An Econometric Forecasting for the Social Security Trust Funds

by

H. Shawn Lin
School of Business
University of Wisconsin-Madison
975 University Avenue, #1251
Madison, WI 53705
U.S.A.
hlin@bus.wisc.edu

Abstract

The financing for Social Security is a pay-as-you-go system, rather than the full-reserve system utilized in private insurance, and the financing conditions of the trust funds have more uncertainty than does the reserve in private insurance. Traditionally, the financing conditions have been decided by a deterministic approach using economic and demographic assumptions that rely upon negotiation among experts at various government agencies. In this paper, we propose an econometric approach, called the co-integrated multivariate time series model, for forecasting the financing conditions of the trust funds. This stochastic approach not only establishes the interrelationships but also clarifies the co-movements among the economic assumptions. Furthermore, it helps us analyze the original assumptions rather than the differences of the non-stationary assumption variables adopted by researchers in the past that are more difficult to interpret. Using this approach, we can assess the probability distribution of the future balance of trust funds instead of assessing estimates from the three alternative deterministic assumptions. This distribution will provide policy-makers with a better assessment of stability of the trust funds.

Keywords: Economic assumption variables; Vector autoregressive time series models; Co-integration; Vector error correction models
1. Introduction

Social Security is the major economic security program in the United States. Similar national retirement or disability benefits insurance programs, such as the Canada/Quebec Pension Plan for Canada, are also the foundation for social economic security in most developed countries. In 1996, almost 43.5 million persons received benefits of $347 billion from Social Security. In total, 16.4% of the American population obtained 4.6% of Gross Domestic Product (GDP) payments from the same trust funds. This program not only alleviates economic insecurity for the general population but also impacts the whole country economically because of the tremendous amount of money in its premium contributions and benefits. This program contributes to economic stability as a tool to moderate the economic cycle and redistributes incomes (see, for example, Aaron, 1982). It also has important macroeconomic effects on the level of aggregate demand in the economy, on consumption, and on saving (see, for example, Lesnoy and Leimer, 1985). However, it is also a burden on the whole working population (see, for example, Overview of Entitlement programs 1992 Green Book).

Because of its influential role in the whole economy, this program must be soundly financed to pay promised benefits. As President Clinton said in his State of the Union Address in 1998:

"...What should we do with this projected surplus? I have a simple four-word answer: Save Social Security first."

However, there exists great uncertainty in the financing conditions for Social Security in the future. To have a clearer future outlook on Social Security trust funds, government agencies currently use a deterministic forecast approach that is based on several alternative sets of economic and demographic assumptions made by negotiation among experts at the agencies. The results from this approach are easily understood, but they are restricted to a limited number of estimates rather than providing a more substantial picture of the future. In 1991, a panel of technical experts, appointed by the Quadrennial Advisory Council on Social Security, made recommendations concerning the assumptions and projections. The panel was particularly interested in the development of methods to help quantify the degree of uncertainty of short- and long-range forecasts for both particular assumptions and for projections. However, the panel did not recommend changes in the demographic assumptions that underlie the current projections. Furthermore, the panel also recommended the use of the statistical techniques of time series analysis. The panel suspected that a multivariate approach might ultimately be worthwhile, although the suggested initial focus was on a univariate approach. After Foster's (1994) univariate analysis of the short-range economic assumptions, Frees, Kung, Young, Rosenberg,
and Lai (1997) conducted the pioneering work on the multivariate framework. Several of the economic assumption variables they considered, such as interest rate and unemployment rate, seemed not to stay around the same level over time, so they were assumed to be non-stationary. To explore the multivariate approach, Frees et. al. (1997) took the first order difference of these non-stationary time series variables individually, then they combined these first order differences with the other stationary assumption variables. However, it is known that differencing can twist or even obliterate co-movements between non-stationary variables and can cause heteroscedastic problems on the residuals from the model. Further, Frees et. al.'s (1997) approach did not account for exogenous interventions that may bias the level of the forecast, as Foster (1994) considered through the outlier analysis in the univariate framework. Compared with the fund forecasts by the agencies, Frees et. al.'s (1997) forecasts are more optimistic. In this paper, we propose to introduce an alternative approach for the economic assumption variables, called the co-integrated multivariate time series model, to forecast the financing conditions of the Social Security trust funds. According to the economic literature, because of the consideration of co-movements among the non-stationary economic variables, this model may yield better forecasts in an economic system with similar non-stationary variables, particularly for the long-run time horizon. This model also allows for exogenous interventions. We find that the fund forecasts from this approach are more consistent with the agencies’ forecasts than those proposed by Frees et al. (1997).

This paper is organized as follows. In Section 2, we review the Social Security financing mechanism and its actuarial assumptions. Section 3 introduces the proposed co-integrated multivariate time series model and a related particular model called the vector error correction model. This section also includes the economic interpretation of co-integration, a key theorem called Granger’s Representation Theorem, and illustrated examples. Section 4 discusses the statistical methodologies of model fitting and forecasting. Section 5 presents the short-range (10-years) fund forecasts created by applying this proposed approach to the actuarial economic assumption variables. Conclusion is given in Section 6.

2. Social Security Financing and its Actuarial Assumptions

Two trust funds have been established by law to finance Social Security. The Federal Old-Age and Survivors Insurance (OASI) Trust Fund pays retirement and survivor benefits; the Federal Disability Insurance (DI) Trust Fund pays benefits after a worker becomes disabled. When both OASI and DI are considered together, they are called Social Security, or the OASDI program. The trust funds are accounts in the U.S. Treasury. Social Security taxes and other
income are deposited in these accounts, and benefits are paid from them. The only purposes for which these trust funds can be used are to pay benefits and program administrative costs.

The trust funds hold money not needed to pay benefits and administrative costs and, by law, are invested in special Treasury bonds that are guaranteed by the U.S. Government. A market rate of interest on these bonds is paid to the trust funds. When these bonds reach maturity or are needed to pay benefits, the Treasury redeems them. At the end of 1996, the total assets of these two trust funds were $566.9 billion.

The major financing source for Social Security is payroll taxes on earnings that are paid by employees and their employers and by the self-employed. In 1996, almost $378.9 billion (89 percent) of the total trust income came from payroll taxes. The remaining income is primarily from interest earnings and from taxation of the OASDI benefits. The payroll tax rates are set by laws for OASI and DI and are applied to earnings up to a certain annual amount, called the earnings base, rising as average wages increase.

Short-range (10-year) and long-range (75-year) estimates are reported for trust funds. Because the future cannot be predicted with certainty, three alternative sets of economic and demographic assumptions are used in the annual report of the Board of Trustees to show a range of possibilities. Alternative I, the intermediate assumptions, reflects the Trustees' best estimate of future experience. Alternative II, called the low cost assumptions, is more optimistic; alternative III, called the high cost assumptions, is more pessimistic. The annual report shows how the trust funds would operate under different economic and demographic conditions. The assumptions are made about economic growth, wage growth, inflation, unemployment, fertility, immigration, and mortality, as well as about specific factors relating to disability, hospital, and medical services costs. The assumptions are developed through the collective analysis, judgment, and negotiation of actuaries, economists, and other staff at the Social Security Administration and the Departments of Treasury, Labor, and Health and Human Services. The assumptions are also re-examined each year in light of recent experience and new information about future trends, and are revised if warranted. Based on these assumptions, the results of the Trustees' financial evaluation of the OASDI program are presented in their annual report to Congress.

In 1990, the Trustees Report's assumptions were reviewed by an independent team of actuaries and economists convened by the Quadrennial Advisory Council on Social Security. The panel made a number of recommendations concerning the assumptions and the methodology used in setting the assumptions (Advisory Council on Social Security, 1991). The panel recommended development of methods to help quantify the degree of uncertainty present in the financial projections for the Social Security program and in the key assumptions underlying the
projections. They further recommended use of statistical time series techniques as a part of this effort.

