Comovement of Stock Markets—An Analysis by Nonlinear Cointegration*

Kazumi Asako[1], Zhentao Liu[2]

[1]Faculty of Economics, Rissho University, Tokyo, Japan
[2]Institute for Financial and Accounting Studies, Xiamen University, Xiamen, China
 Email: kasako@ris.ac.jp

Abstract

This paper proposes and estimates a statistical model of nonlinear cointegration, with applications to the stock markets of Japan and the United States. We define nonlinear cointegration as a long-run stable relationship between two time series variables even in the presence of temporary nonlinear divergence from this long-run relationship. More concretely, extending the bubble model of Asako and Liu (2013) [1] to stock price ratio variables, both upward and downward divergent bubble processes are estimated at a time. We conclude that, although two stock price indexes are not linearly cointegrated, they are considered to be cointegrated nonlinearly.

Keywords

Booms and Busts, Stock Prices, Nonlinear Cointegration, Bayesian Estimation

1. Introduction

In this paper we propose and develop the recursive estimation method of a nonlinear statistical model of speculative bubbles and utilize this model in establishing an idea of nonlinear cointegration. We then apply this idea to the stock market indexes of Japan and the United States, and we detect how these indexes commove in the long run although they deviate from the long-run relationship nonlinearly in the short run.

So far, whether stock markets of different countries commove together has mainly been tested by utilizing the linear cointegration relationship *à la* Engle and Granger (1987) [2]. We owe the main idea of cointegration to this line of research. However, what we propose in this paper is a statistical model that incorporates latent cointegration relationship not linearly but nonlinearly. The nonlinearity here stems from the consideration of booms and busts in stock price indexes (and thereby the ratio of indexes of different markets). When bubbles are born

---

and boom for certain periods only to crash in due course, time series of these events are hardly captured by linear models.

As empirical investigation of comovement of stock markets, there have been a number of research and the results vary depending on the countries and sample periods. Asako, Zhang and Liu (2014) [3] conduct linear cointegration test among any pair of Japan, the United States and China and reach the conclusion that the linear cointegration is rejected. This is the origin of our analysis here because our daily observation suggests that the worldwide stock markets commove at any rate.

The construction of the present paper is as follows. In Section 2, we propose a time series model of the boom and bust and develop its recursive estimation method. Section 3 modifies this basic model to apply for a ratio variable, which has more restrictive feature within the model of booms and busts. In Section 4, we apply the modified model to detect the nonlinear cointegration relationship between the stock price indexes of Japan and the United States. Section 5 conclude the paper.

2. Model of Nonlinear Cointegration

In this section, we develop a model of nonlinear cointegration and explain how to estimate the relevant parameters.

2.1. The Basic Model

As an extended model to Asako and Liu (2013) [1], which in turn has its origin in Asako, Kanoh and Sano (1990) and Liu, Asako and Kanoh (2011) [4] [5], we propose a model of bubble booms and busts by, for $x_t > 0$,

$$x_t = \begin{cases} \beta_t x_{t-1} + u_t & \text{with} \quad \pi_t \quad (A) \\ \theta_t + u_t & \text{with} \quad 1 - \pi_t \quad (B) \end{cases} \tag{1}$$

where x_t denotes a sequence of variables measured as the ratio of stock prices in different countries and π_t denotes a probability that x_t follows model (A) depending on x_{t-1}. A newly arisen bubble u_t is a serially independent and normally distributed random variable with mean 0 and constant variance σ_u^2 which is unknown to us. The coefficient β_t is a time dependent parameter whose variation is given by the following random walk process:

$$\beta_t = \beta_{t-1} + v_t, \qquad v_t \sim N\left(0, \sigma_v^2\right) \tag{2}$$

Like σ_u^2, the constant variance of innovations σ_v^2 is unknown to us. Since we assume $x_t > 0$, the probability that u_t and v_t happen to bring about $x_t \leq 0$ is assumed virtually nil.

Let us consider briefly the implication of this model. Our basic model consists of two regimes or models (A) and (B). At period t, x_t is expressed by a divergent time series model when a speculative bubble continues. We describe this phenomenon by the autoregressive model (A) with parameter β_t exceeding unity. As implied by a speculative bubble, the divergent sequence will suddenly crash at a certain unknown time. We formulate this event by a systematic and probabilistic switch from model (A) to model (B). In model (B), irrespective of the position x_{t-1} at the previous period, x_t on average returns at period t to the fundamental value θ_t.

More concretely, we assume that the probability of bubble continuation π_t can be expressed as

$$\pi_t = e^{-\gamma - \alpha m_{t-1}} \tag{3}$$

$$m_t = |x_t - \theta_t| \tag{4}$$

where α and γ are positive unknown parameters. This formulation implies that π_t decreases as the deviation between x_t and θ_t becomes grater in its absolute value. To put it another way, the probability of a bubble crash, 1-π_t, is an increasing function of how distant the observed bubble deviates from market fundamentals. When $\alpha = 0$, π_t is independent of x_{t-1} and therefore the probability of crash is constant, which corresponds to the formulation given by Blanchard and Watson (1982) [6]. When $\alpha = \gamma = 0$, the whole process is described by the autoregressive process (A) and when γ is large or $\beta_t = 0$, the process reduces to a simple white noise process and there is no speculative bubble. Thus, by investigating the parameter estimates, we may statistically test the properties of the process.

In principle, we can generalize our formulation by considering a broader class of stochastic models for u_t such as ARMA process or by introducing the fundamental values into the functional form of the transition probability

(3). However, we have tried to keep our model as simple as possible because this paper is only meant to be a first step in this research direction. The specification, (3), of the probability turns out to be one of the few analytically tractable formulations in the following analyses.

When the probability structure of crashes is taken into consideration, we see that the bubble cannot continue forever. As it grows, the probability of a crash approaches unity and x_t will sooner or later be pulled back to the fundamental value θ_t. In this way, the time series of x_t never diverges, but exhibits more or less stable behavior in the longer run.

