Title	A survey on the estimation of CCAPMs via moment restrictions : the case of Japan
Sub Title	
Author	野田, 顕彦(Noda, Akihiko)
Publisher	Keio Economic Society, Keio University
Publication year	2013
Jtitle	Keio economic studies Vol.49, (2013.) ,p.69- 91
JaLC DOI	
Abstract	The purpose of this paper is to provide brief review of estimation methods for the standard consumption-based asset pricing model (CCAPM) and some of its serious empirical problems, namely, the two puzzles in the Japanese financial market. First, we introduce Hansen's (1982) generalized method of moments (GMM) estimator for estimating the parameter of the standard CCAPM. Second, we show the superiority of alternative GMM estimator, generalized empirical likelihood (GEL), by measuring the difference of the higher order bias on the standard CCAPM and indicate the GEL estimator suggest a possibility for solving the puzzles. Last, we suggest a few methods to examine the standard CCAPM for future research.
Notes	
Genre	Journal Article
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AA00260492-20130000-0069

KEIO ECONOMIC STUDIES **49**, 69–91 (2013)

A SURVEY ON THE ESTIMATION OF CCAPMS VIA MOMENT RESTRICTIONS: THE CASE OF JAPAN

Akihiko NODA

Faculty of Economics, Wakayama University, Wakayama, Japan
&
Keio Economic Observatory, Keio University, Tokyo, Japan

First version received August 2011; final version accepted September 2012

Abstract: The purpose of this paper is to provide brief review of estimation methods for the standard consumption-based asset pricing model (CCAPM) and some of its serious empirical problems, namely, the two puzzles in the Japanese financial market. First, we introduce Hansen's (1982) generalized method of moments (GMM) estimator for estimating the parameter of the standard CCAPM. Second, we show the superiority of alternative GMM estimator, generalized empirical likelihood (GEL), by measuring the difference of the higher order bias on the standard CCAPM and indicate the GEL estimator suggest a possibility for solving the puzzles. Last, we suggest a few methods to examine the standard CCAPM for future research.

Key words: GMM, GEL, CCAPM, Equity Premium Puzzle, Risk-free Rate Puzzle.
JEL Classification Number: G12, E21, C13.

1. INTRODUCTION

In economics, it is expected that an investor will change his/her intertemporal consumption allocation in response to changes in real assets returns. This relationship can be explained by the subjective discount factor and the degree of relative risk aversion (and/or intertemporal elasticity of substitution). Especially in finance, many studies examine the relationship using the standard consumption-based asset pricing model (CCAPM) developed by Rubinstein (1976), and Lucas (1978) and Breeden (1979), and estimate the parameters of the standard CCAPM using the generalized method of moments (GMM) estimator.

There are, however, some serious empirical problems that arise in the standard

Acknowledgments. The author would like to thank Mikio Ito, Colin McKenzie, Taisuke Otsu, Naoya Sueishi, Shunsuke Sugiyama, Tatsuma Wada, Makoto Yano, and an anonymous referee for their helpful comments and suggestions. The author would also like to thank the financial assistance provided by the Japan Society for the Promotion of Science Grant in Aid for Scientific Research No.24530364 and the Keio Gijuku Academic Development Funds. All data and programs used for this paper are available on request.
E-mail: noda@eco.wakayama-u.ac.jp; noda@sanken.keio.ac.jp

CCAPM framework. Initially, Mehra and Prescott (1985) point out a puzzle, namely, the inability of standard CCAPM to rationalize the statistics that have characterized U.S. financial markets over the past century. Specifically, they show that the models fail to explain the difference between the average returns of risky and risk-free assets in U.S. financial markets. This puzzle, called the equity premium puzzle, comes from an equation describing the intertemporal rational behavior of participants in financial markets. It is easy to verify the puzzle using several statistics calculated from financial data and estimates of the subjective discount factor and the degree of risk aversion. Inspired by this puzzle, Weil (1989) points out another puzzle that is a variant of the equity premium puzzle. In turn, economists confront the inability of these models to explain the average return of safe assets. The puzzles remain unsolved for the U.S. and other industrialized countries, including Japan (see also Kocherlakota (1996) and Mehra and Prescott (2003) for details).

In order to resolve the discrepancy between the model's predictions and empirical data, some economists have modified their theoretical models. To date, not so much attention has been paid to the methods of estimating the parameters in the standard CCAPM. As Stock and Wright (2000) report, the analysis of conventional GMM procedures of the standard CCAPM and linear instrumental variable regressions breaks down when some or all of the parameters are weakly identified. In an effort to improve the poor performance of GMM in small samples, a number of alternative estimators have been suggested. One of the alternatives is the class of generalized empirical likelihood (GEL) estimators. GEL estimators are attracting the attention of many econometricians because of their better performance compared to the GMM estimator. This paper argues that instead of using the GMM estimator, we should employ GEL estimators to estimate the parameters of the standard CCAPM. We introduce earlier studies and discuss alternative methods to resolve the puzzles on the standard CCAPM framework in more detail below.

The outline of this paper is as follows. In section 2, we introduce Hansen's (1982) GMM estimator, and how the GMM estimator is applied to the standard CCAPM. In section 3, we give a brief review of the puzzles and several existing attempts in the context of the standard CCAPM framework in Japan. In section 4, we discuss small sample properties of the GMM estimator and introduce an alternative estimator, the GEL estimator, to avoid poor small sample properties of the GMM estimator. In section 5, we present an application to the standard CCAPM using Japanese financial data and suggest a possibility for solving the puzzles. In the last section, we summarize this paper and suggest a few methods to examine the standard CCAPM for future research.

2. GMM AND CCAPM[1]

In this section, we present the traditional empirical method, the GMM estimator of Hansen (1982), to estimate the parameters of the standard CCAPM. First, we introduce

[1] This section and section 4 rely on Otsu (2007), which contains excellent lecture notes on the GMM and GEL estimators.

the Hansen's (1982) GMM estimator, which provides us with a way to estimate parameters, and undertake hypotheses testing. We then indicate how this estimator is used to estimate the parameters of the standard CCAPM.

2.1. Hansen's (1982) GMM Estimator

Here, we introduce a brief review of the estimators of a moment restriction model. Before we review the estimators of the parameters in such a model, let us define some notation. $\{x_i\}_{i=1}^n$ denotes a sequence of i.i.d. observations of the data, and p denotes the number of parameters to be estimated. We sometimes write $g(\cdot)$ as a m vector of functions of the data and the parameters. We assume that $m \geq p$, and the model has a true parameter θ_0 that satisfies the following condition:

$$E[g(x_i, \theta_0)] = 0 , \tag{1}$$

where $E[\cdot]$ denotes expectations taken with respect to the distribution of x_i's. First, we briefly present Hansen's (1982) two-step GMM (2S-GMM) estimator. Let

$$\hat{g}(\theta) := \frac{1}{n}\sum_{i=1}^{n} g(x_i, \theta) , \tag{2}$$

where n is the sample size. The GMM estimator, $\hat{\theta}_{GMM}$, minimizes the quadratic form:

$$\hat{\theta}_{GMM} = \arg\min_{\theta\in\Theta} n\hat{g}(\theta)'\hat{W}\hat{g}(\theta) . \tag{3}$$

where $\hat{W}$ is an $m \times m$ weighting matrix such that $\hat{W} \xrightarrow{p} W$. Denoting $G := E\left[\frac{\partial g(x_i,\theta_0)}{\partial\theta'}\right]$ and $\Omega := E[g(x_i, \theta_0)g(x_i, \theta_0)']$, then the asymptotic properties of the GMM estimator are as follows (see Newey and McFadden (1994) for details):

- Consistency: $\hat{\theta}_{GMM} \xrightarrow{p} \theta_0$.
- Asymptotic normality:
  $\sqrt{n}\big(\hat{\theta}_{GMM} - \theta_0\big) \xrightarrow{d} \mathcal{N}(0, (G'WG)^{-1}G'W\Omega WG(G'WG)^{-1})$.
- Asymptotic efficiency:
  if $W = \Omega^{-1}$, $\sqrt{n}\big(\hat{\theta}_{GMM} - \theta_0\big) \xrightarrow{d} \mathcal{N}(0, (G'\Omega^{-1}G)^{-1})$.

