Time Series Factor Analysis with an Application to Measuring Money

Paul D. Gilbert
Department of Monetary and Financial Analysis, Bank of Canada,
234 Wellington Street, Ottawa, Canada, K1A 0G9
pgilbert@bank-banque-canada.ca

Erik Meijer
Department of Econometrics, University of Groningen,
PO Box 800, 9700 AV Groningen, The Netherlands
e.meijer@eco.rug.nl

November 24, 2004

Abstract

Technological innovations in the financial industry pose major problems for the measurement of monetary aggregates. This paper examines an estimation methodology for a new measure of money that has a more satisfactory means to identify and remove the effects of financial innovations. It uses the general ideas laid out in Gilbert and Pichette (2003), but explores improved estimation methods.

We use a time series factor analysis (TSFA) model to link the measured data (currency and deposit balances) to the underlying phenomena of interest (the intended use of money for transactions and savings). This TSFA model obviates the need for explicitly modeling the dynamic process that governs the underlying phenomena, unlike dynamic factor analysis (DFA). It is also different from standard factor analysis in important respects: the data does not need to be covariance stationary as long as differenced data satisfies a weak boundedness condition, the factor model has a nontrivial mean structure, and the observations are allowed to be dependent over time. We discuss how this affects the estimation of parameters and prediction of the factors.

The measures of money are then estimated by using an adapted form of a factor score predictor. The statistical properties of this predictor are studied, both over repeated samples and within a given sample. Some

*This paper was prepared in part while the second author was at the Bank of Canada. The views expressed in this paper are those of the authors. No responsibility for them should be attributed to the Bank of Canada.
apparent anomalies are found in a simulation experiment and explained analytically.

The main result is that, contrary to what is usually expected in cross-sectional factor analysis, the sampling variability in the (time-invariant) parameters, which is $O_p(T^{-1/2})$, accounts for most of the prediction errors, and the fundamental inability to estimate the factors consistently, which accounts for an $O_p(1)$ term in the prediction error, turns out to have only a very small impact for this type of data.

In the example application it is shown that these techniques provide improvements over results in Gilbert and Pichette (2003) and what was considered an important difficulty in that paper is now understood and shown not to be a serious problem. The approach has considerable promise for replacing the monetary aggregates.

1 Introduction

Standard multivariate factor analysis (see, for example, Wansbeek and Meijer 2000) usually makes assumptions which cause difficulties estimating “underlying factors” or “latent variables” in macro-economic time series. One is that the data is mean zero, or the mean can be subtracted, so the factor means are not important. In typical factor analysis applications, the data mean is of little interest as the “co-movements” are the real concern. While this is sometimes true in a macro-economic context, there are two important reasons the mean is of interest. First, the intuition of the interpretation is clearer: if the factors are zero the explained phenomena are also zero. Second, macro-economic variables are often interpreted in growth rates, and the mean affects the magnitude of growth rates. Another assumption is that the data are covariance stationary. Macro-economic data are typically growing, so this assumption limits applications.

Dynamic factor analysis (DFA) is related to TSFA. The name TSFA has been adopted to distinguish the techniques described here because of important aspects in which the emphasis is different from most DFA literature. First, we are interested in the interpretation of the factors, not just their ability to fit the measured data. That is, we are hoping to discover underlying structure that suggests an economic interpretation of the data, not just a reduced dimensional statistical description of the data. Also, DFA often uses principal components analysis, which imposes that the factors are uncorrelated, but uncorrelated factors are not expected in the application motivating this research. Furthermore, in TSFA there is no attempt to model the dynamics of the factors, as is usually done in DFA. The intention is to separate, as clearly as possible, the data measurement problem from the economic modeling of the process dynamics. While factor dynamics will eventually be of interest, that is a future economic modeling problem, not a data measurement problem.

The application motivating this research is that of measuring transactions, money and savings. Traditional measures of these are called narrow and broad monetary aggregates. Problems with the traditional measures are discussed in Gilbert and Pichette (2002 2003), and in many of the references cited in
those papers. However, the techniques developed are applicable to most macroeconomic data and so should be of much broader interest.

This paper explores a factor estimation methodology for integrating time series data. Corrections are also included to accommodate non-zero means. Section 2 of the paper gives an intuitive description of TSFA and important estimation issues. Section 3 develops theory for parameter estimation of factor models in the time series context. Section 4 develops theory for factor score prediction. Section 5 gives a Monte Carlo analysis of the techniques, with a sample size and data as might be expected in the transactions money problem (and which would typically be encountered in many macro-economics problems). Section 6 gives an example using the application motivating this research, extracting factors from Canadian money data. Sections 7 discusses the sensitivity of the results to the selected sample. Finally, section 8 summarizes the results and discusses outstanding issues.

2 Time series factor analysis (TSFA)

In TSFA the observed variables $y_{ti}$ ($i = 1, 2, \ldots, M$) at each period $t$ are expressed in terms of $k$ factors (or latent variables) $\xi_{tj}$, where $k < M$, and idiosyncratic terms $\varepsilon_{ti}$ which are mean zero and uncorrelated. In the factor analysis literature the observed variables are often called indicators. The measurement model is given by the equation:

$$y_{ti} = B\xi_{tj} + \varepsilon_{ti},$$

at each period $t$, where $B$ is a matrix called the factor loadings or simply loadings.

This resembles the usual factor analysis model except that here the observations are indexed by time. More importantly, the factor analysis model usually either includes an explicit intercept term or implicitly does so, by assuming that all variables are centered and thus all intercepts and means are zero. The “simplification” that the intercept term is omitted implies that the factor mean must explain the data mean. This actually complicates the estimation from the usual case and the theory is developed in the next section. The fact that the data is time series is important mainly because economic time series data are typically growing and thus not covariance stationary. Other than this, the sequential order of the data is irrelevant in TSFA as opposed to DFA. Theory is developed in following sections which allows standard factor analysis techniques to be applied.

In DFA a Kalman filter is sometimes used to estimate latent variables. This requires estimating a dynamic (state space) model which is then used to predict the factors. The TSFA procedure described here does not estimate a dynamic model. TSFA is intended to focus on measurement and make few assumptions about the dynamic economic models that might result.

The term factor analysis is often used to describe several techniques, including principal components analysis (PCA), but TSFA does not use a PCA based approach. Most importantly, there is no constraint that the factors should be
uncorrelated. Also, in contrast to PCA in which the most important component explains the most variance, TSFA estimates factors to make $\text{Cov}(\varepsilon_t)$ diagonal, that is, to explain co-movements.