Note that letting $\theta_t = 0$ and assuming away the constraint $x_t > 0$ leads us to the models of Asako, Kanoh and Sano (1990) [4], Liu, Asako and Kanoh (2011) [5] and Asako and Liu (2013) [1]. In those models, x_t is not a ratio variable but is a stock price bubble measured as deviations from their fundamental values. The model of nonlinear cointegration, which is developed in Section 4, adds to this basic model the property that ratio bubbles are symmetric between upwards and downwards.

2.2. On Recursive Estimation

In Liu, Asako and Kanoh (2011) [5] and Asako and Liu (2013) [1], the entire Bayesian recursive estimation process is described for the periods from 0 to 1 and from period t-1 to period t, thus establishing by way of mathematical induction the validity of the recursive estimation method. We develop here only the recursive way of estimating parameters at period t conditioned on the available data up to period t-1. For more in detail of the entire estimation, refer to Liu, Asako and Kanoh (2011) [5] or Asako and Liu (2013) [1].

One notable difference between the present model (1)-(4) and the earlier ones is that Liu, Asako and Kanoh (2011) [5] and Asako and Liu (2013) [1] assume $\theta_t = 0$. Once we allow for $\theta_t > 0$, whether θ_t is known or unknown causes a big difference in the Bayesian recursive estimation. If it is unknown and to be estimated in the same way as the other parameters of the model, the estimation process becomes too complicated for us to manipulate the model explicitly. On the other hand, if θ_t is known and treated as a predetermined parameter even though we have to somehow "estimate" it eventually, this estimation can be separated from the estimation of the entire model and its recursive estimation process remains, in terms of hardness, almost at the same level as Asako and Liu (2013) [1]. In fact, we let θ_t be known and propose its two candidates in Section 3.

2.3. Recursive Estimation at Period t

In this section, we describe a Bayesian recursive technic to estimate the parameters of our model. Before proceeding to this task, we put $\mathbf{X}^t = \{x_t, x_{t-1}, \ldots, x_0\} = \{x_t, \mathbf{X}^{t-1}\}$ the set of data observations up to period t, and by $\mathbf{I}^t = \{i_1, i_2, \ldots, i_t\}$, we denote the set of ordered integer indices where each i_s ($s = 1; 2; : : : : ; t$) is either 1, 2, or 3.

With these new notations, we write down the joint density for α, β_{t-1}, γ conditional on $\mathbf{X}^{t-1}$:

$$P(\alpha, \beta_{t-1}, \gamma | \mathbf{X}^{t-1}) = \sum_{\mathbf{I}^{t-1}} f(\mathbf{I}^{t-1}) g(\mathbf{I}^{t-1}) P_{\mathbf{I}^{t-1}} \left(\alpha, \gamma | \mathbf{X}^{t-1}\right) N_{\beta_{t-1}} \left(\hat{\mu}(\mathbf{I}^{t-1}), \sigma^2(\mathbf{I}^{t-1})\right) \quad (5)$$

where $f(\mathbf{I}^t)$ and $g(\mathbf{I}^t)$ are certain deterministic functions of $\mathbf{I}^t$ that are to be determined in the sequel so as to satisfy the recursive pattern, whereas $P(.)$ and $N(.)$ denote density functions; $P_{\mathbf{I}^t} (\alpha, \gamma | \mathbf{X}^t)$ is the joint prior density function for constant α and γ^1 over time conditioned on $\mathbf{X}^t$ and $N_{\beta_t} \left(\hat{\mu}(\mathbf{I}^t), \sigma^2(\mathbf{I}^t)\right)$ is the density function of the normal distribution with mean $\hat{\mu}(\mathbf{I}^t)$ and variance $\sigma^2(\mathbf{I}^t)$. Their detailed functional forms as well as the definition of the other factors on the right-hand-side of (4) are given immediately below. Note that the summation is over the entire combination of indices which amount to 3^{t-1} terms at stage t. Then, in view of (2), the joint prior density function for α, β_t, and γ conditioned on $\mathbf{X}^{t-1}$ is

$$P(\alpha, \beta_t, \gamma | \mathbf{X}^{t-1}) = \sum_{\mathbf{I}^{t-1}} f(\mathbf{I}^{t-1}) g(\mathbf{I}^{t-1}) P_{\mathbf{I}^{t-1}} \left(\alpha, \gamma | \mathbf{X}^{t-1}\right) N_{\beta_t} \left(\hat{\mu}(\mathbf{I}^{t-1}), \sigma^2(\mathbf{I}^{t-1}) + \sigma_v^2\right) \quad (6)$$

Now our main task is to calculate the updated posterior density (6) by utilizing the Bayes' theorem:

$$P(\alpha, \beta_t, \gamma | \mathbf{X}^t) = \frac{P(\alpha, \beta_t, \gamma | \mathbf{X}^{t-1}) P(x_t | \alpha, \beta_t, \gamma, \mathbf{X}^{t-1})}{P(x_t | \mathbf{X}^{t-1})}. \quad (7)$$

---

[1]Even if we instead allow for time dependent α and γ, the computational burden remains the same as α and γ are at any rate estimated differently over time.

Introducing a new parameter

$$\phi_t = \theta_t \tag{8}$$

for the sake of later convenience in notation, from (1) and the normality of u_t, we have

$$P(x_t|\alpha, \beta_t, \gamma, \mathbf{X}^{t-1}) = e^{-\gamma - \alpha m_{t-1}} N_{x_t}\left(\beta_t x_{t-1}, \sigma_u^2\right) + (1 - e^{-\gamma - \alpha m_{t-1}}) N_{x_t}\left(\phi_t, \sigma_u^2\right) \tag{9}$$

Therefore, in view of (7), the multiplication of (6) and (9) yields the updated formula of (6) for period t if and only if we have, to begin with

$$P_{\mathbf{I}^t}(\alpha, \gamma|\mathbf{X}^t) = \begin{bmatrix} \left[\left(\hat{c}_0 + \sum_{s=0}^{t-1} \delta(i_s)\right) e^{-\left(\hat{c}_0 + \sum_{s=0}^{t-1} \delta(i_s)\right)\gamma}\right] \\[2em] \left[\left(\hat{a}_0 + \sum_{s=0}^{t-1} |x_s|\delta(i_s)\right) e^{-\left(\hat{a}_0 + \sum_{s=0}^{t-1} |x_s|\delta(i_s)\right)\alpha}\right] \end{bmatrix} \tag{10}$$

where the first and second terms within the large brackets represent, respectively, the probability density function of exponentially and mutually independently distributed α and γ^2. The integer function $\delta(i_t)$

$$\delta(i_t) = \begin{cases} 1 & \text{for } i_t = 1, 2 \\ 0 & \text{for } i_t = 3 \end{cases}$$

is introduced to simplify the mathematical expression.

