KESDP 07-8

A New Method for Estimating Economic Models with General Time-varying Structures *

Mikio Ito†

March 18, 2009

Abstract

Shown is a new method for estimating linear models with general time-varying structures such as the State Space Model based on the idea that the models can be represented as a classical regression model. The parameters are all estimated by OLS or GLS. An application of the smoothing to a time-varying AR model is presented.

1 Introduction

The State Space Model has been familiar to researchers who study control theory and signal extraction in engineering since Kalman (1960) published a pioneering paper. In his paper Kalman presented the estimation methods, which we call Kalman filtering, Kalman smoothing and Kalman forecasting, to solve the problem of signal extraction for non stationary time series data; the problem had been regarded as too tough to solve since Wiener (1949) and Kolmogorov (1941) had attacked. His methods, estimating states in each period of the system iteratively, could implement accurate signal extraction subject to computers in his days, with much poorer memory and MPU power than today. Kalman developed his methods crucially on the State Space Model.

Few statisticians had contributed to theories based on the State Space Model from 1960 to early 1980's: an exception, Duncan and Horn (1972)'s paper providing another proof of

*This draft is based on a working paper circulated as "A Direct Method of Smoothing for State Space Models(*Joutai-Kuukan Moderu no Chokusetsuhou niyoru Heikatsuka* in Japanese)", Keio Economic Society, Discussion Paper Series, KESDP 06-5.

†Faculty of Economics, Keio University, Email: ito@econ.keio.ac.jp

Kalman filtering in respect of ordinary linear regression theory. Sorenson (1970) reviews filtering theories from Gauss to Kalman, pointing that any filtering theory has its root as Gauss's least square method applied to linear models. This article owes much to their works.

As to econometrics, few econometricians has developed the theory of Kalman filter so as to be easy to use for economists. As an exception, Cooley and Prescott (1973), who developed a model to depict continuous structural changes, applied a State Space Model 40 years ago.

One should never jump into conclusion that Kalman's method based on the State Space Model has been useless for most econometric analysis; econometricians could have used the State Space Model in many situations since it is just an extended linear regression model.

The author lists the reasons why few applications of the State Space Model to econometric analyses have been reported for long time as follow: (1)most economists have supposed that structural changes in real economy are not gradual but discontinuous, (2)the statistical model in which hypothesis tests are unavailable has no interest for most econometricians, (3)economists have thought that Kalman smoothing, which corresponds to usual analysis in econometrics, is relatively complex, (4) the assumption that covariance matrices are known could be implausible in econometric analysis when the State Space Models were applied in econometric analysis.

The author shows that Kalman smoothing is easy to apply to usual econometric analysis because of the recent development of computers. He represents a class of general linear models with time-varying structure, which includes the conventional State Space Models, and shows that whether a time-varying structure is gradual or discontinuous, parameters in such a model can be estimated by OLS or GLS. Since the essence of Kalman smoothing with a fixed interval for the State Space Model is a orthogonal projection of observable variables in observation equations to the data space (the sample space), which is interpreted as the given information set, we naturally adopt OLS and GLS if it is possible.

The power of recent computers would change our traditional idea on the method for smoothing; few econometricians have imagined that recent personal computers can compute the inverse of a nonsigular matrix with 2000×2000 components or the generalized inverse of a matrix with several million components in a minute with practical precision. There is no reason to adopt conventional iterative techniques such as Kalman smoothing with a fixed interval or Levinson method in discrete signal extraction problems; the techniques have helped researchers to save memory of their computers for over 40 years.

This article is organized as follows. In Section 2, the author illustrates how the State Space Model, which we can interpret as a linear model with time-varying parameters, can be transformed to a linear model. Furthermore, the author generalizes the linear regression model so as to be one allowing more general time-varying structure than the usual State Space Model. In Section 3, the author shows theorems guaranteeing the generality of our model. Section 4 shows the power of our smoothing method for the usual State Space Model, presenting an application of our method to estimate parameters of time-varying AR model. Section 5 is located for conclusion.

2 Linear Model with Time-Varying Parameters

This section shows how we represent the State Space Model as a linear regression model with time-varying parameters and introduces a linear model allowing more general time-varying structures.

2.1 State Space Model and smoothing with a fixed interval

Kalman (1960) used the State Space Model to solve the tough problem of signal extraction, which two great mathematicians in the twentieth century, Wiener and Kolmogorov, failed to achieve a complete solution. Their approach by using Fourier analysis, a typical frequencydomain analysis, was inappropriate to the problem with non stationary data; Kalman's one, which was categorized as a time-domain analysis, could successfully treat non stationary data by using the State Space Model.

The State Space Model consists of two equations as follows:

$$\boldsymbol{y}_t = X_t \boldsymbol{\beta}_t + \boldsymbol{u}_t, \quad \boldsymbol{u}_t \overset{iid}{\sim} \mathcal{N}(\boldsymbol{0}, R_t) \tag{1}$$

$$\boldsymbol{\beta}_{t+1} = \Phi_{t+1,t} \boldsymbol{\beta}_t + \boldsymbol{v}_t, \quad \boldsymbol{v}_t \overset{iid}{\sim} \mathcal{N}(\boldsymbol{0}, Q_t), \tag{2}$$

where the supscript t expresses a period, each $\boldsymbol{y}_t$ is n-dimensional vector, each $\boldsymbol{\beta}_t$ is a m-dimensional vector, and X_t is a $n \times m$ matrix. We call the equation (1) the observation equation, and the equation (2) the transition equation or the state equation. The matrices, $X_t, \Phi_{t+1,t}, R_t, Q_t$ are assumed to be known. We assume

$$(\forall t_1)(\forall t_2) \; E[\boldsymbol{u}_{t_1} \boldsymbol{v}'_{t_2}] = \boldsymbol{O}.$$

Kalman considered the problem of estimating the state variables, $\boldsymbol{\beta}_\tau$, $(\tau = 1, 2, \cdots, t)$, at each period based on the information $\mathcal{Y}_t$ and the prior distribution, $\boldsymbol{\beta}_0 \overset{iid}{\sim} \mathcal{N}(\boldsymbol{\beta}_0, P_0)$, both of which are given, where $\mathcal{Y}_t = \sigma(\{\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_t\})$ is the σ-field generated by the stochastic process, $Y_t = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_t\}$.

We call this problem the State Space estimation. There are three kinds of the estimation: (1) forecasting, (2) filtering and (3) smoothing. Each estimation corresponds to what information is based for estimating $\boldsymbol{\beta}_t$: (1) $\mathcal{Y}_\tau, \tau < t$, (2) $\mathcal{Y}_t$ and (3) $\mathcal{Y}_T$, where T denotes the terminal period.

The following remarks might be useful.

- the equations (1) and (2) generate a Gaussian process, $\{\boldsymbol{y}_t\}$, which is not always stationary.
- The covariance matrices are time-varying in general.
- Under the assumption
$$(\forall t)[\Phi_{t+1,t} = I \;\&\; \boldsymbol{v}_t = \boldsymbol{0})],$$
the State Space estimation is just an OLS(ordinary least square) estimation for a conventional linear regression model.
- When the signal is generated by the equations (1) and (2), the solution of the filtering problem formulated by Wiener and Kolmogorov, $\boldsymbol{s}_t$, is
$$\hat{\boldsymbol{y}}_{t|t} = X_t \hat{\boldsymbol{\beta}}_{t|t},$$
where $\hat{\boldsymbol{\beta}}_{t|t}$ is the solution of Kalman filtering, more generally, $\hat{\boldsymbol{\beta}}_{t_1|t_2}$ denotes the optimal estimator at period t_1 based on the information available at period t_2.
- Modified procedures are taken for forecasting and smoothing.