Identification is an issue. Any invertible matrix $G$ defines new factors $(G\xi)_t$ and weights $BG^{-1}$ and the equation

$$y_t = (BG^{-1})(G\xi_t) + \varepsilon_t$$

(2)

gives identical indicator variables $y_t$ and idiosyncratic terms $\varepsilon_t$ as in Equation (1). Thus these factors cannot be distinguished statistically and some otherwise motivated objective or constraint must be imposed. This is typically called the rotation and allowing the possibility that the factors are correlated gives an oblique rotation.

The most controversial aspect of factor analysis is the imposed rotation. This must give factor measurements that are economically interesting, without imposing (potentially controversial) theory. In other words, the resulting factors should be compatible with a wide range of economic theory. The intention is to find a simple, intuitive, interpretable structure. Rotation objectives are a tool to help do this. Rotation changes the factor loadings and simultaneously changes the meaning of the factors. The measurement model is not changed statistically but recast in a form that should be substantively easier to interpret.

Suppose for the moment that “savings money” and “transactions money” really are important fundamental underlying phenomena influencing the population behaviour for the measured variables, but we do not know exactly how. (If we thought we knew how the problem would be cast as confirmatory factor analysis rather than exploratory factor analysis.) An “unrotated” estimate would likely result in two measured factors that are each some linear combination of the fundamental phenomena. However, these combinations would not usually give such a useful interpretation as the fundamental phenomena, and so could not be incorporated into economic models in an important interpretive way.

The above supposes that “savings money” and “transactions money” are the real phenomena. If this is not the case then the results of rotation tend to be ambiguous and one is pushed toward looking for other interpretations of the phenomena influencing the measured variables. Decades of experience with factor analysis in psychology (intelligence, personality) shows that the “unrotated” solution often gives an important “general” factor and a couple of secondary factors that hamper interpretation. After rotation, the various factors represent clearly interpreted sub-concepts such as “verbal intelligence” and “mathematical intelligence”.

3 Estimation theory

As discussed in the previous section, we consider the factor analysis model

$$y_t = \alpha + B\xi_t + \varepsilon_t,$$

(3)
where \( y_t \) is a random vector of \( M \) observed indicators, \( \xi_t \) is a random vector of \( k \) unobserved factors, \( \varepsilon_t \) is a random vector of error terms (measurement errors, disturbances, unique factors), \( \alpha \) is an intercept vector and \( B \) is a matrix of factor loadings. Note that, unlike ordinary factor analysis, we explicitly state the model in terms of the uncentered data, with possibly nonzero means and intercepts. In our particular application, however, we will assume that \( \alpha = 0 \), cf. (1).

Factor analysis is most frequently applied to cross-sectional data, in which it is reasonable to assume that the data are independently and identically distributed. Therefore, the means and covariance matrix are the same for every observation and the observations are independent. This is particularly convenient for estimation of the parameters.

The i.i.d. assumption obviously does not apply to a typical time series. However, this is not necessary either. For example, if the series \( \{ \xi_t \} \) and \( \{ \varepsilon_t \} \) are serially dependent, but \( \xi_t \) and \( \varepsilon_t \) are uncorrelated (at the same time point) with zero means and constant covariance matrices \( \Gamma \) and \( \Psi \), then the mean and covariance matrix of \( y_t \) are \( \mu_y \equiv \alpha \) and \( \Sigma_y \equiv BB' + \Psi \), respectively. Note that \( \xi_t \) and \( \varepsilon_s \) may even be correlated for \( t \neq s \). Then, under some regularity conditions, the sample mean and covariance matrix of the observed variables \( y \) will be consistent estimators of \( \mu_y \) and \( \Sigma_y \) and therefore the usual estimators of the parameters (such as ML) are consistent. The necessary regularity conditions include assumptions about the diminishing dependence of observations when they are further apart in time (e.g., mixing conditions). Statistical theory for the estimators, such as standard errors and test statistics, can then be based on GMM theory.

This applies in particular to stationary time series. For stationary time series, state space models are often used, see, e.g. Watson and Engle (1983) and Hamilton (1994). A state space model contains both a specification of how the observed variables are related to the factors (the states) and a specification of a dynamic model for the relation between the states at different time points, typically an autoregressive model. A DFA model is a state space model without observed explanatory (input) variables. Efficient estimators for the parameters of a state space model have been described in the literature. The factor scores can then be predicted with relatively high precision by using a Kalman filter or Kalman smoother.

The drawback of using a state space model is that a (substantive) model for the development of the factors has to be specified. It is our intention to refrain from assumptions about this as much as possible, concentrating on the measurement part. Therefore, we will investigate the possibilities of estimating parameters and predicting factor scores under minimal assumptions about the factors. However, it will turn out that we will have to make a few assumptions.

We begin our investigation by assuming a slightly more general variant of (3):

\[
y_t = \alpha_t + B\xi_t + \varepsilon_t, \tag{4}
\]

where \( \alpha_t \) is a possibly time-varying intercept vector. Note that we assume that the factor loadings are time-invariant. Many time series are integrated of order
1. If this is the case, the variances of the observed variables are increasing with time and are not converging to a finite limit. This poses problems for standard estimation methods, because these typically assume that parameters are constant over observations, or at least that certain moments converge in probability to a finite limit (see, e.g., Wansbeek and Meijer 2000, p. 234).

However, \( \{y_t\} \) can be integrating but have a stationary first difference. Therefore, we will study the consequences of the model (4) for the differenced data. As will appear below, we do not even need stationarity of the differenced series. A weaker form of boundedness will suffice. The equation for the differenced series is

\[
Dy_t \equiv y_t - y_{t-1} = (\alpha_t - \alpha_{t-1}) + B(\xi_t - \xi_{t-1}) + (\varepsilon_t - \varepsilon_{t-1})
\]

or

\[
Dy_t = \tau_t + B D\xi_t + D\varepsilon_t. \tag{5}
\]

The latter is again an equation with a factor structure, with the same loadings matrix \( B \). If we assume that \( \tau_t = \tau \) is a vector of constants, this implies that \( \alpha_t = \alpha_0 + t\tau \) for some vector of constants \( \alpha_0 \). Analogously, if we assume that \( E(D\xi_t) = \kappa \) is a vector of constants, this implies that \( E(\xi_t) = \gamma_0 + t\kappa \) for some vector of constants \( \gamma_0 \). Of course, we assume that the means of the errors are zero, so that then

\[
E(y_t) = \alpha_t + B E(\xi_t) = \alpha_0 + t\tau + B\gamma_0 + tB\kappa = (\alpha_0 + B\gamma_0) + t(\tau + B\kappa)
\]

Thus, a factor analysis model for the differenced data allows for a linear deterministic trend in the mean of the original data as well as for a stochastic trend.