Since 1990, the panel’s recommendations have been considered by several researchers. Foster (1994) evaluated four of the key short-range economic assumption variables through use of univariate time series models. These four variables are inflation rate, unemployment rate, real interest rate, and real wage increase. Frees et. al. (1997) considered slightly different and more recent data. They first examined the univariate case and obtained the results comparable to Foster’s work. Additionally, they developed the multivariate time series model to capture the contemporaneous correlations. However, they did not include any restrictions on the parameters, such as co-integration, to improve the forecasting and to investigate the common trends or the economic short-run or long-run interrelationship among these series. In this paper, we consider the implementation of co-integrated multivariate time series models to further study this issue.

3. Multivariate Time Series Models

3.1. Basics

The class of univariate time series models has been popularized by Box and Jenkins (see, for example, Box, Jenkins, and Reinsel, 1994). This class of models has been applied extensively in actuarial science, risk management, and insurance studies. Multivariate time series models have been formulated over two decades (see, for example, Reinsel, 1993 or Lutkepohl, 1993). However, insurance researchers have only recently started to apply these advanced modeling techniques. A particular class of multivariate time series models, called the vector autoregressive (VAR) time series model, has been advocated most notably by Sims (1980) as a way to estimate the dynamic relationships among jointly endogenous variables without imposing strong apriori restrictions. This class of models is the center of our discussion in this paper.

Let \( X_t = (X_{1t}, \ldots, X_{kt}) \)' denote a \( k \)-dimensional time series vector of random variables of interest. In the discussion of time series modeling, the very important concept of stationarity first needs to be considered. The vector time series process \( \{X_t\} \) is strictly stationary if the joint probability distribution of the random vectors \( (X_{1t}, \ldots, X_{kt}) \) and \( (X_{1t+l}, \ldots, X_{kt+l}) \) are the same for arbitrary times \( t_1, \ldots, t_n \), all \( n \), and all lags or leads \( l = 0, \pm 1, \pm 2, \ldots \). The vector process \( \{X_t\} \) is weakly stationary if \( X_t \) processes finite first and second moments and satisfies the conditions that the expectations \( E(X_t) = \mu_X \) does not depend on \( t \), and \( E[(X_t - \mu_X)(X_{t+l} - \mu_X)'] \) depends only on \( l \), where \( \mu_X = (\mu_{1t}, \ldots, \mu_{kt})' \) is the mean vector of the process. In this paper, the term
stationarity will be used for weakly stationarity. Weak stationarity is a less restrictive requirement than strict stationarity. There is a relationship between strict stationarity and weak stationarity: If the vector time series process is weakly stationary and follows the multivariate normal distribution, then it is strictly stationary.

A basic vector process \( \{e_t\} \) is called a vector white noise process if \( \{e_t\} \) has \( E(e_t) = 0 \) and \( E(e_t e'_t) = \Sigma \), which is a \( k \times k \) covariance matrix assumed to be positive-definite, and \( E(e_t e'_t) = 0 \) for \( t \neq 0 \). Further, a vector time series \( X_t \) is a vector autoregressive process of order \( p \), VAR\((p)\), if \( X_t \) can be represented as

\[
X_t = \sum_{j=1}^{p} \Phi_j X_{t-j} + \varepsilon_t, \quad \text{or} \quad \Phi(B)X_t = \varepsilon_t, \tag{3.1}
\]

where \( \Phi(B) = I - \Phi_1 B - \ldots - \Phi_p B^p \) is a matrix polynomial in the backshift operator \( B \) with \( B^t X_t = X_{t-j} \), \( \Phi_j \) is a \( k \times k \) matrix, and \( \varepsilon_t \) is a vector white noise process. We will pay more attention to this model in the following sections. Another frequently discussed representation of a multivariate time series is the vector moving average time series model. A vector time series \( X_t \) is a vector moving average process of order \( q \), VMA\((q)\), if \( X_t \) can be represented as

\[
X_t = \varepsilon_t - \sum_{j=1}^{q} \Theta_j \varepsilon_{t-j}, \quad \text{or} \quad X_t = \Theta(B)\varepsilon_t, \tag{3.2}
\]

where \( \Theta(B) = I - \Theta_1 B - \ldots - \Theta_q B^q \) is a matrix polynomial in \( B \), \( \Theta_j \) is a \( k \times k \) matrix, and \( \varepsilon_t \) is a vector white noise process.

An important result called Wold's Theorem, which decomposes the vector weakly stationary process into the white noise processes, is stated as follows. If a vector time series \( X_t \) is a weakly stationary vector time series with mean vector \( \mu_X \) and is purely non-deterministic (i.e., \( X_t \) does not contain any purely deterministic component process by which future values can be perfectly predicted from past values), then \( X_t \) can be represented as an infinite order vector moving average (VMA) representation,

\[
X_t = \mu_X + \sum_{j=0}^{\infty} \Theta_j \varepsilon_{t-j} = \mu_X + \Theta(B)\varepsilon_t, \tag{3.3}
\]

where the coefficients \( \Theta_j \) are not necessarily absolutely summable but do satisfy \( \sum_{j=1}^{\infty} ||\Theta_j|| < \infty \).

A stationary VAR\((p)\) process \( X_t \) is said to be causal if it can be represented as the equation (3.3)
with \( \sum_{j=1}^{\infty} \| \Theta_j \| < \infty \). A VMA\((q)\) process \( X_t \) is said to be invertible if it can be represented in the equation (3.1) with \( \sum_{j=1}^{\infty} \| \Phi_j \| < \infty \). If all the roots of \( \det{\Theta(B)} = 0 \) are greater than one in absolute value, then a VAR\((p)\) process \( X_t \) will be stationary, with \( X_t \) processing the casual infinite MA representation as in the equation (3.2) with \( \Theta(B) = \Phi_j B \). Such process \( X_t \) is denoted by \( X_t \sim I(0) \). If all the roots of \( \det{\Theta(B)} = 0 \) are greater than one in absolute value, then a VMA\((q)\) process \( X_t \) is invertible. See Reinsel (1993) or Lutkephol (1993) for further discussion of these multivariate time series properties.

### 3.2. Vector Error Correction Models (VECM) and Co-Integration

Error correction mechanisms have been used widely in economics to model an economic system where agents learn from the past when making their plans for the future, so that a proportion of dis-equilibrium from one period is corrected in the next period. For example, the change in price of a particular good in one period may depend upon the degree of excess demand in the previous period. Early versions of error correction models are Sargan (1964), Phillips (1957), Davidson, Hendry, Srba, and Yeo (1978), and Salmon (1982). A number of papers have been devoted to the analysis of the concept of co-integration and its relationship with the error correction models (Granger, 1981; Granger and Weiss, 1983; and Engle and Granger, 1987). In particular, Johansen (1988 and 1991) and Johansen and Juselius (1990) considered the vector error correction model for the co-integrated vector autoregressive time series process.

According to Engle and Granger (1987) and Johansen (1988, 1991), a VAR\((p)\) process \( X_t \), defined in the equation (3.1), can be rewritten in a form called the vector error correction model (VECM) as

\[
\Delta X_t = \Pi X_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta X_{t-j} + \varepsilon_t, \text{ or } \Gamma(B) \Delta X_t = \Pi X_{t-1} + \varepsilon_t, \tag{3.4}
\]

where the notations \( \Delta \) is the first order difference and \( \Gamma(B) = I - \sum_{j=1}^{p-1} \Gamma_j B^j \), with \( \Gamma_j = -\sum_{j=1}^{\infty} \Phi_j \), and \( \Pi = -I + \sum_{j=1}^{\infty} \Phi_j \). The matrix \( \Pi \) is called the impact matrix, and there are three possible cases of \( \Pi \) :
(1) $\text{rank}(H) = k$, i.e., the matrix $H$ has full rank, indicating that the vector time series process $X_t$ is stationary;

(2) $\text{rank}(H) = 0$, i.e., the matrix $H$ is the null matrix, and the VECM corresponds to a VAR ($p-I$) model for the differenced vector time series $\Delta X_t$; and

(3) $\text{rank}(H) = r$ with $0 < r < k$, implying that there exist $k \times r$ matrices $\alpha$ and $\beta$ of rank $r$ such that $HH' = \alpha \beta'$. Here matrices $\alpha$ and $\beta$ are called the adjustment coefficient and the cointegrating matrix, respectively.