The author here gives a brief summary of Kalman's theory. Kalman (1960) used the orthogonal projection to solve the filtering problem in the same way as Wiener did; Wiener did not consider the data generating process (DGP), corresponding to the equation (2). Kalman successfully derived the famous iterative procedure, Kalman filtering.

One can derive the solution of filtering problem as follows. Since the orthogonal projection assures that the solution, $\hat{\boldsymbol{\beta}}_{t|t}$, is a linear combination of the forecasting without using the new information $\boldsymbol{y}_t$,

$$\hat{\boldsymbol{\beta}}_{t|t-1} = \Phi_{t,t-1} \hat{\boldsymbol{\beta}}_{t-1|t-1} \tag{3}$$

and the residual,

$$\boldsymbol{r}_t = \boldsymbol{y}_t - X_t \hat{\boldsymbol{\beta}}_{t|t-1} = \boldsymbol{y}_t - X_t \Phi_{t,t-1} \hat{\boldsymbol{\beta}}_{t-1|t-1},$$

we have

$$\hat{\boldsymbol{\beta}}_{t|t} = \Phi_{t,t-1}\hat{\boldsymbol{\beta}}_{t-1|t-1} + K_t(\boldsymbol{y}_t - X_t\Phi_{t,t-1}\hat{\boldsymbol{\beta}}_{t-1|t-1}). \tag{4}$$

The matrix K_t, Kalman gain, is chosen in such a way as

$$E[(\boldsymbol{\beta}_t - \hat{\boldsymbol{\beta}}_{t|t})'(\boldsymbol{\beta}_t - \hat{\boldsymbol{\beta}}_{t|t})]$$

is minimized. Through tedious algebraic operations, one derives

$$K_t = P_{t|t-1}X_t'(X_tP_{t|t-1}X_t' + R_t)^{-1}, \tag{5}$$

where $P_{t|t-1}$ is the covariance matrix of $\hat{\boldsymbol{\beta}}_{t|t-1}$, that is,

$$P_{t|t-1} := E[(\boldsymbol{\beta}_t - \hat{\boldsymbol{\beta}}_{t|t-1})(\boldsymbol{\beta}_t - \hat{\boldsymbol{\beta}}_{t|t-1})'].$$

By using the equation (3) we derive

$$P_{t|t-1} = \Phi_{t|t-1}P_{t-1|t-1}\Phi'_{t|t-1} + Q_{t-1}. \tag{6}$$

$P_{t|t}$ denotes the covariance matrix of $\hat{\boldsymbol{\beta}}_{t|t}$;

$$P_{t|t} := E[(\boldsymbol{\beta}_t - \hat{\boldsymbol{\beta}}_{t|t-1})(\boldsymbol{\beta}_t - \hat{\boldsymbol{\beta}}_{t|t-1})'].$$

Finally, we have an equation, which we intreprete as a revising procedure:

$$P_{t|t} = P_{t|t-1} - K_tX_tP_{t|t-1}. \tag{7}$$

The Kalman filter is the revising algorithm which consists of (4), (5), (6), (7).

There are few cases where Kalman filtering is solely needed for econometric analysis. Since most estimations of parameters of the statistical model in the usual econometric analyses correspond to the smoothing, in the sense that any estimation is made on the base of the observation, $Y_T = \{\boldsymbol{y}_0, \boldsymbol{y}_1, \cdots, \boldsymbol{y}_T\}$ in a sample period, $\mathcal{T} = \{0, 1, \cdots, T\}$, that is, the estimation of $\hat{\boldsymbol{\beta}}_t$, $(t < T)$, the smoothing for a fixed interval. If we use familiar notations in the filtering theory, $\hat{\boldsymbol{\beta}}_t = \hat{\boldsymbol{\beta}}_{t|T}$.

For the readers' convenience, we present the Kalman smoothing algorithm for a fixed interval. It consists of the following two steps:

1. Derive the estimates, by Kalman filtering, $\hat{\boldsymbol{\beta}}_{t|t-1}$, $\hat{\boldsymbol{\beta}}_{t|t}$ and each of the estimates of their covariance matrices, $P_{t|t-1}$, $P_{t|t}$ for $t \in \{0, 1, \cdots, T\}$.

2. Derive $\hat{\boldsymbol{\beta}}_{T-1|T}, \cdots, \hat{\boldsymbol{\beta}}_{0|T}$ from $\hat{\boldsymbol{\beta}}_{T|T}$, in the reverse direction, by iterative substitutions in the following way:

$$\hat{\boldsymbol{\beta}}_{t|T} = \hat{\boldsymbol{\beta}}_{t|t} + C_t(\hat{\boldsymbol{\beta}}_{t+1|T} - \hat{\boldsymbol{\beta}}_{t+1|t}) \tag{8}$$

$$C_t = P_{t|t}\Phi_{t+1|t}P^{-1}_{t+1|t} \tag{9}$$

$$P_{t|T} = P_{t|t} + C_t(P_{t+1|T} - P_{t+1|t})C'_T \tag{10}$$

We stress that one must assume the covariance matrices, R_t and Q_t in each period, are known when one applies the above smoothing algorithm. However, such an assumption has been supposed implausible in real econometric analyses. The assumption as well as the complex procedure above of the Kalman smoothing makes the State Space Model unfamiliar to economists. A practical smoothing method is needed, if possible, on the basis of the classical regression theory.

2.2 Smoothing as a regression: Durbin-Koopman Regression

Following Durbin and Koopman (2001), we show a smoothing method for the State Space Model, (1) and (2), They expand the equations in each period to a linear equation with a matrix of huge size. All the matrices $X_t, \Phi_{t+1,t}, R_t, Q_t$ are assumed to be known as before. We assume

$$(\forall t_1)(\forall t_2)\ E[\boldsymbol{u}_{t_1}\boldsymbol{v}'_{t_2}] = \boldsymbol{O}.$$

The sample periods are $\mathcal{T} = \{0, 1, \cdots, T\}$ and the state variables at the initial period, $\boldsymbol{\beta}_0$, follow a normal distribution given:

$$\boldsymbol{\beta}_0 \overset{iid}{\sim} \mathcal{N}(\bar{\boldsymbol{\beta}}_0, P_0).$$

For the convenience, we summarize the dimensions of the vectors and matrices in our State Space Model in Table 1.

Table 1: dimensions of State Space Model

vector		matrix	
$\boldsymbol{y}_t$	n	X_t	$n \times m$
$\boldsymbol{\beta}_t$	m	$\Phi_{t+1,t}$	$m \times m$
$\boldsymbol{u}_t$	n	R_t	$n \times n$
$\boldsymbol{v}_t$	m	Q_t	$m \times m$
$\boldsymbol{\beta}_0$	m	P_0	$m \times m$

By defining the following matrices and vectors, we represent a linear equation with a matrix of huge size in place of the equation (1).

$$\boldsymbol{y} = \begin{pmatrix} \boldsymbol{y}_0 \\ \vdots \\ \boldsymbol{y}_T \end{pmatrix}, \quad X = \begin{pmatrix} X_0 & \cdots & \boldsymbol{O} & \boldsymbol{O} \\ \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{O} & \cdots & X_T & \boldsymbol{O} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_0 \\ \vdots \\ \boldsymbol{\beta}_T \\ \boldsymbol{\beta}_{T+1} \end{pmatrix},$$

$$\boldsymbol{u} = \begin{pmatrix} \boldsymbol{u}_0 \\ \vdots \\ \boldsymbol{u}_T \end{pmatrix}, \quad R = \begin{pmatrix} R_0 & \cdots & \boldsymbol{O} \\ \vdots & \ddots & \vdots \\ \boldsymbol{O} & \cdots & R_T \end{pmatrix}$$