The optimal weighting matrix Ω^{-1} is often unknown. Therefore, we calculate the first-step estimator of θ, $\hat{\theta}_{1S}$, by using a known weighting matrix (for example, choosing W as the identity matrix I). Then, using $\hat{\theta}_{1S}$ we can obtain an estimate of the optimal weighting matrix Ω^{-1} as

$$\hat{\Omega} := \frac{1}{n}\sum_{i=1}^{n} g(x_i, \hat{\theta}_{1S})g(x_i, \hat{\theta}_{1S})' . \tag{4}$$

The feasible two-step GMM (2S-GMM) is defined as

$$\hat{\theta}_{2S-GMM} := \arg\min_{\theta\in\Theta} n\hat{g}(\theta)'\hat{\Omega}^{-1}\hat{g}(\theta) . \tag{5}$$

If $\hat{\Omega}$ is a consistent estimator of Ω, $\hat{\theta}_{2S-GMM}$ is an asymptotically efficient estimator,[2] and its asymptotic distribution is given by

$$\sqrt{n}\big(\hat{\theta}_{2S-GMM} - \theta_0\big) \xrightarrow{d} \mathcal{N}(0, (G'\Omega^{-1}G)^{-1}) .$$

In order to analyze the specification of the estimated model, we can adopt Hansen's (1982) J test of overidentifying restrictions which has the following property under the null hypothesis that equation (1) is true:

$$J := n\hat{g}(\hat{\theta}_{2S-GMM})'\hat{\Omega}^{-1}\hat{g}(\hat{\theta}_{2S-GMM}) \xrightarrow{d} \chi^2_{m-p} . \tag{6}$$

In the next subsection, we present the Hansen's (1982) GMM estimator to estimate the parameters of the standard CCAPM in moment restriction models.

2.2. *Moment Restriction Model for CCAPM*

Let us consider the following Euler equation in the standard CCAPM, in which we assume that the utility function is in the constant relative risk aversion (CRRA) class:

$$E_t\left[\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}(1 + R_{i,t+1}) - 1\right] = 0 , \quad i = 1, 2, \ldots, N , \tag{7}$$

where $R_{i,t+1}$ is the real return of the ith asset at time $t + 1$, which is defined as

$$R_{i,t+1} = \frac{p_{i,t+1} + d_{i,t+1}}{p_{i,t}} - 1 , \quad i = 1, 2, \ldots, N , \tag{8}$$

where the subscript t indicates time, C_t is real per capita consumption, $p_{i,t}$ is the price of the ith asset, $d_{i,t}$ is the dividend of the ith asset, N is the number of assets, β is the subjective time discount factor, γ is the relative risk aversion (RRA), and $E_t[\cdot]$ is the expectation operator conditional on the information available. In most empirical research, equation (7) is used to estimate the subjective time discount factor and the degree of relative risk aversion (or its reciprocal, the intertemporal elasticity of substitution) and to test the restrictions imposed by equation (7).

We derive unconditional moment restrictions for the standard CCAPM with each N error vectors $\mathbf{u}_{t+1}(\theta)$ depending on the underlying parameter vector θ and the instruments, $\mathbf{z}_t$ be a $K \times 1$ vector of instruments known at time t. We define a $m(= N \cdot K)$ moment indicator vector $g_t(\theta)$ as

$$g_t(\theta) = \mathbf{u}_{t+1}(\theta) \otimes \mathbf{z}_t . \tag{9}$$

The error vector is defined as

$$\mathbf{u}_{t+1}(\theta) = \left[\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} \cdot (\mathbf{1} + \mathbf{R}_{t+1})\right] - \mathbf{1} , \tag{10}$$

where $\mathbf{R}_{t+1} = (R_{1,t+1}, R_{2,t+1}, \ldots, R_{N,t+1})'$, $\mathbf{1} = (1, 1, \ldots, 1)'$ and $\theta = (\beta, \gamma)'$. Given equations (7) and (10), we can derive

$$E_t[\mathbf{u}_{t+1}(\theta)] = \mathbf{0} , \tag{11}$$

[2] We must employ a heteroskedasticity and auto-correlation consistent (HAC) estimate of $\hat{\Omega}$ like the one proposed by Newey and West (1987) when we use time-series data. This estimator depends on a kernel and its bandwidth that can be chosen using the procedure suggested by Andrews (1991).

where $\mathbf{0} = (0, 0, \ldots, 0)'$. Let $\mathbf{z}_t$ be an K vector of instruments which is a subset of the information set available at time t, I_t, and define the $NK(= m)$ vector, $g_t(\theta)$, as

$$g_t(\theta) = \mathbf{u}_{t+1}(\theta) \otimes \mathbf{z}_t \,. \tag{12}$$

Then the Euler equation implies

$$\begin{aligned} E[g_t(\theta)] &= E[E[g_t(\theta)|I_t]] \\ &= E[E[\mathbf{u}_{t+1} \otimes \mathbf{z}_t|I_t]] \\ &= E[E[\mathbf{u}_{t+1}|I_t] \otimes \mathbf{z}_t] \\ &= E[\mathbf{0}] \\ &= \mathbf{0}\,, \end{aligned} \tag{13}$$

where $E[\cdot]$ is the unconditional expectation operator. We call equation (13) a moment restriction model.

Generally, let $\{y_t\}_{t=1}^{T}$, denote observations on a finite dimensional process, which is usually assumed to be stationary and strongly mixing (see Smith (2011)). The right-hand side of equation (12), $g_t(\theta)$, is called a moment indicator, which is a function of the parameters, but also depends on the data, y_t, and potentially the instruments, $\mathbf{z}_t$. When our attention is focused on the parameters to be estimated, y_t and $\mathbf{z}_t$ are usually omitted.

3. TWO PUZZLES AND EARLIER EMPIRICAL STUDIES IN JAPAN

Since Mehra and Prescott (1985) and Weil (1989), many studies indicate that there are puzzles, the equity premium puzzle and the risk-free rate puzzle, in the financial market when the standard CCAPM framework is used. First, we first introduce two puzzles in the financial market: the equity premium puzzle and the risk-free rate puzzle. Second, we present a brief review of Japanese earlier empirical studies in the context of the standard CCAPM.

3.1. The Equity Premium Puzzle

In order to analyze the equity premium, Hansen and Singleton (1983) assume that the joint conditional distribution of asset returns and the stochastic discount factor is lognormal and homoskedastic. These assumptions simplify the later discussion of how the equity premium is determined. When a random variable X is conditionally lognormally distributed,

$$\log E_t X = E_t \log X + \frac{1}{2}\mathrm{Var}_t \log X \,, \tag{14}$$

where $\mathrm{Var}_t \log X := E_t[(\log X - E_t \log X)^2]$. If, in addition X, is conditionally homoskedastic, we obtain

$$\mathrm{Var}_t \log X = E[(\log X - E_t \log X)^2] = \mathrm{Var}(\log X - E_t \log X) \,. \tag{15}$$

Thus, with joint conditional lognormality and the homoskedasticity of asset returns and consumption, we can take logs of equation (7) to obtain

$$0 = E_t[r_{i,t+1}] + \log\beta - \gamma E_t[\Delta c_{t+1}] + \frac{\sigma_i^2 + \gamma^2\sigma_c^2 - 2\gamma\sigma_{ic}}{2}\,, \tag{16}$$

where $c_t = \log(C_t)$, $r_{i,t} = \log(1 + R_{i,t})$, σ_i^2 denotes the unconditional variance of log return innovations $\mathrm{Var}(r_{i,t+1} - E_t(r_{i,t+1}))$, σ_c^2 denotes the unconditional variance of log consumption innovations $\mathrm{Var}(c_{t+1} - E_t(c_{t+1}))$, and σ_{ic} denotes the unconditional covariance of log-return and log-consumption innovations $Cov(r_{i,t+1} - E_t(r_{i,t+1}), c_{t+1} - E_t(c_{t+1}))$.

We consider first an asset with a risk-free real return r_{t+1}^f. For this asset the return innovation variance σ_f^2 and the covariance σ_{fc} are both zero, so we obtain from equation (16) the following expression for the risk-free real return:

$$r_{t+1}^f = -\log\beta + \gamma E_t[\Delta c_{t+1}] - \frac{\gamma^2\sigma_c^2}{2}\,. \tag{17}$$

Equation (17) indicates that the risk-free real return is linear in expected consumption growth, with a slope coefficient equal to the degree of relative risk aversion. Subtracting equation (17) from equation (16) yields an expression for the expected excess return on risky assets over the return on the risk-free asset:

$$E_t[r_{i,t+1} - r_{t+1}^f] + \frac{\sigma_i^2}{2} = -\gamma\sigma_{ic}\,. \tag{18}$$

Equation (18) shows that risk premia is determined by the degree of relative risk aversion times the covariance with consumption growth. This indicates that an asset with a high covariance of consumption tends to have low returns when consumption is low, that is, when the marginal utility of consumption is high.