Moreover, for the unspecified coefficient functions, we must have

$$\begin{cases} f(\mathbf{I}^{t-1}, 1) = \dfrac{f(\mathbf{I}^{t-1})}{P(x_t|\mathbf{X}^{t-1})} N_{x_t}\left(\mu(\mathbf{I}^{t-1})x_{t-1}, (\hat{\sigma}^2(\mathbf{I}^{t-1}) + \sigma_v^2)x_{t-1}^2 + \sigma_u^2\right) \\[1em] f(\mathbf{I}^{t-1}, 2) = -f(\mathbf{I}^{t-1}, 3) = -\dfrac{f(\mathbf{I}^{t-1})}{P(x_t|\mathbf{X}^{t-1})} N_{x_t}(\phi_t, \sigma_u^2) \end{cases} \tag{11}$$

and

$$\begin{aligned} g(\mathbf{I}^t) &= g(\mathbf{I}^{t-1}, i) \\[1em] &= g(\mathbf{I}^{t-1}) \left[\frac{\hat{a}_0 + \sum_{s=0}^{t-2} m_s\delta(i_s)}{\hat{a}_0 + \sum_{s=0}^{t-1} m_s\delta(i_s)}\right] \left[\frac{\hat{c}_0 + \sum_{s=0}^{t-2}\delta(i_s)}{\hat{c}_0 + \sum_{s=0}^{t-1}\delta(i_s)}\right] \\[1em] &= \left[\frac{\hat{a}_0}{\hat{a}_0 + \sum_{s=0}^{t-1} m_s\delta(i_s)}\right] \left[\frac{\hat{c}_0}{\hat{c}_0 + \sum_{s=0}^{t-1}\delta(i_s)}\right]. \end{aligned} \tag{12}$$

Also for means and variances of the normal distributions, it must be

$$\begin{cases} \hat{\mu}(\mathbf{I}^{t-1}, 1) = \dfrac{(\hat{\sigma}^2(\mathbf{I}^{t-1}) + \sigma_v^2)x_{t-1}x_t + \sigma_u^2\hat{\mu}(\mathbf{I}^{t-1})}{(\hat{\sigma}^2(\mathbf{I}^{t-1}) + \sigma_v^2)x_{t-1}^2 + \sigma_u^2} \\[1em] \hat{\mu}(\mathbf{I}^{t-1}, 2) = \hat{\mu}(\mathbf{I}^{t-1}, 3) = \hat{\mu}(\mathbf{I}^{t-1}) \end{cases} \tag{13}$$

and

$$\begin{cases} \hat{\sigma}^2(\mathbf{I}^{t-1}, 1) = \dfrac{(\hat{\sigma}^2(\mathbf{I}^{t-1}) + \sigma_v^2)\sigma_u^2}{(\hat{\sigma}^2(\mathbf{I}^{t-1}) + \sigma_v^2)x_{t-1}^2 + \sigma_u^2} \\[1em] \hat{\sigma}^2(\mathbf{I}^{t-1}, 2) = \hat{\sigma}^2(\mathbf{I}^{t-1}, 3) = \hat{\sigma}^2(\mathbf{I}^{t-1}) \end{cases} \tag{14}$$

Finally, it must be recalled, that by making use of the relationship that applies for conditional density functions

---

[2]We assume that α and γ are exponentially distributed in accordance with the exponential probability of the bubble continuation (3).

$$P(x_t, \alpha, \beta_t, \gamma | \mathbf{X}^{t-1}) = P(\alpha, \beta_t, \gamma | \mathbf{X}^{t-1}) \cdot P(x_t | \alpha, \beta_t, \gamma, \mathbf{X}^{t-1}) \tag{15}$$

and knowing that α, β_t, γ are mutually independent in (6), we immediately obtain

$$P(x_t | \mathbf{X}^{t-1}) = \iiint P(x_t, \alpha, \beta_t, \gamma | \mathbf{X}^{t-1}) d\alpha \, d\beta_t \, d\gamma = \sum_{\mathbf{I}^t} f(\mathbf{I}^t) g(\mathbf{I}^t) \tag{16}$$

which appears in the denominators of (7) and (11). This establishes all requirement that enable Bayesian recursive estimation to update consistently.

2.3.1. Parameter Estimates

The estimates of α, β_t, γ at period t are the conditional expectations on $\mathbf{X}^t$. Thus, referring to period t by suffix t, we have

$$\hat{\alpha}_t \;=\; \mathrm{E}(\alpha | \mathbf{X}^t) = \sum_{\mathbf{I}^t} \frac{1}{\hat{a}_0 + \sum_{s=0}^{t-1} m_s \delta(i_s)} f(\mathbf{I}^t) g(\mathbf{I}^t) \tag{17}$$

$$\hat{\beta}_t \;=\; \mathrm{E}(\beta_t | \mathbf{X}^t) = \sum_{\mathbf{I}^t} f(\mathbf{I}^t) g(\mathbf{I}^t) \hat{\mu}(\mathbf{I}^t) \tag{18}$$

and

$$\hat{\gamma}_t = \mathrm{E}(\gamma | \mathbf{X}^t) = \sum_{\mathbf{I}^t} \frac{1}{\hat{c}_0 + \sum_{s=0}^{t-1} m_s \delta(i_s)} f(\mathbf{I}^t) g(\mathbf{I}^t). \tag{19}$$