The VAR model in differenced variables is incompatible with these representations, because it omits the error correction term $\Pi X_t$. The VAR in levels of the series ignores cross equation constraints and some roots of $\det[\Phi(B)] = 0$ are equal to one in absolute value. In the following, we will concentrate on the case (3), i.e., $0 < \text{rank}(H) = r < k$.

As defined before, a process $X_t$ is said to be stationary, denoted $X_t \sim I(0)$, if all the roots of $\det[\Phi(B)] = 0$ are greater than one in absolute value. However, the roots of the autoregressive operator, i.e., $\det[\Phi(B)] = 0$, can be equal to one, the same as the case (2) in the VECM defined above. First, we consider the univariate autoregressive time series, such as

$$X_t = \sum_{j=1}^{\infty} \phi_j X_{t-j} = \varepsilon_t \quad \text{or} \quad \Phi(B)X_t = \varepsilon_t, \quad (3.5)$$

If one root in the AR operator $\Phi(B)$ is equal to one and the remaining roots are all greater than one in absolute value for a univariate AR($p$) process $X_t$, then $X_t$ is called integrated of order one, denoted $X_t \sim I(1)$. The first order difference $\Delta X_t = X_t - X_{t-1}$ is a stationary time series, i.e., $\Delta X_t \sim I(0)$ for a univariate time series $X_t$ integrated of order one. For the multivariate case, a VAR($p$) process $X_t$ is defined to be integrated of order one, denoted $X_t \sim I(1)$, if $\Delta X_t = X_t - X_{t-1}$ is stationary and $X_t$ is not stationary (Lutkepohl, 1993). This definition differs from the one given by Engle and Granger (1987) in that they do not exclude $X_t$ with some stationary components. For instance, if there is just one stationary component and all other components are $I(1)$ in a VAR($p$) process $X_t$, then $X_t \sim I(1)$ according to this definition.

However, for a VAR($p$) process $X_t$ integrated of order one, there could exist non-zero $k$-dimensional vectors $\beta_i$'s so that linear combinations of the components of $X_t$, $Y_t = \beta'_i X_t \sim I(0)$. If such linear combinations exist, then the process $X_t$ is said to be cointegrated of order $(1,1)$, denoted $X_t \sim CI(1,1)$, where the vectors $\beta_i$'s are called the co-
The co-integrating matrix $\beta = (\beta_1, \ldots, \beta_r)$, which consists of the $r$ linearly independent co-integrating vectors such that $Y_t = \beta' X_t \sim I(0)$. The co-integrating rank $r$ is the rank of the co-integrating matrix and the number of linearly independent co-integrating vectors, with $r \leq k - 1$. This definition is a special case of Engle and Granger (1987), who introduced $CI(d, b)$ for $d \geq 1, b \geq 1$, but we also include the $I(1)$ process with some stationary components. This definition simplifies the terminology because it avoids distinguishing between variables with different orders of integration.

### 3.3. Economic Interpretation of Co-Integration

According to economic literature, the concept of an co-integration mimics the existence of equilibrium relationship in which an economic system with a non-stationary time series converges over time. Equilibrium relationships are suspected between many economic variables within an economic system, such as household income and expenditures or prices of the same commodity in different markets. The long-run equilibrium relationship is a state of equilibrium where there is no inherent tendency to change since economic forces are in balance, while the short-run equilibrium relationship depicts the dis-equilibrium state. In most time periods, an economic system will not be in a long-run equilibrium relationship. It is also not necessary to achieve a long-run equilibrium relationship at any point in time, even as time goes to infinity. All that is required is that economic forces move the system toward the equilibrium defined by the long-run equilibrium relationship. When considering long-run equilibrium relationships, it becomes necessary to consider the underlying properties of the processes that generate time series variables of interest in that economic system. That is, we must distinguish between stationary and non-stationary variables since failure to do so can lead to a problem of spurious results. If some variables of interest are non-stationary and the linear combinations of these variables become stationary, then there exist co-integration among the variables in this economic system.

If we only consider that the data generating process in an economic system follows the co-integrated VAR process $X_t \sim CI(1,1)$, then the long-run equilibrium relationship is defined as $Y_{t,t} = \beta' X_t = 0$ with the co-integrating vector $\beta, \beta$. The long-run equilibrium relationship is not necessarily unique, and thus $Y_{t,t} = \beta' X_t = 0$ denotes the set of long-run equilibrium relationships, where $\beta = (\beta_1, \ldots, \beta_r)$ consists of the $r$ linearly independent co-integrating vectors. The disequilibrium error is defined as $Y_{t,t} = \beta' X_t$, and it is the discrepancy between the outcome and the long-run equilibrium relationship. The vector $Y_t = \beta' X_t$ denotes the set of equilibrium errors. If
a long-run equilibrium really exists, then it seems plausible to assume that the variables of $X$, move together and that $Y$, is stationary. The vector error correction model (VECM) is particularly useful for co-integration, because this formulation contains information on both the short-run and long-run properties of the model, with dis-equilibrium as a process of adjustment to the long-run equilibrium relationship.

### 3.4. Granger’s Representation Theorem

A fundamental theorem called *Granger’s Representation Theorem* (Engle and Granger, 1987; and Johansen, 1991) verifies the equivalence of three vector time series representations. These three representations are vector autoregression models, vector error correction models, and vector moving average models. This key theorem of co-integration is stated as follows.

**Theorem 1 (Granger’s Representation Theorem)**

If a co-integrated VAR($p$) process $X_t \sim CI(1,1)$ satisfies

1. **Reduced Rank Condition:** $\text{rank}(\Pi) = r$, where the impact matrix is $\Pi = \alpha \beta'$, the $k \times r$ co-integrating matrix is $\beta$, and the $k \times r$ matrix $\alpha$ is called the adjustment coefficient; and
2. **Full Rank Condition:** $\text{rank}(\alpha'_i \Gamma \beta'_i) = k - r$, where $\alpha_i$ and $\beta_i$ are $k \times (k - r)$ matrices of full rank $k - r$ such that $\alpha'_i \alpha_i = 0$ and $\beta'_i \beta_i = 0$,

then

1. $\Delta X_t$ and $\beta' X_t$ are stationary, and $X_t$ is non-stationary;
2. there exists VECM($p$) representation as in equation (3.4);
3. the first order difference $\Delta X_t$ has the VMA representation $\Delta X_t = \Theta(B) \varepsilon_t$, where $\Theta$ is defined implicitly through the relation $\Theta(B) = \Theta(1) + (1 - B) \tilde{\Theta}(B)$, and
4. $\Theta(1) = \Theta = \beta, (\alpha'_i \Gamma \beta'_i) \alpha'_i$, with $\Gamma = I - \sum_{j=1}^{k} \Gamma_j$, and

$$
\Theta(1) = \Theta = \beta, (\alpha'_i \Gamma \beta'_i) \alpha'_i, \text{ with } \Gamma = I - \sum_{j=1}^{k} \Gamma_j, \text{ and }
$$

$$
\Theta(1) = \Theta = \beta, (\alpha'_i \Gamma \beta'_i) \alpha'_i, \text{ with } \Gamma = I - \sum_{j=1}^{k} \Gamma_j, \text{ and }
$$

$$
\Theta(1) = \Theta = \beta, (\alpha'_i \Gamma \beta'_i) \alpha'_i, \text{ with } \Gamma = I - \sum_{j=1}^{k} \Gamma_j, \text{ and }
$$

$$
\Theta(1) = \Theta = \beta, (\alpha'_i \Gamma \beta'_i) \alpha'_i, \text{ with } \Gamma = I - \sum_{j=1}^{k} \Gamma_j, \text{ and }
$$

It is noted that the VAR of differences in co-integrated VAR process omits the co-integration constraints among the variables of interest. This theorem connects VAR and VECM representations for the co-integrated vector time series, and VMA representation for the
difference of these vector time series. The co-integrated VAR process can also be written in the form of VMA representation as result (r4) with a singular coefficient matrix over time $\Theta$, plus a stationary vector time series $S$. From this representation, the non-stationarity in the process $X$, is created by the cumulative sum of the vector white noise processes $\varepsilon_j$'s. However, from the expression for $\Theta$, only the combinations $a_i \sum_{j=1}^{i} \varepsilon_j$ enter the process and contribute the non-stationarity. Thus, the term $a_i \sum_{j=1}^{i} \varepsilon_j$ is called the common trend of the co-integrated VAR process $X$, (Stock and Watson, 1988; and Johansen, 1994).

