Online Appendix
The Evolution of Stock Market Efficiency in the U.S.: A Non-Bayesian Time-Varying Model Approach

Mikio Ito[a], Akihiko Noda[b*] and Tatsuma Wada[c]

[a] *Faculty of Economics, Keio University, 2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan*

[b] *Faculty of Economics, Kyoto Sangyo University, Motoyama, Kamigamo, Kita-ku, Kyoto 603-8555, Japan*

[c] *Faculty of Policy Management, Keio University, 5322 Endo, Fujisawa, Kanagawa, 252-0882, Japan*

Abstract: This note provides detailed computations in the main text.

Keywords: Market Efficiency; Non-Bayesian Time-Varying AR Model; Degree of Market Efficiency; The Adaptive Market Hypothesis.

JEL Classification Numbers: C22; G14.

A.1 State Space Regression Model

This section shows a framework allowing us to deal with a state space models as a linear regression model with time-varying parameters. Such a framework provides an alternative of the Kalman smoother. The alternative estimator takes a crucial role in this article. The idea is so simple that we reduce a state space model to more general regression models, usually called random parameter regression models. Our framework covers wide range of state space models, which have random parameters with occasional jumps.

We first reexamine the structure of the state space model as a linear regression model, equation (10), which we call state space regression model.

When we stand on the regression theory, $\boldsymbol{\gamma} = W\boldsymbol{\beta} + \boldsymbol{v}$ in equation (10) illustrates how the parameters are randomized, given some prior vector $\boldsymbol{\gamma}$. To simplify the discussion, we temporally suppose that the matrix W is non-singular and that we have the following representation of $\boldsymbol{\beta}$:

$$\boldsymbol{\beta} = W^{-1}\boldsymbol{\gamma} - W^{-1}\boldsymbol{v}. \tag{A.1}$$

We regard the first term of RHS in equation (A.1) as the expected values of the *randomized* parameters and the second one the random effects of the disturbance terms. Notice that the vital supposition in our discussion is not the invertibility of W but the existence of such a decomposition of randomized parameters.

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 equation (9) in the following simple form:

$$\boldsymbol{\beta} = \bar{\boldsymbol{\beta}} + D\boldsymbol{w}, \tag{A.2}$$

*Corresponding Author. E-mail: noda@cc.kyoto-su.ac.jp, Tel. +81-75-705-1510, Fax. +81-75-705-3227.

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

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

where $\boldsymbol{w}$ might have smaller dimension than that of $\boldsymbol{v}$ and $\mathcal{WS}(\mathbf{0}, \Sigma_{\boldsymbol{w}})$ is any distribution depending only on mean $\mathbf{0}$ and variance $\Sigma_{\boldsymbol{w}}$ such as the normal distribution. We will use this notation to represent a distribution with wider sense assumptions.

We assumes that the rank of D is $\ell > 0$ and that $\bar{\boldsymbol{\beta}}$ is known. Note that when rank $D = \ell = 0$, equation (A.2) has no parameter to estimate. Notice that the matrix D reflects the time-varying structure in place of W.

This class of random parameter regression models covers a conventional state space model:

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

$$\boldsymbol{\beta}_{t+1} = \Phi_{t+1,t} \boldsymbol{\beta}_t + G_t \boldsymbol{w}_t, \quad \boldsymbol{w}_t \overset{iid}{\sim} \mathcal{WS}(\mathbf{0}, \Sigma_{\boldsymbol{w}}), \tag{A.3}$$

where the matrix G_t in equation (A.3) 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 $\text{cov}(G_t \boldsymbol{w}_t) = G_t G_t'$ and thus G_t has all information about the covariance matrix of the state equation. We make 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.

A.2 Random Parameter Regression

In this section, we demonstrate that OLS and GLS assure the MMSLE estimator in the random parameter regression model defined above. To simplify our notations and discussion, we present the random parameter regression model as follows. Let M denote the number of unknown parameters and N the dimension of observation vector $\boldsymbol{y}$ is N.

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{A.4}$$

and

$$\bar{\boldsymbol{\beta}} = \boldsymbol{\beta} - D\boldsymbol{w}, \tag{A.5}$$

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

$$\boldsymbol{\varepsilon} \overset{iid}{\sim} \mathcal{WS}(\mathbf{0}, \Sigma_{\boldsymbol{\varepsilon}}) \text{ and } \boldsymbol{w} \overset{iid}{\sim} \mathcal{WS}(\mathbf{0}, \Sigma_{\boldsymbol{w}}),$$

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

We suppose some regularity conditions for the least square estimation.