Based on the considerations mentioned, we propose to estimate an ordinary factor analysis model from the differenced data. We will now give a list of sufficient conditions (assumptions) such that this leads to consistent estimators of relevant parameters. First, we will assume the model (4) and hence (5). Second, we will assume that \( \tau_t = \tau \) is a vector of constants in the latter. It is conceivable that useful models with time-varying \( \tau_t \) exist, and that our procedure gives consistent estimators of some parameters (\( B \) in particular) if \( \tau_t \) is time-varying but bounded in some sense, but this leads to some additional problems with the factor score predictors that will be proposed in the next section. In our application, we assume that \( \alpha_t = 0 \) and \( \tau_t = 0 \) for all \( t \), and therefore, we restrict ourselves to time-constant \( \tau \) and leave extensions to those who need them.

Third, we will assume that the following conditions are met:

1. \( \kappa \equiv \operatorname{plim}_{T \to \infty} \sum_{t=1}^T D\xi_t/T \) exists and is finite.
2. \( \operatorname{plim}_{T \to \infty} \sum_{t=1}^T D\varepsilon_t/T = 0 \).
3. \( \Phi \equiv \text{plim}_{T \to \infty} \sum_{t=1}^{T} (D\xi_t - \kappa)(D\xi_t - \kappa)' / T \) exists and is finite and positive definite.

4. \( \Omega \equiv \text{plim}_{T \to \infty} \sum_{t=1}^{T} D\varepsilon_t D\varepsilon'_t / T \) exists and is finite and positive definite.

5. \( \text{plim}_{T \to \infty} \sum_{t=1}^{T} (D\xi_t - \kappa) D\varepsilon'_t / T = 0. \)

Note that no explicit assumptions are made about the possible serial dependence of the differenced data, although some sorts of serial dependence (unit roots) imply violation of some of the assumptions. But generally, these assumptions allow a considerable amount of serial dependence in the variables. Furthermore, it is not assumed that means and variances are constant over time, only that they are bounded in such a way that the required probability limits exist.

Although we tend to think of \( \xi_t \) as a random vector, we may even consider \( \{\xi_t\} \) as a series of given constants, so that the measurement model is interpreted as a functional model and not as a structural model, in which case “plim” has the same meaning as “lim” (Wansbeek and Meijer 2000, pp. 11–12).

The conditions 2 and 5 are implied by the alternative condition \( E(D\varepsilon_t \mid D\xi_t) = 0 \) combined with the finiteness of \( \Phi \) and \( \Omega \). This is a substantively more meaningful assumption than 2 and 5 and therefore we will assume that this is satisfied as well.

Let the sample mean and covariance matrix of the differenced series \( Dy_t \) be denoted by \( \bar{D}y \) and \( S_{Dy} \), respectively. That is,

\[
\bar{D}y \equiv \frac{1}{T} \sum_{t=1}^{T} Dy_t
\]

and

\[
S_{Dy} \equiv \frac{1}{T} \sum_{t=1}^{T} (Dy_t - \bar{D}y)(Dy_t - \bar{D}y)'.
\]

From the stated assumptions, it follows that

\[
\text{plim}_{T \to \infty} \bar{D}y = \mu \equiv \tau + B\kappa
\]  \hspace{1cm} (6)

and

\[
\text{plim}_{T \to \infty} S_{Dy} = \Sigma \equiv B\Phi B' + \Omega.
\]  \hspace{1cm} (7)

Conventional estimators for factor analysis (such as ML) use the sample covariance matrix to estimate the factor loadings \( B \), the factor covariance matrix \( \Phi \), and the error covariance matrix \( \Omega \). From (7) it follows that these estimators must also be consistent in the current context, provided that \( B, \Phi, \) and \( \Omega \) are identified from this equation, i.e., can be uniquely found if \( \Sigma \) is known. Therefore, we will assume that this is the case. In most applications, \( \Omega \) is assumed to
be diagonal. We will also use this assumption in our application. Then, if the Ledermann bound

$$(M - k)^2 \geq M + k$$

is satisfied, $\Omega$ is generally identified (Wansbeek and Meijer 2000, pp. 169–170). However, if $B$ and $\Phi$ are not restricted, these parameters are not identified. Ideally, economic theory implies a sufficient number of restrictions on these parameters to render them identified. The resulting restricted model is then a confirmatory factor analysis (CFA) model.

In many cases, however, there is no theory that implies prior restrictions and the data analysis is performed to investigate the possible structure. This is called exploratory factor analysis (EFA). In EFA, the parameters are not identified. If we postmultiply a proposed matrix $B$ by a nonsingular matrix $G^{-1}$ and simultaneously replace $\Phi$ by $G\Phi G'$, then the resulting transformed model is statistically indistinguishable from the original one. This is known as the rotation problem. It is typically solved by optimizing a function of $BG^{-1}$ over all allowed candidate transformation matrices $G^{-1}$. This optimization, called rotation, is an attempt to find the solution that has the simplest structure, which is frequently substantively the most meaningful. Various functions have been proposed as objective functions for rotation, each one operationalizing “simple structure” in a slightly different way and leading to a different rotation method. In our application, we use the widely-used direct oblimin (or quartimin) method. In EFA, the rotation method chosen acts as an identifying constraint on the parameters, so that the parameters may be considered identified in this broader sense.

From the discussion above, it follows that neither normality nor serial independence is required for the consistency of the ordinary estimators. However, note that typical factor analysis routines in general purpose statistical packages tend to use the correlation matrix for estimating the parameters. The estimates from such a package must be scaled back if they are to be interpreted on the original scale. That is, the rows of $B$ must be multiplied by the corresponding sample standard deviations and the (diagonal) elements of $\Omega$ must be multiplied by the corresponding sample variances.

Given estimators $\hat{B}$, $\hat{\Phi}$, and $\hat{\Omega}$, estimators for $\tau$ and/or $\kappa$ can be obtained from (6). It is clear that the number of sample means in this equation is smaller than the number of parameters. Therefore, some restrictions must be imposed. In a typical factor analysis model, the intercepts are free parameters, so that the means of the factors can be arbitrarily but conveniently restricted to zero. This gives the restriction $\kappa = 0$ and estimator $\hat{\tau} = \overline{Dy}$. This result illustrates why the means are usually neglected in factor analysis applications. In our application, however, we will assume that $\tau = 0$ and $\kappa$ may not be zero. In this case, a good estimator of $\kappa$ will be the GLS estimator

$$\hat{\kappa} = (\hat{B}'W\hat{B})^{-1}\hat{B}'W\overline{Dy},$$

where $W = S_{Dy}^{-1}$, $W = (\hat{B}\hat{\Phi}\hat{B}' + \hat{\Omega})^{-1}$, or $W = \Omega^{-1}$. The latter two choices give the same estimator and the first gives an asymptotically equivalent one,
because the probability limit of the first two choices is the same.