The observation equation (1) can be expressed as follows:

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{u}, \quad \boldsymbol{u} \overset{iid}{\sim} \mathcal{N}(\boldsymbol{0}, R). \tag{11}$$

We can represent the transition equation, (2), as a single linear equation by defining the matrices and the vectors as follows:

$$\Phi = \begin{pmatrix} I & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & & \boldsymbol{O} & \boldsymbol{O} \\ \Phi_{1,0} & I & \boldsymbol{O} & \boldsymbol{O} & & \boldsymbol{O} & \boldsymbol{O} \\ \Phi_{2,1}\Phi_{1,0} & \Phi_{2,1} & I & \boldsymbol{O} & & \boldsymbol{O} & \boldsymbol{O} \\ \Phi_{3,2}\Phi_{2,1}\Phi_{1,0} & \Phi_{3,2}\Phi_{2,1} & \Phi_{3,2} & I & & \boldsymbol{O} & \boldsymbol{O} \\ \vdots & \vdots & \vdots & \vdots & & \ddots & \vdots \\ \Phi_{T,T-1}\cdots\Phi_{1,0} & \Phi_{T,T-1}\cdots\Phi_{2,1} & \Phi_{T,T-1}\cdots\Phi_{3,2} & \Phi_{T,T-1}\cdots\Phi_{4,3} & \cdots & \Phi_{T,T-1} & I \end{pmatrix},$$

$$\boldsymbol{\beta}_0^* = \begin{pmatrix} \boldsymbol{\beta}_0 \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \end{pmatrix}, \quad J = \begin{pmatrix} \boldsymbol{O} & \boldsymbol{O} & \cdots & \boldsymbol{O} \\ I & \boldsymbol{O} & \cdots & \boldsymbol{O} \\ \boldsymbol{O} & I & \cdots & \boldsymbol{O} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{O} & \boldsymbol{O} & \cdots & I \end{pmatrix},$$

$$\boldsymbol{v} = \begin{pmatrix} \boldsymbol{v}_0 \\ \vdots \\ \boldsymbol{v}_T \end{pmatrix}, \quad Q = \begin{pmatrix} Q_0 & \cdots & \boldsymbol{O} \\ \vdots & \ddots & \vdots \\ \boldsymbol{O} & \cdots & Q_T \end{pmatrix}$$

We have a linear equation with a huge matrix:

$$\boldsymbol{\beta} = \Phi(\boldsymbol{\beta}_0^* + J\boldsymbol{v}), \quad \boldsymbol{v} \overset{iid}{\sim} \mathcal{N}(\boldsymbol{0}, Q) \tag{12}$$

in place of (2).

Durbin and Koopman derive

$$\boldsymbol{y} = X\Phi\boldsymbol{\beta}_0^* + X\Phi J\boldsymbol{v} + \boldsymbol{u} \tag{13}$$

from (12) and (11). This equation shows that the observation vector $\boldsymbol{y}$ is a linear function of the initial vector $\boldsymbol{\beta}_0^*$ and the disturbance vectors, $\boldsymbol{v}$ and $\boldsymbol{u}$. After cumbersome algebraic operations, they derive two fundamental equations:

$$E[\boldsymbol{y}] = X\Phi\bar{\boldsymbol{\beta}}_0^*$$

and

$$Var[\boldsymbol{y}] = X\Phi(P_0^* + JQJ')\Phi' X' + R,$$

where

$$\bar{\boldsymbol{\beta}}_0^* = \begin{pmatrix} \bar{\boldsymbol{\beta}}_0 \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \end{pmatrix}, \quad P_0^* = \begin{pmatrix} P_0 & \boldsymbol{O} & \cdots & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} & \cdots & \boldsymbol{O} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{O} & \boldsymbol{O} & \cdots & \boldsymbol{O} \end{pmatrix}.$$

The author calls the equation (11) Durbin-Koopman Regression. Durbin-Koopman successfully derived the smoother $\hat{\boldsymbol{\beta}}$ by thoroughly examining their matrix representation of the State Space Model. However, we cannot compute $\hat{\boldsymbol{\beta}}$ by using the observation $\boldsymbol{y}$ directly, while Durbin-Koopman Regression (11) apparently has the conventional regression form. This is because they assume that the covariance matrices of both the transition equation and the observation equation are known. Thus, we conclude that Durbin-Koopman Regression cannot give a practical smoothing procedure for most econometric analyses.

2.3 Direct Method of Smoothing in the State Space Model: Ito Regression

In this subsection, we translate the State Space Model, (1) and (2), to a linear equation, in which we can estimate the state vector $\boldsymbol{\beta}$ without any knowledge of the covariance matrices, $R_t, Q_t, (t = 1, 2, \cdots, T)$. We assume that the matrices $X_t, \Phi_{t+1,t}$ are known and that

$$(\forall t_1)(\forall t_2)\ E[\boldsymbol{u}_{t_1}\boldsymbol{v}_{t_2}'] = \boldsymbol{O}.$$

However, we do not assume that the matrices, R_t, Q_t, are known.

The sample periods $\mathcal{T} = \{0, 1, \cdots, T\}$ and the state vector at the initial period, $\boldsymbol{\beta}_0$, follows a normal distribution:

$$\boldsymbol{\beta}_0 \overset{iid}{\sim} \mathcal{N}(\bar{\boldsymbol{\beta}}_0, P_0).$$

Note that $\boldsymbol{y}$ denotes the stacked vector with $\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_T$.

Our representation is simple and straightforward: (1) for the observation equation is the same as Durbin-Koopman Regression, we have

$$\begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_T \end{pmatrix} = \begin{pmatrix} X_1 & & & \boldsymbol{O} \\ & X_2 & & \\ & & \ddots & \\ \boldsymbol{O} & & & X_T \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_T \end{pmatrix} + \begin{pmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \\ \vdots \\ \boldsymbol{u}_T \end{pmatrix} \tag{14}$$

and (2) for the transition equation, we have

$$\begin{pmatrix} -\Phi_{1,0}\bar{\boldsymbol{\beta}}_0 \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} -I & & & \boldsymbol{O} \\ \Phi_{2,1} & -I & & \\ & \ddots & \ddots & \\ \boldsymbol{O} & & \Phi_{T,T-1} & -I \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_T \end{pmatrix} + \begin{pmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \\ \vdots \\ \boldsymbol{v}_T \end{pmatrix}. \tag{15}$$

Now we show that these equations (14) and (15) enable us to treat the State Space Model as the usual classical regression one under the assumptions that the matrices, $X_1, X_2, \cdots, X_t$, $\Phi_{1,0}, \Phi_{2,1}, \cdots, \Phi_{T,T-1}$ and the prior vector $\bar{\boldsymbol{\beta}}_0$ are all known.

We have the following equation (16), which we call Ito regression of the State Space Model, by simply gathering (14) and (15).

$$\begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_T \\ -\Phi_{1,0}\bar{\boldsymbol{\beta}}_0 \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} X_1 & & & \boldsymbol{O} \\ & X_2 & & \\ & & \ddots & \\ \boldsymbol{O} & & & X_T \\ -I & & & \boldsymbol{O} \\ \Phi_{2,1} & -I & & \\ & \ddots & \ddots & \\ \boldsymbol{O} & & \Phi_{T,T-1} & -I \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_T \end{pmatrix} + \begin{pmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \\ \vdots \\ \boldsymbol{u}_T \\ \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \\ \vdots \\ \boldsymbol{v}_T \end{pmatrix} \tag{16}$$

When we estimate the state vectors $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_T$ at one time based on the observation $\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_T$ by OLS or GLS, familiar with economists, we can regard this estimate $\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \cdots, \hat{\boldsymbol{\beta}}_T$ as the the smoother with a fixed interval of the original State Space Model. If we use usual notations in the theory of Kalman filter, $\hat{\boldsymbol{\beta}}_t = \boldsymbol{\beta}_{t|T}$. This is because the Aitken theorem in the theory of linear regression assures that the estimator of OLS or GLS is always an unbiased estimator based on the data $Y_T = \{\boldsymbol{y}_0, \boldsymbol{y}_1, \cdots, \boldsymbol{y}_T\}$. (See Ruud (2000)[p.432].) Note here that this unbiasedness of the estimator by linear regressions is independent of the covariance structure of the State Space Model.

