SAMPLE SENSITIVITY FOR TWO-STEP AND CONTINUOUS UPDATING GMM ESTIMATORS

RIKUTO ONISHI AND TAISUKE OTSU

Abstract. This paper follows up the sensitivity analysis by Andrews, Gentzkow and Shapiro (2017) for biases in GMM estimators due to local violations of identifying assumptions, and proposes complementary bias measures that are sensitive to different choices of GMM weight matrices by considering a specific form of the local perturbation. Our method accommodates the two-step and continuous updating GMM estimators with or without centering. The proposed bias measures are illustrated by a consumption based asset pricing model using Japanese data.

1. Introduction

In a recent paper, Andrews, Gentzkow and Shapiro (2017) (hereafter, AGS) have introduced a novel and practical measure of sensitivity of the GMM estimates for structural parameters against local perturbations of different moments. More precisely, consider the GMM estimator $\hat{\theta} = \arg\min_\theta \hat{g}(\theta)'\hat{W}\hat{g}(\theta)$ for the parameters θ_0, where $\hat{g}(\theta)$ is a vector of sample moments and $\hat{W}$ is a weight matrix. Then AGS's sensitivity of $\hat{\theta}$ is defined as $\Lambda = -(G'WG)^{-1}G'W$, where G and W are probability limits of $\partial\hat{g}(\theta_0)/\partial\theta'$ and $\hat{W}$, respectively.

AGS showed that under local violation of the identifying assumption in the sense of $\sqrt{n}\hat{g}(\theta_0) \xrightarrow{d} \tilde{g}$ for a random vector $\tilde{g}$ with non-zero mean, the asymptotic bias of the GMM estimator is expressed as $\Lambda\mathbb{E}[\tilde{g}]$. Based on this, they advocated to report the estimate of Λ to make the structural parameter estimate $\hat{\theta}$ 'transparent', i.e., researchers can easily assess the potential bias of the parameter estimators for various scenarios of misspecification represented by $\mathbb{E}[\tilde{g}]$. Furthermore, Section 5 of AGS introduced the notion of the sample sensitivity, which is associated with the derivative $\partial\hat{\theta}(0)/\partial\delta$ of the GMM estimates $\hat{\theta}(\delta)$ under the perturbed sample moment $\hat{g}(\theta,\delta) = \hat{g}(\theta) + \delta\eta$ with $\eta = \text{plim}_{n\to\infty}\hat{g}(\theta_0)$.

In this paper, we follow up the analysis in AGS by investigating the effect of different GMM weight matrices $\hat{W}$ on the sample sensitivity. There are several motivations for this study. First, the GMM estimation typically involves the first stage parameter estimates in the construction of the weight matrix. Thus, researchers may want to allow the perturbation parameter δ to affect on the weight matrix. Indeed such analysis is conducted in the online appendix of AGS (Section 2.1), where $\hat{W}$ is replaced with $\hat{W}(\delta)$ and the corresponding sample sensitivity is derived. However, a typical form of the GMM weight matrix is the inverse of $\hat{\Omega}(\hat{\theta}_1) = \frac{1}{n}\sum_{i=1}^n g(X_i,\hat{\theta}_1)g(X_i,\hat{\theta}_1)'$, where $g(X,\theta)$ is the moment functions for observables X and $\hat{\theta}_1$ is a first stage estimate, and thus it is not clear how the perturbation parameter δ in the sample moment $\hat{g}(\theta,\delta)$ will affect on $\hat{\Omega}(\hat{\theta}_1)^{-1}$ and $\hat{\theta}_1$. Second, for the GMM estimation, it is also common to employ (the inverse of) the centered variance $\tilde{\Omega}(\hat{\theta}_1) = \frac{1}{n}\sum_{i=1}^n\{g(X_i,\hat{\theta}_1) - \hat{g}(\hat{\theta}_1)\}\{g(X_i,\hat{\theta}_1) - \hat{g}(\hat{\theta}_1)\}'$ as the weight matrix. Indeed Hall (2000) argued that this centered weight matrix $\tilde{\Omega}(\hat{\theta}_1)^{-1}$ is more stable than $\hat{\Omega}(\hat{\theta}_1)^{-1}$

under global misspecification. Although AGS's sensitivity focuses on local misspecification, it would be useful to see the effect of centering in the weight matrix for the sample sensitivity. See also Hall and Inoue (2003) for fundamental roles of the GMM weight matrix under global misspecification. Third, another popular choice of the GMM weight matrix is the one for the continuous updating GMM estimator (Hansen, Heaton and Yaron, 1996). In this case, the weight matrix involves the unknown parameter θ, say $\hat{W}(\theta, \delta)$, and hence the framework of AGS needs to be adjusted. Finally, in the applied econometrics literature, it is well known that the above variants of the GMM estimators show rather different finite sample properties (see, e.g., the special issue of the *Journal of Business & Economic Statistics*, Vol. 14, No. 3). Therefore, it is of interest to provide bias measures which are also sensitive to different choices of the weight matrix and estimation method.

