matrix under the several null hypotheses

A: Diagonal	L	S	C	CU	FFR	$INFR$
$\boldsymbol{\Omega}_{ii}$	**0.003** (0.000)	**0.041** (0.003)	**0.085** (0.006)	**0.352** (0.026)	**0.248** (0.018)	**0.104** (0.008)

B: No intersection	L	S	C	CU	FFR	$INFR$
L	**0.003** (0.000)					
S	**0.031** (0.004)	**0.055** (0.004)				
C	**0.010** (0.001)	0.001 (0.001)	**0.087** (0.006)			
CU				**0.402** (0.029)		
FFR				**0.143** (0.023)	**0.447** (0.033)	
$INFR$				**0.027** (0.011)	**0.026** (0.011)	**0.104** (0.008)

C: No macro to yield	L	S	C	CU	FFR	$INFR$
L	**0.003** (0.000)					
S	**0.031** (0.004)	**0.055** (0.004)				
C	**0.010** (0.001)	0.001 (0.001)	**0.087** (0.006)			
CU				**0.352** (0.026)		
FFR				**0.049** (0.016)	**0.248** (0.018)	
$INFR$				**0.022** (0.010)	0.015 (0.009)	**0.104** (0.008)

D: No yield to macro	L	S	C	CU	FFR	$INFR$
L	**0.003** (0.000)					
S	**0.031** (0.004)	**0.041** (0.003)				
C	**0.010** (0.001)	**0.002** (0.001)	**0.085** (0.006)			
CU				**0.352** (0.026)		
FFR				**0.049** (0.016)	**0.447** (0.034)	
$INFR$				**0.022** (0.010)	**0.026** (0.011)	**0.104** (0.008)

and there is no links from yields to the macroeconomy. Therefore, we conduct the LM test of the null hypothesis as follows:

$$H_0 : \mathbf{\Gamma}_{21} = \mathbf{\Omega}_{12} = \mathbf{0}. \tag{4.4.2}$$

Specification test for no links between yields and the macroeconomy

Further, in our setup, there is no relation between $\boldsymbol{\beta}$ and $\boldsymbol{x}$ when $\mathbf{\Gamma}_{12} = \mathbf{\Gamma}_{21} = \mathbf{\Omega}_{12} = \mathbf{0}$, and there is no link between the macroeconomy and yields. Therefore, we conduct the LM test of the null hypothesis as follows:

$$H_0 : \mathbf{\Gamma}_{12} = \mathbf{\Gamma}_{21} = \mathbf{\Omega}_{12} = \mathbf{0}. \tag{4.4.3}$$

We report in Table 4.7[5] the results of several specification tests based on LM procedure. All four tests overwhelmingly reject the null hypotheses. Therefore, we conclude that there is clear evidence in favor of a bidirectional link between the macroeconomy and the yield curve.

Table 4.7: Specification tests for links between the macroeconomy and the yield curve and diagonality of $\mathbf{\Omega}$

Case	Restricted parameters	*LM* statistic	*df*	*P*-value
Diagonality of $\mathbf{\Omega}$ matrix	$\mathbf{\Omega} = \text{diag}(\sigma_{11}, \sigma_{22}, \ldots, \sigma_{66})$	**323.443**	15	0.000
No intersection	$\mathbf{\Gamma}_{12} = \mathbf{\Gamma}_{21} = \mathbf{\Omega}_{12} = \mathbf{0}$	**249.600**	27	0.000
No macro to yield	$\mathbf{\Gamma}_{12} = \mathbf{\Omega}_{12} = \mathbf{0}$	**45.213**	18	0.000
No yield to macro	$\mathbf{\Gamma}_{21} = \mathbf{\Omega}_{12} = \mathbf{0}$	**43.423**	18	0.001

4.5 Summary

We have specified and estimated the yield curve model that incorporates both yield curve factors (level, slope, and curvature) and macroeconomic factors (manufacturing capacity utilization, the federal funds rate, and annual price inflation) based on Diebold et al. [2006]. This model's convenient state-space representation facilitates estimation, the extraction of latent yield-curve factors. However, this model is highly complicated and the parameter space is quite large and the optimization problem of maximization problem of the likelihood function is non-trivial and time consuming. Therefore, we propose specification tests based on the LM procedure in application of Chiba [2007]'s framework. This framework does not require the estimation of alternative and eases numerical task for analyzing dynamic interactions between the macroeconomy and the yield curve. From the results of specification

[5] *Note*: Bold entries denote test statistics significant at the 5 percent level. All test statistics is a Chi-squared.

testing, we find evidence of macroeconomic effects on the future yield curve and evidence of yield curve effects on future macroeconomic development Hence, the results indicate that the bidirectional causality between them is present.