When the Gauss-Markov theorem holds, that is, when any column vector of the regressor,

$$Z = \begin{pmatrix} X_1 & & & \boldsymbol{O} \\ & X_2 & & \\ & & \ddots & \\ \boldsymbol{O} & & & X_T \\ -I & & & \boldsymbol{O} \\ \Phi_{2,1} & -I & & \\ & \ddots & \ddots & \\ \boldsymbol{O} & & \Phi_{T,T-1} & -I \end{pmatrix}$$

is statistically independent of $\boldsymbol{y}$, one can use conventional hypothesis tests such as t-test, F-test and so on. However, note that there is no consistent estimator of $\boldsymbol{\beta}$ since the number of parameters in the State Space Model goes to infinity as the sample size goes to inifinity.

Our estimator of the state vector $\hat{\boldsymbol{\beta}}$ is a linear function of the data $\boldsymbol{y}$. When we use OLS, for example, the estimator is expressed as

$$\hat{\boldsymbol{\beta}} = (Z'Z)^{-1}Z'\boldsymbol{y}. \tag{17}$$

Note that each column of the matrix $(Z'Z)^{-1}Z'$ in RHS of the above equation (17) is jut the filter gain of the smoother. Specifically, τ'th row of the matrix signifies the weights of $\boldsymbol{y}_t$'s in calculating the estimate $\hat{\boldsymbol{\beta}}_\tau$ at a sepecific period τ. When $n = 1$, by plotting t'th row of the matrix as time series data, we know how the observation $\boldsymbol{y}_t$ in each period contributes to the estimate of $\hat{\boldsymbol{\beta}}_\tau$ in a specific period as will be illustrated in Section 4.

Our smoothing method have several advantages while the basic idea is quite simple and straightforward. The main concern with our method is whether it is practical or not because the regressor Z is of huge size in most cases. In fact, few researchers have attempt to use this method for smoothing for a fixed interval and most researchers have never imagined that the newest personal computers in 2007 with more than 2GB main memory are capable of calculating the inverse of a nonsigular matrix of its size, 8000×8000, in several minutes.

So far, we know that methods for classical linear regression models are useful for smoothing for a fixed interval in the State Space Model, which one can interprete as a linear model with time-varying parameters. However, when we suppose the normal distribution for the disturbance terms in the State Space Model, the *structural changes* should be gradual. In the following subsection, the author shows an extended linear model, called a random parameter regression model, to handle more general time-varying structure.

2.4 An Extension of Ito Regression

In this section, we extend the State Space Model as a linear regression model with gradual time-varying parameters to more general regression models, which we call random parameter regression models. Such models can treat many kinds of structures of time-varying parameters. Furthermore, we can regard the usual random effects model for panel data as a typical example of our model.

We first reexamine the structure of the State Space Model as a linear regression model, the equations (14) and (15), which we called Ito regression. One regards the equation (14) as an ordinary linear regression model; the author regards the equation (15) as the one specifying a time-varying structure of parameters. We can regard the equation (15) as an difference equation with disturbance terms.

In respect of the regression theory, the equation (15) illustrates how randomized are the parameters within the model. Now we rewrite the equation (15) as follows:

$$\boldsymbol{\alpha} = \boldsymbol{\Phi}\boldsymbol{\beta} + \boldsymbol{v}. \tag{18}$$

To simplify the discussion, we temporally assume that the matrix $\boldsymbol{\Phi}$ is non-singular. Thus, we have

$$\boldsymbol{\beta} = \boldsymbol{\Phi}^{-1}\boldsymbol{\alpha} - \boldsymbol{\Phi}^{-1}\boldsymbol{v}. \tag{19}$$

We regard the first term of RHS in the equation (19) as the expected values of the *randomized* parameters and the second one the random effects of the disturbance terms.

Considering the case where some parameters might not be random or the one where the disturbance terms affect the parameters in degenerated ways, we generalize the equation (15) in the following simple form:

$$\boldsymbol{\beta} = \bar{\boldsymbol{\beta}} + D\boldsymbol{w}, \tag{20}$$

where D is a matrix with the size of $N \times \ell, \ell \leq N, N := nT$ and

$$\boldsymbol{w} \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \Sigma_{\boldsymbol{w}}), \tag{21}$$

where $\boldsymbol{w}$ might have smaller dimension than that of $\boldsymbol{v}$.

We assumes that the rank of D is $\ell > 0$ and that $\bar{\boldsymbol{\beta}}$ is known. Note that when rank $D = \ell = 0$, the model (20) has no parameter to estimate. The author stresses that the matrix D reflects the time-varying structure in place of $\boldsymbol{\Phi}$. This random parameter regression model can treat the following extended the State Space Model:

$$\boldsymbol{y}_t = X_t\boldsymbol{\beta}_t + \boldsymbol{u}_t, \quad \boldsymbol{u}_t \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, R_t) \tag{22}$$

$$\boldsymbol{\beta}_{t+1} = \Phi_{t+1,t}\boldsymbol{\beta}_t + G_t\boldsymbol{w}_t, \quad \boldsymbol{w}_t \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \Sigma_{\boldsymbol{w}}), \tag{23}$$

where the equation (22) is identical with (1) and the matrix G_t in (23) has the size $m \times \ell$, $\ell \leq m$ and its rank is ℓ. It is natural to assume that $\Sigma_{\boldsymbol{w}} = I$ because this assumption implies

$Cov(G_t\boldsymbol{w}_t) = G_tG_t'$ and thus G_t has all information about the covariance matrix of the state equation. The author makes additional remark that the class of the random parameter regression model defined above covers quite wide range of linear models such as linear models for panel data.

In the next section, the author give a necessary and sufficient condition assuring that the conventional least square methods such as OLS or GLS bring us the MMSLE(minimum mean square linear estimator).

3 Random Parameter Regression

In this section the author shows that methods based on OLS or GLS assure the MMSLE estimator in random parameter regression model defined above. For convenience of the readers, we rewrite the random parameter regression model. Note that we suppose the number of unknown parameters in M and the dimension of observation vector $\boldsymbol{y}$ is N for simplicity since we consider static linear model.

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{24}$$

and

$$\bar{\boldsymbol{\beta}} = \boldsymbol{\beta} - D\boldsymbol{w}, \tag{25}$$

where D is a matrix known with the size of $M \times \ell, \ell \leq M$,

$$\boldsymbol{w} \overset{iid}{\sim} \mathcal{N}(\boldsymbol{0}, \Sigma_{\boldsymbol{w}}), \tag{26}$$

and $\bar{\boldsymbol{\beta}}$ is known.

We pose the following assumption.

Assumption 1

$$rank\ D = \ell > 0$$

and there is a generalized inverse D^- of D such that

$$D^-D = I.$$

This assumption excludes the case where there is no parameter to be estimated.