We also obtain the probability estimate of bubble continuation from period t-1 to t as

$$\hat{\pi}_t = e^{-\hat{\gamma}_t - \hat{\alpha}_t m_{t-1}} \tag{20}$$

or we can directly obtain the conditional expectation as

$$\hat{\pi}_t^* = \mathrm{E}(e^{-\gamma - \alpha m_{t-1}} | \mathbf{X}^t) = \sum_{\mathbf{I}^t} \frac{\hat{a}_0 f(\mathbf{I}^t) g(\mathbf{I}^t)}{\hat{a}_0 + g(\mathbf{I}^t) m_{t-1}}. \tag{21}$$

Finally, the estimate of the variance of β_t is given by

$$\hat{\sigma}_{\beta_t}^2 = \mathrm{Var}(\beta_t | \mathbf{X}^t) = \sum_{\mathbf{I}^t} f(\mathbf{I}^t) g(\mathbf{I}^t) \hat{\sigma}^2(\mathbf{I}^t). \tag{22}$$

2.3.2. Maximum Likelihood Estimates of Variances

In carrying out the recursive procedure explained above, two variance parameters are to be specified. These are the dispersions of the random terms in (1) and (2), *i.e.*, σ_u^2 and σ_v^2. The likelihood function for these parameters can be obtained in the following way.

Let us put $\sigma^2 = (\sigma_u^2, \sigma_v^2)$ for simplicity. The likelihood function for σ^2 with T periods of data is defined as

$$\begin{aligned} P_T(x_1, x_2, \ldots, x_T | x_0, \sigma^2) \;&=\; P_T(x_T | \mathbf{X}^{T-1}, \sigma^2) P_{T-1}(x_{T-1} | \mathbf{X}^{T-2}, \sigma^2) \cdots P_1(x_1 | x^0, \sigma^2) \\ &=\; \prod_{t=1}^{T} P_t(x_t | \mathbf{X}^{t-1}, \sigma^2). \end{aligned} \tag{23}$$

On the other hand, since

$$P(x_t, \alpha, \beta_t, \gamma | \mathbf{X}^{t-1}, \sigma^2) = P(\alpha, \beta_t, \gamma | \mathbf{X}^{t-1}, \sigma^2) P(x_t | \alpha, \beta_t, \gamma, \mathbf{X}^{t-1}, \sigma^2) \tag{24}$$

and

$$P_t(x_t | \mathbf{X}^{t-1}, \sigma^2) = \iiint P(x_t, \alpha, \beta_t, \gamma | \mathbf{X}^{t-1}, \sigma^2) d\alpha \, d\beta_t \, d\gamma \tag{25}$$

we have, like (16)

$$P_t(x_t | \mathbf{X}^{t-1}, \sigma^2) = \sum_{\mathbf{I}^t} f(\mathbf{I}^t) g(\mathbf{I}^t) \tag{26}$$

Therefore, the log likelihood function of σ^2 can be expressed by

$$\log P_T(\mathbf{X}^T | x_0, \sigma^2) = \sum_{t=1}^{T} \log \left\{ \sum_{\mathbf{I}^t} f(\mathbf{I}^t) g(\mathbf{I}^t) \right\} \tag{27}$$

and the resulting set of variances $\hat{\sigma}^2 = (\hat{\sigma}_u^2, \hat{\sigma}_v^2)$ which maximize (27) are the desired estimates.

2.3.3. Condensation of Recursive Estimation

So far is the complete and mathematically rigorous description of the Bayesian recursive estimation and we can estimate parameters for any length of sample periods. However, the number of terms we need to compute in equations from (11) to (14) and others increases at a rate of 3^t to exceed a standard capacity of computer as the number of time series data increases. For this reason and to reduce the computational burden, we introduce the so-called condensation procedure first proposed by Harrison and Stevens (1981) [7] and applied for the estimation of the basic model by Liu, Asako and Kanoh (2011) [5] and Asako and Liu (2013) [1]. By condensation, we update the parameters of the next period's prior distribution by utilizing the first and second moments of the approximated marginal posterior distribution. This enables the computational burden to remain at a constant level over time.

What we have to do in practice is to approximate the posterior density (5) at period t or the left hand side of (7) by a joint density of the following form

$$
\begin{aligned}
P(\alpha, \beta_{t-1}, \gamma | \mathbf{X}^{t-1}) &= P(\alpha | \mathbf{X}^{t-1}) P(\beta_{t-1} | \mathbf{X}^{t-1}) P(\gamma | \mathbf{X}^{t-1}) \\
&= \hat{a}_{t-1} e^{-\hat{a}_{t-1}\alpha} \hat{c}_{t-1} e^{-\hat{c}_{t-1}\gamma} N_{\beta_{t-1}} \left(\hat{\beta}_{t-1}, \hat{\sigma}_{t-1}^2 \right),
\end{aligned}
\tag{28}
$$

where we utilize the fact that α, β_t, and γ are mutually independent. Then the first and second moments of the marginal densities for each parameter are equated. That is, (5) at period t is approximated by

$$\hat{c}_t e^{-\hat{c}_t \gamma} \hat{a}_t e^{-\hat{a}_t \alpha} N_{\beta_t}(\hat{\beta}_t, \hat{\sigma}_t^2) \tag{29}$$

so that the joint prior density at period $t + 1$ can be written as

$$P(\alpha, \beta_{t+1}, \gamma | \mathbf{X}^t) = \hat{c}_t e^{-\hat{c}_t \gamma} \hat{a}_t e^{-\hat{a}_t \alpha} N_{\beta_{t+1}}(\hat{\beta}_t, \hat{\sigma}_t^2 + \sigma_v^2) \tag{30}$$

where $\hat{a}_t$ and $\hat{c}_t$ are equated, respectively, to the reciprocal of the mean estimates (17) and (19)

$$\hat{a}_t = 1/\hat{\alpha}_t \tag{31}$$

$$\hat{c}_t = 1/\hat{\gamma}_t \tag{32}$$

whereas $\hat{\beta}_t$ and $\hat{\sigma}_{\hat{\beta}_t}^2$ are estimates given by (18) and (22). This procedure can be repeated at each stage.