Mehra and Prescott (1985) show that the standard CCAPM fails to explain the observed difference between the average returns of risky and risk-free assets in U.S. financial markets. This puzzle, called the equity premium puzzle, comes from an equation (18) describing the intertemporal rational behavior of participants in financial markets. This puzzle we can easily verify using several statistics calculated from financial data together with estimates of the subjective discount factor and the degree of relative risk aversion. Unless investors are extremely risk averse, the observed equity premium is too high to be consistent with observed consumption behavior.

3.2. The Risk-free Rate Puzzle

Weil (1989) indicates that even if we resolve the equity premium puzzle by using larger values for the degree of relative risk aversion, γ, this leads to a second puzzle. Equation (17) implies that the unconditional mean risk-free return is given by

$$E[r_{t+1}^f] = -\log\beta + \gamma g - \frac{\gamma^2\sigma_c^2}{2}\,, \tag{19}$$

where g is the mean growth rate of consumption. Weil (1989) shows that there is another puzzle called the "risk-free rate puzzle" which is that if investors are risk-averse (γ is too high), then with power utility they must also be extremely unwilling to substitute intertemporally (ψ is too small). Given positive average consumption growth, a low risk-free rate and a high rate of time preference, such investors would have a strong

Table 1. Estimation results in earlier studies

	$\hat{\beta}$	$\hat{\gamma}$	HJD	$\mathcal{N}$	CONS	Puzzles
Hamori (1992)	0.997	0.216	–	96	M	No
Hamori (1994)	0.997	0.153	–	240	M	No
Tanigawa (1994)	0.998	0	Reject	108	M	Yes
Hori (1996)	0.997	0.049	Reject	179	M	Yes
Nakano and Saito (1998)	0.983	0	Reject	65	SA	Yes
Baba (2000)	0.996	0.721	Accept	76	Q	No

Notes:

(1) "$\hat{\beta}$" denotes the estimate of the subjective discount rate, "$\hat{\gamma}$" denotes the estimate of degree of the relative risk aversion, "HJD" denotes the result of Hansen and Jagannathan's (1991) volatility bound test, and "$\mathcal{N}$" denotes the number of observations.

(2) "CONS" denotes the frequency of the gross per capita consumption growth: "M" for monthly, "SA" for semi-annual, and "Q" for quarterly.

desire to borrow from the future to reduce their average consumption growth rate. A low risk-free rate is possible in equilibrium if investors have a low or negative rate of time preference that reduces their desire to borrow.

If the degree of relative risk aversion γ is high enough then the negative quadratic term $-\gamma^2\sigma_c^2/2$ in equation (19) dominates the linear term and pushes the risk-free rate down again. The quadratic term reflects precautionary savings; risk-averse agents with uncertain consumption streams have a precautionary desire to save, which can work against their desire to borrow. But a reasonable rate of time preference is obtained only as a knife-edge case. In the next section, we introduce several approaches in earlier studies that have sort to resolve these two puzzles.

3.3. *Earlier Empirical Studies in Japan*

Although there is a consensus that the puzzles exists in the U.S. stock market, the same statement is not true when we use Japanese data. In other words, several earlier studies attempted to resolve the puzzles in the Japanese financial markets, there is not yet a general consensus. For example, Hamori (1992, 1994) and Baba (2000) conclude that the puzzles does not exist, while Tanigawa (1994), Hori (1996) and Nakano and Saito (1998) show that there exists the puzzles in Japan. Therefore, we only present a brief review of Japanese earlier empirical studies below to avoid confusion.

Table 1 presents the estimation results using the GMM estimator in earlier studies. It shows that (1) the estimates of parameters of the standard CCAPM is almost from zero to 0.2 except with Baba (2000), (2) there is no consensus about the results of Hansen and Jagannathan's (1991) volatility bound test, and (3) the sample sizes are all less than 300 in these papers.

In particular, the estimates by Hamori (1992) lead to the conclusion that the puzzles do not exist, while Nakano and Saito (1998) report quite opposite results: the puzzles exists. We can confirm that the Hamori's (1992) 2S-GMM estimates of β and γ lead to the conclusion. However, we argue that his 2S-GMM estimates is unreliable empirically in small sample case as will be shown in next section. In turn, Nakano and Saito

(1998) assert that their 2S-GMM estimates of β and γ by a single asset CCAPM with stock data lead to contradiction among the sample moments in three markets: stock, real estate, and call money, suggesting the existence of the puzzles. However, Nakano and Saito's analysis has two drawbacks. Their estimates are as unreliable as those of Hamori (1992), and estimates of a multiple assets CCAPM cannot produce a contradiction among several financial markets to lead to the puzzles.

In the context of the previous empirical work on the puzzles, other earlier studies introduce additional assumptions or extremely modify the standard CCAPM to improve their ability to explain the puzzles in the Japanese financial markets: For example, Hamori (1994) employs the non-expected utility model proposed by Epstein and Zin (1989, 1991). Bakshi and Naka (1997) and Hamori and Tokunaga (1999) use an asset pricing model with habit formation. Iwaisako (2001) applies Jagannathan and Wang's (1996) cross-section approach. Maki and Sonoda (2002) consider trading costs, and Fuse (2004) uses Ogaki and Reinhart's (1998) canonical cointegrating approach. Basu and Wada (2006) estimate a CCAPM taking into account the international risk sharing between the U.S. and Japan, and Kubota et al. (2008) employ both a limited participation model suggested by Mankiw and Zeldes (1991) and a persistent income shock model proposed by Constantinides and Duffie (1996) to examine the puzzles using micro-data from the Japanese *Family Income and Expenditure Survey*.

All of these Japanese use the GMM estimator to estimate the parameters of the various CCAPMs. With the exception of Kubota et al. (2008), the sample sizes are all less than 300. However, it is obvious that the sample sizes used in these papers is "small" sample (see Hansen et al. (1996)). As described in next section, the GMM estimator has some serious small sample properties. Thus, we will discuss those properties of the GMM estimator in more detail in next section.

4. ALTERNATIVE GMM ESTIMATORS

In this section, we first introduce poor small sample properties of the GMM estimator, and then introduce some alternatives to the GMM estimator. In particular, we present the recently developed generalized empirical likelihood (GEL) estimator and show the difference of the higher order bias between the GMM estimator and the GEL ones.

4.1. Small Sample Properties of GMM Estimator

There are many studies which point out that the GMM estimator has some serious small sample properties (for example, the special issue of the *Journal of Business & Economic Statistics* Vol.14 No.3 1996 is devoted to the "Small-Sample Properties of Generalized Method of Moments"). We can classify these problems concerning the poor finite sample performance of the GMM estimator as follows.

1. Small sample biases of the GMM estimator.
2. Over-rejection of the null hypothesis when GMM-based tests are used.
3. Poor power properties of GMM-based tests.

The first problem, the "small sample biases of the GMM estimator," is due to: weight estimation errors, many (possibly weak or uninformed) instruments, and weak identification. Altonji and Segal (1996) indicate that the correlation between the weight estimation errors of the 2S-GMM estimation and the sample moments leads to a serious small sample bias. Han and Phillips (2006), Chao and Swanson (2007) and Newey and Windmeijer (2009) show that the asymptotic approximations for the GMM estimator and tests become imprecise when there are many moments are used to estimate parameters. Stock and Wright (2000) show that even if only one parameter is weakly identified, all the other parameters do not satisfy the consistency and asymptotic normality properties of the usual 2S-GMM estimator.[3] They point to the parameters of the standard CCAPM as an example of weak identification in non-linear GMM estimation.

The second problem is the "over-rejection of the null hypothesis when GMM-based tests are used." The tests refer to the GMM-based Wald test and the overidentifying restrictions test. Burnside and Eichenbaum (1996) conclude that there is some tendency for GMM-based Wald tests to overreject the null hypothesis. Hansen et al. (1996) also conclude that tests of the overidentifying restrictions lead to overrejections of the model in small samples when the 2S-GMM and iterative GMM (IT-GMM) estimators used.

The third issue concerns the "poor power properties of GMM-based tests." For example, Ahn and Gadarowski (2004) show that the Hansen and Jagannathan's (1997) specification error test is quite unreliable when the sample size is small. We can understand that the volatility bound test of Hansen and Jagannathan (1991) lacks power because their test statistic depend in a large part on GMM estimates.

Thus, we know that there are many problems with GMM-based estimation. Recently, many studies attempt to improve on poor small sample properties of the GMM estimator. Hall (2000) applies a centered covariance matrix to modify the estimation error of associated with estimating Ω. Stock and Wright (2000) and Kleibergen (2005) propose reliable methods of statistical inference when researchers are confronted by weak identification problems. Han and Phillips (2006), Chao and Swanson (2007), and Newey and Windmeijer (2009) propose some modified asymptotic distributions using usual Taylor expansion to avoid the many moment conditions problem. However, Donald et al. (2009) report that the performance of the bias-corrected GMM estimator is far from rosy. Therefore, recent several studies propose alternative estimators to overcome poor small sample properties of the GMM estimator.