We stack the equations (24) and a translated equation of (25) by D^-.

$$\mathcal{Y} = \mathcal{X}\boldsymbol{\beta} + \boldsymbol{\xi} \tag{27}$$

for

$$\begin{bmatrix} D^-\bar{\boldsymbol{\beta}} \\ \boldsymbol{y} \end{bmatrix} = \begin{bmatrix} D^- \\ X \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} -\boldsymbol{w} \\ \boldsymbol{\varepsilon} \end{bmatrix}.$$

The equation (27) can be written as

$$(\mathcal{Y} - \mathcal{X}\boldsymbol{\beta}) \sim \mathcal{N}(\boldsymbol{0}, \Sigma_{\boldsymbol{\xi}}), \tag{28}$$

where

$$\Sigma_{\boldsymbol{\xi}} = \begin{pmatrix} \Sigma_{\boldsymbol{w}} & \boldsymbol{O} \\ \boldsymbol{O} & \Sigma_{\boldsymbol{\varepsilon}} \end{pmatrix}. \tag{29}$$

In Table 2, we summarize the dimensions of the vectors and matrices in this section.

The above framework enables us to cope with the random parameter regression theory as we do with the familiar regression theory. In the usual regression theory, an important estimator is the WLSE(weighted least squares estimator) of $\boldsymbol{\beta}$. The rest of this section owes to the main result of Duncan and Horn (1972); they did not concern the case where the probabilistic relation between $\boldsymbol{w}$ and $\boldsymbol{\beta}$ be degenerated. This article extends their framework to the one in which we can treat more general randomization of the parameters.

Table 2: dimensions of Random Parameter Regression Model

vector		matrix	
$\boldsymbol{y}$	N	$\mathcal{X}$	$(N+\ell)\times M$
$\boldsymbol{\beta}$	M	$\Sigma_{\boldsymbol{w}}$	$\ell\times\ell$
$\boldsymbol{\varepsilon}$	N	$\Sigma_{\boldsymbol{\varepsilon}}$	$N\times N$
$\boldsymbol{w}$	ℓ	$\Sigma_{\boldsymbol{\xi}}$	$(N+\ell)\times(N+\ell)$
$\mathcal{Y}$	$N+\ell$	D	$M\times\ell$

Definition 1 (WLSE(weighted least squares estimator)) *$\boldsymbol{b}$ is the WLSE of $\boldsymbol{\beta}$ for (28) if and only if*

$$(\mathcal{Y}-\mathcal{X}\hat{\boldsymbol{\beta}})'\Sigma_{\boldsymbol{\xi}}^{-1}(\mathcal{Y}-\mathcal{X}\hat{\boldsymbol{\beta}}) \tag{30}$$

is minimized when $\boldsymbol{b}=\hat{\boldsymbol{\beta}}$, that is,

$$\boldsymbol{b}=argmin\{(\mathcal{Y}-\mathcal{X}\hat{\boldsymbol{\beta}})'\Sigma_{\boldsymbol{\xi}}^{-1}(\mathcal{Y}-\mathcal{X}\hat{\boldsymbol{\beta}}):\hat{\boldsymbol{\beta}}\in\mathbf{R}^M\}$$

The following theorem is easy to prove.

Theorem 1 (Normal Equation) *The WLSE of $\boldsymbol{\beta}$ is given as the solution of the normal equation*

$$(\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1}\mathcal{X})'\boldsymbol{b}=\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1}\mathcal{Y} \tag{31}$$

Proof: Since the quadratic form (31) to be minimized has the same form as that in the usual linear regression, the minimum is given by the solution of (31). □

We could express the WLSE estimation, $\boldsymbol{\beta}$, based on the givens in (24) and (25) through simple algebraic operations.

$$\boldsymbol{b}=V[D^{-\prime}\Sigma_{\boldsymbol{w}}D^{-}\bar{\boldsymbol{\beta}}+X'\Sigma_{\boldsymbol{\varepsilon}}\boldsymbol{y}], \tag{32}$$

where

$$V=[D^{-\prime}\Sigma_{\boldsymbol{w}}D^{-}+X'\Sigma_{\boldsymbol{\varepsilon}}X]^{-1}. \tag{33}$$

Since the estimation $\boldsymbol{\beta}$ in our random parameter regression theory is of our interest, our main concern is how the distribution of $\boldsymbol{b}-\boldsymbol{\beta}$ is. The following lemma is useful for the rest of this section.

Lemma 1 (the distribution of $\boldsymbol{b}-\boldsymbol{\beta}$) *If*

$$(\mathcal{Y}-\mathcal{X}\boldsymbol{\beta})\sim\mathcal{N}(\mathbf{0},\Sigma_{\boldsymbol{\xi}})$$

and

$$\boldsymbol{b}=(\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1}\mathcal{X})^{-1}\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1}\boldsymbol{y}$$

then

(1) $(\boldsymbol{b}-\boldsymbol{\beta})\sim\mathcal{N}(\mathbf{0},(\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1}\mathcal{X})^{-1})$

(2) $E(\boldsymbol{b}-\boldsymbol{\beta})(\mathcal{Y}-\mathcal{X}\boldsymbol{\beta})'=\mathbf{0}$

(3) $E(\boldsymbol{b}-\boldsymbol{\beta})\boldsymbol{y}'=\mathbf{0}$

Proof: We first note that

$$\boldsymbol{b}-\boldsymbol{\beta}=(\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1}\mathcal{X})^{-1}\mathcal{X}\Sigma_{\boldsymbol{\xi}}^{-1}(\mathcal{Y}-\mathcal{X}\boldsymbol{\beta}).$$

Since $\mathcal{Y} = \mathcal{X}\boldsymbol{\beta} + \boldsymbol{\xi}$,

$$\boldsymbol{b} - \boldsymbol{\beta} = (\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1}\mathcal{X})^{-1}\mathcal{X}\Sigma_{\boldsymbol{\xi}}^{-1}\boldsymbol{\xi}. \tag{34}$$

On the other hand,

$$\begin{aligned}\mathcal{Y} - \mathcal{X}\boldsymbol{\beta} &= (I - \mathcal{X}(\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1}\mathcal{X})^{-1}\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1})\mathcal{Y} \\ &= (I - \mathcal{X}(\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1}\mathcal{X})^{-1}\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1})(\mathcal{X}\boldsymbol{\beta} + \boldsymbol{\xi}) \\ &= (I - \mathcal{X}(\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1}\mathcal{X})^{-1}\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1})\boldsymbol{\xi}\end{aligned} \tag{35}$$

Thus the assertions (1) and (2) follow directly.

Considering (24) and (25),

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} = X\bar{\boldsymbol{\beta}} + XD\boldsymbol{w} + \boldsymbol{\varepsilon}. \tag{36}$$

From (34) and (36),

$$\begin{aligned}E(\boldsymbol{b} - \boldsymbol{\beta})\boldsymbol{y}' &= (\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1}\mathcal{X})^{-1} \cdot E\left[\begin{pmatrix} -\boldsymbol{w} \\ \boldsymbol{\varepsilon} \end{pmatrix} D^{-\prime}\bar{\boldsymbol{\beta}}'X' + \begin{pmatrix} -\boldsymbol{w} \\ \boldsymbol{\varepsilon} \end{pmatrix}(\boldsymbol{w}'D'X' + \boldsymbol{\varepsilon}')\right] \\ &= \mathbf{0} + (\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1}\mathcal{X})^{-1}(-\Sigma_{\boldsymbol{w}}^{-1}\Sigma_{\boldsymbol{w}}X' + X'\Sigma_{\boldsymbol{\varepsilon}}^{-1}\Sigma_{\boldsymbol{\varepsilon}}) \\ &= \mathbf{0}.\end{aligned}$$

Since $E(\boldsymbol{b} - \boldsymbol{\beta})\bar{\boldsymbol{\beta}} = \mathbf{0}$ also, (3) follows. □

Now we define several properties of an estimator of $\boldsymbol{\beta}$ relative to a matrix to characterize the condition under which $\boldsymbol{\beta}$ is the MMSLE (minimum mean square linear estimator).