2.4. Nonliner Cointegration

The basic bubble model (1)-(4) formulates the feature that a ratio variable returns to its fundamental value in the long run as the probability that a bubble crashes reaches 100% insofar as the divergent bubble continues. In other words, although short-run bubbles generate explosive discrepancies between x_t and θ_t, divergent booms would bust eventually and in this sense there is a stable relationship in the long run. This phenomenon is what we call the nonlinear cointegration.

Unlike the definition of linear cointegration, the definition of nonlinear relationship is model-specific. There may be other models of nonlinear cointegration and our nonlinear cointegration should more restrictively be named speculative bubble nonlinear cointegration or boom and bust nonlinear cointegration.

Such being the case, there is no established method to test the nonlinear cointegration relationship. Instead, we are obliged to accept the existence of the nonlinear relationship only passively. We especially put emphasis on the bubble process in (2) and thereby we detect whether $\beta_t > 1$ and how often switches occur between two models or how high is the probability of bubble continuation π_t.

In the empirical analysis in Section 4, we compute the pseudo-t statistics:

$$\hat{t} = (\hat{\beta}_t - 1)/\hat{\sigma}_{\beta_t} \tag{33}$$

in order to sense the "significance" regarding the validity of $\beta_t > 1$. Since the present estimation technic is Bayesian in the sense that we utilize prior information besides the information extracted from the data, statistics like (33) may not obey Student's t-distribution. Nonetheless, we would presume that $t = 1.65$, which is one sided 5%

significant for a standard t test, is a critical level to rely on.

In detecting the validity of the nonlinear cointegration, we may as well examine into the probability of bubble continuation π_t. We check in Section 4 the probability of bubble crash, $1 - \pi_t$, and see its movement over time.

3. Nonlinear Cointegration: Modification of the Basic Model

The basic model we developed in Section 2 is applicable to any series of x_t. In this section, we modify the basic model to deal with a ratio variable $x_t > 0$. A ratio variable may exhibit both upwards and downwards bubble processes with $\theta_t > 0$, which necessitates certain nontrivial revision in recursive estimation.

3.1. Modification of the Basic Model

We alter the basic model into a double regime switching model. One regime switching is that the basic model is of the boom-and-bust type. The other regime switching is that a ratio variable has both upwards (or positive) and downwards (or negative) bubble processes. On the other hand, we maintain (2) or the transition equation of β_t as it is.

Then, we can naturally regard it a bubble by $\beta_t > 1$ once x_t keeps increasing over time. But even when x_t keeps decreasing by a downwards bubble, estimates may end up with $\beta_t < 1$ for certain periods of time. In such a case, we may misunderstand what is really happening because $\beta_t < 1$ is usually a case for a stationary autoregressive process. This is quite embarrassing and we may as well be advised to treat the upwards and downwards bubbles asymmetrically. For this aim, we take the reciprocal of the original ratio when the ratio itself is smaller than θ_t as in (3), thus resulting in a drastic regime switch for negative downwards bubbles.

Let y_t represent an original ratio variable of two stock prices, and let us redefine x_t by

$$x_t = \begin{cases} y_t & \text{when} \quad y_t \geq \theta_t \\ \frac{1}{y_t} & \text{when} \quad y_t < \theta_t \quad \left(\text{or } \frac{1}{y_t} > \frac{1}{\theta_t} \right) \end{cases} \tag{34}$$

With this new x_t, we assume that every aspect of the basic model (1)-(4) is valid, *i.e.*,

$$x_t = \begin{cases} \beta_t x_{t-1} + u_t & \text{with} \quad \pi_t \\ \phi_t + u_t & \text{with} \quad 1 - \pi_t \end{cases} \tag{35}$$

except that

$$\phi_t = \begin{cases} \theta_t & \text{when} \quad y_t \geq \theta_t \\ \frac{1}{\theta_t} & \text{when} \quad y_t < \theta_t \quad \left(\text{or } \frac{1}{y_t} > \frac{1}{\theta_t} \right) \end{cases} \tag{36}$$

replaces (8).

Note that integrating artificially two regimes most likely causes heteroscedasticity in innovation term u_t in (1) or (35). In fact, we will introduce proportional variance of u_t to x_{t-1} squared in our empirical analysis in Section 4:

$$Var(u_t) = x_{t-1}^2 \sigma_u^2 \tag{37}$$

Lastly, we need to revise the probability of bubble continuation. That is, in (3), we have

$$m_t = \max \left\{ |x_t - \theta_t|, \left| x_t - \frac{1}{\theta_t} \right| \right\} \tag{38}$$

or

$$m_t = \max \left\{ |y_t - \theta_t|, \left| \frac{1}{y_t} - \frac{1}{\theta_t} \right| \right\} \tag{39}$$

that replaces (4). In (38) or (39), the greater deviation is $x_t - \theta_t$ for the positive upwards bubble and $x_t - 1/\theta_t = 1/y_t - 1/\theta_t$ for the negative downwards bubble.

3.2. Known θ_t

As we have already noted, the fundamental stock prices ratio θ_t is assumed known and given to us exogenously

at period t. There may be several candidates for θ_t. Here we propose two alternative ones[3].

3.2.1. Past Average

The first candidate is the simple arithmetic average of all the past data:

$$\theta_t = \frac{\sum_{i=1}^{t} y_t}{t} \tag{40}$$

Although we put equal weight on each data, the informational role of the current data decreases over time as (40) by definition is rewritten as $\theta_t = \{(t-1)\,\theta_{t-1} {}^{+} y_t\}/t$, which in turn is rewritten as

$$\theta_t = \theta_{t-1} + \frac{y_t - \theta_{t-1}}{t} \tag{41}$$

Equation (41) implies that θ_t follows a random-walk type sticky movement except that the drift term is not stochastic but is given deterministically. As t increases, the contribution of the second term on the right hand side of (41) decreases over time.

3.2.2. Fixed Period Moving Average

The second candidate approximates the fundamental value by the fixed period (say 12 months) moving average up to the current one. Thus in place of (40) we have

$$\theta_t = \frac{\sum_{t-11}^{t} y_t}{12}. \tag{42}$$

And thereby in place of (41), we have

$$\theta_t = \theta_{t-1} + \frac{y_t - \theta_{t-1}}{12} \tag{43}$$

for $t > 12$. As for the first 12 months, we use the simple average (40).