In this paper, we address the above issues by considering a more specific form of the local perturbation, that is $g(X_i, \theta, \delta) = g(X_i, \theta) + \delta\eta$ for each i. Although this perturbation is less general than the perturbed moment $\hat{g}(\theta, \delta)$ studied by AGS, it allows us to derive explicit bias formulae for the above variants of the GMM estimators. Also the derived bias estimators are all consistent for AGS's (population) bias $\Lambda\eta$. Therefore, our bias estimates may be considered as characterizations of the higher-order terms to estimate $\Lambda\eta$ under the particular form of local perturbation, and provide useful complements to AGS's sensitivity analysis for researchers who are also concerned with the choice of the GMM estimation method in practice.

This paper is organized as follows. Section 2 presents our main result on the bias measures of the different GMM estimates. Section 3 illustrates our methodology by a consumption based asset pricing model using Japanese data.

2. Main results

We first fix the notation. Let $g_i(\theta) = g(X_i, \theta)$ be a d-dimensional vector of moment functions for a k-dimensional vector of parameters θ with $d > k$, and data $\{X_i\}_{i=1}^n$. Define the sample moment function $\hat{g}(\theta) = n^{-1} \sum_{i=1}^n g_i(\theta)$, uncentered sample second moments matrix $\hat{\Omega}(\theta) = n^{-1} \sum_{i=1}^n g_i(\theta) g_i(\theta)'$, and centered sample second moments matrix $\tilde{\Omega}(\theta) = n^{-1} \sum_{i=1}^n \{g_i(\theta) - \hat{g}(\theta)\}\{g_i(\theta) - \hat{g}(\theta)\}'$. We consider the GMM estimators solving

$$\min_{\theta} \hat{g}(\theta)' W \hat{g}(\theta), \tag{1}$$

with five different choices for W:

(i): $\hat{W}$ which does not involve $g_i(\theta)$ (first stage GMM, denoted by $\hat{\theta}_1$),
(ii): $\hat{\Omega}(\hat{\theta}_1)$ (two-step GMM, denoted by $\hat{\theta}_2$),
(iii): $\tilde{\Omega}(\hat{\theta}_1)$ (two-step GMM with centering, denoted by $\tilde{\theta}_2$),
(iv): $\hat{\Omega}(\theta)$ (continuous updating GMM, denoted by $\hat{\theta}_c$),
(v): $\tilde{\Omega}(\theta)$ (continuous updating GMM with centering, denoted by $\tilde{\theta}_c$).

If the model is correctly specified (i.e., $\mathbb{E}[g_i(\theta_0)] = 0$ for a unique true value θ_0), all these estimators are consistent and asymptotically normal under certain regularity conditions, and the estimators (ii)-(v) are asymptotically efficient. This paper is concerned with bias measures of these estimators against local misspecification from the model assumption.

In particular, we consider the following perturbations for the moment functions

$$g_i(\theta, \delta) = g_i(\theta) + \delta\eta, \tag{2}$$

for each $i = 1, \dots, n$, where δ is a scalar perturbation tuning parameter and $\eta = \text{plim}_{n\to\infty}\hat{g}(\theta_0)$ is a d-dimensional constant vector. Researchers can choose η for each specific scenario of violation of the identifying assumptions.

To begin with, let us introduce the bias formula for the first stage GMM estimate $\hat{\theta}_1$ derived in Section 5 of AGS. Let $\hat{g}(\theta, \delta) = n^{-1}\sum_{i=1}^n g_i(\theta, \delta)$ and $\hat{\theta}_1(\delta) = \arg\min_\theta \hat{g}(\theta, \delta)'\hat{W}\hat{g}(\theta, \delta)$ be the first stage GMM estimator using the perturbed moment function $g_i(\theta, \delta)$. Then we obtain (see, Section 2.1 in the online appendix of AGS)

$$\frac{\partial \hat{\theta}_1(0)}{\partial \delta} = -\{\hat{G}(\hat{\theta}_1)'\hat{W}\hat{G}(\hat{\theta}_1) + \hat{A}_1\}^{-1}\hat{G}(\hat{\theta}_1)'\hat{W}\eta, \tag{3}$$