The idea of common trend is easily interpreted in the co-integrated VAR(1) process $X$, in which VECM(1) does not have short-run dynamics, i.e., $\Delta X_t = \mu D X_{t-1} + \varepsilon_t$ (see, for example, Johansen, 1994). The space $sp(\beta_j)$ is called the attractor set, which represents the long-run equilibrium relationship. The process $X_t$ is pushed along the attractor set $sp(\beta_j)$ by the common trend $a_i \sum_{j=1}^{i} \varepsilon_j$, and it reacts to the dis-equilibrium error $\beta'X$, by being pushed back toward the attractor set through the direction of the adjustment coefficient $\alpha$ and by being pushed away by the shocks to the system. If we let $\varepsilon_{i,t} = 0$ for $i=1,2,\ldots$, then the process $X_t$ will converge to a point on the attractor set and stay there forever. If there are short-run dynamics, i.e., the autoregression order of the model is more than one, then the adjustment will be diffused by the short-run matrices $f_i$'s.

### 3.5. Examples of Co-Integration

For illustration, we consider the following two examples. In Example 1, we consider a co-integrated 2-dimensional VAR(1) process where both components are $I(1)$ processes. In Example 2, we extend Example 1 to include one stationary variable in addition to the two $I(1)$ processes.

**Example 1** *(The Co-Integrated VAR(1) process with all $I(1)$ components)*

We consider the bivariate time series $X_t = (X_{1,t}, X_{2,t}) = \left( \sum_{j=1}^{i} u_{1,j}, \sum_{j=1}^{i} u_{1,j} + u_{2,t} \right)$, where $u_{1,j}$ and $u_{2,t}$ are white noise processes. Clearly, both $X_{1,t}$ and $X_{2,t}$ are $I(1)$ processes, because
the first variable is a random walk and the second variable is the random walk plus a white noise process. This bivariate time series follows a VAR(1) model \( \Phi(B)X_t = \epsilon_t \), with

\[
\Phi(B) = \begin{bmatrix}
1-B & 0 \\
-B & 1 \\
\end{bmatrix} = I - \begin{bmatrix}
1 & 0 \\
1 & 0 \\
\end{bmatrix} B \quad \text{and} \quad \epsilon_t = \begin{bmatrix}
\epsilon_{t,1} \\
\epsilon_{t,2} \\
\end{bmatrix} = \begin{bmatrix}
u_{t,1} \\
u_{t,2} + u_{t,1} \\
\end{bmatrix}.
\]

The process \( X_t \) has a VECM(1) representation \( \Delta X_t = \Pi X_{t-1} + \epsilon_t \), with

\[
\Pi = -I + \Phi = \begin{bmatrix}
0 & 0 \\
1 & -1 \\
\end{bmatrix} \quad \text{and} \quad \Gamma = I.
\]

It satisfies the reduced rank condition \( (1) \) in Granger's Representation Theorem such that \( \text{rank}(\Pi) = 1 < 2 \). There exists a co-integrating matrix \( \beta = (-1, 1)' \) such that \( Y_t = \beta X_{t-1} = X_{t-1} - X_{t-1} = u_{t-1} \) is a white noise process, and also a stationary time series, i.e., a \( I(0) \) process. Thus, \( X_t - CI(1,1) \) with a co-integrating vector \( \beta = (-1, 1)' \).

Because of \( \Pi = \alpha \beta' \), we have the adjustment coefficient \( \alpha = (0, -1)' \) and the matrices \( \alpha_1 = (1, 0)' \), \( \beta_1 = (1, 1)' \).

This process also satisfies the full rank condition \( (c) \) in the Granger's Representation Theorem, such that \( \alpha_1' \Gamma \beta_1 = 1 \) and \( \text{rank}(\alpha_1' \Gamma \beta_1) = 1 \). Because of Granger's Representation Theorem, the first difference of \( X_t \), i.e. \( \Delta X_t \), has the VMA representation as \( \Delta X_t = \Theta(B)\epsilon_t \), with

\[
\Delta X_t = \begin{bmatrix}
\Delta X_{t,1} \\
\Delta X_{t,2} \\
\end{bmatrix} = \begin{bmatrix}
u_{t,1} \\
u_{t,2} + u_{t,1} - u_{t,2} \\
\end{bmatrix}, \quad \Theta(B) = \begin{bmatrix}
1 & 0 \\
B & 1-B \\
\end{bmatrix} = I + \begin{bmatrix}
0 & 0 \\
1 & -1 \\
\end{bmatrix} B, \quad \text{and}
\]

\[
\Theta = \Theta(1) = \beta_1 (\alpha_1' \Gamma \beta_1)' \alpha_1' = \begin{bmatrix}
1 & 0 \\
0 & 1 \\
\end{bmatrix}.
\]

The rank of the matrix \( \Theta(B) \) is only equal to one, rather than a full rank two. One root of \( \det(\Theta(B)) = 0 \) is equal to one. Thus, this representation does not satisfy the invertibility condition, and \( \Delta X_t \) does not have a finite order vector autoregressive (VAR) representation. A finite-order VAR in differences \( \Delta X_t \), only affords a poor approximation to this co-integrated vector time series, because the level of \( X_{t,1} \) contains information that is useful for forecasting \( X_{t,1} \) beyond that contained in a finite number of lagged differences in \( X_{t,1} \) alone.

The process \( X_t \) also has a VMA representation as \( X_t = X_0 + \Theta \sum_{i=0}^{\infty} \epsilon_i + S_t - S_0 \), with

\[
S_t = \bar{\Theta} \epsilon_t = \begin{bmatrix}
0 & u_{t,1} \\
\end{bmatrix}'. \quad \text{The process} \ X_t \ \text{is pushed along the attractor set} \ \text{spanned by} \ \beta_1 = (1, 1)', \ i.e., \ the \ 45-degree \ line \ passing \ the \ origin \ point \ common \ trend \ by \ the
common trend $\alpha_i \sum_{i=1}^{\infty} \varepsilon_i = \sum_{i=1}^{\infty} u_{i,i}$. The process reacts to the dis-equilibrium error $\beta X_i$ by being pushed back toward the attractor set through the direction of the adjustment coefficient $\alpha = (0 \ -1)^T$, and by being pushed away by the shocks to the system.

**Example 2** (The Co-Integrated VAR(1) process with two I(1) components and one I(0) component)

We consider a three-dimensional vector time series $X_i = (X_{1,i}, X_{2,i}, X_{3,i})'$ with $X_{1,i} = \sum_{j=1}^{i} u_{1,j}, \ X_{2,i} = \sum_{j=1}^{i} u_{2,j} + u_{2,i},$ and $X_{3,i} = u_{3,i}$, where all $u_{1,j}, u_{2,j},$ and $u_{3,j}$ are white noise processes. Both $X_{1,i}$ and $X_{2,i}$ are still I(1) processes, but $X_{3,i}$ is an I(0) process. The process $X_i$ follows a VAR(1) model $\Phi(B)X_i = \varepsilon_i$ with $\Phi(B) = \begin{bmatrix} 1 - B & 0 & 0 \\ -B & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I - \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}B$ and

$$
\varepsilon_i = \begin{bmatrix} \varepsilon_{1,i} \\ \varepsilon_{2,i} \\ \varepsilon_{3,i} \end{bmatrix} = \begin{bmatrix} u_{1,i} \\ u_{2,i} + u_{2,i} \\ u_{3,i} \end{bmatrix}.
$$