3.3. Estimation Procedure at Period t

At period t, we compute θ_t once we get a new data y_t and we determine which regime we are in, *i.e.*, whether a positive bubble ($y_t \geq \theta_t$) or a negative bubble ($y_t < \theta_t$). If we are rigorously interested in whether the stock price ratio is in positive upwards phase or in negative downwards phase, we may watch where we have been in the past. For example, we would recognize regime shifts only if the opposite new regime continues at least a few consecutive periods. This will exclude a fake regime shift that occurs unsystematically. The idea of this rule of thumb stems from the Bry-Boschan method in the judgment of the business cycle phase.

Once θ_t and thereby the data x_t of (34) is obtained, we are ready to utilize the recursive estimation technic developed in Section 2. We estimate the basic model as applied to the stock market prices of Japan and the United States.

4. Stock Prices of Japan and the United States

Asako, Zhang and Liu (2014) attempted to apply the nonlinear cointegration to the stock markets of Japan, the United States and China. They first checked whether there is a linear cointegration relationship between these countries and concluded negatively for any pair of countries. Then they estimated the basic model of (1)-(4) and of three ways of the known fundamental stock prices ratio including (40) and (42). Among these, in what follow, we develop the most representative case of the nonlinear cointegration; namely the one between the stock price indexes of Japan and the United State.

4.1. Preparatory Steps

The monthly time series data we have chosen are the Nikkei225 index (hereinafter Nikkei225) for Japan and the

---

[3]We may find some variables Z_t that are to be reflected in the fundamentals θ_t. Then the fitted value of an OLS regression equation of x_t on Z_t appears to be another candidate. However, the estimate of θ_t thus constructed is neither consistent nor efficient, if not unbiased.

Dow-Jones Industrial Average Stock Price Index (hereinafter DJ) for the United States. **Figure 1** plots these stock prices and their ratio (DJ/Nikkei225) from January 1970 to December 2012.

4.1.1. Derivation of Known θ_t

Figure 2 exhibits the fundamental stock prices ratio given by (40) and (42). Not surprisingly, (i) the past average θ_t^1 shows a random-walk type sluggish swing whereas (ii) the fixed period moving average θ_t^2 traces short lived ups and downs around the historical actual path of the ratio y_t.

4.1.2. Artificial Dependent Variable

Next, we construct from the time series y_t that of the artificial variable x_t by (34). Referring to the realized y_t and two fundamental stock prices ratio θ_t, the time series of x_t consists of negative bubble ($y_t < \theta_t$) up to the mid 1990s and thereby, by definition, x_t equals the reciprocal of y_t. On the contrary, during the latter half of the sample period, x_t consists of positive bubble ($y_t > \theta_t$) and x_t is y_t itself. In the case of θ_t^2, however, $y_t > \theta_t$ and $y_t < \theta_t$ interchange with small intervals, as does x_t.

4.1.3. Maximum Likelihood Estimates of Variances

We need to obtain the maximum likelihood estimates for the variances of u_t in (1) and v_t in (2). We also have to set initial values in beginning the recursive estimation. The effect of the initial conditions turns out to be minimal as we tried several combinations to result in little difference in the main feature of estimation except for several initial periods. The final choice was $\hat{\beta}_0 = 1$, $\hat{\sigma}_0 = 0.01$, and $\hat{a}_0 = \hat{c}_0 = 0.01$ and denoting by $\hat{\sigma} = (\hat{\sigma}_u, \hat{\sigma}_v)$ the pair of standard deviations, the maximum likelihood estimates were (0.0536, 0.0000) for θ_t^1 and (0.0456, 0.0000) for θ_t^2. The resultant log likelihoods were 377.9 and 600.7, respectively.

Judging on the log likelihood, between the two fundamental stock prices ratio, θ_t^2 fits the data better than θ_t^1 does. Knowing this consequence, we yet report those alternative fundamental values as these yield really comparable estimation results as we explain in the sequel[4].
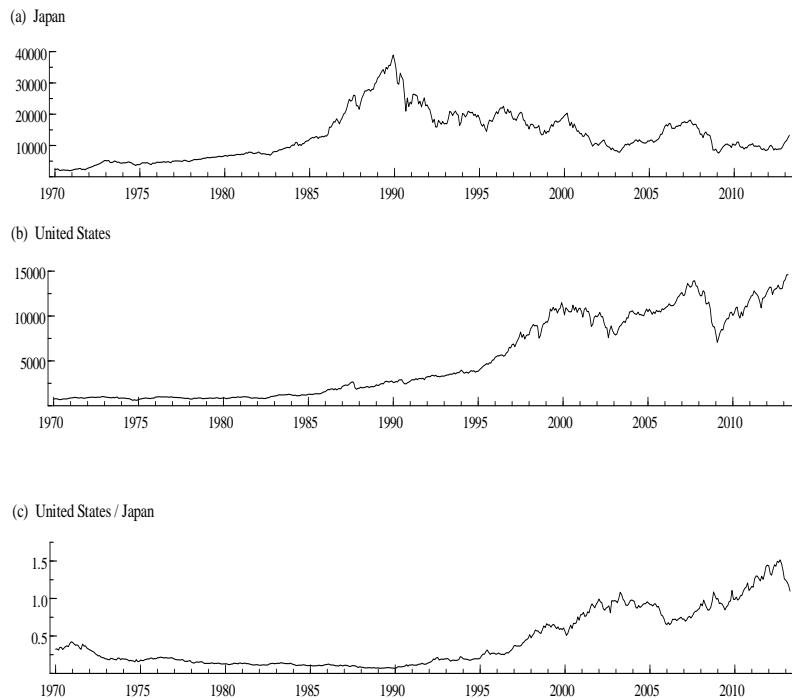

Figure 1. Stock Price Indexes: Japan and the US. Note) The Nikkei225 for Japan and DJ for the United States.