References

- Ang, A. and Piazzesi, M. (2003), "A no-arbitrage vector autoregression of term atructure dynamics with macroeconomic and latent variables," *Journal of Monetary Economics* 50, 745-787.
- Ang, A., Piazzesi, M., and Wei, M. (2003), "What does the yield curve tell us about GDP growth?," *Manuscript*, Columbia University, UCLA.
- Chiba, M. (2007), "Specification testing in dynamic factor models," *Manuscript*.
- Dai, Q. and Singleton, K. (2000), "Specification analysis of affine term structure models," *Journal of Finance* 55, 1943-1978.
- Diebold, F.X. and Li, C. (2006), "Forecasting the term structure of goverment bond yields," *Journal of Econometrics* 130, 337-364.
- Diebold, F.X., Rudebusch, G. D., and Aruoba, S. B. (2006), "The macroeconomy and the yield curve: a dynamic latent factor approach," *Journal of Econometrics* 131, 309-338.
- Duffie, D. and Kan, R. (1996), "An yield-factor model of interests rate," *Mathematical Finance* 6, 379-406.
- Estrella, A. and Hardouvelis, G. A. (1991), "The term structure as a predictor of real economic activity," *Journal of Finance* 46, 555-576.
- Estrella, A. and Mishkin, F. S. (1998), "Predicting US recessions: financial variables as leading indicators," *Review of Economics and Statistics* 80, 45-61.
- Evans, C. L. and Marshall, D. (1998), "Monetary policy and the term structure of nominal interest rates: Evidence and Theory," *Carnegie-Rochester Conference Series on Public Policy* 49, 53-111.
- Evans,C. L. and Marshall. D. (2001), "Economic determinants of the nominal treasury yield curve," *Manuscript*, Federal Reserve Bank of Chicago.
- Fama, E. and Bliss. R. (1987), "The information in long-term maturity forward rate," *American Economic Review* 77, 680-692.
- Hördahl, P., Tristani, O., and Vestin, D. (2002), "A joint econometric model of macroeconomic and term structure dynamics," *Manuscript*, European Central Bank.
- Rudebusch, G. D. (1995), "Federal reserve interest rate targeting, rational expectations, and the term structure," *Journal of Monetary Economics* 35, 245-274.

- Rudebusch, G. D. (1998), "Do measures of monetary policy in a VAR make sense?," *International Economic Review* 39, 907-931.

- Rudebusch, G. D. (2002), "Term structure evidence on interest rate smoothing and monetary policy inertia," *Journal of Monetary Economics* 49, 1161-1187.

- Rudebusch, G. D. and Wu, T. (2003), "A macro-finance model of the term structure, monetary policy, and the economy," *Manuscript*, Fedreral Reserve Bank of San Franscisco.

- Stock, J. H. and Watson, M.,W. (2000), "Forecasting output and inflation: the role of asset prices," *Manuscript*, Kennedy School of Goverment (revised Januay 2003).

- Wu, T. (2002), "Monetary policy and the slope factors in empirical term structure estimations," *Federal Reserve Bank of San Francisco Working Paper* 2-7.

Chapter 5

Conclusion

State-space models are important for macroeconomic and financial analysis, for example, the modeling of the business cycle, the characterization of the yield curve, and the mimicking portfolios in the APT model. However, we rarely know priori structure of the exact form. In Chapter 2, we attempted to fill this gap, by proposing specification tests for weak exogeneity, linear dependency, and omitted explanatory variables based on the Lagrange Multiplier principle.

In Chapter 3, we proposed the new APT framework composed of unobserved factors and observed macroeconomic factors, based on the state-space system. Employing our framework, we estimated risk exposure profiles and associated risk premiums, and extract unobserved factors simultaneously. From the estimation result, we found that the large proportion of the expected excess return depends on the unobserved factor. Further, we examine the adequacy of macroeconomic factors as systematic variables. From the result of specification tests, we also found that the effect of macroeconomic factors to the stock return is decreasing in recent years.