where $\hat{G}(\theta) = \frac{\partial \hat{g}(\theta)}{\partial \theta'}$, $\hat{A}_1 = \left[\left(\frac{\partial \hat{G}(\hat{\theta}_1)}{\partial \theta^{(1)}}\right)' \hat{W}\hat{g}(\hat{\theta}_1), \dots, \left(\frac{\partial \hat{G}(\hat{\theta}_1)}{\partial \theta^{(k)}}\right)' \hat{W}\hat{g}(\hat{\theta}_1)\right]$, and the matrix $\hat{G}(\hat{\theta}_1)'\hat{W}\hat{G}(\hat{\theta}_1) + \hat{A}_1$ is assumed to be non-singular.

As clarified in AGS, the component $-\{\hat{G}(\hat{\theta}_1)'\hat{W}\hat{G}(\hat{\theta}_1) + \hat{A}_1\}^{-1}\hat{G}(\hat{\theta}_1)'\hat{W}$ in (3) is a consistent estimator for the sensitivity Λ. However, as shown in AGS, if the weight also depends on δ, then the derivative $\frac{\partial \hat{\theta}_1(0)}{\partial \delta}$ may not take a linear form in η without an intercept. Thus, we hereafter present analogous derivatives for the GMM estimators (ii)-(v).

Define $\hat{\Omega}(\theta, \delta) = n^{-1}\sum_{i=1}^n g_i(\theta, \delta)g_i(\theta, \delta)'$, and

$$\begin{aligned}
\hat{\theta}_2(\delta) &= \arg\min_\theta \hat{g}(\theta, \delta)'\hat{\Omega}(\hat{\theta}_1(\delta), \delta)^{-1}\hat{g}(\theta, \delta), \qquad \tilde{\theta}_2(\delta) = \arg\min_\theta \hat{g}(\theta, \delta)'\tilde{\Omega}(\hat{\theta}_1(\delta))^{-1}\hat{g}(\theta, \delta), \\
\hat{\theta}_c(\delta) &= \arg\min_\theta \hat{g}(\theta, \delta)'\hat{\Omega}(\theta, \delta)^{-1}\hat{g}(\theta, \delta), \qquad \tilde{\theta}_c(\delta) = \arg\min_\theta \hat{g}(\theta, \delta)'\tilde{\Omega}(\theta)^{-1}\hat{g}(\theta, \delta).
\end{aligned}$$

Note that the weight matrices of $\tilde{\theta}_2(\delta)$ and $\tilde{\theta}_c(\delta)$ are written by using $\tilde{\Omega}(\theta)$ (because $g_i(\theta, \delta) - \hat{g}(\theta, \delta) = g_i(\theta) - \hat{g}(\theta)$). The derivatives of these estimators with respect to δ are summarized in the following proposition.

Proposition. *Under the setup of this section, it holds*

(i): *Bias of the first stage GMM:*

$$\frac{\partial \hat{\theta}_1(0)}{\partial \delta} = -\{\hat{G}(\hat{\theta}_1)'\hat{W}\hat{G}(\hat{\theta}_1) + \hat{A}_1\}^{-1}\hat{G}(\hat{\theta}_1)'\hat{W}\eta.$$

(ii): *Bias of the two step GMM:*

$$\frac{\partial \hat{\theta}_2(0)}{\partial \delta} = -\{\hat{G}(\hat{\theta}_2)'\hat{\Omega}(\hat{\theta}_1)^{-1}\hat{G}(\hat{\theta}_2) + \hat{A}_2\}^{-1}\{\hat{G}(\hat{\theta}_2)'\hat{\Omega}(\hat{\theta}_1)^{-1}\eta + \hat{B}_2\},$$