[4]This ordering is not robust. When we assume away the heteroscedastic variance (37), log likelihood becomes larger with θ_t^1 than the one with θ_t^2.
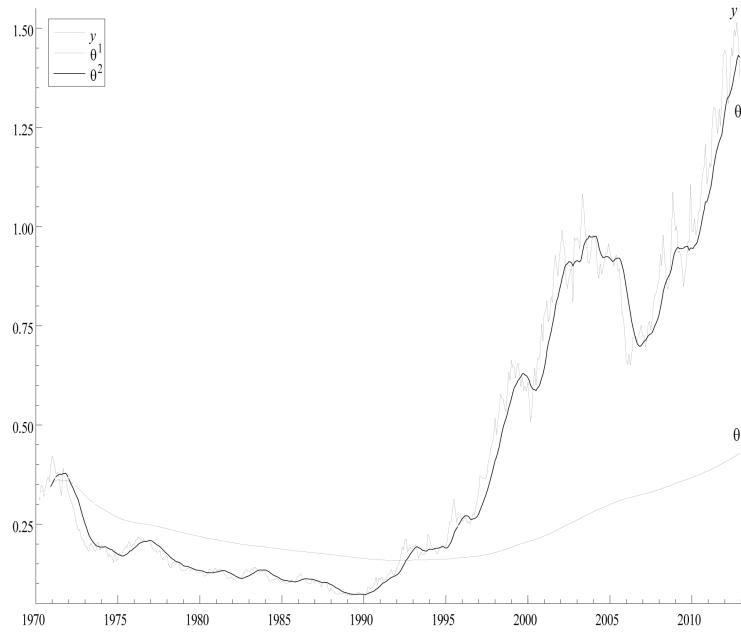
Figure 2. Fundamental value θ_t.

4.2. Necessary Condition for the Bubble

With the above preparation, **Figure 3** exhibits the estimate of the key parameter β_t. The percentage of samples that satisfies the necessary requirement for bubbles $\beta_t > 1$ amounts to 82.9% for θ_t^1 and 100% for θ_t^2 among the entire 43 years' sample periods (516 months from 1970; 1 to 2012;12). Namely with both θ_t^1 and θ_t^2, samples with $\beta_t > 1$ exceed more than 80%. These observations may as well support the view that the model (1)-(4) with reasonable modification fits the data and the stock markets of Japan and the United States are cointegrated nonlinearly in the long run. But how reliable is this result?

To answer to this question, we checked the pseudo t t-statistic (33) and found, as summarized in **Table 1**, that $\beta_t > 1$ is one sided 5% "pseudo-significant" is nil for θ_t^1 and 93.0% for θ_t^2 (similarly the nonstationarity condition $\beta_t < 1$ is not significant). These suggest that the standard deviation of β_t is relatively large, and the reliability of the estimates is limited. Note, on the contrary, that $\beta_t > 1$ is 93.3% pseudo-significant for θ_t^2.

A clue to this is that the maximum likelihood variance estimate σ_v^2 is extremely small and is virtually the corner solution at zero. In this case, the key parameter β_t is theoretically regarded constant in (2). But, like the parameters α and γ, the estimate of β_t does not have to stay unchanged over time. Even if the variance of v_t is 0 in (2), we have

$$\beta_t = \beta_{t-1} + \text{constant},$$

and β_t can be different from β_{t-1}. Moreover, even if the constant term is 0 and $\beta_t = \beta_{t-1}$, in theory, because β_t is estimated as the expected value of the posterior distribution *à la* Bayesian, it can differ from β_{t-1} once the data increases information in the posterior distribution in (18).

4.3. Probability of Bubble Crash

In **Figure 4**, we plot the probability of bubble crash, 1-π_t. As π_t, the conditional expectation (21), rather than the point estimate (20), is chosen[5]. With θ_t^1 the crash probability remains small except for the early 1970's, which seems to be a transitional feature incorporating specific initial conditions, whereas with θ_t^2 the probability repeatedly rises and falls depending on the state of bubbles.

4.4. Other Cases

Asako, Zhang and Liu (2014) [3] estimate several other cases including exchange rate adjusted stock prices, the