Assumption 1

$$\text{rank } D = \ell > 0,$$

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

$$D^- D = I.$$

We stack equations (A.4) and (A.5) multiplied by D^-.

$$\mathcal{Y} = \mathcal{X}\boldsymbol{\beta} + \boldsymbol{\xi}, \tag{A.6}$$

for

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

Equation (A.6) can be written as

$$(\mathcal{Y} - \mathcal{X}\boldsymbol{\beta}) \sim \mathcal{WS}(\mathbf{0}, \Sigma_{\boldsymbol{\xi}}), \tag{A.7}$$

where

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

Table A.1 summarizes the dimensions of the vectors and matrices.

(Figure A.1 here)

The above framework enables us to deal with the random parameter regression theory as we do with the familiar regression theory. In the usual regression theory, an important estimator is the weighted least squares estimator (WLSE) of $\boldsymbol{\beta}$.

Definition 1 (WLSE (weighted least squares estimator)) $\boldsymbol{b}$ *is the WLSE of* $\boldsymbol{\beta}$ *for equation (A.7) if and only if*

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

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

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

where $\mathcal{Y}$ *and* $\mathcal{X}$ *are given and* $\Sigma_{\boldsymbol{\xi}}$ *is known.*

Note that WLSE $\boldsymbol{b}$ does not always hold unbiasedness and consistency when $\mathcal{Y}$ and any column of $\mathcal{X}$ are statistically correlated.

A.3 The Relation to the Moving-Window Method

To highlight the difference between our method and the moving-window method (for example, Kim et al. (2011) and Lim et al. (2013)), let us rewrite the estimated p-th coefficient of the time-varying AR(q) model at time-t is as follows:

$$\hat{\alpha}_{p,t} = \sum_{\tau=1}^{T} \omega_{p,\tau,t} x_\tau + \omega_{p,0,t}, \quad (p = 1, \cdots, q);$$

where $\omega_{p,\tau,t}$'s are weights. Our method allows us to readily compute the weights by the following matrices: By a $(1 + qT) \times (1 + qT)$ permutation matrix P, we have a modified coefficient vector

$$\underbrace{\boldsymbol{b}}_{(1+qT)\times 1} \equiv {}^{t}(\alpha_0 \mid \alpha_{1,1}\ \alpha_{1,2}\ \cdots\ \alpha_{1,T} \mid \alpha_{2,1}\ \cdots\ \alpha_{2,T} \mid \cdots\ \cdots\ \cdots \mid \alpha_{q,1}\ \cdots\ \alpha_{q,T})$$
$$= P\boldsymbol{\beta}.$$

Then, defining a $(1+qT) \times (T+qT)$ matrix Ω such that:

$$\underbrace{\Omega}_{(1+qT)\times(T+qT)} = \underbrace{P}_{(1+qT)\times(1+qT)} \underbrace{Q}_{(1+qT)\times(T+qT)}$$
$$= \begin{bmatrix} \boldsymbol{r_x} & \boldsymbol{r}_\gamma \\ \Omega_{1,\boldsymbol{x}} & \Omega_{1,\gamma} \\ \Omega_{2,\boldsymbol{x}} & \Omega_{2,\gamma} \\ \vdots & \vdots \\ \Omega_{q,\boldsymbol{x}} & \Omega_{q,\gamma} \end{bmatrix},$$

we arrive at the least squares estimator for $\boldsymbol{b}$:

$$\widehat{\boldsymbol{b}} = \Omega \begin{bmatrix} \boldsymbol{x} \\ \cdots \\ \boldsymbol{\gamma} \end{bmatrix}.$$

Thus, $\omega_{p,\tau,t}$ is the (t,τ)-th element of $\Omega_{p,\boldsymbol{x}}$; and $\omega_{p,0,t}$ is a linear combination of t-th row of $\Omega_{p,\boldsymbol{\gamma}}$.

It is important to point out that in the case of the moving-window method, the band-width (size) of the window for t is fixed and smaller than the whole sample size, T. In contrast, our method is to find the orthogonal projection onto the space spanned by all the information, $(x_1, \ldots, x_T)$, as shown in equation (10). Because of this, in fact, our estimator is the minimized mean squared error estimator (MMSE).