The process $X_i$ has a VECM(1) representation $\Delta X_i = \Pi X_{i-1} + \varepsilon_i$, with $\Pi = -I + \Phi = \begin{bmatrix} 0 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$ and $\Gamma = I$. Because of $\text{rank}(\Pi) = 2 < 3$, the process satisfies the reduced rank condition (c1) in Granger’s Representation Theorem. The co-integrating matrix is $\beta = [\beta_1, \beta_2] = \begin{bmatrix} 1 & 0 \\ -1 & 0 \end{bmatrix}$ such that $Y_i = \beta X_i = \begin{bmatrix} Y_{1,i} \\ Y_{2,i} \end{bmatrix}$ with $Y_{1,i} = \beta_1 X_{1,i} \sim I(0), i = 1,2$, where $\gamma$ and $\delta$ are arbitrary nonzero real values. These two co-integrating vectors are linearly independent, and other co-integrating vectors are the linearly dependent combinations of these two vectors. Thus the co-integrating rank is two. Hence, $X_i \sim CI(1,1)$ with co-integrating matrix $\beta = (\beta_1, \beta_2)$.
Because of \( \Pi = \alpha \beta' = \begin{bmatrix} 0 & 0 \\ 1 - \gamma / \delta & 0 \\ 0 - \gamma / \delta \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 - 1 & \gamma \\ 1 - 1 & 0 \\ 1 - 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \), we have the adjustment coefficient \( \alpha' = \begin{bmatrix} 0 & 0 \\ 1 - \gamma / \delta \\ 0 - \gamma / \delta \end{bmatrix} \) and the vectors \( \alpha_i = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \) and \( \beta_i = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \). Further, \( \alpha'_i I \beta_i = 1 \) and \( \text{rank}(\alpha_i I \beta_i) = 1 \), so this process also satisfies the full rank condition (c2).

Thus, the VMA representation for the first difference of \( X_t \), is \( \Delta X_t = \Theta(B)c_t \), with

\[
\begin{bmatrix}
\Delta X_{ts} \\
\Delta X_{ts, 1} \\
\Delta X_{ts, 2}
\end{bmatrix} =
\begin{bmatrix}
u_{ts, 1} \\
u_{ts, 1} + u_{t+1} - u_{t-1} \\
u_{ts, 2} - u_{t+1} - u_{t-1}
\end{bmatrix},
\Theta(B) =
\begin{bmatrix}
1 & 0 & 0 \\
B & 1 - B & 0 \\
0 & 0 & 1 - B
\end{bmatrix} = I - \begin{bmatrix}
0 & 0 & 0 \\
1 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix} B,
\]

and \( \Theta = \Theta(1) = \beta_1 (\alpha'_i I \beta_i)^{-1} \alpha'_i \).

Two roots of \( \det(\Theta(B)) = 0 \) equal one. Thus, this representation does not satisfy the invertibility condition, and \( \Delta X_t \) does not have a finite order vector autoregressive (VAR) representation. A finite-order VAR in differences \( \Delta X_t \) only affords a poor approximation to this cointegrated vector time series, because the level of \( X_t \) contains information that is useful for forecasting \( X_t \), beyond that contained in a finite number of lagged differences in \( X_t \) and \( X_{t-1} \).

The process \( X_t \) also has a VMA representation as \( X_t = X_0 + \sum_{j=1}^t \epsilon_j + S_t - S_0 \), with \( S_t = \tilde{\varepsilon}_t = (0 \quad u_{ts, 1} \quad u_{ts, 2})' \). The process \( X_t \) is pushed along the attractor set \( \text{sp}(\beta_i) \), the space spanned by \( \beta_i = (1 \quad 1 \quad 0)' \), i.e. the 45-degree line passing the origin point common trend by the common trend \( \alpha'_i \sum_{j=1}^t \epsilon_j = \sum_{j=1}^t u_{ij} \). The process reacts to the dis-equilibrium error \( \beta'X_t \), by being pushed back toward the attractor set through the direction of the adjustment coefficient

\[
\alpha = \begin{bmatrix}
0 & 0 \\
1 - \gamma / \delta \\
0 - \gamma / \delta 
\end{bmatrix}
\]

and by being pushed away by the shocks to the system.
3.6. Vector Error Correction Models (VECM) with Linear Trend and Co-Integration

In this subsection, we extend the vector error correction model to include linear trends. The vector time series \( X_t \) can be decomposed into two components as

\[
X_t = Z_t + T_t,
\]

where \( Z_t \) is the irregular or random component discussed in previous subsections, and \( T_t \) is the trend component. We assume that the trend component \( T_t \) only includes the linear trend. To include the linear trend, we rewrite the VAR(\( p \)) process defined in the equation (3.1) as

\[
X_t = \sum_{j=1}^{\infty} \Phi_j X_{t-j} + \mu + \epsilon_t,
\]

where \( \mu \) is a \( k \)-dimensional vector of intercept coefficients and represents the linear trend.

The VAR(\( p \)) process with linear trend in the equation (3.7) can also be rewritten in the vector error correction model with autoregression order \( p \), VECM(\( p \)):

\[
\Delta X_t = \Pi \Delta X_{t-1} + \sum_{j=1}^{\infty} \Gamma_j \Delta X_{t-j} + \mu + \epsilon_t,
\]

where \( \Gamma(B) = I - \sum_{j=1}^{\infty} \Gamma_j B^j \), with \( \Gamma_j = - \sum_{j=1}^{\infty} \Phi_j \), and \( \Pi = -I + \sum_{j=1}^{\infty} \Phi_j \). This VECM(\( p \)) in the equation (3.8) is the same as the VECM(\( p \)) in the equation (3.4), except for a linear trend \( \mu \). We also concentrate our discussion on the case in which the impact matrix \( \Pi \) is reduced rank, i.e., \( \text{rank}(\Pi) = r < k \). We also keep the same definition of co-integration, regardless of the linear trend. To clarify the role of the linear trend, we restate Granger's Representation Theorem with linear trend, which has been proven by Johansen (1991).

**Theorem 1A (Granger's Representation Theorem with Linear Trend)**

If a co-integrated VAR(\( p \)) process as the equation (3.7) also satisfies the conditions (c1) and (c2) in Theorem 1, and a \( k \times k \) matrix \( \Theta \) is defined as \( \Theta = \beta^(-1) \Gamma \beta^{-1} \) \( \alpha \), then

- (r1a) \( \Delta X_t \) and \( \beta X_t \) are stationary, and \( X_t \) is non-stationary with linear trend \( \tau t = \Theta \mu t \);
- (r2a) there exists VECM(\( p \)) representation as the equation (3.8);
(r3a) the first order difference $\Delta X_t$ has a VMA representation as $\Delta X_t = \Theta(B)(\mu + \epsilon_t)$ where

$$\Theta(B) = \Theta(1) + (1-B)\Theta(B)$$

and $\Theta(1) = \Theta = \beta \{ (\alpha', I\beta')^{-1} \alpha' \}$, with $\Gamma = I - \sum_{j=1}^{\infty} \Gamma_j$; and

(r4a) $X_t$ also has the VMA representation as $X_t = X_0 + \sum_{j=1}^{\infty} \epsilon_j + \Theta \mu + S_t - S_0$, where

$S_t = \Theta(B)e_t$ and $\beta X_0 = \beta S_0$.

4. Statistical Methodologies for Model Fitting and Forecasting

To fit a particular co-integrated multivariate time series model, we use the maximum likelihood estimation with the technique of reduced rank regression in the regression context (Anderson, 1951). The details can be found, for example, in Johansen (1988, 1991). The maximum likelihood estimator of the co-integrating matrix $\beta$ is found as

$$\hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_r),$$

where $\hat{\beta}_1, \ldots, \hat{\beta}_r$ are the co-integrating vectors and the eigenvectors corresponding to the $r$ largest eigenvalues $\hat{\lambda}_1 > \ldots > \hat{\lambda}_r$ from solving the eigenvalue problem

$$\hat{\lambda}_i S_i - S_0 S_0' S_i = 0$$

The residual cross moment matrices are defined as

$$S_{ij} = T^{-1} \sum_{t=1}^{T} R_{it} R_{jt}', \quad i, j = 0, 1$$

where $R_{it}$ and $R_{jt}$ are the residuals of regression of $\Delta X_t$ and $X_{t-1}$, on $Z_t = (\Delta X'_t, \ldots, \Delta X'_{t-p+1})'$, respectively. The adjustment coefficient $\alpha$ and other parameters in the model will be related to the estimated co-integrating matrix $\hat{\beta}$.