where

$$\begin{aligned}
\hat{A}_2 &= \left[\left(\frac{\partial \hat{G}(\hat{\theta}_2)}{\partial \theta^{(1)}}\right)' \hat{\Omega}(\hat{\theta}_1)^{-1}\hat{g}(\hat{\theta}_2), \dots, \left(\frac{\partial \hat{G}(\hat{\theta}_2)}{\partial \theta^{(k)}}\right)' \hat{\Omega}(\hat{\theta}_1)^{-1}\hat{g}(\hat{\theta}_2)\right], \\
\hat{B}_2 &= -\hat{G}(\hat{\theta}_2)'\hat{\Omega}(\hat{\theta}_1)^{-1}\hat{C}_2\hat{\Omega}(\hat{\theta}_1)^{-1}\hat{g}(\hat{\theta}_2), \\
\hat{C}_2 &= \frac{1}{n}\sum_{i=1}^{n}\left[\left(G_i(\hat{\theta}_1)\frac{\partial \hat{\theta}_1(0)}{\partial \delta}\right) g_i(\hat{\theta}_1)' + g_i(\hat{\theta}_1)\left(G_i(\hat{\theta}_1)\frac{\partial \hat{\theta}_1(0)}{\partial \delta}\right)'\right] \\
&\quad +\hat{g}(\hat{\theta}_1)\eta' + \eta\hat{g}(\hat{\theta}_1)'.
\end{aligned}$$

(iii): *Bias of the two-step GMM with centering for the weight*

$$\frac{\partial \tilde{\theta}_2(0)}{\partial \delta} = -\{\hat{G}(\tilde{\theta}_2)'\tilde{\Omega}(\hat{\theta}_1)^{-1}\hat{G}(\tilde{\theta}_2) + \tilde{A}_2\}^{-1}\{\hat{G}(\tilde{\theta}_2)'\tilde{\Omega}(\hat{\theta}_1)^{-1}\eta + \tilde{B}_2\},$$

where

$$\begin{aligned}
\tilde{A}_2 &= \left[\left(\frac{\partial \hat{G}(\tilde{\theta}_2)}{\partial \theta^{(1)}}\right)' \tilde{\Omega}(\hat{\theta}_1)^{-1}\hat{g}(\tilde{\theta}_2), \dots, \left(\frac{\partial \hat{G}(\tilde{\theta}_2)}{\partial \theta^{(k)}}\right)' \tilde{\Omega}(\hat{\theta}_1)^{-1}\hat{g}(\tilde{\theta}_2)\right], \\
\tilde{B}_2 &= -\hat{G}(\tilde{\theta}_2)'\tilde{\Omega}(\hat{\theta}_1)^{-1}\tilde{C}_2\tilde{\Omega}(\hat{\theta}_1)^{-1}\hat{g}(\tilde{\theta}_2), \\
\tilde{C}_2 &= \frac{1}{n}\sum_{i=1}^{n}\left[\left(G_i(\hat{\theta}_1)\frac{\partial \hat{\theta}_1(0)}{\partial \delta}\right) g_i(\hat{\theta}_1)' + g_i(\hat{\theta}_1)\left(G_i(\hat{\theta}_1)\frac{\partial \hat{\theta}_1(0)}{\partial \delta}\right)'\right] \\
&\quad -\left(\hat{G}(\hat{\theta}_1)\frac{\partial \hat{\theta}_1(0)}{\partial \delta}\right)\hat{g}(\hat{\theta}_1)' + \hat{g}(\hat{\theta}_1)\left(\hat{G}(\hat{\theta}_1)\frac{\partial \hat{\theta}_1(0)}{\partial \delta}\right)'.
\end{aligned}$$

(iv): *Bias for the continuous updating GMM:*

$$\frac{\partial \hat{\theta}_c(0)}{\partial \delta} = -\left[\{\hat{G}(\hat{\theta}_c)' - \hat{\xi}(\hat{\theta}_c)'\}\hat{\Omega}(\hat{\theta}_c)^{-1}\hat{G}(\hat{\theta}_c) + \hat{A}_c\right]^{-1}[\{\hat{G}(\hat{\theta}_c)' - \hat{\xi}(\hat{\theta}_c)'\}\hat{\Omega}(\hat{\theta}_c)^{-1}(\eta - \hat{D}_c) + \hat{B}_c],$$