4.2. *Generalized Empirical Likelihood*

In this subsection, we introduce one alternative to the GMM estimator, the generalized empirical likelihood (GEL) estimator. The GEL estimator has many more desirable properties than the GMM estimator. Let us define some notation. $\{x_i\}_{i=1}^{n}$ denotes i.i.d. observations of the data, and p denotes the number of parameters to be estimated. We

[3] Earlier studies of the problems associated with many weak instruments and weak identification include Nelson and Startz (1990), Bekker (1994), Bound et al. (1995), and Staiger and Stock (1997) which focus on the instrumental variable and two-stage least square estimators (see also Stock et al. (2002) for details).

sometimes write $g(\cdot)$ as a m-dimensional vector of functions of the data and the parameters. We assume that $m \geq p$, and that the model has a true parameter θ_0 satisfying the following moment condition:

$$E[g(x_i, \theta_0)] = \int g(x, \theta_0) dF(x) = 0\,, \tag{20}$$

where the moment function $g(\cdot)$ is known, but the distribution function F of x_i is not specified. We consider the multinominal probability function $p_i = P\{x = x_i\}$ of the random variable x which is based on the data $\{x_i\}_{i=1}^n$. Then, using p_i we can express the log-likelihood function as $\sum_{i=1}^n \log p_i$. If there is no information about $\{x_i\}_{i=1}^n$, the likelihood function maximization problem is given by

$$\max_{\{p_i\}_{i=1}^n} \sum_{i=1}^n \log p_i\,,$$

$$\text{s.t. } 0 \leq p_i \leq 1\,, \quad \sum_{i=1}^n p_i = 1\,.$$

To solve this restricted maximization problem, we define the Lagrangian function as

$$\mathcal{L} = \sum_{i=1}^n \log p_i - \lambda\left(\sum_{i=1}^n p_i - 1\right),$$

and obtain the following first-order conditions,

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial p_1} &= \frac{1}{p_1} - \lambda = 0\,,\\
&\vdots\\
\frac{\partial \mathcal{L}}{\partial p_n} &= \frac{1}{p_n} - \lambda = 0\,,\\
\frac{\partial \mathcal{L}}{\partial \lambda} &= 1 - \sum_{i=1}^n p_i = 0\,.
\end{aligned}$$

Then we also obtain

$$\lambda = \frac{1}{p_1} = \frac{1}{p_2} = \cdots = \frac{1}{p_n} \Leftrightarrow p_1 = p_2 = \cdots = p_n\,,$$

and derive the following maximum likelihood estimator (MLE) of $\tilde{p}_i$

$$p_1 + p_2 + \cdots + p_n = n\tilde{p}_i = 1 \Leftrightarrow \tilde{p}_i = \frac{1}{n} = n^{-1}\,. \tag{21}$$

Therefore, the MLE of the distribution function F is the empirical distribution function (EDF), and we can show that $\tilde{F}(x) := n^{-1} \sum_{i=1}^n 1\{x_i \leq x\}$ where $1\{\cdot\}$ is an indicator function. We incorporate the moment condition model (equation (20)) into the likelihood function. Then, we can write the moment condition as $\sum_{i=1}^n p_i g(x_i, \theta_0) = 0$ by using the multinominal probability p_i, and solve the likelihood function maximization problem associated with p_i given as follows:

$$\max_{\{p_i\}_{i=1}^n} \sum_{i=1}^n \log p_i \,,$$

$$\text{s.t. } 0 \le p_i \le 1, \quad \sum_{i=1}^n p_i = 1 \,, \quad \sum_{i=1}^n p_i g(x_i, \theta) = 0 \,,$$

$$\mathcal{L} = \sum_{i=1}^n \log p_i - n\lambda\Big(\sum_{i=1}^n p_i g(x_i, \theta)\Big) + \gamma\Big(\sum_{i=1}^n p_i - 1\Big), \tag{22}$$

where we multiply the second term of right-hand of equation (22) by n to simplify the calculations. Then, we also obtain the following first-order conditions:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial p_1} &= \frac{1}{p_1} - n\lambda' g(x_1, \theta) + \gamma = 0 \,, \\ &\vdots \\ \frac{\partial \mathcal{L}}{\partial p_n} &= \frac{1}{p_n} - n\lambda' g(x_n, \theta) + \gamma = 0 \,, \end{aligned} \tag{23}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \sum_{i=1}^n p_i g(x_i, \theta) = 0 \,. \tag{24}$$

Multiplying all the elements of equation (23) by p_i gives,

$$\begin{aligned} p_1 \frac{\partial \mathcal{L}}{\partial p_1} &= 1 - n\lambda' p_1 g(x_1, \theta) + \gamma p_1 = 0 \,, \\ &\vdots \\ p_n \frac{\partial \mathcal{L}}{\partial p_n} &= 1 - n\lambda' p_n g(x_n, \theta) + \gamma p_n = 0 \,. \end{aligned} \tag{25}$$

The equations in (25) can be rewritten concisely as:

$$p_i \frac{\partial \mathcal{L}}{\partial p_i} = 1 - n\lambda' p_i g(x_i, \theta) + \gamma p_i = 0 \,, \quad \forall i \,. \tag{26}$$

Summing over all elements of equation (25), we also derive that

$$\sum_{i=1}^n p_i \frac{\partial \mathcal{L}}{\partial p_i} = n - n\lambda' \sum_{i=1}^n p_i g(x_i, \theta) + \gamma \sum_{i=1}^n p_i = n + \gamma = 0 \,. \tag{27}$$

Substituting this expression into equation (26), we also obtain the MLE of $\hat{p}_i$

$$1 = np_i(1 + \lambda'(\theta) g(x_i, \theta)) \Leftrightarrow \hat{p}_i = \frac{1}{n(1 + \lambda'(\theta) g(x_i, \theta))} \,, \tag{28}$$

so that $\hat{p}_i$ is explicitly determined if $\lambda(\theta)$ is given. Multiplying both sides of equation (28) by $g(z_i, \theta)$ and summing over i gives

$$\sum_{i=1}^n \frac{g(x_i, \theta)}{1 + \lambda'(\theta) g(x_i, \theta)} = 0 \,. \tag{29}$$

Given $g(x_i, \theta)$, we can derive the λ which satisfies equation (29). We can define the empirical likelihood ratio (ELR) and the empirical likelihood (EL) estimator of θ, $\hat{\theta}_{EL}$, as follows:

$$\begin{aligned} ELR(\theta) &= \sum_{i=1}^{n} \log \hat{p}_i - \sum_{i=1}^{n} \log \tilde{p}_i \\ &= -\sum_{i=1}^{n} \log n(1 + \lambda'(\theta) g(x_i, \theta)) + \sum_{i=1}^{n} \log n \\ &= -\sum_{i=1}^{n} \log n - \sum_{i=1}^{n} \log (1 + \lambda'(\theta) g(x_i, \theta)) + \sum_{i=1}^{n} \log n \\ &= -\sum_{i=1}^{n} \log (1 + \lambda'(\theta) g(x_i, \theta)) , \end{aligned} \tag{30}$$

$$\hat{\theta}_{EL} := \arg\max_{\theta \in \Theta} ELR(\theta) , \tag{31}$$

$ELR(\theta)$ can also be obtained using the dual problem of equation (22)

$$ELR(\theta) = \min_{\lambda} - \sum_{i=1}^{n} \log (1 + \lambda' g(x_i, \theta)) . \tag{32}$$

Therefore, we can consider that the empirical likelihood is an expansion of likelihood function using a semi-parametric approach.[4] Newey and Smith (2004) generalize the duality between equation (22) and equation (32), and define the generalized empirical likelihood (GEL) estimator as

$$\hat{\theta}_{GEL} := \arg\min_{\theta \in \Theta} \ell(\theta) := \arg\min_{\theta \in \Theta} \sup_{\lambda} \sum_{i=1}^{n} \rho(\lambda' g(x_i, \theta)) . \tag{33}$$

This has a dual problem as follows

$$\hat{\theta}_{GEL} = \arg \min_{\theta \in \Theta, \{p_i\}_{i=1}^{n}} \sum_{i=1}^{n} \log h(p_i) , \tag{34}$$

$$\text{s.t. } 0 \le p_i \le 1 , \quad \sum_{i=1}^{n} p_i = 1 , \quad \sum_{i=1}^{n} p_i g(x_i, \theta) = 0 .$$

By specifying ρ and h in equations (33) and (34), we can derive following estimators.