It is also possible to estimate all parameters jointly from the mean and covariance structure, i.e., use (6) and (7) jointly. We have experimented a little with this, but this appeared not to lead to improved estimators. Therefore, we will restrict our attention to a standard covariance-based estimator of the free parameters in $B$, $\Phi$, and $\Omega$. In particular, we will use the maximum likelihood estimator, which is found by minimizing the function

$$L \equiv \log \det \Sigma + \text{tr}(\Sigma^{-1} S_{Dy})$$

(8)

as a function of the parameters, where $\Sigma$ is implicitly a function of the parameters, as given in (7). Because of the restriction in (6), and because of the possibility of serial dependence, the resulting estimators will not be full maximum likelihood. However, as noted above, the estimators are consistent.

Under fairly standard assumptions, a form of the central limit theorem implies that the elements of the sample covariance matrix $S_{Dy}$ are jointly asymptotically normally distributed: Let $s_{Dy}$ be the vector consisting of all unique (nonduplicated) elements of $S_{Dy}$, and let $\sigma$ be the vector consisting of the corresponding elements of $\Sigma$. Then

$$\sqrt{T}(s_{Dy} - \sigma) \xrightarrow{d} N(0, \Upsilon)$$

(9)

for some finite positive definite matrix $\Upsilon$. This matrix can be estimated consistently by a heteroskedasticity and autocorrelation consistent (HAC) covariance matrix estimator, such as the well-known Newey-West estimator (Newey and West 1987). See De Jong and Davidson (2000) for a very general consistency result for HAC estimators.

Let the parameters in $B$, $\Phi$, and $\Omega$ be stacked in the parameter vector $\theta$ with population value $\theta_0$. Evidently, the estimator $\hat{\theta}$ of $\theta$ is a function of $s_{Dy}$. By combining the implicit function theorem and the delta method with (9), we obtain

$$\sqrt{T} (\hat{\theta} - \theta_0) \xrightarrow{d} N(0, V),$$

where, if the model is identified (with CFA),

$$V = \left( \frac{\partial^2 L}{\partial \theta \partial \theta'} \right)^{-1} \left( \frac{\partial^2 L}{\partial s_{Dy} \partial \theta'} \right) \Upsilon \left( \frac{\partial^2 L}{\partial s_{Dy} \partial \theta'} \right) \left( \frac{\partial^2 L}{\partial \theta \partial \theta'} \right)^{-1}$$

and the partial derivatives of $L$ are evaluated in $s_{Dy} = \sigma$ and $\theta = \theta_0$.

In EFA, where rotation is used as a means to identify the parameters, the formula for $V$ becomes more complicated. The method to obtain standard errors in this situation was derived by Archer and Jennrich (1973). See also Ogasawara (2002) for similar formulas as the one presented here.

Let $h(\theta)$ be the objective function for the rotation. Then $V$ is the upper-left
block of the partitioned matrix

\[
\begin{pmatrix}
\frac{\partial^2 L}{\partial \theta \partial \theta'} & \frac{\partial^2 h}{\partial \theta \partial \theta'} \\
\frac{\partial^2 h}{\partial \theta \partial \theta'} & 0
\end{pmatrix}

-1 \cdot
\begin{pmatrix}
\frac{\partial^2 L}{\partial \theta \partial s_{D_y}} & 0 \\
0 & 0
\end{pmatrix}

\Upsilon \left( \frac{\partial^2 L}{\partial s_{D_y} \partial \theta}, 0 \right)

\begin{pmatrix}
\frac{\partial^2 L}{\partial \theta \partial \theta'} & \frac{\partial^2 h}{\partial \theta \partial \theta'} \\
\frac{\partial^2 h}{\partial \theta \partial \theta'} & 0
\end{pmatrix}

-1,
\]

where all partial derivatives are again evaluated in \( s_{D_y} = \sigma \) and \( \theta = \theta_0 \). Note that the expression cannot be simplified, because the blocks in the matrices that are inverted are singular themselves. Only the whole partitioned matrix is invertible.

In both cases, \( V \) can be consistently estimated by evaluating the partial derivatives in \( s_{D_y} \) (the elements of the observed sample covariance matrix) and \( \theta = \hat{\theta} \), and using a HAC estimator for \( \Upsilon \). Standard errors of the parameter estimators are then straightforwardly obtained and Wald and LM tests can be routinely applied if desired.

### 4 Predicting factor scores

In many cases, such as our substantive application, one is not only interested in the model parameters, such as \( B \), but also, or even primarily, in the values of the factors themselves. The factors (also called factor scores) \( \xi_t \) are unobserved and generally cannot be estimated consistently. We can, however “predict” them.

Once the parameters are estimated, the model (4) can be viewed as an ordinary factor analysis model. For prediction of the factor scores, the time series nature of the data is irrelevant, unless we are willing to make stronger assumptions about the serial dependence of the factors. For example, if we assume (and estimate) a state space model, the Kalman smoother/filter gives better results (Watson and Engle 1983). However, we are not willing to do that for the reasons already mentioned.

Given these considerations, the prediction error is due to two sources: (1) the estimation error in the parameters, and (2) the inability to obtain perfect predictions of the factors even if the parameters are known. The estimation error in the parameters diminishes with increasing sample size, because the estimators are consistent. More precisely, the estimation error is of the order \( O_p(T^{-1/2}) \). Therefore, in large samples, this error should be small.

The second source of error does not diminish with sample size. The reason for this is that, once the parameters are known, the observations \( y_s, s \neq t \), do not provide information about \( \xi_t \) unless we are willing to make assumptions about the dynamic dependencies, which we are not. Therefore, all usable information about \( \xi_t \) is contained in \( y_t \) and this information does not increase with sample size. It follows that this error is \( O_p(1) \).

Consequently, in large samples, the prediction error is dominated by the second source of error. Therefore, all analytical analysis of factor score predictors has been asymptotic, which means that the parameters are assumed known.
Compared to discussions of factor score predictors in the literature, (4) is different in one respect: means and intercepts are allowed to be nonzero. The standard factor score predictor formulas have to be slightly adapted to accommodate this. Although our model will imply a zero intercept, we will give the general formulas here, with possibly nonzero means and intercepts.

Analogous to most of the literature on factor score prediction with zero means, we study linear predictors. Nonlinear predictors have been proposed (Meijer and Wansbeek 1999), but linear factor score predictors are dominant in the literature. Therefore, we will also restrict ourselves to linear factor score predictors.