where

$$\begin{aligned}
\hat{A}_c &= \left[\frac{\partial G(\hat{\theta}_c)'}{\partial \theta^{(1)}} - \frac{\partial \hat{\xi}(\hat{\theta}_c)'}{\partial \theta^{(1)}} - \{\hat{G}(\hat{\theta}_c)' - \hat{\xi}(\hat{\theta}_c)'\}\hat{\Omega}(\hat{\theta}_c)^{-1}\frac{\partial \hat{\Omega}(\hat{\theta}_c)}{\partial \theta^{(1)}}\hat{\Omega}(\hat{\theta}_c)^{-1}\hat{g}(\hat{\theta}_c), \dots\right. \\
&\quad \left.\left(\frac{\partial G(\hat{\theta}_c)'}{\partial \theta^{(k)}} - \frac{\partial \hat{\xi}(\hat{\theta}_c)'}{\partial \theta^{(k)}} - \{\hat{G}(\hat{\theta}_c)' - \hat{\xi}(\hat{\theta}_c)'\}\hat{\Omega}(\hat{\theta}_c)^{-1}\frac{\partial \hat{\Omega}(\hat{\theta}_c)}{\partial \theta^{(1)}}\right)\hat{\Omega}(\hat{\theta}_c)^{-1}\hat{g}(\hat{\theta}_c)\right] \\
\hat{B}_c &= \frac{\partial \hat{\xi}(\theta, 0)'}{\partial \delta}\hat{\Omega}(\hat{\theta}_c)^{-1}\hat{g}(\hat{\theta}_c), \qquad \hat{D}_c = \{\hat{g}(\hat{\theta}_c)\eta' + \eta\hat{g}(\hat{\theta}_c)'\}\hat{\Omega}(\hat{\theta}_c)^{-1}\hat{g}(\hat{\theta}_c), \\
\hat{\xi}(\hat{\theta}_c)' &= \begin{bmatrix} \hat{g}(\hat{\theta}_c)'\hat{\Omega}(\hat{\theta}_c)^{-1}\frac{\partial \hat{\Omega}(\hat{\theta}_c)}{\partial \theta^{(1)}} \\ \vdots \\ \hat{g}(\hat{\theta}_c)'\hat{\Omega}(\hat{\theta}_c)^{-1}\frac{\partial \hat{\Omega}(\hat{\theta}_c)}{\partial \theta^{(k)}} \end{bmatrix}, \qquad \frac{\partial \hat{\Omega}(\hat{\theta}_c)}{\partial \theta^{(j)}} = \frac{1}{n}\sum_{i=1}^{n} g_i(\hat{\theta}_c)G_i^{(j)}(\hat{\theta}_c)' + \frac{1}{n}\sum_{i=1}^{n} G_i^{(j)}(\hat{\theta}_c)g_i(\hat{\theta}_c)', \\
\frac{\partial \hat{\xi}^{(j)}(\theta, 0)'}{\partial \delta} &= \left(\eta' - \hat{g}(\hat{\theta}_c)'\hat{\Omega}(\hat{\theta}_c)^{-1}\{\hat{g}(\hat{\theta}_c)\eta' + \eta\hat{g}(\hat{\theta}_c)'\}\right)\hat{\Omega}(\hat{\theta}_c)^{-1}\frac{\partial \hat{\Omega}(\hat{\theta}_c)}{\partial \theta^{(j)}} \\
&\quad +\hat{g}(\hat{\theta}_c)'\hat{\Omega}(\hat{\theta}_c)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\eta\hat{G}^{(j)}(\hat{\theta}_c)' + \frac{1}{n}\sum_{i=1}^{n}\hat{G}^{(j)}(\hat{\theta}_c)\eta'\right),
\end{aligned}$$

where the superscript j refers to the j-th row of an object.

(v): *Bias of the continuous updating GMM with centering for the weight:*

$$\frac{\partial\tilde{\theta}_c(0)}{\partial\delta} = -\left[\{\hat{G}(\tilde{\theta}_c)' - \tilde{\xi}(\tilde{\theta}_c)'\}\tilde{\Omega}(\tilde{\theta}_c)^{-1}\hat{G}(\tilde{\theta}_c) + \tilde{A}_c\right]^{-1}\left[\{\hat{G}(\tilde{\theta}_c)' - \tilde{\xi}(\tilde{\theta}_c)'\}\tilde{\Omega}(\tilde{\theta}_c)^{-1}\eta + \tilde{B}_c\right],$$