Definition 2 *Let C be a $q \times M$ matrix and $\hat{\boldsymbol{\gamma}}$ be a linear function of $\boldsymbol{\beta}$ such that $\boldsymbol{\gamma} = C\boldsymbol{\beta}$,*

1. *$\hat{\boldsymbol{\gamma}}$ is linear if and only if $\hat{\boldsymbol{\gamma}} = A_1\bar{\boldsymbol{\beta}} + A_2\boldsymbol{y}$ for some matrices A_1 and A_2.*
2. *$\hat{\boldsymbol{\gamma}}$ is unbiased if and only if $E(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) = \mathbf{0}$.*
3. *$\hat{\boldsymbol{\gamma}}$ is minimum mean square if and only if $(\forall i)E(\hat{\gamma}_i - \gamma_i)^2$ is minimized.*
4. *$\hat{\boldsymbol{\gamma}}$ is minimum variance if and only if $(\forall i)Var(\hat{\gamma}_i - \gamma_i)$ is minimized.*
5. *$\hat{\boldsymbol{\gamma}}$ is the MVLUE (minimum variance linear unbiased estimator) if and only if (i) $\hat{\boldsymbol{\gamma}}$ is linear and unbiased and (ii) $(\forall i)(\forall \boldsymbol{\gamma}')\left[\boldsymbol{\gamma}' \text{ is linear and unbiased} \Longrightarrow Var(\gamma_i' - \gamma_i) \geq Var(\hat{\gamma}_i - \gamma_i)\right]$.*
6. *$\hat{\boldsymbol{\gamma}}$ is the MMSLE (minimum mean square linear estimator) if and only if (i) $\hat{\boldsymbol{\gamma}}$ is linear and (ii) $(\forall i)(\forall \boldsymbol{\gamma}')\left[\boldsymbol{\gamma}' \text{ is linear} \Longrightarrow E(\gamma_i' - \gamma_i)^2 \geq E(\hat{\gamma}_i - \gamma_i)^2\right]$.*

Remark 1 *A linear estimator $\boldsymbol{\gamma}$ of $\boldsymbol{\beta}$ is linear in $\bar{\boldsymbol{\beta}}$ and $\boldsymbol{y}$.*

Irrespective of the rank condition of the matrix D, the following lemma by Duncan and Horn (1972) [Lemma 2.6] holds.

Lemma 2 (Equivalence of MVLUE and MMSLE in random parameter regression) *Assume*

$$(\mathcal{Y} - \mathcal{X}\boldsymbol{\beta}) \sim \mathcal{N}(\mathbf{0}, \Sigma_{\boldsymbol{\xi}}).$$

Then the MVLUE of $\boldsymbol{\gamma} = C\boldsymbol{\beta}$ of $\boldsymbol{\beta}$ for some C is the MMSLE of $\boldsymbol{\gamma}$.

The next theorem completely characterizes the MMSLE estimator based on the observation $\boldsymbol{y}$, the prior $\bar{\boldsymbol{\beta}}$ and the matrix D reflecting the time-varying structure; it corresponds to the classical Gauss-Markov theorem in a sense.

Theorem 2 (Modified Gauss-Markov theorem) *Assume*

$$(\mathcal{Y} - \mathcal{X}\boldsymbol{\beta}) \sim \mathcal{N}(\mathbf{0}, \Sigma_{\boldsymbol{\xi}}),$$

and that $\hat{\boldsymbol{\gamma}}$ is the MMSLE based on $\mathcal{Y} = \begin{bmatrix} D^{-}\bar{\boldsymbol{\beta}} \\ \boldsymbol{y} \end{bmatrix}$ of any linear vector function $\boldsymbol{\gamma} = C\boldsymbol{\beta}$ of $\boldsymbol{\beta}$ for some C if and only if

(1) *$\hat{\gamma}$ is unbiased estimator of $\boldsymbol{\gamma}$,*

(2) *$\hat{\gamma}$ is linear vector function such that*

$$\hat{\gamma} = M\boldsymbol{g},$$

where

$$\boldsymbol{g} = \mathcal{X}'\Sigma_{\boldsymbol{\xi}}\mathcal{Y} = [D^{-\prime}\bar{\boldsymbol{\beta}}; X]\begin{pmatrix} \Sigma_{\boldsymbol{w}} & \boldsymbol{O} \\ \boldsymbol{O} & \Sigma_{\boldsymbol{\varepsilon}} \end{pmatrix}^{-1}\begin{bmatrix} D^{-}\bar{\boldsymbol{\beta}} \\ \boldsymbol{y} \end{bmatrix}.$$

Proof: [Necessity] The necessity of the condition (1) follows from Lemma 2. For the condition (2) one can easily prove it by contradiction.

[Sufficiency] Assume that the conditions (1) and (2) hold. To prove the sufficiency of them, it is sufficient to show that $\hat{\gamma}$ satisfying (1) and (2) is the unique MVLUE of $\boldsymbol{\gamma}$, since the MVLUE is the MMSLE from Lemma 2. Assume that

$$\boldsymbol{\gamma}^* = A\mathcal{Y} = [A_1; A_2]\begin{bmatrix} D^{-}\bar{\boldsymbol{\beta}} \\ \boldsymbol{y} \end{bmatrix} = A_1 D^{-}\hat{\boldsymbol{\beta}} + A_2\boldsymbol{y}$$

for some $A = [A_1; A_2]$ is an unbiased estimator of $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}^* \neq \hat{\boldsymbol{\gamma}}$. Then

$$E[\boldsymbol{\gamma}^*] = E[\gamma] = E[\gamma],$$

since

$$E[\boldsymbol{\gamma}^*] - E[\gamma] = \mathbf{0}$$

and

$$E[\hat{\gamma}] - E[\gamma] = \mathbf{0}.$$

Thus, considering $E(\boldsymbol{\beta}) = \bar{\boldsymbol{\beta}}$,

$$A\mathcal{X}\bar{\boldsymbol{\beta}} = C\bar{\boldsymbol{\beta}} = M\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1}\bar{\boldsymbol{\beta}}. \tag{37}$$

Since (37) is identity in $\bar{\boldsymbol{\beta}}$, we have

$$A\mathcal{X} - M\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1}\mathcal{X} = \mathbf{0} \tag{38}$$

and

$$A\mathcal{X} - C = \mathbf{0}. \tag{39}$$

Then

$$M\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1}\mathcal{X} - C = \mathbf{0}. \tag{40}$$

From (39) we have

$$\begin{aligned}
Var(\boldsymbol{\gamma}^* - \gamma) &= Var(A\mathcal{Y} - C\boldsymbol{\beta}) \\
&= Var(A(\mathcal{X}\boldsymbol{\beta} + \boldsymbol{\xi}) - C\boldsymbol{\beta}) \\
&= Var(A\{\mathcal{X}(\bar{\boldsymbol{\beta}} + D\boldsymbol{w}) + \boldsymbol{\xi}\} - C\boldsymbol{\beta}) \\
&= Var(A\mathcal{X}\bar{\boldsymbol{\beta}} + A\mathcal{X}D\boldsymbol{w} + A\boldsymbol{\xi} - C\bar{\boldsymbol{\beta}} - CD\boldsymbol{w}) \\
&= Var((A\mathcal{X} - C)D\boldsymbol{w} + (A\mathcal{X} - C)\bar{\boldsymbol{\beta}} + A\boldsymbol{\xi}) \\
&= Var((A\mathcal{X} - C)D\boldsymbol{w} + A\boldsymbol{\xi}) \\
&= Var(A\boldsymbol{\xi}) \\
&= A\Sigma_{\boldsymbol{\xi}}A'
\end{aligned}$$