The easiest way to find suitable factor score predictors when intercepts and means are possibly nonzero is to transform the model such that the means and intercepts are zero, then apply the standard predictors to the transformed model and transform back. Starting from the model (4) with possibly nonzero means and intercepts, the transformed model is

$$y_t - \alpha_t - B\gamma_t = B(\xi_t - \gamma_t) + \epsilon_t, \quad (10)$$

where $\gamma_t \equiv E(\xi_t)$ and we use the assumption that $E(\epsilon_t) = 0$. Following the literature on factor scores, we assume that the parameters are known in this analysis, which is therefore asymptotic. Therefore, the left-hand side of (10) can be viewed as a vector of observed indicators with mean zero, and $\xi_t - \gamma_t$ in the right-hand side of (10) can be viewed as a vector of factors with mean zero.

The two most frequently used factor score predictors for a model with zero means and intercepts are the regression predictor and the Bartlett predictor. See, e.g., Wansbeek and Meijer (2000, pp. 164–166) for their derivation.

The regression predictor is equal to the conditional mean of the factors given the observations if joint normality is assumed. If normality is not assumed, it is still the predictor with the smallest (unconditional) mean squared prediction error. However, the regression predictor is not unbiased in the sense that its expectation conditional on the factor is not equal to the value of the factor. The Bartlett predictor (after Bartlett 1937) is unbiased in this sense and is sometimes preferred for this reason.

Applied to (10), the regression predictor for the zero-mean factor $(\xi_t - \gamma_t)$ is

$$\hat{\xi}^R_t = \gamma_t + \Gamma_t \Sigma^{-1}_y (y_t - \alpha_t - B\gamma_t),$$

so that the resulting regression predictor for $\xi_t$ is

$$\hat{\xi}^R_t = \gamma_t + \Gamma_t \Sigma^{-1}_y (y_t - \alpha_t - B\gamma_t),$$

where $\Gamma_t \equiv \text{Cov}(\xi_t)$ and $\Sigma_y \equiv \text{Cov}(y_t)$. Similarly, the Bartlett predictor for $(\xi_t - \gamma_t)$ is

$$(B' \Psi^{-1}_t B)^{-1}B' \Psi^{-1}_t (y_t - \alpha_t - B\gamma_t)$$

so that the resulting Bartlett predictor for $\xi_t$ is

$$\hat{\xi}^B_t = (B' \Psi^{-1}_t B)^{-1}B' \Psi^{-1}_t (y_t - \alpha_t),$$

11
where $\Psi_t \equiv \text{Cov}(\varepsilon_t)$.

The (unconditional) means and covariance matrices of the predictors are

$$
E(\hat{\xi}_R^t) = \Gamma_t B' \Sigma_y^{-1} (\alpha_t + B\gamma_t) - \Gamma_t B' \Sigma_y^{-1} \alpha_t - \left( \Gamma_t B' \Sigma_y^{-1} B - I_k \right) \gamma_t
$$

$$
\text{Cov}(\hat{\xi}_R^t) = \Gamma_t B' \Sigma_y^{-1} (\Sigma_y)^{-1} B \Gamma_t
$$

$$
= \Gamma_t - \Lambda^{-1} < \Gamma_t
$$

and

$$
E(\hat{\xi}_B^t) = (B' \Psi_t^{-1} B)^{-1} B' \Psi_t^{-1} (\alpha_t + B\gamma_t) - (B' \Psi_t^{-1} B)^{-1} B' \Psi_t^{-1} \alpha_t
$$

$$
= \gamma_t
$$

$$
\text{Cov}(\hat{\xi}_B^t) = (B' \Psi_t^{-1} B)^{-1} B' \Psi_t^{-1} (\Sigma_y)^{-1} B (B' \Psi_t^{-1} B)^{-1}
$$

$$
= \Gamma_t + (B' \Psi_t^{-1} B)^{-1}
$$

$$
> \Gamma_t
$$

where $\Lambda = \Gamma_t^{-1} + B' \Psi_t^{-1} B$ and $\Lambda^{-1} = \Gamma_t - \Gamma_t B' \Sigma_y^{-1} B \Gamma_t$, cf. Wansbeek and Meijer (2000, pp. 164–165), and matrix inequalities are in the sense of Löwner (1934), i.e., $A < B$ means that $B - A$ is a positive definite matrix.

Apparently, the means of the predictors are the same as the mean of $\xi_t$, but the variances are different. If it is desirable that the predictors have the same covariance matrix as the factors themselves, so-called covariance preserving predictors can be developed. See Ten Berge et al. (1999) for these predictors with zero means. We will not use these, however.

Both predictors require knowledge of $\alpha_t$, the intercept, and $\Psi_t$, the covariance matrix of $\varepsilon_t$. In addition to this, the regression predictor requires knowledge of $\gamma_t$ and $\Gamma_t$, the mean and covariance matrix of $\xi_t$, whereas the Bartlett predictor does not. This is an important difference, because we generally do not know these and we have thus far attempted to minimize the assumptions made about $\xi_t$. Moreover, with an integrated series $\gamma_t$ and $\Gamma_t$ will be increasing in $t$ and any assumptions made will become more problematic when time progresses. In the functional model view mentioned earlier, it is even more questionable whether the MSE optimality of the regression predictor is meaningful. In this case, the Bartlett predictor is a perfectly natural estimator of an unknown vector.

However, it would seem that knowledge of $\Psi_t$ is still needed. It is substantively and interpretationally convenient to assume that $\Psi_t \equiv \text{Cov}(\varepsilon_t) = \Psi$ is a time-invariant diagonal matrix and $\varepsilon_t$ and $\varepsilon_s$ are independent for $t \neq s$. This assumption about $\{\varepsilon_t\}$ implies that the covariance matrix of $D\varepsilon_t$ is $\Omega = 2\Psi$. It now follows immediately that $\Psi_t$ may be replaced by $\Omega$ in the definition of the Bartlett predictor. Because $\Omega$ is estimated from the differenced data, it may be assumed known asymptotically. If the i.i.d. assumption about $\varepsilon$ is not met, and
a positive definite weight matrix \( W \) is inserted for \( \Psi^{-1}_t \) in the computation of the Bartlett predictor, it is still unbiased, although not optimally efficient. This follows from standard GLS regression theory. Therefore, the Bartlett predictor with \( \Omega^{-1} \) inserted for \( \Psi^{-1}_t \) generally makes sense and will have relatively good properties even in the non-i.i.d. case.