The co-integrating rank $r$ will be determined by a likelihood ratio test statistic, called the trace statistic,

$$\lambda_{max} = -2 \log Q(H(r) | H(k)) = -T \sum_{r=1}^{k} \log (1 - \hat{\lambda}_r), \quad r = 0, \ldots, k - 1$$

The specification of co-integrating rank is based on a sequence of trace tests of $H_0$: $\text{rank}(\Pi) = r$ for values of $r = 0, \ldots, k - 1$, and an appropriate value of $r$ can be chosen as the smallest value such that hypothesis $H_0$ is not rejected. The asymptotic distribution of the trace statistic is not given by the usual $\chi^2$ distribution but as the multivariate version of Dickey-Fuller distribution.
The distribution is conveniently described by certain stochastic integrals and can be tabulated by simulation, for example, in Johansen (1988, 1991) and Johansen and Juselius (1990).

The selected model should avoid too many lags, since the number of parameters grows very fast with the lag length. The information criteria strike a compromise between the lag length and the number of parameters by minimizing a linear combination of the residual sum of squares and the number of parameters. There are many different kinds of information criteria available, and we consider the three prominent criteria.

\[ AIC_p = \log|\hat{\Sigma}(p)| + \frac{2n}{T - p} \]  
\[ HQ_p = \log|\hat{\Sigma}(p)| + \frac{2n \log(\log(T - p))}{T - p} \]  
\[ BIC_p = \log|\hat{\Sigma}(p)| + \frac{n \log(T - p)}{T - p} \]

Here, \( \hat{\Sigma}(p) \) is the estimate of the residual covariance matrix with respect to the lag length \( p \), and \( n \) denotes the number of parameters estimated. If the model includes the linear trend, then the parameter number is \( n = pk^2 + k \). These information criteria are normalized by \( T - p \) instead of \( T \) for order \( p \). We select the model by minimizing either criterion. The AIC criterion (Akaike, 1974, 1976) is the most popular, and it imposes a smaller penalty for the number of estimated parameters than does the BIC criterion (Schwarz, 1978). The HQ is the intermediate criterion in the penalty factor, which was suggested by Hannan and Quinn (1979).

It is difficult to compute the prediction intervals of the co-integrated multivariate time series model. We adopt the simulation approach, rather than approximate simultaneous prediction intervals through the methods of Scheffe and Bonferroni. As Frees et al. (1997) argued, the Monte Carlo simulation approach not only allows convenient computation of nearly exact prediction intervals but also is readily amendable to computing prediction intervals for nonlinear transformation of the economic variables of interest.

To simulate the forecast distribution, we first generate the random errors \( \tilde{\varepsilon}_{t,i,r} \)'s from standard multivariate normal distribution \( N(0, I) \) for \( l = 1, \ldots, Q \), and \( r = 1, \ldots, R \), where \( Q \) is the number of forecasting quarters and \( R \) is the random simulation replications. To have the random errors \( \varepsilon_{t,i,r} \)'s for a particular model, we multiply these random errors \( \tilde{\varepsilon}_{t,i,r} \)'s by the Cholesky decomposed matrix of residual covariance matrix \( \hat{\Sigma}_{1/2} \). We can obtain a forecast simulation distribution of \( X_{t,i,r} \) for \( l = 1, \ldots, Q \), and \( r = 1, \ldots, R \), by the model equation.
\[ \hat{X}_{t+1} = \sum_{j=1}^{J} \Phi_j \hat{X}_{t+1-j} + \varepsilon_{t+1} \]  

(4.8)

In particular, \( \hat{X}_{t+1-j} = X_{t+1-j} \) for \( j > 1 \). Thus, we are able to have a multivariate simulation distribution of the economic assumption variables \( \hat{X}_{t+1} = (\hat{X}_{3,t+1}, \hat{X}_{2,t+1}, \hat{X}_{1,t+1}, \hat{X}_{4,t+1})' \) with the inflation rate \( \hat{X}_{1,t+1} \), the logarithmic investment return rate \( \hat{X}_{2,t+1} \), the nominal wage rate \( \hat{X}_{3,t+1} \), and the unemployment rate \( \hat{X}_{4,t+1} \) for the \( r \)th replication at the time \( T+1 \).

In the operations of the Social Security trust funds, income sources are premium contributions, interest from investment in Treasury bonds, taxes on benefits, and payments from the general fund of the Treasury. The expenditures are benefit payments, administrative expenses, and transfers to the Railroad Retirement program. Compared with the major items, a relatively smaller amount of money comes from taxes on benefits and payments from the Treasury and outflows to administrative expenses or transfers to the Railroad Retirement program. From a forecasting prospective, these items are very dependent upon political negotiation and bargaining so that they become more unpredictable from the statistical and econometric analyses. Therefore, we only consider the major and more predictable items for the fund forecasting.

We use \( \hat{N}_{t+1} \) for the number of beneficiaries, \( \hat{B}_{t+1} \) for their average benefits amount, \( \hat{M}_{t+1} \) for the number of covered workers, and \( \hat{C}_{t+1} \) for their average contribution amount. We also need to transform the logarithmic assumption variables back to the original scale, so the principal plus the interest rate will be equal to \( \exp(\hat{X}_{2,t+1} / 100) \) and the acting work force will be \( 2 - \exp(\hat{X}_{4,t+1} / 100) \). If we assume that the incomes of the fund are only from the covered workers’ contributions and the interest and that the expenditures are only paid to benefits, then the forecasting fund \( \hat{F}_{t+1} \) follows the recursive equation

\[ \hat{F}_{t+1} = \hat{F}_{t+1-j} \exp(\hat{X}_{2,t+1} / 100) - \hat{B}_{t+1} \hat{N}_{t+1} + \hat{C}_{t+1} \hat{M}_{t+1} [2 - \exp(\hat{X}_{4,t+1} / 100)] \]  

(4.9)

with \( \hat{B}_{t+1} = \hat{B}_{t+1-j} \left( 1 + \frac{\hat{X}_{3,t+1}}{100} \right) \) and \( \hat{C}_{t+1} = \hat{C}_{t+1-j} \left( 1 + \frac{\hat{X}_{4,t+1}}{100} \right) \).
5. Fund Forecasting by Modeling of the Actuarial Economic Assumption Variables

In the Social Security program, the principal economic assumption variables are real Gross Domestic Product (GDP), average annual wage in covered employment, the consumer price index, real wage differential, average annual interest rate, average annual unemployment rate, and average annual percentage increase in labor force. Similar to Foster (1994) and Frees et. al. (1997), we use the economic variables which seem to be good approximations to those assumption variables used in the Social Security program. For the purpose of comparison, we adopt the same data used in Frees et. al. (1997) which are from the Citibase database of macroeconomic time series from FAME Information Service in New York. These data are the following four key economic assumption variables made on a quarterly basis: the inflation rate, the investment return rate on five-year Treasury notes, the wage rate, and the unemployment rate. The time frame of data for model fitting is from the third quarter of 1953 through the fourth quarter of 1992, and the data in the eight quarters in 1993 and 1994 are reserved for model validation. Based on the information in the 1995 Annual Report of the Board of Trustees of the Social Security, the short-range (10-year) forecast for the Social Security trust fund is from 1995 through 2004.