From (40) similarly, by replacing A with $M\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1}$, we have

$$\begin{aligned}
Var(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}) &= Var(M\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1}\mathcal{Y}-C\boldsymbol{\beta}) \\
&= Var(M\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1}(\mathcal{X}\boldsymbol{\beta}+\boldsymbol{\xi})-C\boldsymbol{\beta}) \\
&= Var(M\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1}\{\mathcal{X}(\bar{\boldsymbol{\beta}}+D\boldsymbol{w})+\boldsymbol{\xi}\}-C\boldsymbol{\beta}) \\
&= Var(M\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1}\mathcal{X}\bar{\boldsymbol{\beta}}+M\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1}\mathcal{X}D\boldsymbol{w}+M\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1}\boldsymbol{\xi}-C\bar{\boldsymbol{\beta}}-CD\boldsymbol{w}) \\
&= Var((M\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1}\mathcal{X}-C)D\boldsymbol{w}+(M\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1}\mathcal{X}-C)\bar{\boldsymbol{\beta}}+M\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1}\boldsymbol{\xi}) \\
&= Var((M\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1}\mathcal{X}-C)D\boldsymbol{w}+M\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1}\boldsymbol{\xi}) \\
&= Var(M\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1}\boldsymbol{\xi}) \\
&= M\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1}\Sigma_{\boldsymbol{\xi}}(M\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1})' \\
&= M\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1}\mathcal{X}M.
\end{aligned}$$

Using the equation (38) with a tricky operation, we have

$$\begin{aligned}
Var(\boldsymbol{\gamma}^*-\boldsymbol{\gamma}) &= (A-M\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1}+M\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1})\Sigma_{\boldsymbol{\xi}}(A'-\Sigma_{\boldsymbol{\xi}}^{-1}M'+\Sigma_{\boldsymbol{\xi}}^{-1}M') \\
&= (A-M\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1})\Sigma_{\boldsymbol{\xi}}(A'-\Sigma_{\boldsymbol{\xi}}^{-1}M')+2(A-M\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1})\mathcal{X}M'+M\mathcal{X}'\Sigma_{\boldsymbol{\xi}}\mathcal{X}M' \\
&= (A-M\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1})\Sigma_{\boldsymbol{\xi}}(A'-\Sigma_{\boldsymbol{\xi}}^{-1}M')+Var(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}). \qquad (41)
\end{aligned}$$

Since the diagonal components of the first term in RHS of (41) are non-negative,

$$(\forall i)\ Var(\gamma_i^*-\gamma_i)\geq Var(\hat{\gamma}_i-\gamma_i).$$

Thus $\hat{\boldsymbol{\gamma}}$ is a MVLUE of $\boldsymbol{\gamma}$. Uniqueness follows from (41), for if $Var(\boldsymbol{\gamma}^*-\boldsymbol{\gamma})=Var(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})$ we necessarily have $\boldsymbol{\gamma}^*=\hat{\boldsymbol{\gamma}}$, which contradicts our premises. This completes the proof of sufficiency. □

The following corollary of Theorem 2 guarantees good properties of our method such as Ito Regression.

Corollary 1 (MMSLE of $\boldsymbol{\beta}$ in random parameter regression) *Assume*

$$(\mathcal{Y}-\mathcal{X}\boldsymbol{\beta})\sim\mathcal{N}(\mathbf{0},\Sigma_{\boldsymbol{\xi}}).$$

Then

$$\boldsymbol{b}=(\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1}\mathcal{X})^{-1}\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1}\mathcal{Y}.$$

Proof: From Lemma 1

$$E(\boldsymbol{b}-\boldsymbol{\beta})=\mathbf{0}$$

and

$$\boldsymbol{b}=M\boldsymbol{g},$$

where

$$M=(\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1}\mathcal{X})^{-1},\quad \boldsymbol{g}=\mathcal{X}'\Sigma_{\boldsymbol{\xi}}^{-1}\mathcal{Y}.$$

The conclusion follows directly from Theorem 2. □

The above Theorem 2 and Corollary 1 assures that usual least square methods take crucial roles even in our random parameter regression theory.

In the next section, we give an application of Ito Regression, a typical example of the random parameter regression.

4 An Application

This section shows an application of the method in Section 2 to a time-varying AR model. The model is useful for treating non-stationary time series; it cope with such time series as locally stationary ones. The coefficients estimated in each period would be used to analyze local behavior of the time series. This section owes much to Ito and Sugiyama (2007).

Since the model is a linear time series model, it has a representation as the State Space Model. At first, we represent a time-varying AR model as the State Space Model as follows:

observation equation

$$x_t = \begin{pmatrix} x_{t-1} & x_{t-2} & \cdots & x_{t-k} \end{pmatrix} \begin{pmatrix} \beta_{1,t} \\ \beta_{2,t} \\ \vdots \\ \beta_{k,t} \end{pmatrix} + u_t, \qquad u_t \sim \mathcal{N}(0, \sigma^2_{u_t}) \tag{42}$$

state equation

$$\begin{pmatrix} \beta_{1,t} \\ \beta_{2,t} \\ \vdots \\ \beta_{k,t} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} \beta_{1,t-1} \\ \beta_{2,t-1} \\ \vdots \\ \beta_{k,t-1} \end{pmatrix} + \begin{pmatrix} v_{1,t} \\ v_{2,t} \\ \vdots \\ v_{k,t} \end{pmatrix}, \quad \boldsymbol{v}_t \sim \mathcal{N}_k(\mathbf{0}, \sigma^2_{v_t}\mathbf{I}) \tag{43}$$

Note $\boldsymbol{v}_t \equiv (v_{1,t}\ v_{2,t}\ \cdots\ v_{k,t})'$ in (43).

Next, we have to derive Ito Regression Model as in the equation (16). Applying *Ito Regression* to time-varying AR(k) model, we can rewrite the model,

$$\boldsymbol{y} = \boldsymbol{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{44}$$

where

$$\boldsymbol{y} \equiv \left(\begin{array}{c} x_1 \\ x_2 \\ \vdots \\ x_T \\ \hline\hline -\beta_{1,0} \\ -\beta_{2,0} \\ \vdots \\ -\beta_{k,0} \\ \hline 0 \\ 0 \\ \vdots \\ 0 \\ \hline \vdots \\ \hline 0 \\ 0 \\ \vdots \\ 0 \end{array}\right), \quad \boldsymbol{\beta} \equiv \left(\begin{array}{c} \beta_{1,1} \\ \beta_{1,2} \\ \vdots \\ \beta_{1,T} \\ \hline \beta_{2,1} \\ \beta_{2,2} \\ \vdots \\ \beta_{2,T} \\ \hline \vdots \\ \hline \beta_{k,1} \\ \beta_{k,2} \\ \vdots \\ \beta_{k,T} \end{array}\right), \quad \boldsymbol{\varepsilon} \equiv \left(\begin{array}{c} u_1 \\ u_2 \\ \vdots \\ u_T \\ \hline\hline v_{1,1} \\ v_{2,1} \\ \vdots \\ v_{k,1} \\ \hline v_{1,2} \\ v_{2,2} \\ \vdots \\ v_{k,2} \\ \hline \vdots \\ \hline v_{1,T} \\ v_{2,T} \\ \vdots \\ v_{k,T} \end{array}\right)$$

$$
\boldsymbol{Z} \equiv \left(\begin{array}{cccc|cccc|c|cccc}
x_0 & & & & x_{-1} & & & & & x_{-k} & & & \\
& x_1 & & & & x_0 & & & \cdots & & x_{-k+1} & & \\
& & \ddots & & & & \ddots & & & & & \ddots & \\
& & & x_{T-1} & & & & x_{T-2} & & & & & x_{T-k} \\
\hline\hline
-1 & & & & & & & & & & & & \\
& & & & -1 & & & & \cdots & & & & \\
& & & & & & & & nbsp; & & & & \\
& & & & & & & & & -1 & & & \\
\hline
1 & -1 & & & & & & & & & & & \\
& & & & 1 & -1 & & & \cdots & & & & \\
& & & & & & & & & 1 & -1 & & \\
\hline
& \vdots & & & & \vdots & & & \ddots & & \vdots & & \\
\hline
& & 1 & -1 & & & & & & & & & \\
& & & & & & 1 & -1 & \cdots & & & & \\
& & & & & & & & & nbsp; & & & \\
& & & & & & & & & & & 1 & -1
\end{array}\right)
$$

(Blanks are all 0 in $\boldsymbol{Z}$.)