In Chapter 4, we specified and estimated the yield curve model that incorporate both yield curve factors (level, slope, and curvature) and macroeconomic factors (manufacturing capacity utilization, the federal funds rate, and annual price inflation) based on Diebold's paper. This model, however, was highly complicated and the parameter space is quite large. Hence, the optimization problem for maximization problem of the likelihood function is non-trivial and time consuming. We attempted to ease this difficulty, by proposing formal tests of macro and yield curve inteactions based on our testing framework. From the result of specification testing, we found strong evidence of macroeconomic effects on the future yield curve and evidence of yield curve effects on future macroeconomic development.

It should be noted that it is open to question whether the test work under the autocorrelation or heteroskedasticity of the disturbances. Try to expand the specification tests to a more general case is our future work.

The abstract of the doctoral dissertation

Chiba Masaru

Title

Specification Testing in State-Space Models and Its Applications to Financial Data

1 State-space models

State-space models are not new in the statistics and econometric literatures. But, a growing number of published papers that employ them demonstrates their usefulness and widening application. Harvey [1981] introduced to economists Kalman filter for obtaining maximum likelihood estimates of parameters through prediction error decomposition was introduced. It became clear from Harvey's work and others' that a surprising range of econometric models, including regression models with time-varying parameters, ARIMA models and unobserved componenents time series models, could be cast in state-space form and thus be rendered amenable to Kalman machinery for parameter estimation and extraction of state variables.

State-space models have a wide range of potential applications in econometrics - for example, permanent income, expectations, the ex ante real rate of interest, and the reservation wage. Engle and Watson [1981] apply it to modeling the behavior of wage rate; Garbage and Wachetel [1978] and Antonicic [1986] apply it to modeling the behavior of ex ante real interests; Burmeister and Wall [1982] and Burmeister, Wall, and Hamilton [1986] apply it to modeling a time-varying monetary reaction function of the Federal Reserve. Stock and Watson's [1991] dynamic factor model of coincident economic indicators is a recent application of state-space models. Thus, state-space models have highly productive paths for research in econometrics and finance.

However, we rarely know priori structure of an exact model. In fact, investigators estimate several models but may not undertake comprehensive testing of the adequacy of their preffered model. Thus, there are some requirements for specification testing.

2 Purpose of this study

This paper has two purposes. One is to propose specification tests for the dynamic factor model. The weak exogeneity, linear dependency, and omitted explanatory variables tests will be presented in this paper.

Another is to apply these tests to the factor augumented arbitrage pricing theory (APT) model and the term structure model of yield curve. In the APT model, we will examine the adequacy of macroeconomic factors on systematic variables to the stock return. In the term structure model of yield curve, we will examine the nature of the linkage between factors driving the yield curve and macroeconomic factors.

Thus, we have grouped this paper into three categories: (1) specification testing in dynamic factor models, (2) an asset pricing model, with links to macroeconomy, and (3) an yield curve model, with links to macroeconomy.

3 Structure of this study

This paper is arranged as follows. Chapter 1 is introduction of this paper. Chapter 2 is "Specification Testing in Dynamic Factor Models." Chapter 3 is "Joint Estimation of Factor Sensitivities and Risk Premia in the Factor Augumented APT Model." Chapter 4 is "The Macroeconomy and the Yield Curve: Specification Testing based on Lagrange Multiplier Approach." The conclusion is given in Chapter 5.

4 The abstract

4.1 The abstract of Chapter 2

Recently, researchers have been interested in economic and financial models in which the dynamics of large scale economic variables can be specified by a smaller number of indices or "common factors." When the dynamic factor model proposed by Stock and Watson [1989] is applied to a time series model, the result is a model based on the assumption that one latent variable causes the co-movement of four observed variables. This approach has provided some new perspectives on several economic analysis. However, we rarely know priori structure of an exact model. In fact, investigators estimate several models but may not undertake comprehensive testing of the adequacy of the preffered model. Thus, there are some requirements for specification testing.

Because common factors are generally unobserved, we need to extract them using statistical techniques. Estimation procedure, such as Principal component analysis and Kalman filter, are used to extract them. In the method of the former technique, a number of past studies considered the problem of verifying the adequacy of the model. However, in the method of the latter technique, little work has been done for the hypothesis testing or the model selection on the dynamic factor model. Thus, in Chapter 2, we attempts to fill this gap, by proposing specification tests for weak exogeneity, linear dependency, and omitted explanatory variables based on the Lagrange Multiplier principle.