The measure of the inflation rate is the proportional change in the Consumer Price Index for Urban Wage Earners and Clerical Workers, without seasonal adjustment. The investment return rate on five-year maturity Treasury notes closely approximates investment return rate of the Social Security trust funds, because currently these trust funds are only allowed to buy federal government bonds which maturities primarily are at least four years. The wage rate is gauged by the proportional change in the wage index, which is the ratio of the wage and salary disbursements of all industries to the number of wage and salary workers. The unemployment rate is measured by the percent of the civilian labor force, seasonally adjusted. Both the investment return rate on five-year Treasury notes and the wage rate are nominal rates, that is, they include the effects of inflation. We rescale the return rate and the unemployment rate by the natural logarithm of the addition of the original series plus one, because this logarithmic transformation dampens fluctuation and explains the forces of return and unemployment respectively. Table 1 summarizes the descriptive statistics for these four major economic assumption variables. More details can be found in Frees et. al. (1997).
Table 1. Descriptive Statistics for Four Major Economic Assumption Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>No. of Obs.</th>
<th>Mean</th>
<th>Median</th>
<th>Standard Deviation</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inflation Rate</td>
<td>158</td>
<td>1.0547%</td>
<td>0.8987%</td>
<td>0.8809%</td>
<td>-0.5530%</td>
<td>3.9267%</td>
</tr>
<tr>
<td>Investment Return Rate</td>
<td>158</td>
<td>1.6260%</td>
<td>1.6107%</td>
<td>0.6988%</td>
<td>0.4705%</td>
<td>3.6155%</td>
</tr>
<tr>
<td>Wage Rate</td>
<td>158</td>
<td>1.3118%</td>
<td>1.3332%</td>
<td>0.7307%</td>
<td>-1.3858%</td>
<td>3.2323%</td>
</tr>
<tr>
<td>Unemployment Rate</td>
<td>158</td>
<td>5.7600%</td>
<td>5.6850%</td>
<td>1.4780%</td>
<td>2.6640%</td>
<td>10.256%</td>
</tr>
</tbody>
</table>

The time series plot of these four major economic assumptions is presented in Figure 1. In that figure, we use the logarithmic transformation of the investment return rate and the unemployment rate. As shown in the investigation in Frees et al. (1997), these economic assumption variables are likely to have not only autocorrelation but also interrelationships, either contemporaneous or time-lagged. These relationships are consistent with the characteristics of vector autoregressive time series models discussed in the previous sections and in the literature of macroeconomics. Multivariate modeling has also been identified as a worthwhile approach in the recommendation of the Social Security technical panel in 1991. Although some series seem to be non-stationary, particularly the investment return rate for which the economic turmoil in the mid-1970s to early 1980s possibly accounts, it is too arbitrary to make the distinction between the stationary and the non-stationary variables. Fortunately, the common attributes of non-stationarity seem to exist among the variables, so the non-stationary variables could be simply decomposed into the linear combination of the common non-stationary components among those variables and the stationary components that deviate from this common non-stationarity. It is more conservative to study the characteristics of non-stationarity through the decomposition of common non-stationary features and various remaining stationary features, rather than arbitrarily differencing any of the variables. This is the principle of co-integration, which we introduce in this paper.

We use the model fitting techniques discussed in the Section 4. If we fit the data into the model with a larger autoregression order, then the number of parameters increases dramatically. The model becomes too complicated and inefficient to explore the underlying economics system. Thus, we only consider the VECM models with autoregression order one and two. In Table 2, we compare the trace statistics with 99% percentiles of its asymptotic distribution. The specification of co-integrating rank is based on a sequence of trace tests of \( H_0: \text{rank}(\Pi) = r \) for values of \( r = 0, \ldots, k - 1 \), and an appropriate value of \( r \) can be chosen as the smallest value such that hypothesis \( H_0 \) is not rejected. We use * to denote the appropriate co-integrating rank for each
Thus, we suggest that the co-integrating ranks are three for both models VECM(1) and VECM(2).

Figure 1. Time Series Plot of Four Major Economic Assumption Variables

Table 2. Co-Integrating rank and trace statistics for VECM(1) and VECM(2)

<table>
<thead>
<tr>
<th>$r$</th>
<th>Trace(99%)</th>
<th>$\lambda_{\max}(p=1)$</th>
<th>$\lambda_{\max}(p=2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>53.91</td>
<td>224.29</td>
<td>159.15</td>
</tr>
<tr>
<td>1</td>
<td>34.87</td>
<td>64.20</td>
<td>71.80</td>
</tr>
<tr>
<td>2</td>
<td>19.09</td>
<td>17.88</td>
<td>26.10</td>
</tr>
<tr>
<td>3</td>
<td>6.64</td>
<td>*3.22</td>
<td>*5.29</td>
</tr>
</tbody>
</table>

The information criteria in Table 3 help us select the model with the appropriate autoregression order. The model with autoregression order two and co-integrating rank three,
that is, the VECM(2) with $r = 3$, has smaller values of all three information criteria than the model with only autoregression order one and co-integrating rank three, so we prefer to select VECM(2) with $r = 3$. In other words, we suggest that there is only one common non-stationary characteristic in the system of these four economic variables that all the non-stationary properties in the system come from.

Table 3. Information Criteria for VECM(1) with $r=3$ and VECM(2) with $r=3$

<table>
<thead>
<tr>
<th>Information Criteria</th>
<th>$p=1$</th>
<th>$p=2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$AIC$</td>
<td>-7.999</td>
<td>-8.292</td>
</tr>
<tr>
<td>$HQ$</td>
<td>-8.167</td>
<td>-8.597</td>
</tr>
<tr>
<td>$BIC$</td>
<td>-7.974</td>
<td>-8.248</td>
</tr>
</tbody>
</table>

If we use $X_t = (X_{1,t}, X_{2,t}, X_{3,t}, X_{4,t})'$ for the inflation rate, the investment return rate, the wage rate, and the unemployment rate respectively, then the suggested model is VECM(2) with co-integrating rank three: $\Delta X_t = \Pi X_{t-1} + \Gamma_1 \Delta X_{t-1} + \mu + \epsilon_t$, where the parameter estimates are

$$\hat{\Pi} = \hat{\alpha} \hat{\beta}' = \begin{bmatrix} -0.316 & 0.223 & 0.146 & -0.024 \\ 0.032 & -0.084 & 0.086 & 0.020 \\ 0.407 & -0.064 & -0.980 & 0.053 \\ 0.162 & 0.158 & -0.233 & -0.125 \end{bmatrix}$$

$$\hat{\Gamma}_1 = \begin{bmatrix} -0.088^* & -0.111^* & -0.074^* \\ -0.026^* & 0.038^* & 0.027^* \\ 0.368^* & -0.169^* & 0.103^* \\ 0.114^* & 0.081^* & -0.101^* \end{bmatrix} \begin{bmatrix} 1.385 & 0.079 & -2.417 & -0.191 \\ 1.151 & -0.702 & 0.568 & -0.205 \\ 0.889 & -2.048 & 0.043 & 0.856 \end{bmatrix}, \quad (5.1)$$

$$\hat{\mu} = \begin{bmatrix} -0.290^* & 1.157^* & -0.101 & -0.024 \\ -0.058^* & 0.046 & -0.044^* & -0.033 \\ -0.189^* & 0.593 & -0.022 & -0.420^* \\ -0.052 & -0.237 & 0.041 & 0.310^* \end{bmatrix}, \quad (5.2)$$

We use * to denote the statistically significant parameter estimates. With three co-integrating vectors, the interpretation in terms of long-run equilibria is no longer straightforward as in Examples 1 and 2 in subsection 3.5. The problem is that any linear combination of two co-
integrating relations above will preserve the stationarity property. In the co-integration literature this problem is often emphasized by noting that only the space spanned by \( \beta \) is uniquely determined. In the sense of economics, the relation \( \beta \mathbf{X} = 0 \) is defined as the long-run equilibrium relationship, that is the underlying economic relationships. The economic agents in this system react to the disequilibrium error \( \beta \mathbf{X}_{t-1} \) through the adjustment coefficient \( \alpha \), to bring back the variables on the right track, that is, such that they satisfy the economic relationships. We suggest that there exist underlying long-run equilibrium relationships of these four economic assumption variables of the Social Security trust funds:

\[
\begin{align*}
1.385X_1 + 0.079X_2 - 2.417X_3 - 0.191X_4 &= 0 \\
1.151X_1 - 0.702X_2 + 0.568X_3 - 0.205X_4 &= 0 \\
0.889X_1 - 2.048X_2 + 0.043X_3 + 0.856X_4 &= 0.
\end{align*}
\]

Any other long-run equilibrium relationships among these economic assumption variables are able to be decomposed into linear combinations of these three economic relationships. Statistically speaking, they will fall into the space spanned by these three co-integrating vectors. To bring back the variables to the long-run equilibrium relationships, the economic agents in this system react to the disequilibrium error \( \beta \mathbf{X}_{t-1} \) through the adjustment coefficient

\[
\hat{\alpha} = \begin{bmatrix}
-0.088^* & -0.111^* & -0.074^* \\
-0.026^* & 0.038^* & 0.027^* \\
0.368^* & -0.169^* & 0.103^* \\
0.114^* & 0.081^* & -0.101^* \\
\end{bmatrix}
\]

but diffused by the short-run matrix \( \hat{I}_t \).