- Empirical likelihood (EL):
  $\rho(v) = \log(1 - v)$ or $h(p) = -\log(p)$.
- Exponential tilting (ET):
  $\rho(v) = -\exp(v)$ or $h(p) = p \log(p)$, proposed by Imbens et al. (1998).
- Continuous updating GMM (CU-GMM):
  $\rho(v) = -(1 + v)^2/2$ or $h(p) = [(np)^2 - 1]/2n$, proposed by Hansen et al. (1996).

[4] The empirical likelihood incorporates the moment condition using the multinominal probability p_i.

We can thus confirm that the GEL estimator is more a general formulation because it contains the GMM estimator as a special case. The asymptotic distribution of the GEL estimator is given as

$$\sqrt{n}\big(\hat{\theta}_{GEL} - \theta_0\big) \xrightarrow{d} \mathcal{N}(0, (G'\Omega^{-1}G)^{-1}) . \tag{35}$$

That is, the GEL estimator is asymptotically equivalent to the optimal GMM estimator, and is thus asymptotically efficient. Similar to the J-statistic used when models are estimated using the GMM estimator, we can use a J-statistic to test the overidentifying restrictions when the parameters are estimated using the GEL estimator. The J-statistic for the GEL estimator is computed using the kernel-smoothed moment indicator (see Section 4 in Newey and Smith (2004) for details).

$$-2n \min_{\theta\in\Theta} \ell(\theta) = -2n\ \ell(\hat{\theta}_{GEL}) . \tag{36}$$

Under the null hypothesis that equation (20) is true, this statistic is asymptotically distributed as χ^2_{m-p}.

4.3. *Higher Order Biases of GMM and GEL*

In section 4.2, we confirmed that the GEL estimator is asymptotically equivalent to the GMM estimator. However, this theoretical result is confined to the first-order term in the asymptotic expansion for both estimators. In the case of small samples, we must analyze the properties of the estimators in more detail using higher order expansions. Newey and Smith (2004) apply the stochastic expansion of Nagar (1959) to the GMM and GEL estimators.

They decomposite the higher order bias of the 2S-GMM estimator, $\hat{\theta}_{2S-GMM}$, as follows:

$$\begin{aligned} Bias(\hat{\theta}_{GMM}) &= B_I + B_G + B_\Omega + B_W , \\ B_I &= H(-a + E[GHg(x_i,\theta)])/n , \\ B_G &= -\Sigma E[G'Pg(x_i,\theta)]/n , \\ B_\Omega &= HE[g(x_i,\theta)g(x_i,\theta)'Pg(x_i,\theta)]/n , \\ B_W &= -H\sum_{j=1}^{p} \bar{\Omega}_{\theta_j}(H_W - H)'e_j/n , \end{aligned} \tag{37}$$

where $G(\theta) = \partial g(x_i,\theta)/\partial\theta$, and $G = E[G(\theta_0)]$ are both $p \times m$ matrices. And $\Omega = E[g(x_i,\theta)g(x_i,\theta)']$ is a $m \times m$ positive definite matrix. Now we define

$$\Sigma = (G'\Omega^{-1}G)^{-1} , \quad H = \Sigma G'\Omega^{-1} , \quad P = \Omega^{-1} - \Omega^{-1}G\Sigma G'\Omega^{-1} ,$$
$$H_W = (G'W^{-1}G)^{-1}G'W^{-1} .$$

The matrix W appearing in the definition of H_W depends on the initial weighting matrix $\hat{W}$ used for the GMM estimator. Assumption 4 in Newey and Smith (2004) states that there exists a W and $\xi(x)$ such that $\hat{W} = W + \sum_{i=1}^{n}\xi(x_i)/n + O_p(n^{-1})$, W is positive definite, $E[\xi(x_i)] = 0$, and $E[||\xi(x_i)||^6] < \infty$. We should note that the higher order bias depends on the preliminary estimator $\tilde{\theta}$ only through the limit of W and the functions $\xi(x_i)$. Let

$$\bar{\Omega}_{\theta_j} = E[\partial(g(x_i,\theta_0)g(x_i,\theta_0)')/\partial\theta_j]\,, \quad i=1,2,\ldots,n\,, \quad j=1,2,\ldots,p\,,$$

and let a be a m vector such that

$$a_j = \mathrm{tr}(\Sigma E[\partial^2 g_j(x_i,\theta_0)/\partial\theta\partial\theta'])/2\,, \quad i=1,2,\ldots,n\,, \quad j=1,2,\ldots,m\,,$$

where $g_j(x_i,\theta)$ is the jth element of $g(x_i,\theta)$, and e_j is a unit vector.

As Newey and Smith (2004) explain, the terms of $Bias(\hat{\theta}_{GMM})$ have the following interpretation. B_I denotes the asymptotic biases for an optimal GMM estimator with known Ω and G, B_G denotes the estimation biases associated with estimating G, B_Ω denotes the estimation biases associated with estimating the second moment matrix, and B_W denotes the estimation biases associated with estimating the preliminary weighting matrix.

In contrast, Newey and Smith (2004) show the higher order bias of the GEL estimator, $\hat{\theta}_{GEL}$, is given by

$$Bias(\hat{\theta}_{GEL}) = B_I + \left(1+\frac{\rho_3}{2}\right)B_\Omega\,, \tag{38}$$

where $\rho_j(v) = \partial^j\rho(v)/\partial v^j$ and $\rho_j = \rho_j(0)$ for each j. In the special case where $\rho_3 = -2$ (in the EL case), the bias expression simplifies to:

$$Bias(\hat{\theta}_{GEL}) = Bias(\hat{\theta}_{EL}) = B_I\,. \tag{39}$$

That is, we find that the higher order bias of the GEL estimator is less than the higher order bias of the 2S-GMM estimator because the GEL estimator is free from the estimation biases associated with estimating G (B_G) and the preliminary weighting matrix (B_W). In particular, the small sample biases of the 2S-GMM estimator is serious when there are many moment conditions because B_G and B_Ω depend heavily on the number of moment conditions. Newey and Smith (2004) also show that the higher order mean squared error (MSE, Donald and Newey (2001)) of the GEL estimator is smaller than that of the GMM estimator. As a result, we can consider that the GEL estimator is preferable compared with the 2S-GMM estimator in terms of its higher order properties (see Newey and Smith (2004) and Anatolyev (2005) for details).

5. AN APPLICATION TO THE STANDARD CCAPM[5]

In this section, we apply the GEL estimators to estimate the parameters of the standard CCAPM and calculate the higher order bias on the estimations . In particular, we first show how to compute the higher order bias of the GMM and GEL estimates for the standard CCAPM. Second, we estimate the parameters of the standard CCAPM using the GMM and GEL estimators and calculate the higher order bias of our estimates. Last, we show the nonexistence of the well-known risk-free rate puzzle in the Japanese financial markets by using the GEL estimates.

[5] This section relies on Ito and Noda (2012a).

5.1. Higher Order Bias on the Standard CCAPM

Let us apply the higher order bias formula of the GMM estimator in equation (37) and the GEL estimator in equation (38) to the standard CCAPM. The moment restrictions can be written as

$$g(x_t,\theta) = \left[\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}\cdot(\mathbf{1}+\mathbf{R}_{t+1})-\mathbf{1}\right]\otimes\mathbf{z}_t\,.$$

Then, we derive the following gradient of $g(x_t,\theta)$ as

$$G(\theta) = \frac{\partial g(x_t,\theta)}{\partial\theta} = \begin{bmatrix} \left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}\cdot(\mathbf{1}+\mathbf{R}_{t+1}) \\ -\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}\cdot(\mathbf{1}+\mathbf{R}_{t+1})\cdot\log\left(\frac{C_{t+1}}{C_t}\right)\end{bmatrix}\otimes\mathbf{z}_t\,.$$

We define

$$G = E[G(\theta_0)] = E\begin{bmatrix} \left(\frac{C_{t+1}}{C_t}\right)^{-\gamma_0}\cdot(\mathbf{1}+\mathbf{R}_{t+1}) \\ -\beta_0\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma_0}\cdot(\mathbf{1}+\mathbf{R}_{t+1})\cdot\log\left(\frac{C_{t+1}}{C_t}\right)\end{bmatrix}\otimes\mathbf{z}_t\,,$$

where θ_0 is the true value of θ.