In our application to the measurement of money, we assume that \( \alpha = \alpha_t = 0 \). In this case, the regression predictor becomes

\[
\hat{\xi}_t^R = \gamma_t + \Gamma_t B' \Sigma_y^{-1} (y_t - B \gamma_t)
\]

and the Bartlett predictor becomes

\[
\hat{\xi}_t^B = (B' \Psi^{-1}_t B)^{-1} B' \Psi^{-1}_t y_t
\]

or

\[
\hat{\xi}_t^B = (B' \Omega^{-1} B)^{-1} B' \Omega^{-1} y_t
\]

Note that the latter formula does not contain any possibly time-varying parameters. This further enhances the advantage of the Bartlett predictor over the regression predictor, because the Bartlett predictor can be computed by using just the estimation results from the factor analysis of the differenced series, in particular \( B \) and \( \Omega \).

From the above, it is now clear that we have a strong preference for the Bartlett predictor of the factors in a time-series context. Hence, we will use this in our attempts to predict the amount of money.

## 5 Simulation study

To assess the precision with which the factor scores are predicted, we performed a small simulation study. In order to get a realistic view, we first estimated the model parameters from a real data set and subsequently predicted the factor scores. These results are described in detail in the next section and will therefore not be discussed here. Here, we only note that we have \( M = 6 \) indicators and \( k = 2 \) factors, and data for \( T = 215 \) consecutive months. The predicted scores and estimated parameters were then used to generate new samples. In this way, the new samples should be similar to data encountered in a real application. We simulated data for 100 replications (samples).

The parameters were estimated as discussed in section 3, with standard ML applied to the differenced data. We used the direct oblimin (quartimin) rotation method with Kaiser normalization (Wansbeek and Meijer 2000, pp. 168–169). The factor score predictor we used was the Bartlett predictor, as argued in the previous section. However, as shown there, the covariance matrix of these predictors is larger than the covariance matrix of the actual factors. Therefore, we rescaled the scores such that the (sample) covariance matrix of their first differences was exactly equal to the estimated covariance matrix \( \tilde{\Phi} \) of the first differences of the factors. Otherwise, the true value of the covariance matrix of the differenced factor scores in the simulation would not be \( \Phi \).
For the generation of the data in the simulation, the factor scores \( \tilde{\xi}_t \) thus obtained were used as the true values of the factors. Hence, all replications had the same values of the factor scores and the results of the simulation are conditional upon these. New values \( \varepsilon_t^{(r)} \) of the errors, where \( r \) is the replication number, were drawn from a normal distribution with mean zero and covariance matrix \( \hat{\Omega} \), the corresponding estimated covariance matrix from the original data. New sample data were subsequently obtained from

\[
y_t^{(r)} = \hat{B}\tilde{\xi}_t + \varepsilon_t^{(r)},
\]

where \( \hat{B} \) is the estimated loadings matrix from the original data.

The estimation and factor score prediction from the simulated data were done in the same way as used for obtaining the parameters and factor scores from the original data, except that the rescaling of the factor scores was not performed, because this would make the predictors biased and thus “less optimal”. (In the substantive application described in the next section, this has also not been done.)

The computations were performed in the software package R (R Development Core Team 2004), using the gpa functions of Bernaards and Jennrich (2004) for the rotations.\(^1\)

The first replication

To show what can be expected in an empirical situation, we first generated one sample. This sample was subsequently used to estimate the parameters (in particular \( B \) and \( \Omega \)) and predict the factor scores based on this. The true factor scores (solid line) and the Bartlett predictions thereof (dashed line), using sample estimates of the parameters, are depicted in figure 1.

The most striking feature of this figure is that there appears to be a large and systematic bias in the factor score predictions. To analyze this bias, we can study the prediction error \( \hat{\xi}_t - \xi_t \):

\[
\hat{\xi}_t - \xi_t = \hat{L}'y_t - \xi_t = (\hat{B}'\hat{\Omega}^{-1}\hat{B})^{-1}\hat{B}'\hat{\Omega}^{-1}y_t - \xi_t = (\hat{B}'\hat{\Omega}^{-1}\hat{B})^{-1}\hat{B}'\hat{\Omega}^{-1}(B\xi_t + \varepsilon_t) - \xi_t = -\hat{L}'(\hat{B} - B)\xi_t + \hat{L}'\varepsilon_t,
\]

(11)

where \( \hat{L}' = (\hat{B}'\hat{\Omega}^{-1}\hat{B})^{-1}\hat{B}'\hat{\Omega}^{-1} \). The first term on the right-hand side of (11) is \( O_p(T^{-1/2}) \). It is zero if \( \hat{B} = B \). The second term is \( O_p(1) \) and does not converge to zero with increasing sample size (unless \( L'\varepsilon_t = 0 \), where \( L' = (B'\Omega^{-1}B)^{-1}B'\Omega^{-1} \), which has a zero probability of happening).

Because the \( \{\varepsilon_t\} \) are i.i.d. with mean zero in our simulation, the second term in (11) is nonsystematic with a zero mean. Therefore, the systematic error

\(^1\)We would like to thank Coen A. Bernaards and Robert I. Jennrich for extensive discussions about their code.
cannot be due to this term and must be due to the first one. This is illustrated by figure 2, which is analogous to figure 1, except that in figure 2 the factor scores are computed using the true values of the parameters.

In this figure, there is no noticeable bias. Moreover, the predictions are extremely good, much better than the predictions based on the sample estimates of the parameters. We can conclude that the prediction errors are largely due to estimation errors of the parameters. Apparently, the sample size is (much) too small for the asymptotic analysis of the prediction error. Asymptotically, the estimation error is a negligible part of the prediction error, because the former is $O_p(T^{-1/2})$ and the latter is $O_p(1)$, but here we find the opposite: estimation error explains almost all of prediction error.

The systematic nature of the prediction errors can now be explained by the positivity of $\xi_t$ and the fact that $L'(\hat{B} - B)$ is a constant matrix in a given sample. Therefore, the errors tend to have the same sign for all $t$. Moreover, if the $\xi_t$ are serially dependent, the prediction errors will also be serially dependent and prediction errors of consecutive time points tend to be approximately the same.

Following this reasoning, the effect should be much smaller when comparing first differences of the predicted factor scores with the first differences of the true factor scores, because the first difference can be negative as well and are considerably less serially dependent. This is indeed the case, as illustrated in figure 3.
Figure 2: True factor scores vs. Bartlett factor score predictions (using true values of the parameters) in the first replication.

Figure 3: First differences of true factor scores vs. first differences of Bartlett factor score predictions (using sample estimates of the parameters) in the first replication.
Finally, because $\xi_t$ is generally increasing with $t$, it then follows from (11) that the prediction errors become larger for larger $t$. That is, the prediction errors are larger for more recent time points.