(Figure A.1 here)

Figure A.1 exhibits estimated weights $\omega_{\tau,t}$ following equation (10). The smoothed estimate for α_{1942}, for example, requires approximately 25 years of observations x_t, from 1930 to 1954, with variant weights on each observation. Another way to interpret this finding is that the band-width for α_{1942} is about 25 years. Since these weights are the result of the least squares estimation, our implicitly-defined band-width can be seen as the "optimally-chosen band-width," in the sense of the MMSE, as opposed to an arbitrarily selected band-width in the moving-window method.

References

Kim, J. H., Shamsuddin, A., and Lim, K. P. (2011), "Stock Return Predictability and the Adaptive Markets Hypothesis: Evidence from Century-Long U.S. Data," *Journal of Empirical Finance*, 18, 868–879.

Lim, K. P., Luo, W., and Kim, J. H. (2013), "Are US Stock Index Returns Predictable? Evidence from Automatic Autocorrelation-Based Tests," *Applied Economics*, 45, 953–962.

Figure A.1: Optimal Weights for the Smoother

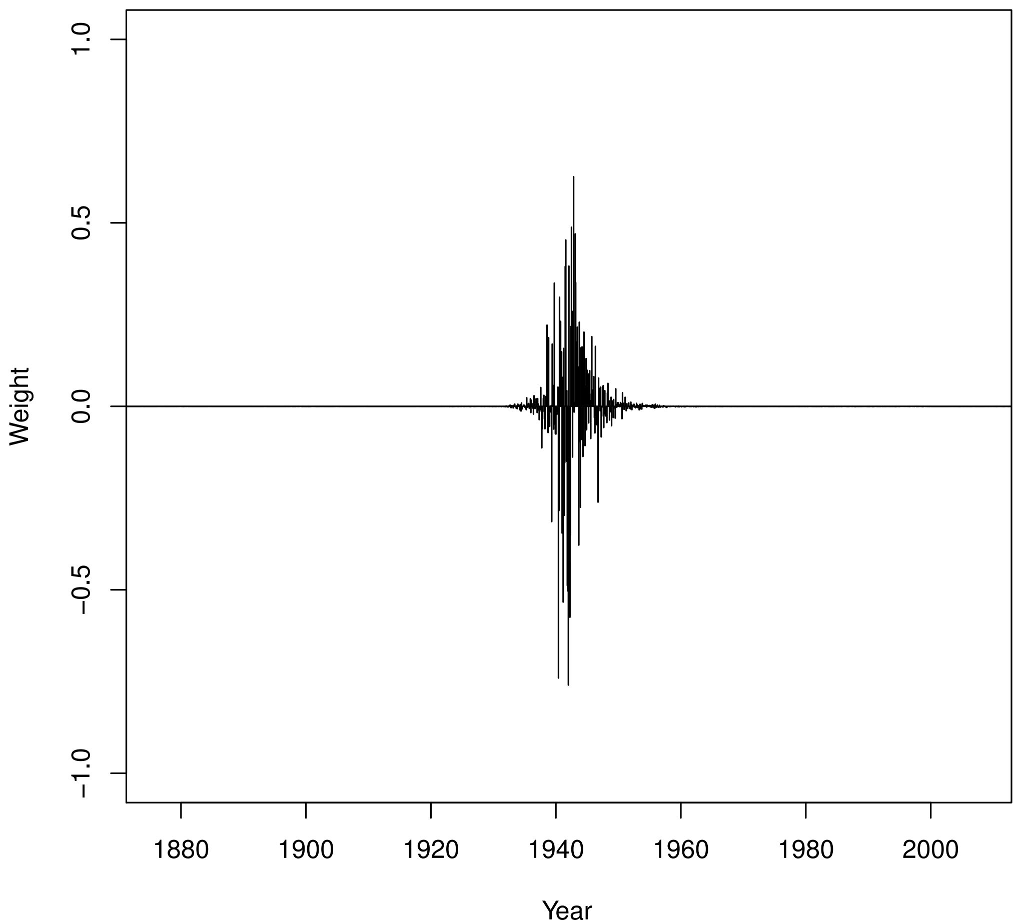

Table A.1: Dimensions of Random Parameter Regression Model

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	$M\times\ell$