$$\begin{aligned}\Omega &= E\left[g(x_t,\theta_0)g(x_t,\theta_0)'\right]\\ &= E\left[\left[\left(\beta_0\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma_0}\cdot(\mathbf{1}+\mathbf{R}_{t+1})-\mathbf{1}\right)\otimes\mathbf{z}_t\right]\right.\\ &\quad\left.\times\left[\left(\beta_0\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma_0}\cdot(\mathbf{1}+\mathbf{R}_{t+1})-\mathbf{1}\right)\otimes\mathbf{z}_t\right]'\right],\end{aligned}$$

Ω denotes the covariance matrix of "the residuals" of the model. We substitute Ω and G into the right-hand sides of definitions of Σ, H, P, and H_W. In this context, Σ denotes the covariance matrix of the parameters, H reflects the marginal changes of the parameters preliminary estimated affecting the covariance matrix Ω, P also reflects the marginal changes of the parameters of the model, and H_W corresponds H to the secondary estimates.

$$\begin{aligned}\bar{\Omega}_{\theta_j} = E\left[\frac{\partial}{\partial\theta_j}\left(\left(\beta_0\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma_0}\cdot(\mathbf{1}+\mathbf{R}_{t+1})-\mathbf{1}\right)\otimes\mathbf{z}_t\right)\right.\\ \left.\left(\left(\beta_0\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma_0}\cdot(\mathbf{1}+\mathbf{R}_{t+1})-\mathbf{1}\right)\otimes\mathbf{z}_t\right)'\right],\end{aligned}$$

$\bar{\Omega}_{\theta_j}$ denotes the marginal change of the jth parameter θ_j independently affecting the covariance matrix Ω.

$$a_k = \mathrm{tr}\left[\Sigma E \left[\begin{array}{cc} 0 & -\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma_0} \cdot (\mathbf{1}+\mathbf{R}_{t+1}) \cdot \log\left(\frac{C_{t+1}}{C_t}\right) \\ -\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma_0} \cdot (\mathbf{1}+\mathbf{R}_{t+1}) \cdot \log\left(\frac{C_{t+1}}{C_t}\right) & \beta_0 \left(\frac{C_{t+1}}{C_t}\right)^{-\gamma_0} \cdot (\mathbf{1}+\mathbf{R}_{t+1}) \cdot \left\{\log\left(\frac{C_{t+1}}{C_t}\right)\right\}^2 \end{array} \right] \right.$$

$$\left. \times \otimes \mathbf{z}_{k,t} \right] / 2,$$

a_k denotes that the second order effect of the marginal change of the parameter on the kth moment restriction g_k. k runs 1 through m, where m is the number of asset returns. We can compute the higher order biases for specifications using Newey and Smith's (2004) theoretical framework. Following the section 5 in Newey and Smith (2004), we use the empirical distribution to estimate the above bias formula. Specifically, we replace the expectation operators by the averages.

5.2. *Data*

In this paper, quarterly data from 1980Q3 to 2009Q4 are used. The per capita consumption is computed as "Nondurable goods plus service consumption (benchmark year 2000)" divided by the estimates of the total population reported in the *Annual Report on National Accounts* in Japan. The per capita consumption data are seasonally adjusted using the X-12 ARIMA procedure. The returns on short-term instruments are employed as the return on the risk-free asset and these are obtained from Nikko Financial Intelligence. The Fama-French's market portfolio returns are treated as the returns on the risky asset and these are obtained from Nikkei Portfolio Master.[6]

To deflate all series, the "Nondurable plus service consumption" deflator published in the *Annual Report on National Accounts* is used.[7] Lagged values of the real consumption growth rate, the real return on the risk-free asset, and the real return on the market portfolio are used as instruments. For the GMM and GEL estimator, all variables that appear in the moment conditions should be stationary. To check whether the variables satisfy the stationarity condition, we use the ADF test of Dickey and Fuller (1981). Table 2 provides some descriptive statistics and the results of the ADF tests. For all the variables, the ADF test rejects the null hypothesis that the variable contains a unit root at conventional significance levels.

5.3. *GMM and GEL Estimates*

We estimate the two basic parameters in the standard CCAPM. To confirm the accuracy of our estimates, we compare GMM and GEL estimates for the standard CCAPM. Table 3 shows the empirical results with GMM estimators (2S-GMM and CU-GMM). In GMM estimations, we employ an appropriate HAC covariance matrix of Andrews (1991) to reduce estimation biases, which is the asymptotically optimal lag truncation/bandwidth for the quadratic spectral kernel estimator we used.

[6] Fama-French's market factors in Japan are calculated by following Kubota and Takehara (2007).

[7] The "Nondurable plus service consumption" deflator is a weighted inflation rate using "Nondurable goods" and "Service" deflators that are also published in the *Annual Report on National Accounts*.

Table 2. Descriptive statistics and unit root tests

Variables	Mean	SD	Min	Max	ADF	Lag	$\mathcal{N}$
CG_t	1.0037	0.0091	0.9770	1.0312	-10.5587	1	
R_t^f	0.0050	0.0063	-0.0143	0.0207	-5.7382	1	118
R_t^m	0.0121	0.1041	-0.3335	0.2331	-7.2208	1	

Notes:

(1) "CG_t" denotes the gross real per capita consumption growth, "R_t^f" denotes the real return on risk-free asset, and "R_t^m" denotes the real return on market portfolio.
(2) "ADF" denotes the Augmented Dickey-Fuller (ADF) test statistics, "Lag" denotes the lag order selected by the Bayesian information criterion, and "$\mathcal{N}$" denotes the number of observations.
(3) In computing the ADF test, a model with a time trend and a constant is assumed. The critical values at the 1% significance level for the ADF test is "-3.99".

Table 3. Empirical results

	GMM		GEL		
	2S-GMM	CU-GMM	CUE	EL	ET
$\hat{\beta}$	0.9972	0.9984	0.9985	0.9978	0.9981
	[0.0007]	[0.0008]	[0.0008]	[0.0008]	[0.0008]
$\hat{\gamma}$	0.5601	0.8685	0.8502	0.7526	0.8026
	[0.1836]	[0.2343]	[0.2084]	[0.2067]	[0.2087]
p_J	0.9179	0.3947	0.3352	0.5890	0.4845
DF	6	6	6	6	6

Notes:

(1) "$\hat{\beta}$" denotes the estimate of the subjective discount rate, and "$\hat{\gamma}$" denotes the estimate of degree of the relative risk aversion.
(2) "p_J" denotes the p-value for Hansen's J, and "DF" denotes the degrees of freedom for the Hansen's J test.
(3) The Andrews (1991) adjusted standard errors for each of the estimates are reported in brackets.
(4) R version 2.15.1 was used to compute the estimates, the starting values of the parameters are set equal to $\beta = 1$ and $\gamma = 1$.

All estimates of β and γ are statistically significant at conventional levels. The estimates of β range from 0.9972 to 0.9984, which is plausible, but the estimates of γ range from 0.5601 to 0.8685, which implies lack of robustness. The p-values for Hansen's J test are large enough that we cannot reject the null that the moment conditions hold.

Table 3 also shows the empirical results with GEL estimators (CUE, EL, and ET). In GEL estimations, we choose the truncated kernel proposed by Kitamura and Stutzer (1997) and Smith (1997) to smooth the moment function (that is, equation (33) in our case) because Anatolyev (2005) demonstrates that, in the presence of correlation in the moment function, the smoothed GEL estimator is efficient.[8] In addition, we employ an

[8] We employ the smoothed GEL estimator, but the optimal kernel weights do not exceed one. This suggests that the kernel smoothing has no effect.

appropriate HAC covariance matrix of Andrews (1991) to reduce estimation biases. The estimates of β and γ are statistically significant at conventional levels. The estimates of β range from 0.9978 to 0.9985; the estimates of γ range from 0.7526 to 0.8502. In contrast to the GMM estimators, the GEL estimates are very stable. The p-values for the Hansen's J test are large enough that we cannot reject the null that the moment conditions hold.

In addition, we calculate the higher order biases of Newey and Smith (2004) to investigate the asymptotic higher order properties of our estimates.[9] When the sample size is not so large that we cannot rely on the GMM estimates, we should suspect the reliability of the estimates. Table 4 shows higher order biases and MSEs for each estimate.

We find that higher order biases of the 2S-GMM estimates is are more than 20 times larger than those of the CU-GMM and the GEL estimates (CUE, EL, and ET). In particular, biases of the preliminary weighting matrix estimator are huge in the 2S-GMM estimates. This suggests that both the GMM estimates are unreliable because of potential biases. We also find that the higher order MSEs of these estimates are also more than 20 times larger than those of the CU-GMM and the GEL estimates. We should note that the value of B_W in case of CU-GMM is zero, verifying the independence of the preliminary estimate of the covariance matrix W. For the same reason as in case of higher order biases, the GMM estimates other than CU-GMM are unreliable. This result corresponds to that of Noda and Sugiyama (2010), who compare the shapes of objective functions to be minimized for 2S-GMM and CU-GMM (see Figures 3 and 4 in Noda and Sugiyama (2010) for details).