**Observable bias and bias correction**

Once we have obtained a factor score predictor $\hat{\xi}_t$, we can compute predicted values of the indicators just like predicted values of the dependent variable are computed in a linear regression analysis:

$$\hat{y}_t = \hat{B}\hat{\xi}_t.$$  

Then we study whether these predicted values are close to the observed values of the indicators. Note that the factor analysis model does not imply that the errors $\varepsilon_t$ should be small. It only implies that they are uncorrelated (at the same time point) and have mean zero. However, the comparison of the predicted values with the observed values could give us a rough idea about the fit of the model, in addition to likelihood ratio statistics and other fit indexes. For our first replication, the predicted values of the indicators are plotted with the observed values in figures 4 and 5.

It appears that the systematic prediction errors in the factor scores largely carry over to the prediction errors in the observed indicators. But, because
Figure 5: Observed vs. predicted values (using sample estimates of the parameters) of indicators 4–6 in the first replication.

we have observed the latter, this suggests that we could do some kind of bias correction so that the prediction errors for the indicators vanish. This could then reduce the systematic errors in the factor score predictors as well. Therefore, let us study in greater detail whether we can obtain better predictors of the observed indicators. The differences between the measured values of the indicators and their predicted values are

\[ y_t - \hat{y}_t = y_t - \hat{B}\hat{\xi}_t. \]

It now seems natural to define the optimal predictors as the ones that minimize a generalized least squares function

\[ F \equiv (y_t - \hat{y}_t)'W(y_t - \hat{y}_t), \]

where \( W \) is a weight matrix. It is now a straightforward consequence of standard regression theory that the optimal factor score predictor is

\[ \hat{\xi}_t = (\hat{B}'W\hat{B})^{-1}\hat{B}'Wy_t. \]

Analogous to GLS regression theory, it is natural to choose \( W = \hat{\Omega}^{-1} \) here. Consequently, the optimal factor score predictor is the Bartlett predictor. Given that the analysis did not involve the other observations, this predictor is optimal for each observation separately and therefore for all observations jointly as well. Although there seems to be a systematic bias in the predictors and this
Table 1: Bias and variability in the factor loadings estimates.

<table>
<thead>
<tr>
<th>True values</th>
<th>Bias</th>
<th>Standard dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.8</td>
<td>5.2</td>
<td>-0.4</td>
</tr>
<tr>
<td>23.8</td>
<td>-12.6</td>
<td>-2.5</td>
</tr>
<tr>
<td>5.2</td>
<td>-2.0</td>
<td>-0.4</td>
</tr>
<tr>
<td>36.8</td>
<td>16.9</td>
<td>-1.3</td>
</tr>
<tr>
<td>-2.8</td>
<td>31.0</td>
<td>1.0</td>
</tr>
<tr>
<td>2.6</td>
<td>47.6</td>
<td>2.5</td>
</tr>
</tbody>
</table>

systematic bias seems to be carried over to the predicted indicators, it is not possible to obtain a better (bias-corrected) factor score predictor that reduces the bias in the observed indicators.

More replications

Now, let us move away from studying only the first replication and consider all 100 replications. Table 1 shows the bias and variability of the estimates of the factor loadings. This table shows that, with the possible exception of the bottom-left element, there are no substantial biases. There is, however, a large variability in the estimates.

In figure 6, the simulation results for the factor score predictors are summarized. In this figure, the true factor scores (solid line) and the mean (across replications) of the predictions (dotted line, middle) are plotted, as well as the means plus 2 times the standard deviations (dashed line, upper) and the means minus 2 times the standard deviations (dash-dotted line, lower). The latter two give an impression of the variability of the factor score predictions around the means.

From this figure, we see that the systematic prediction errors we encountered before are not due to a bias, because across replications, there is little sign of bias, except perhaps for a small bias in the second factor in later time points. This corroborates our earlier analysis of this phenomenon. There is, however, a considerable amount of variability around the means. Based on our earlier analysis, we think that this is mainly due to the large variability in the parameter estimates, in particular the factor loadings.

6 Example application to money data

This section updates the application described in Gilbert and Pichette (2003) using the estimation techniques developed above. Relative to Gilbert and Pichette (2003) the estimation algorithm has been improved to take better advantage of known factor analysis techniques, augmented in the ways described in previous sections. Also, a different rotation criterion is used, but rotation is not the primary focus of the current work.
In brief, the hope is to replace traditional monetary aggregates with index measures using the techniques discussed above. That is, aggregation will be replaced with the measurement of latent variables that describe the fundamental underlying shifts in the financial assets of the population. These are estimated from the component data used to construct the monetary aggregates. In the TSFA context the term component is misleading and so the terminology observed variables or the factor analysis terminology indicators is used. It is anticipated that one factor will represent transactions money (assets intended for imminent transactions), and another factor will represent savings (assets intended for longer term investment). It is possible there are other factors, for example, corporate finances may be arranged differently than personal finances, but that possibility is not examined here.

The indicators are organized into six categories. Data is measured in a larger number of categories, but market shifts between indicators for reasons other than the fundamental economic behaviour of interest will interfere with extracting the factors of interest. Thus it is necessary to combine asset categories among which there have been important shifts for other reasons. For example, Canadian Savings Bonds are no longer such a widely used savings instrument and people have shifted much of this type of investment into mutual funds. All savings type instruments have been grouped together in a single indicator and it is expected that this indicator will load heavily on the savings factor. The six final indicator groupings are currency, personal chequing deposits, non-bank
chequing deposits, non-personal demand and notice deposits, non-personal term deposits, and investment. Non-personal accounts mostly belong to businesses. In Canada, trust companies, credit unions, and some other financial institutions are not grouped with banks, so the term “non-bank” refers to deposits at these institutions. Considerably more detail about the data is provided in Gilbert and Pichette (2003) and in Kottaras (2003). To remove factors that are not the main interest, the data is per capita and real Canadian dollars.

The data is not seasonally adjusted. It is possible that seasonal adjustment could introduce an artificial common factor that affects the estimation. The actual seasonal influence is thought to affect transactions more than savings (currency has marked seasonal peaks at Christmas and in September), so this may help distinguish the factors of interest. For economic modeling at a future stage of research the estimated factors could be seasonally adjusted.

The Bartlett predictor as developed in previous sections is used to illustrate the application. This is combined with an oblimin (quartimin) rotation. Rotation criteria will be examined more carefully in future work. The oblimin objective is to give loadings that weight heavily on one factor or the other. This may be appropriate for currency and investment, at the two ends of the scale, but is probably not appropriate for some deposit types in between. Modern deposit accounts can both pay good interest rates and allow the convenience of point-of-sale payments, so they may be used for both savings and transactions. The data explained by two factors, which has the obvious shortcomings discussed earlier, is shown in figures 7 and 8. The differenced version, which suggests the factors capture movements fairly well, is shown in figures 9 and 10.