We provides an expression for the score in this model, defined as the derivative of the conditional log likelihood of the tth observation with respect to the parameter vector. This permits calculation of all of the necessary test statistic as well as an intuitive interpretation of what each test is based on. The score turns out to be a natural byproduct of the routine used to calculate the expected value with respect to smoothed density. In addition, from the same calculation we construct asymptotic standard errors for the parameter vector and specification tests. Therefore, proposed tests can be calculated together from a single pass through the data.

We also investigate finite sample properties of the tests. Monte Carlo results show that tests are reliable in terms of both size and power performance.

4.2 The abstract of Chapter 3

There is a long tradition of factor or multi-index models in finance where they were originally introduced to simplify the computation of the covariance of returns in a mean-variance portfolio allocation framework. In this context, two major theories provide a rigorous foundation for computing the trade-off between risk and return: the capital asset pricing model (CAPM) and the arbitrage pricing theory (APT).

The APT takes the view that there need not be any single way to measure systematic risk. There are two alternative approaches to estimate them. The first approach relies in statistical techniques such as factor analysis or principal component to estimate risk exposure profiles and associated risk premiums. The second approach estimates them from available macroeconomic and financial data. However, each of these two approaches has its merits and demerits.

Then, in Chapter 3, we unify these two approaches and overcome the weakness of them. We propose a framework composed of observed macroeconomic factors and unobserved factors, based on the state-space system. Employing our framework, we estimate risk exposure profiles and associated risk premiums, and extract unobserved factors simultaneously. In addition, we examine the adequacy of the macroeconomic factors on the systematic variables based on the framework of Chapter 2.

In addition, we use simple arguments to choose a set of macroeconomic factors that were candidates as sources of systematic asset risk. Macroeconomic factors we use are monthly and annual growth of industrial production, unexpected inflation, the change in expected inflation, and the spread between long and short interest. From the estimation result, we find that the large proportion of the expected excess return depends on the unobserved factor. Therefore, we conclude that the contribution by the total one of macroeconomic factors are relatively decreasing.

Further, from the result of specification tests, we find that three macroeconomic factors have significant influence on the stock return. They are monthly growth of industrial production, unexpected inflation, and the change in expected inflation. We also find that the number of unobserved factor which has a significant effect on the stock return are decreasing as time goes by. Therefore, we conclude that the effect of macroeconomic factors to the stock return is decreasing in recent years.

4.3 The abstract of Chapter 4

Understanding the term structure of interest rate has been a topic on the agenda of both financial economists and macroeconomists, albeit for different reasons. Therefore, they all have attempted to build good and tractable models of the yield curve, yet unusually large gap is apparent between the yield curve model developed by financial economists and the model developed by macroeconomists. Thus, combining these two lines of research seems fruitful, in that there are potential gains both ways.

Recently, a number of papers take a step toward bridging this gap by formulating and estimating a yield curve model that integrates macroeconomic and financial factors. However, these literatures focus only one-way yields-to-macro or macro-to-yields links, so the interactions between macroeconomic and term-structure dynamics have also been left unexplored.

In order to redress these shortcomings, Diebold et al. [2006] constructs a dynamic term structure model entirely based on macroeconomic factors, which allows for an explicit feedback from the yield curve to macroeconomic outcomes. Therefore, we employ Diebold's framewok for analyzing the potential bidirectional feedback from the yield curve to the economy and back again.

Whereas, Diebold et al. [2006] estimates many parameters by numerical optimization and examine the interactions between the macroeconomy and the yield curve based on Likelihood Ratio and Wald methods. This approach needs high numerical task. The Lagrange Multiplier procedure is regarded as the most suitable for that situation because, unlike Likelihood Ratio and Wald procedure, it does not require the estimation of the alternative. Therefore, in Chapter 4, we attempt to ease this difficulty, by proposing formal test of macro and yield curve inteactions based on the framework of Chapter 2.

From the result of specification testing, we find evidence of macroeconomic effects on the future yield curve and evidence of yield curve effects on future macroeconomic development. Hence, the result indicates that the bidirectional causality between them is still present.