Table 4. Higher order biases for each estimates

Estimates	B_I	B_G	B_Ω	B_W	B_T	MSE
$\hat{\beta}_{2S-GMM}$	0.0001	−0.0002	0.0001	0.0044	0.0044	1.3124
$\hat{\gamma}_{2S-GMM}$	0.0258	−0.0501	0.0075	1.1476	1.1307	
$\hat{\beta}_{CU-GMM}$	0.0002	−0.0003	0.0003	–	0.0001	0.0549
$\hat{\gamma}_{CU-GMM}$	0.0463	−0.0839	0.0406	–	0.0029	
$\hat{\beta}_{CUE}$	0.0002	–	0.0003	–	0.0005	0.0506
$\hat{\gamma}_{CUE}$	0.0418	–	0.0427	–	0.0845	
$\hat{\beta}_{EL}$	0.0001	–	–	–	0.0001	0.0439
$\hat{\gamma}_{EL}$	0.0341	–	–	–	0.0341	
$\hat{\beta}_{ET}$	0.0001	–	0.0003	–	0.0003	0.0467
$\hat{\gamma}_{ET}$	0.0381	–	0.0368	–	0.0564	

Notes:

(1) B_T denotes the total higher order biases, and MSE denotes the higher order mean squared error for each estimation.

(2) To compute the estimates, R version 2.15.1 was used.

[9] In the case of stationary time series, we should employ the formula in Anatolyev (2005). However, because there are no kernel-smoothing effects in our estimations, we employ the formula in Newey and Smith (2004).

Therefore, we conclude that the CU-GMM and the GEL estimates are indisputably better than the 2S-GMM estimates in asymptotic higher order properties when the sample size is small. We obtain the economically realistic parameters of the standard CCAPM when we employ the CU-GMM estimator and GEL estimators.

5.4. *A Solution to the Risk-Free Rate Puzzle in Japan*

Although several earlier studies attempted to resolve the risk-free rate puzzle in the Japanese financial markets, there is not yet a general consensus. For example, the estimates by Hamori (1992) lead to the conclusion that the risk-free rate puzzle does not exist, while Nakano and Saito (1998) report quite opposite results: the puzzle exists, as does the equity premium puzzle. We confirm that the Hamori's (1992) 2S-GMM estimates of β and γ lead to the conclusion when we substitute them into the formula of Kandel and Stambaugh (1991). However, we argue that his estimates by the 2S-GMM estimator is unreliable as we show that the higher order biases of the 2S-GMM estimates is quite large for samples with the size of around 100. Furthermore, he fails to avoid the problem of weak identification as Stock and Wright (2000) point out; Noda and Sugiyama (2010) show that the CU-GMM estimate of the standard CCAPM applied to the Japanese financial data successfully identifies while that of 2S-GMM does not (see Figures 3 and 4 in Noda and Sugiyama (2010) for details). In turn, Nakano and Saito (1998) assert that their 2S-GMM estimates of β and γ by a single asset CCAPM with stock data lead to contradiction among the sample moments in three markets: stock, real estate, and call money, suggesting the existence of the risk-free rate puzzle. However, their analysis has two drawbacks. Their estimates are as unreliable as those of Hamori (1992), and estimates of a single asset CCAPM cannot produce a contradiction among several financial markets to lead to the puzzle.

Therefore, we investigate whether there is the risk-free rate puzzle in the Japanese financial markets when we use the CU-GMM and GELs estimates. Under the assumption of joint conditional lognormality and homoskedasticity of asset returns, Hansen and Singleton (1983) deliver a convenient equation:

$$0 = E_t[R_{i,t+1}] + \log\beta - \gamma E_t[\Delta C_{t+1}] + \frac{1}{2}\left(\sigma_i^2 + \gamma^2\sigma_C^2 - 2\gamma\sigma_{ic}\right), \tag{40}$$

where σ_i and σ_C are the standard deviations of the ith asset and consumption, respectively, and σ_{ic} is the covariance between them. This equation implies the following equation shown by Kandel and Stambaugh (1991):

$$E[R_t^f] = -\log\beta + \gamma g - \frac{\gamma^2\sigma_c^2}{2}, \tag{41}$$

where $E[R_t^f]$ is the unconditional expectation of the risk-free interest rate, g is the mean growth rate of real consumption, and σ_c^2 is the variance of g.[10] When we substitute the estimates of β and γ on the standard CCAPM into this equation and employ the CU-GMM and the GEL estimates (CUE, EL, and ET), we derive $E[R_t^f]$ in the range 0.0047

[10] The equation of Kandel and Stambaugh (1991) is a special case of the "mean-variance" representation of interest rates derived by Breeden (1986).

to 0.0050, which is close to 0.0050, the sample mean of the returns on the risk-free asset (see Table 2 for details). Thus, we conclude that the risk-free rate puzzle does not exist in Japan when if one adopts the appropriate empirical method.

6. CONCLUDING REMARKS

In this paper, we provide a brief review of estimation methods for the standard CCAPM and some of its serious empirical problems, namely, the two puzzles in the financial market. First, we introduce the standard CCAPM via moment restriction models. Second, we present a brief review of the two puzzles, the equity premium puzzle and the risk-free rate puzzle, and Japanese earlier empirical studies. In particular, although there is a consensus that the estimates of the standard CCAPM adequately explain movements in the U.S. financial market, the same statement cannot be made for Japanese financial data. Third, we expound poor small sample properties of GMM estimators and the alternative GMM estimators, the GEL estimator, to overcome those problems. Last, we measure the difference of the higher order bias on the standard CCAPM between the GMM estimator and the GEL ones and suggest a possibility for solving the puzzles. And we show the nonexistence of the risk-free rate puzzle in the Japanese financial markets using the GEL estimator (see also Ito and Noda (2012a)).

Therefore, we suggest two alternative methods to investigate whether the standard CCAPM performs well in the Japanese financial markets. First, we should use the GEL estimator to improve on poor small sample properties of GMM estimator in the standard CCAPM. We can then measure the asymptotic higher order biases in the GMM and GEL estimators to empirically reveal the differences between the two estimators as shown in section 5.3. Second, we should verify the parameter instability of the standard CCAPM because the equity premium puzzle is still found to exist even when we use the GEL estimator. We can consider that the parameter instability of the standard CCAPM may arise from the failure of the standard CCAPM. Therefore, we assume time-varying parameters in the standard CCAPM framework by applying the GEL estimator to the random parameter regression model of Ito (2007). This is a very recent issue and has been subject to very little research. For example, using a rolling method for the 2S-GMM estimator, Kim (2009) indicates that the degree of risk aversion has been changing during the post-war period. However, given poor small sample properties of the 2S-GMM estimator, it is likely that a rolling 2S-GMM method will also have poor small sample properties. We address these two issues is relation to the estimates of the standard CCAPM in future research (see Ito and Noda (2012b) for details).

REFERENCES

AHN, S. C. AND C. GADAROWSKI (2004): "Small Sample Properties of the GMM Specification Test Based on the Hansen-Jagannathan Distance," *Journal of Empirical Finance*, 11, 109–132.

ALTONJI, J. G. AND L. M. SEGAL (1996): "Small-Sample Bias in GMM Estimation of Covariance Structures," *Journal of Business & Economic Statistics*, 14, 353–366.

ANATOLYEV, S. (2005): "GMM, GEL, Serial Correlation, and Asymptotic Bias," *Econometrica*, 73, 983–1002.

ANDREWS, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

BABA, N. (2000): "Exploring the Role of Money in Asset Pricing in Japan: Monetary Considerations and Stochastic Discount Factors," *Monetary and Economic Studies*, 18, 159–198.

BAKSHI, G. S. AND A. NAKA (1997): "An Empirical Investigation of Asset Pricing Models using Japanese Stock Market Data," *Journal of International Money and Finance*, 16, 81–112.

BASU, P. AND K. WADA (2006): "Is Low International Risk Sharing Consistent with a High Equity Premium? A Reconciliation of Two Puzzles," *Economics Letters*, 93, 436–442.

BEKKER, P. A. (1994): "Alternative Approximations to the Distributions of Instrumental Variable Estimators," *Econometrica*, 62, 657–681.