---

[5]Two estimates are very close and are the same to three or four decimal places.
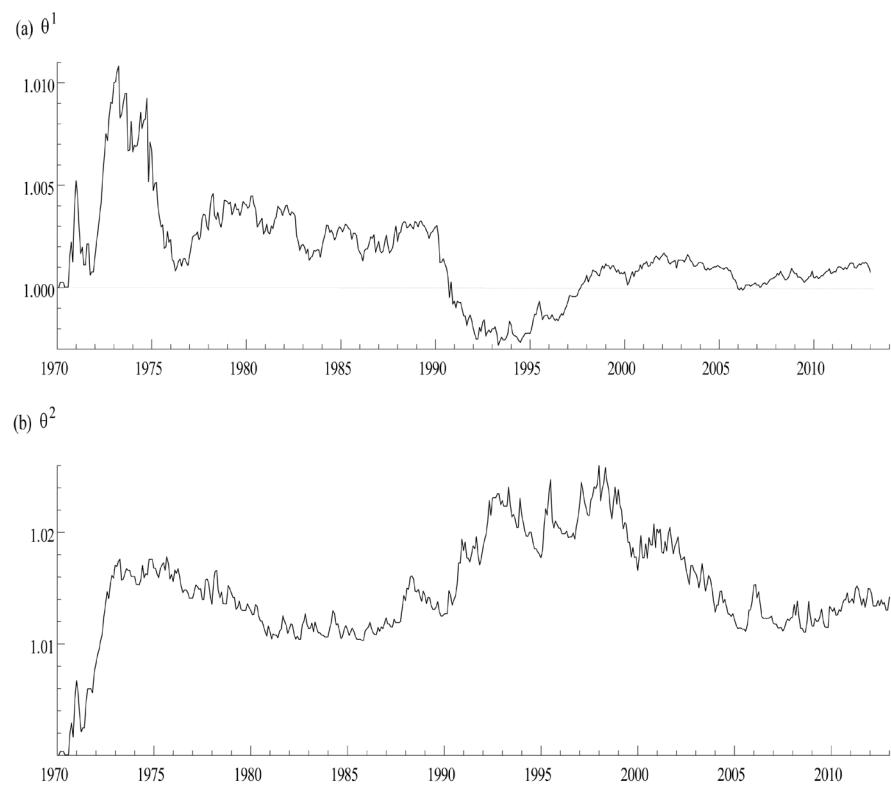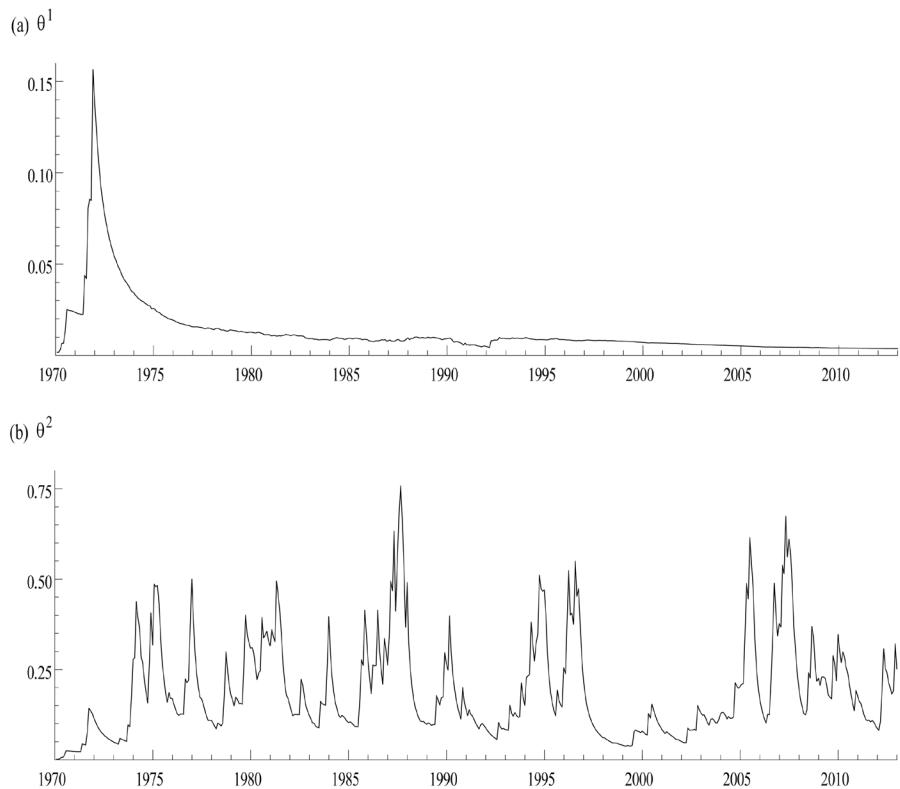
(a) θ^1

(b) θ^2

Figure 3. Estimate of β_t.

(a) θ^1

(b) θ^2

Figure 4. Probability of bubble crash.

Table 1. Number of months of the estimated β_t.

		1970-72	1973-76	1977-80	1981-84	1985-88	1989-92	1993-96	1997-2000	2001-04	2005-08	2009-12
θ_t^1	$\beta_t \geq 1$	35	48	48	48	48	21	0	40	48	44	48
	significant at 5%	0	0	0	0	0	0	0	0	0	0	0
	$\beta_t < 1$	0	0	0	0	0	27	48	8	0	4	0
	significant at 5%	0	0	0	0	0	0	0	0	0	0	0
θ_t^2	$\beta_t \geq 1$	35	48	48	48	48	48	48	48	48	48	48
	significant at 5%	9	48	47	40	48	48	48	48	48	48	48
	$\beta_t < 1$	0	0	0	0	0	0	0	0	0	0	0
	significant at 5%	0	0	0	0	0	0	0	0	0	0	0

case of Var (u_t) = $x_{t-1} \sigma_u^2$ instead of (37), stock prices ratio of Japan and China, and that of China and the United States. The estimation results vary case by case but reaches the conclusion that the basic model (1)-(4) and its modification with $\beta_t > 1$ fits the data reasonably well, thus establishing the latent nonlinear boom and bust relationship between relevant stock prices.

5. Concluding Remarks

In this paper we proposed and developed the recursive estimation method of the nonlinear cointegration. The purpose of this attempt has been to show the usefulness of introducing the idea of nonlinear cointegration. By applying this idea to the stock market indexes of Japan and the United States, we have seen that these indexes commove in the long run although they deviate from this relationship in the short run.

References

[1] Asako, K. and Liu, Z.T. (2013) A Statistical Moddel of Speculative Bubbles, with Applications to the Stock Markets of the United States, Japan, and China. *Journal of Banking & Finance*, **37**, 2639-2651. http://dx.doi.org/10.1016/j.jbankfin.2013.02.015

[2] Engle, R.F. and Granger, C.W.J. (1987) Cointegration and Error Correction: Representation, Estimation and Testing. *Econometrica*, **55**, 251-276. http://dx.doi.org/10.2307/1913236

[3] Asako, K., Zhang, Y. and Liu, Z.T. (2014) The Comovement in Stock Price Indexes of Japan, United States, and China: Estimation of a Nonlinear Cointegration. *Economic Review*, **65**, 56-85 (in Japanese).

[4] Asako, K., Kanoh, S. and Sano, H. (1990) Stock Price and Bubble. In: Nishimura, K. and Miwa, Y., Eds., *Stock and Land Prices in Japan—Mechanism of Price Determination*, University of Tokyo Press, 57-86 (in Japanese).

[5] Liu, Z.T., Asako, K. and Kanoh, S. (2011) Estimation of Speculative Bubble—Application to the Stock Markets of Japan, the United States and China. In: Asakoand, K. and Watanabe T., Eds., *Econometrical Analyses of Finance and Business Cycle*, Minerva Shobou, 9-34 (in Japanese).

[6] Blanchard, O.J. and Watson, M. (1982) Bubbles, Rational Expectations and Financial Markets. In: Wachtel, P., Ed., *Crises in the Economic and Financial Structure*, Lexington Books, 295-315.

[7] Harrison, P.J. and Stevens, C.F. (1976) Bayesian Forcasting (with Discussion). *Journal of the Royal Statistical Society*: *Series B*, **38**, 205-247.