where

$$\begin{aligned}
\tilde{A}_c &= \left[\left(\frac{\partial G(\tilde{\theta}_c)'}{\partial\theta^{(1)}} - \frac{\partial\tilde{\xi}(\tilde{\theta}_c)'}{\partial\theta^{(1)}} - \{\hat{G}(\tilde{\theta}_c)' - \tilde{\xi}(\tilde{\theta}_c)'\}\hat{\Omega}(\tilde{\theta}_c)^{-1}\frac{\partial\tilde{\Omega}(\tilde{\theta}_c)}{\partial\theta^{(1)}}\right)\tilde{\Omega}(\tilde{\theta}_c)^{-1}\hat{g}(\tilde{\theta}_c), \ldots\right. \\
&\quad \left.\left(\frac{\partial G(\tilde{\theta}_c)'}{\partial\theta^{(k)}} - \frac{\partial\tilde{\xi}(\tilde{\theta}_c)'}{\partial\theta^{(k)}} - \{\hat{G}(\tilde{\theta}_c)' - \tilde{\xi}(\tilde{\theta}_c)'\}\hat{\Omega}(\tilde{\theta}_c)^{-1}\frac{\partial\tilde{\Omega}(\tilde{\theta}_c)}{\partial\theta^{(k)}}\right)\tilde{\Omega}(\tilde{\theta}_c)^{-1}\hat{g}(\tilde{\theta}_c)\right], \\
\tilde{B}_c &= \frac{\partial\tilde{\xi}(\tilde{\theta}_c, 0)'}{\partial\delta}\tilde{\Omega}(\tilde{\theta}_c)^{-1}\hat{g}(\tilde{\theta}_c), \\
\tilde{\xi}(\tilde{\theta}_c)' &= \begin{bmatrix} \hat{g}(\tilde{\theta}_c)'\tilde{\Omega}(\tilde{\theta}_c)^{-1}\frac{\partial\tilde{\Omega}(\tilde{\theta}_c)}{\partial\theta^{(1)}} \\ \vdots \\ \hat{g}(\tilde{\theta}_c)'\tilde{\Omega}(\tilde{\theta}_c)^{-1}\frac{\partial\tilde{\Omega}(\tilde{\theta}_c)}{\partial\theta^{(k)}} \end{bmatrix}, \qquad \frac{\partial\tilde{\Omega}(\tilde{\theta}_c)}{\partial\theta^{(j)}} = \frac{\partial\hat{\Omega}(\tilde{\theta}_c)}{\partial\theta^{(j)}} - \hat{G}^{(j)}(\tilde{\theta}_c)\hat{g}(\tilde{\theta}_c)' - \hat{g}(\tilde{\theta}_c)\hat{G}^{(j)}(\tilde{\theta}_c)', \\
\frac{\partial\tilde{\xi}^{(j)}(\tilde{\theta}_c, 0)'}{\partial\delta} &= \eta'\tilde{\Omega}(\tilde{\theta}_c)^{-1}\frac{\partial\tilde{\Omega}(\tilde{\theta}_c)}{\partial\theta^{(j)}} \\
&\quad + \hat{g}(\tilde{\theta}_c)'\tilde{\Omega}(\tilde{\theta}_c)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} G_i^{(j)}(\tilde{\theta}_c)\eta' + \frac{1}{n}\sum_{i=1}^{n}\eta G_i^{(j)}(\tilde{\theta}_c)' - \hat{G}^{(j)}(\tilde{\theta}_c)\eta' - \eta\hat{G}^{(j)}(\tilde{\theta}_c)'\right),
\end{aligned}$$

The derivations are tedious but analogous to the one for the online appendix Proposition 2 of AGS by using the implicit function theorem. We note that under local perturbations in the sense of the online appendix Proposition 1 of AGS, it holds $\hat{g}(\hat{\theta}) \xrightarrow{p} 0$ for $\hat{\theta} = \hat{\theta}_2, \tilde{\theta}_2, \hat{\theta}_c, \tilde{\theta}_c$, and thus all these bias terms converge to $\Lambda\eta$ under mild regularity conditions that guarantee the convergence of the sample moments, such as $\hat{\Omega}(\cdot)$, $\tilde{\Omega}(\cdot)$, $\hat{G}(\cdot)$, and their derivatives.

The derivatives for the two-step and continuous updating GMM estimators take somewhat analogous forms. For example, the terms $\hat{A}_2, \tilde{A}_2, \hat{A}_c, \tilde{A}_c$ emerge by the same reason as $\hat{A}_1$ in the first stage GMM, i.e., the second term in the derivative of the first-order condition

$$\begin{aligned}
&\frac{\partial}{\partial\delta}\{\hat{G}(\hat{\theta}_1(\delta), \delta)'\hat{W}\hat{g}(\hat{\theta}_1(\delta), \delta)\} \\
= \;& \hat{G}(\hat{\theta}_1(\delta), \delta)'\hat{W}\left\{\frac{\partial}{\partial\delta}\hat{g}(\hat{\theta}_1(\delta), \delta)\right\} + \left\{\frac{\partial}{\partial\delta}\hat{G}(\hat{\theta}_1(\delta), \delta)\right\}'\hat{W}\hat{g}(\hat{\theta}_1(\delta), \delta).
\end{aligned}$$

On the other hand, the terms $\hat{B}_2, \tilde{B}_2, \hat{B}_c, \tilde{B}_c$ are additional terms due to dependence of the weight matrices on δ. Observe that the terms $\hat{A}_2, \tilde{A}_2, \hat{A}_c, \tilde{A}_c$ and $\hat{B}_2, \tilde{B}_2, \hat{B}_c, \tilde{B}_c$ involve the components $\hat{g}(\hat{\theta})$ for the corresponding estimators. Therefore, under local perturbations with $\hat{g}(\hat{\theta}) \xrightarrow{p} 0$, these terms are asymptotically negligible. However, as is illustrated in the next section, these terms are useful to distinguish bias properties of the different GMM estimation methods in finite samples.