Figure 2 provides the selected simulated percentiles for each considered economic assumption variable in the panels, respectively. The dashed lines are the 97.5\(^{th}\), 65\(^{th}\), 35\(^{th}\), and 2.5\(^{th}\) percentiles for the suggested model VECM(2) with the co-integrating rank \( r = 3 \), and the dotted lines are those percentiles for the model VECM(1) with the co-integrating rank \( r = 3 \). The solid lines in 1993 and 1994 are the out-of-sample quarterly experience of these variables. The three solid lines from 1995 through 2004 are the low-cost, the intermediate, and the high-cost economic assumptions from the Social Security projection in the 1995 annual report. These panels show that most of the three Social Security alternative assumptions are within the 30% prediction intervals, except the unemployment rate. Compared with these alternative assumptions, the wage rate seems to be too low and the unemployment rate is likely to be too high. In the future, we hope to investigate the causes of these differences.
Figure 2 Selected Percentiles of Four Major Economic Assumption Variables and Three Alternative Sets of Assumptions.
To forecast the Social Security trust funds, we use the recursive equation in the equation (4.9). First, we make the model validation, from the first quarter of 1993 through the fourth quarter of 1994. In this period of two years, we compare fund forecast simulation distributions based on our suggested model of these four economic assumption variables with the fund forecasts based on out-of-sample experience of these variables. We collect the following key data from the 1995 Annual Report of the Board of Trustees of Social Security. Let \( T \) denote the fourth quarter of 1992. At the end of 1992, the funds have accumulated to \( \hat{F}_T = 331.473 \) billion. In 1992, \( \hat{M}_T = 132.7 \) million covered workers who made \( 311.128 \) billion premium contributions to the fund, so the average quarterly contributions \( \hat{C}_T \) from each worker are \( 586.15 \). At the same time, \( \hat{N}_T = 41.029 \) million beneficiaries who obtained \( 285.995 \) billion and the average quarterly benefits \( \hat{B}_T \) are \( 1742.64 \). The numbers of covered workers and beneficiaries are not related to our multivariate model for the economic assumption variables. We use the interpolation method to obtain the numbers of covered workers \( \hat{\dot{M}}_{T,t} \) and beneficiaries \( \hat{\dot{N}}_{T,t} \) from the data available in the Social Security annual report for the eight quarters of the model validation period.

In the model validation period, Figure 3 shows that the simulated prediction intervals of VECM(2) are narrower than the intervals of VECM(1), which gives us more information to support the suggested model, other than the information criteria, such as AIC. We also find that the forecasts by the held-out experience on the economic assumption variables are approximately at the 10th percentile of the simulated forecasts, and this is relatively lower than our simulated forecasts. This could be either a warning sign of under-funding or the over-optimism of this forecasting model.

Later, we make a comparison of 10 years (1995-2004) fund forecast simulation distributions based on our suggested model of the four economic assumption variables and the fund forecasts by the three Social Security alternative sets of economic assumption variables. We use the 1994 data available from the 1995 Social Security annual report. Let \( T_i \) denote the end of 1994. The key data are accumulated assets \( \hat{F}_{T_i} = 436.385 \) billion, covered workers \( \hat{\dot{M}}_{T_i} = 138.786 \) million, average annual contributions \( \hat{C}_{T_i} = 2483.64 \), beneficiaries \( \hat{\dot{N}}_{T_i} = 42.517 \) million, and the average annual benefits \( \hat{B}_{T_i} = 7451.42 \). The future numbers of covered workers \( \hat{\dot{M}}_{T,i} \) and beneficiaries \( \hat{\dot{N}}_{T,i} \) are not related to our multivariate model for the
economic assumption variables and are interpolated from the estimates in the 1995 Social Security annual report.

Figure 3. Simulated Percentiles of Fund Forecasts for Model Validation

![Figure 3](image)

Figure 4 presents comparisons among the simulated forecasts and the forecasts based on intermediate demographic assumptions and three alternative economic assumptions. Similar to the results of the model validation period in Figure 3, the prediction intervals of the model VECM(2) are also narrower than those of the model VECM(1) in panel (a) of Figure 4. This evidence is consistent with the model selection based on the information criteria. In panel (b) of Figure 4, the forecasts by three alternative sets of economic assumptions are close to the 25th, 50th, and 75th percentiles of our forecasts, but higher in the first five years and lower in the second five years. These results are not only in accordance with the projection of government agencies but also provides more conservative and narrower prediction intervals than those in Frees et al. (1997), which are too optimistic and wider. Further, the projections are closer to the estimates shown in the 1995 Annual Report of the Board of the Trustees of the Social Security than the forecasts in Frees et al. (1997). For example, in year 2004, the estimate in 1995 Annual Report is about $1,400 billion, and the 95% confidence interval of our fund projection ranges from $200 to $1,600 billion instead of the $5 to $35 million range proposed by Frees et al. (1997).
Figure 4. Comparison of Assumption Alternatives and Fund Forecasts

Panel (a) Assumption Alternatives and Simulated 2.5th and 97th Percentiles of Forecasts

- 97.5th of VECM(1) with \( r = 3 \)
- 2.5th of VECM(1) with \( r = 3 \)
- 97.5th of VECM(2) with \( r = 3 \)
- 2.5th of VECM(2) with \( r = 3 \)
- Low-Cost Economic Assumptions
- Intermediate Economic Assumptions
- High-Cost Economic Assumptions

Panel (b) Assumption Alternatives and Simulated 25th, 50th, and 75th Percentiles of Forecasts

- 75th of VECM(1) with \( r = 3 \)
- 50th of VECM(1) with \( r = 3 \)
- 25th of VECM(1) with \( r = 3 \)
- 75th of VECM(2) with \( r = 3 \)
- 50th of VECM(2) with \( r = 3 \)
- 25th of VECM(2) with \( r = 3 \)
- Low-Cost Economic Assumptions
- Intermediate Economic Assumptions
- High-Cost Economic Assumptions
6. Conclusion

In the earlier sections of this paper, we forecast the Social Security trust funds by a co-integrated multivariate time series model of the four major economic assumptions. This approach not only establishes the interrelationships but also clarifies the co-movements among the economic assumptions. Because the co-integrating rank is identified as three, we suggest that there is only one common non-stationary figure in the system of these four major economic assumption variables. The autoregression order is two, so even the system of these economic variables will be adjusted back to the long-run equilibrium relationship but diffused by the short-run matrix from the second autoregression order. We find that most of the three Social Security alternative sets of economic assumptions are within the 30% prediction intervals except the unemployment rate. Under the intermediate population assumption, we also find that the forecasts by these alternative economic assumptions are very close to our 25th, 50th, and 75th percentiles of forecasts, but higher in the first five years and lower in the second five years. The results not only confirm the expert opinions in the Social Security projections but also provide a better prediction than those in Frees et. al. (1997) because of the narrower prediction intervals.

The economic assumption variables are easily affected by special events or circumstances, such as oil crises, which are other than the variability apprehended by the modeling tools which we use in this paper. These special events or circumstances could be considered as exogenous shocks to the economic system that we study. The inclusion of exogenous shocks in the model fitting not only would improve the statistical significance and parsimony of the parameter estimates but also help us understand how the exogenous shocks affect the economic assumption variables. In their discussions of Frees et. al. (1997), Foster and Savord pointed out that the model of the economic assumption variables should account for exogenous shocks. In the future, we intend to investigate the model incorporated with exogenous shocks.

References


Johansen, S. and Juselius, K. (1990). The full information maximum likelihood procedure for Inf-