BOUND, J., D. JAEGER, AND R. BAKER (1995): "Problems with Instrumental Variables Estimation When the Correlation between the Instruments and the Endogenous Explanatory Variable Is Weak," *Journal of the American Statistical Association*, 90, 443–450.

BREEDEN, D. T. (1979): "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities," *Journal of Financial Economics*, 7, 265–296.

——— (1986): "Consumption, Production, Inflation and Interest Rates: A Synthesis," *Journal of Financial Economics*, 16, 3–39.

BURNSIDE, C. AND M. EICHENBAUM (1996): "Small-Sample Properties of GMM-Based Wald Tests," *Journal of Business & Economic Statistics*, 14, 294–308.

CHAO, J. C. AND N. R. SWANSON (2007): "Consistent Estimation with a Large Number of Weak Instruments," *Econometrica*, 73, 1673–1692.

CONSTANTINIDES, G. M. AND D. DUFFIE (1996): "Asset Pricing with Heterogeneous Consumers," *Journal of Political Economy*, 104, 219–240.

DICKEY, D. A. AND W. A. FULLER (1981): "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," *Econometrica*, 49, 1057–1072.

DONALD, S. G., G. W. IMBENS, AND W. K. NEWEY (2009): "Choosing Instrumental Variables in Conditional Moment Restriction Models," *Journal of Econometrics*, 152, 28–36.

DONALD, S. G. AND W. K. NEWEY (2001): "Choosing the Number of Instruments," *Econometrica*, 69, 1161–1191.

EPSTEIN, L. G. AND S. E. ZIN (1989): "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework," *Econometrica*, 57, 937–969.

——— (1991): "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: An Empirical Analysis," *Journal of Political Economy*, 99, 263–286.

FUSE, M. (2004): "Estimating Intertemporal Substitution in Japan," *Applied Economics Letters*, 11, 267–269.

HALL, A. R. (2000): "Covariance Matrix Estimation and the Power of the Overidentifying Restrictions Test," *Econometrica*, 68, 1517–1527.

HAMORI, S. (1992): "Test of C-CAPM for Japan: 1980-1988," *Economics Letters*, 38, 67–72.

——— (1994): "The Non-Expected Utility Model and Asset Returns: Some Evidence from Japan," *Japanese Journal of Financial Economics*, 1, 89–99.

HAMORI, S. AND T. TOKUNAGA (1999): "Habit Formation and Durability and Consumption: Some Evidence from Income Quintile Groups in Japan," *Applied Economics Letters*, 6, 397–402.

HAN, C. AND P. C. B. PHILLIPS (2006): "GMM with Many Moment Conditions," *Econometrica*, 74, 147–192.

HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

HANSEN, L. P., J. HEATON, AND A. YARON (1996): "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business & Economic Statistics*, 14, 262–280.

HANSEN, L. P. AND R. JAGANNATHAN (1991): "Implications of Security Market Data for Models of Dynamic Economies," *Journal of Political Economy*, 99, 225–262.

HANSEN, L. P. AND K. SINGLETON (1983): "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns," *Journal of Political Economy*, 91, 249–265.

HORI, K. (1996): "C-CAPM in Japanese Asset Market: Reexamination (in Japanese)," *Osaka Economic Papers*, 44, 76–90.

IMBENS, G., R. SPADY, AND P. JOHNSON (1998): "Information Theoretic Approaches to Inference in Moment Condition Models," *Econometrica*, 66, 333–357.

ITO, M. (2007): "A New Method for Estimating Economic Models with General Time-Varying Structures," Keio Economic Society Discussion Paper Series, KESDP 07-8.

ITO, M. AND A. NODA (2012a): "The GEL Estimates Resolve the Risk-free Rate Puzzle in Japan," *Applied Financial Economics*, 22, 365–374.

——— (2012b): "A Random Regression for Moment Restriction Models," Keio University.

IWAISAKO, T. (2001): "Consumption and Asset Pricing: A Survey (in Japanese)," *Gendai Finance*, 9, 3–39.

JAGANNATHAN, R. AND Z. WANG (1996): "The Conditional CAPM and the Cross-Section of Expected Returns," *Journal of Finance*, 51, 3–53.

KANDEL, S. AND R. F. STAMBAUGH (1991): "Asset Returns and Intertemporal Preferences," *Journal of Monetary Economics*, 27, 39–71.

KIM, K. (2009): "Time-Varying Risk Aversion," University of Chicago.

KITAMURA, Y. AND M. STUTZER (1997): "An Information-Theoretic Alternative to Generalized Method of Moments Estimation," *Econometrica*, 65, 861–874.

KLEIBERGEN, F. (2005): "Testing Parameters in GMM without Assuming That They Are Identified," *Econometrica*, 73, 1103–1123.

KOCHERLAKOTA, N. R. (1996): "The Equity Premium: It's Still a Puzzle," *Journal of Economic Literature*, 34, 42–71.

KUBOTA, K. AND H. TAKEHARA (2007): "Farther Validation of Effectiveness of Fama-French Factor Model (Fama-French Model no Yukousei no Saikensho, in Japanese)," *Gendai Finance*, 22, 3–23.

KUBOTA, K., T. TOKUNAGA, AND K. WADA (2008): "Consumption Behavior, Asset Returns, and Risk Aversion: Evidence from the Japanese Household Survey," *Japan & The World Economy*, 20, 1–18.

LUCAS, R. E. (1978): "Asset Prices in an Exchange Economy," *Econometrica*, 46, 1429–1445.

MAKI, A. AND T. SONODA (2002): "A Solution to the Equity Premium and Riskfree Rate Puzzles: An Empirical Investigation Using Japanese Data," *Applied Financial Economics*, 12, 601–612.

MANKIW, N. G. AND S. P. ZELDES (1991): "The Consumption of Stockholders and Non-Stockholders," *Journal of Financial Economics*, 40, 677–689.

MEHRA, R. AND E. C. PRESCOTT (1985): "The Equity Risk Premium: A Puzzle," *Journal of Monetary Economics*, 15, 145–161.

——— (2003): "The Equity Premium Puzzle in Retrospect," in *Handbook of the Economics of Finance*, ed. by G. M. Constantinides, M. Harris, and R. Stulz, North Holland, vol. 1B, chap. 14, 887–936.

NAGAR, A. L. (1959): "The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations," *Econometrica*, 27, 575–595.

NAKANO, K. AND M. SAITO (1998): "Asset Pricing in Japan," *Journal of the Japanese and International Economies*, 12, 151–166.

NELSON, C. R. AND R. STARTZ (1990): "Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator," *Econometrica*, 58, 967–976.

NEWEY, W. K. AND D. I. MCFADDEN (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, ed. by R. F. Engle and D. I. McFadden, Elsevier, vol. 4, chap. 36, 2111–2245.

NEWEY, W. K. AND R. J. SMITH (2004): "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, 72, 219–255.

NEWEY, W. K. AND K. D. WEST (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

NEWEY, W. K. AND F. WINDMEIJER (2009): "Generalized Method of Moments with Many Weak Moment Conditions," *Econometrica*, 77, 687–719.

NODA, A. AND S. SUGIYAMA (2010): "Measuring the Intertemporal Elasticity of Substitution for Consumption: Some Evidence from Japan," *Economics Bulletin*, 30, 524–533.

OGAKI, M. AND C. M. REINHART (1998): "Measuring Intertemporal Substitution: The Role of Durable Goods," *Journal of Political Economy*, 106, 1079–1098.

OTSU, T. (2007): "Generalized Method of Moments, Empirical Likelihood and Their Recent Developments," Lecture Note at Cowles Foundation and Department of Economics, Yale University.

RUBINSTEIN, M. (1976): "The Valuation of Uncertain Income Streams and the Pricing of Options," *Bell Journal of Economics*, 7, 407–425.

SMITH, R. J. (1997): "Alternative Semi-Parametric Likelihood Approaches to Generalised Method of Moments Estimation," *Economic Journal*, 107, 503–519.

——— (2011): "GEL Criteria for Moment Condition Models," *Econometric Theory*, 27, 1192–1235.

STAIGER, D. AND J. H. STOCK (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557–586.

STOCK, J. H. AND J. H. WRIGHT (2000): "GMM with Weak Identification," *Econometrica*, 68, 1055–1096.

STOCK, J. H., J. H. WRIGHT, AND M. YOGO (2002): "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business & Economic Statistics*, 20, 518–529.

TANIGAWA, Y. (1994): "Japanese Evidence on Consumption-Based Capital Asset Pricing Model (in Japanese)," *Okayama Economic Review*, 25, 315–332.

WEIL, P. (1989): "The Equity Premium Puzzle and the Risk-Free Rate Puzzle," *Journal of Monetary Economics*, 24, 401–421.