3. Illustration

In this section, we apply our proposed sensitivity measures to a consumption based asset pricing model. Our illustrating example follows Hansen and Singleton (1982). Consider a representative investor who maximizes his/her expected discounted utility

$$\max_{\{C_{t+i}\}_{i=1}^{\infty}} \mathbb{E}\left[\sum_{i=0}^{\infty} \beta^i \frac{C_{t+i}^{\gamma}-1}{\gamma} \middle| I_t\right],$$

where C_t denotes consumption at time t, and I_t denotes the information set of the investor at time t. There is an asset available to the investor, and we denote the net rate of return on the asset at time t as R_t. Then the first order condition implies

$$\mathbb{E}\left[\beta(1+R_t)\left(\frac{C_{t+1}}{C_t}\right)^{\gamma-1}-1 \middle| I_t\right] = 0.$$

Let $Z_t \subseteq I_t$ be a subset of the information set I_t. Our parameters of interest $\theta = (\beta, \gamma)'$ are the subjective discount factor β and the coefficient in relative risk aversion γ. These parameters can be estimated by using the moment condition $\mathbb{E}[g_t(\theta_0)] = 0$, where

$$g_t(\theta) = Z_t\left\{\beta(1+R_t)\left(\frac{C_{t+1}}{C_t}\right)^{\gamma-1}-1\right\}. \tag{4}$$

For our illustration, we use data on the consumption growth rate and the average rate of return on stocks in Japan. Using the Nikkei 225 stock price index and the seasonally adjusted nondurable consumption data from Yahoo Finance and Statistics Bureau of Japan, we construct the data on the consumption growth rate and the average rate of return on stocks. These time series data are adjusted for inflation by the consumer price index. The time period is from January 1981 to June 2016.

In the moment function (4), there are many possible choices of instruments Z_t. Hansen and Singleton (1982) used lags of consumption growth and lags of the rate of return on assets, and estimated the parameters with a number of different choices of instruments. Here we focus on one set of instruments to simplify the presentation, that is $Z_t = (1, C_t/C_{t-1}, C_{t-1}/C_{t-2}, R_{t-1}, R_{t-2})'$. The parameter estimates are summarized in Table 1. The obtained GMM estimates are reasonable in the empirical literature of this class of consumption based asset pricing model. The estimate of γ for the first stage GMM with $W = I$ takes a lower value compared to the other GMM estimates.

Table 1. Parameter estimates

$\hat{W}$	Coefficient in relative risk aversion γ	Discount factor β
I	0.6187	0.9970
$\hat{\Omega}(\hat{\theta}_1)$	1.0110	0.9966
$\tilde{\Omega}(\hat{\theta}_1)$	1.0116	0.9966
$\hat{\Omega}(\theta)$	1.0030	0.9966
$\tilde{\Omega}(\theta)$	1.0030	0.9966

The present model assumes time-separable utility, and thus the current consumption decision is independent from the past consumption path. There are some situations where this assumption is not appropriate; in models with habit formation, the current consumption choice may depend on the agent's own past consumption through the habit stock, for example. Based on this background, we consider the perturbed model $g_i(\theta, \delta) = g_i(\theta) + \delta\eta$. If the model is correctly specified, η should equal to zero vector, which corresponds to the original model assumption. Alternatively, a researcher may have doubts about the time-separability assumption and believe that lags of consumption growth cannot be valid instruments. Under this alternative view, the second and third elements of η should be non-zero (with the other elements being zeros).

Under such local violations of identifying assumptions, Table 2 reports the bias estimates $\hat{\Lambda}\eta$ based on AGS's proposed measure of sensitivity

$$\hat{\Lambda} = \{\hat{G}(\hat{\theta})'\hat{W}\hat{G}(\hat{\theta})\}^{-1}\hat{G}(\hat{\theta})'\hat{W},$$

with different GMM weight matrices for $\hat{W}$. Except for the case of $\hat{W} = I$, the bias estimates $\hat{\Lambda}\eta$ are quite similar across different choices of weight matrices. The local violation with $\eta = (0, 0.01, 0.01, 0, 0)'$ implies that the estimates of γ are biased upward by roughly 17%, and that bias for β is negligibly small. For $\hat{W} = I$, although the estimated values are different, the proportions of the biases are comparable to the other cases. In this example, the bias estimates $\hat{\Lambda}\eta$ by AGS are not sensitive to the choice of the weights for the two-stage or continuous updating GMM.