The factor loadings are shown in Table 2. The two left columns are the estimated loadings. These are the same as the loadings used previously to generate simulated data, because of the way that those loadings were constructed. The two right columns are the loadings standardized (scaled so that each series in the factors and data has variance 1.0). These numbers are roughly interpreted as the correlation between the data and the factors. (These would be the correlations if the factors were uncorrelated.)

If we are to give the factors the interpretation suggested above, these loadings indicate that the first factor would be best interpreted as transactions money and the second as savings. (There is no way to force this interpretation in the estimation, so this determination needs to be done by examining the estimated loadings matrix.) The weights on the second factor are consistent with this interpretation.

The moderately weak loading of currency on the first factor is less consistent with this interpretation, and its moderately heavy loading on the second factor is a bit surprising. One interpretation of this would be that currency is used for savings much more than we usually think. A more likely conclusion is that there needs to be a third factor, for example, corporate financial behaviour may be very different from personal behavior. This conclusion is also suggested by the fact that non-personal demand and notice deposits also load heavily on the first factor.

Three factors is the Ledermann bound with six data series, which means the
Figure 7: Explained money indicators 1–3 data using 2 factors.
Figure 8: Explained money indicators 4–6 data using 2 factors.
Figure 9: Explained money indicators 1–3 differenced data using 2 factors.
Figure 10: Explained money indicators 4–6 differenced data using 2 factors.
covariance structure would be perfectly explained by the factors, regardless of the data. Therefore, it is difficult to assess the fit of the model with 3 factors. Furthermore, results may not be very meaningful due to overfitting. In future research this might be addressed with more money indicators or by adding credit series. While the estimates are preliminary in several ways, it is still interesting to compare the results with other measures. The factors are shown in figure 11, plotted against real per capita M1 and real per capita M2++. The factors are scaled to have the same average values as M1 and M2++. The savings factor has a much more dynamic path than M2++, and is also roughly consistent with what one might expect for savings given the Canadian business cycle. The first factor has less pronounced growth than M1 in the second half of the sample. If one thinks that transactions money growth should be correlated with inflation then this less pronounced growth is certainly consistent with observed inflation, however the drop of inflation in the late 1980’s is not so clearly reflected. The differenced factors and differenced real per capita M1 and real per capita M2++ are shown in figure 12. The differenced first factor is surprisingly close to differenced M1. One would expect a transaction money measure to be not too unlike M1 but, given that this whole research program was initiated because of dissatisfaction with M1 and other narrow aggregates, one might expect that it should not be too close either. Thus it is hard to pass judgement on this initial result. The second factor shows considerably more volatility than M2++, and again it is difficult to immediately judge whether that is good or bad.

7 Sensitivity to sample selection and size

This section provides an indication of the extent to which sample size and the selected sample period may influence results.

By most accounts, the most recent data, starting in the mid-1990s, has been problematic because of financial innovations. Figures in this section shows the results estimated on different sub-samples, one up to 1990, a second up to 1995, a third after 1995, and a fourth after 2001.

Figures 13 through 16 show that the differenced data is relatively well explained on all sub-samples, and the level data has the problems illustrated previously (and the effect of sample variability is evident). The factors
Figure 11: M1 and M2++ (solid) and scaled Bartlett Predictors computed using ordinary ML parameters (dashed).
Figure 12: Differenced M1 and M2++ (solid) and scaled Bartlett Predictors computed using ordinary ML parameters (dashed).
Figure 13: Explained money indicators 1–3 data using 2 factors full sample and sub-samples
Figure 14: Explained money indicators 4–6 data using 2 factors full sample and sub-samples
Figure 15: Explained money indicators 1–3, differenced data using 2 factors, full sample and sub-samples
Figure 16: Explained money indicators 4–6, differenced data using 2 factors, full sample and sub-samples
Figure 17: Bartlett Predictors based on full sample and sub-samples computed using ordinary ML parameters.
Figure 18: Differenced Bartlett Predictors based on full sample and sub-samples computed using ordinary ML parameters.
and differenced factors are shown in figure 17 and figure 18. The results for the factor levels reflect problems discussed earlier. However, the differenced factors do suggest that the technique is not overly sensitive to the selected sample. This is extremely encouraging, given that so much earlier work with the monetary aggregates over this period has suggested that structural change is an important issue.

8 Discussion and future research

This paper presents a methodology which we have called time series factor analysis (TSFA) to distinguish it from dynamic factor analysis (DFA). In TSFA the dynamics of the factors are not estimated, as is done in DFA. The reason for not doing this is to more clearly separate the measurement problem from the economic modeling problem. It is hoped that data used in TSFA will give factors which are better data for economic modeling. The TSFA and the economic modeling together would then be similar to DFA, but the TSFA step would not pre-suppose an economic model.

The techniques developed for TSFA use standard factor analysis theory but are different in important respects: the data does not need to be covariance stationary as long as the differenced data satisfies a weak boundedness condition, there is a nontrivial mean structure, and the observations are allowed to be dependent over time. The effects of these on parameter estimation and factor prediction have been discussed.

It is possible that the estimation procedures may be further refined, but the variance in the estimate of the parameters and the related difficulty in predicting factor levels has been explained. However, the differenced factors appear to be estimated well, and in many respects this is even more important than estimating the levels precisely. If the indexes are pegged at a reasonable level then growth rates will be in a reasonable range. There may still be improvements in the estimation techniques which give the desirable property that factor levels explain data levels. This is not viewed as critical.

Factor rotation still needs to be examined more closely. Results here are based on an oblimin rotation. This is appropriate if indicator data should load almost exclusively on one factor or the other. For several of the indicator series this may not be appropriate. Other rotation criteria will be considered in future research.

The number of factors has been assumed here to be two, but there are indications that three factors may be needed. With only six indicators available, however, this would lead to a saturated, just-identified model that fits the observed covariance matrix perfectly, regardless of the data. This leads to unreliable results and overfitting. This may best be accommodated by adding credit data. The Ledermann bound for six data series is three factors. Adding six credit series would give 12 in total, and a Ledermann bound between 7 and 8 factors. Since one would expect some factors to explain both credit and money behaviour, this seems like an attractive option.
There are also questions concerning structural breaks: whether they need to be considered, and if so, determining where they occur. Casual consideration of the results presented here suggests that structural breaks may not be a problem, but research on the monetary aggregates over this period has usually been frustrated by model breakdowns that are often attributed to structural breaks, so formal testing would be appropriate.

The application results are still preliminary, as are the estimation techniques. The estimation methodology should be applicable to a much large range of problems than the application considered here, and indications are that it can work very well. The application to measuring money shows much promise and may provide a measurement of a more fundamental underlying economic phenomena than conventional monetary aggregates.

References