Note that the matrix presented above is essentially the same as the one shown in Section 2.3 althogh several rows and columns are exchanged.

(44) is nothing but a linear regression model. Note that, in Ito Regression, a dynamic model represented by the State Space Model apparently becomes a static model with no subscript t indicating time and that Ito Regression requires no filtering step unlike Kalman smoothing. Notice that $(\beta_{1,0}, \cdots, \beta_{k,0})$ is the initial values of the state variable.

Note that the size of the regressor Z is $(Tk) \times (T(k+1))$ in case of a time-varying AR(k) with the sample periods T. Thus, the compute we use should compute the inverse of $Z'Z$, of which size is $(Tk) \times (Tk)$. For example, The author's personal computer with 512MB main memory and 1.5GHz MPU (Pentium) can compute the inverse of 2500×2500 matrix in about 150 seconds when he uses the statistical package R (ver.2.3.).

We can estimate state variables by OLS or GLS in Ito Regression. Furthermore. We do not have to assume normality on a disturbance, as maximum likelihood method do. Of course, you can do hypothesis testing such as Wald to see if the estimation result is statistically significant.

The estimation result which we derive when applying Ito Regression to a time-varying AR(k) model in case $k = 1$ with the monthly data of rate of return (NIKKEI225 (1955/1 - 2006/2)) is illustrated in Figure 1.

Moreover, as suggested in Section 2.3, Ito Regression method allows us to see how wide the range on which a estimation in each period t depends is by plotting the each row of a matrix $(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'$, corresponding to the filter gain, since

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{y}$$

is the OLS estimator of $\boldsymbol{\beta}$ in (44). For example, we plot 300th row of $(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'$ in Figure 2 using full sample of NIKKEI225. Note that columns from 1 to 610 are only related to calculating AR coefficients, since sample size of the data is 610. This Figure shows that to compute the state β_{300}, we only need data from about $t = 200$ to $t = 400$.

For practical usage, some researchers use Moving-Window method, in which they slide subsamples of the data just like windows. One needs to decide a window width *a priori* in Moving-Window method; the model itself decides the optimal window width in Ito Regression.

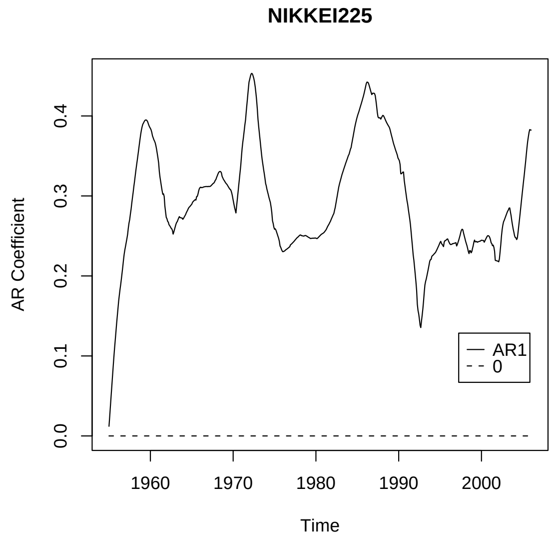

Figure 1: Time Varying AR(1) Coefficient: (rate of return of NIKKEI225 monthly)

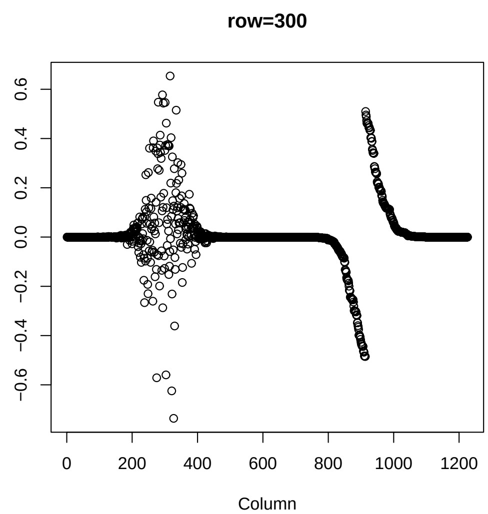

Figure 2: Nikkei225, Weights Used in Smoothing

5 Conclusion

This article presented a linear model which the author call the random parameter regression model. This model covers fairly wide range of linear ones from the State Space Model in time series analysis to a random effects model for panel data analysis. As a linear model allowing flexible time-varying structure, it gives us more general framework to estimate time-varying parameters than the State Space Model.

As shown in Section 3, methods of least square, OLS or GLS, would be still useful for econometric analysis if researchers studied and developed more flexible linear model based on a random parameter regression model here.

The advantages of our model crucially depend on the fact that computers these days are much more powerful than past ones. Who imagined ten years ago that personal computers in 2007 are capable of computing the inverse of 5000×5000 matrix with practical speed and accuracy? A number of ecnometricians are likely to use the method of maximum likelihood even when they estimate parameters in a linear model. They could more depend on recent computer power than they had imagined.

Data Appendix

Monthly data of NIKKEI225 were all downloaded from NIKKEI NEEDS. The author used a statistical package, R (ver. 2.3.1). All scripts of R he used are available for the readers.

References

Cooley, T.F. and E.C. Prescott (1973) “An Adaptive Regression Model,” *International Economic Review*, Vol. 14, No. 2, pp. 364–371.

Duncan, D.B. and S.D. Horn (1972) “Linear Dynamic Recursive Estimation from the Viewpoint of Regresion Analysis,” *Journal of the American Statistical Association*, Vol. 67, No. 340, pp. 815–821.

Durbin, J. and S.J. Koopman (2001) *Time Series Analysis by State Space Methods*: Princeton University Press.

Ito, M. and S. Sugiyama (2007) *Measuring the Degree of Time Varying Market Inefficiency*: Keio Economic Society, Discussion Paper Series, KESDP 07-2.

Kalman, R.E. (1960) “A New Approach to Linear Filtering nad Prediction Problems,” *Journal of Basic Engineering*, Vol. 46, No. 5, pp. 1575–1617.

Kolmogorov, AN (1941) “Interpolation and Extrapolation of Random Sequences,” *Izv. Akad. Nauk SSSR, Ser. Mat*, Vol. 5, pp. 3–14.

Ruud, P.A. (2000) *An introduction to classical econometric theory*: Oxford University Press New York.

Sorenson, H.W. (1970) “Least-Squares estimation: from Gauss to Kalman,” *IEEE Spectrum*, Vol. 7, No. 7, pp. 63–68.

Wiener, N. (1949) *Extrapolation, interpolation, and smoothing of stationary time series*: Wiley.