Table 2. Bias in parameter estimates measured with $\hat{\Lambda}\eta$

$\hat{W}$ and $\hat{\theta}$	Bias in coefficient in relative risk aversionγ	Bias in discount factorβ
I and $\hat{\theta}_1$	0.009113	-1.2425×10^{-6}
$\hat{\Omega}(\hat{\theta}_1)$ and $\hat{\theta}_2$	0.016944	5.0308×10^{-6}
$\tilde{\Omega}(\hat{\theta}_1)$ and $\tilde{\theta}_2$	0.016962	5.0082×10^{-6}
$\hat{\Omega}(\theta)$ and $\hat{\theta}_c$	0.016466	5.1106×10^{-6}
$\tilde{\Omega}(\theta)$ and $\tilde{\theta}_c$	0.016465	5.11×10^{-6}

Notes: This table reports bias in parameter estimates measured with $\Lambda\eta$. We consider a local violation of identifying assumptions with $\eta = (0, 0.01, 0.01, 0, 0)'$.

In Table 3, we now report the estimates of the proposed bias measures. For $\hat{W} = I$, the bias estimates are very similar to AGS's $\hat{\Lambda}\eta$ with $\hat{W} = I$. As expected, our bias measures are sensitive to the choice of the GMM weight. Although the signs of the biases are same as AGS's $\hat{\Lambda}\eta$, their values are different. For example, if we focus on γ, the bias of the continuous updating GMM with centering (i.e., $\partial\tilde{\theta}_c(0)/\partial\delta$) is similar to $\hat{\Lambda}\eta$, and takes the smallest value among the two step and continuous updating GMM methods. On the other hand, the bias of the two step GMM (i.e., $\partial\hat{\theta}_2(0)/\partial\delta$) implies a larger upward bias, around 58% of the estimate of γ. Also, the centering substantially reduces the bias of $\partial\hat{\theta}_2(0)/\partial\delta$. In this example, our bias estimates indicate that the two step GMM methods are more sensitive due to the sensitivity in the first step estimation, and that the continuous updating GMM tends to be less sensitive.

Table 3. Bias in parameter estimates measured with $\partial\hat{\theta}(0)/\partial\delta$

$\frac{\partial\hat{\theta}(0)}{\partial\delta}$	Bias in coefficient in relative risk aversion γ	Bias in discount factor β
$\hat{\theta}_1$	0.0092349	-1.1902×10^{-6}
$\hat{\theta}_2$	0.58207	4.8183×10^{-3}
$\tilde{\theta}_2$	0.13085	6.0069×10^{-4}
$\hat{\theta}_c$	0.024685	2.7922×10^{-7}
$\tilde{\theta}_c$	0.018237	4.0453×10^{-6}

Notes: This table reports bias in parameter estimates measured with $\partial\hat{\theta}(0)/\partial\delta$. We consider a local violation of identifying assumptions with $\eta = (0, 0.01, 0.01, 0, 0)'$.

References

[1] Andrews, I., Gentzkow, M. and J. M. Shapiro (2017) Measuring the sensitivity of parameter estimates to estimation moments, *Quarterly Journal of Economics*, 132, 1553-1592.

[2] Hall, A. R. (2000) Covariance matrix estimation and the power of the overidentifying restrictions test, *Econometrica*, 68, 1517-1527.

[3] Hall, A. R. and A. Inoue (2003) The large sample behaviour of the generalized method of moments estimator in misspecified models, *Journal of Econometrics*, 114, 361-394.

[4] Hansen, L. P., Heaton, J. and A. Yaron (1996) Finite-sample properties of some alternative GMM estimators, *Journal of Business & Economic Statistics*, 14, 262-280.

[5] Hansen, L. P. and K. J. Singleton (1982) Generalized instrumental variables estimation of nonlinear rational expectations models, *Econometrica*, 50, 1269-1286.

Faculty of Economics, Keio University, 2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan.
Email address: rikutoo0802@keio.jp

Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, UK, and Keio Economic Observatory (KEO), 2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan.
Email address: t.otsu@lse.ac.uk