Statistica Sinica **20** (2010), 1725-1747

INSTRUMENTAL VARIABLE AND GMM ESTIMATION FOR PANEL DATA WITH MEASUREMENT ERROR

Zhiguo Xiao[1], Jun Shao[2] and Mari Palta[2]

[1]*Fudan University and* [2]*University of Wisconsin, Madison*

Abstract: Panel data allow correction for measurement error without assuming a known measurement error covariance matrix or using additional validation/replication data to estimate the measurement error covariance matrix. Griliches and Hausman (1986) proposed using the generalized method of moments (GMM) or optimal weighting to efficiently combine instrumental variable (IV) estimators. Wansbeek (2001) applied GMM based on moment conditions expressed in the form of the Kronecker product. This paper studies some issues crucial to applications of these two approaches, including the estimability of the regression parameter under Griliches and Hausman's or Wansbeek's approach, how to choose instruments, what is the optimally weighted IV estimator, how to explicitly construct GMM estimators, how to remove the redundancy of the moment conditions constructed by Wansbeek (2001), and the existence of optimal GMM estimators. We unify Griliches and Hausman's and Wansbeek's approaches by establishing their equivalence. We also consider models with exogenous regressors and models with nonclassical assumptions. We apply the methods in this paper to revisit an investment controversy, viz., whether financially constrained firms respond to internal funds such as cash flow more sensitively than financially unconstrained firms.

Key words and phrases: Equivalence, GMM, instrumental variable, measurement error, panel data, Tobin's q.

1. Introduction

Measurement error (errors-in-variables or errors-in-regressors) leads to the failure of classical estimation methods such as ordinary least squares (OLS). Under standard assumptions, with a single regressor measured with random error, the OLS estimator of the regression coefficient is inconsistent and biased towards zero. Existing remedies for the measurement error problem often require that the measurement error covariance matrix be known or that it can be estimated using additional validation/replication data (see, e.g., Fuller (1987), Zhong, Fung and Wei (2002), Cui, Ng and Zhu (2004), Carroll et al. (2006)). In panel data, each individual has more than one observation, which can be used as "partial" replicates for the purpose of handling measurement error. As shown by Griliches and Hausman (1986) (hereinafter Griliches-Hausman) and Wansbeek (2001) (hereinafter

Wansbeek), under some panel data models, valid estimators can be constructed without the requirement of knowing the measurement error covariance matrix or additional validation/replication data.

In their seminal paper on errors-in-variables in panel data, Griliches-Hausman proposed using either the Generalized Method of Moments (GMM, Hansen (1982)) or weighting to efficiently combine instrumental variable estimators constructed using linear combinations of the observed regressors as instruments. Griliches-Hausman, and later research work by Biørn and Klette (1998) and Biørn (2000), provided a GMM estimator based on instrumental variables derived from difference transformations. However, a general form of Griliches-Hausman's GMM estimator is not available. Griliches-Hausman's idea of using optimal weights to combine instrumental variable estimators has not been pursued by later researchers, possibly because no explicit weighting formula was provided. Wansbeek's GMM approach for panel data with measurement error is based on a set of moment conditions constructed using the structure of measurement error covariance matrix.

The purpose of this paper is to study (i) some issues that are not addressed or not fully addressed in Griliches-Hausman and Wansbeek but are crucial to applications of these two approaches; (ii) the relationship between Griliches and Hausman's and Wansbeek's estimators; (iii) the extensions of the two methods to more complicated models; and (iv) the estimability of the regression parameter under Griliches and Hausman's or Wansbeek's approach. Issues in (i) include how to choose instruments, what the optimally weighted IV estimator is, how to explicitly construct GMM estimators, how to remove the redundancy of the moment conditions constructed by Wansbeek, and the existence of optimal GMM estimators. For (ii), we show that the optimally weighted IV estimator is identical to a GMM estimator when the same set of instruments is used, i.e., Griliches-Hausman's two ways of optimally obtaining an estimator are the same. Furthermore, we show that Griliches-Hausman's and Wansbeek's approaches use equivalent sets of moment conditions and, hence, the efficient GMM estimator under Griliches-Hausman's approach is asymptotically equivalent to the efficient GMM estimator under Wansbeek's approach. For (iii), we extend Griliches-Hausman's method to allow for strictly exogenous covariates in the model; we construct a set of moment conditions that can yield asymptotically more efficient GMM estimators than Wansbeek when strictly exogenous covariates exist; and we extend Griliches-Hausman's and Wansbeek's methods to nonclassical models. For (iv), we show that the estimability in Griliches-Hausman's approach (i.e., the existence of at least one instrumental variable) is the same as the estimability in Wansbeek's approach. We also provide a necessary and sufficient condition for the estimability, which can be easily checked in terms of the correlation structures of the measurement error and the covariate subject to measurement error.

Results related to Griliches-Hausman's approach are given in Section 2, after an introduction of the model and notation. Results related to Wansbeek's GMM and its equivalence to Griliches-Hausman's GMM are presented in Section 3. The estimability issue is studied in Section 4. In Section 5, we generalize the methods to situations where there are exogenous covariates not measured with errors, and where measurement error and covariate are correlated. In Section 6, we apply our methods to a long-debated topic in corporate finance, viz., whether financially constrained firms respond more sensitively to cash flow than unconstrained firms, using the COMPUSTAT database for years 1992-1995. Section 7 contains some concluding remarks. Proofs of lemmas and theorems are given in the Appendix.

2. Griliches-Hausman's GMM and Weighted IV Estimator

We start with the model of Griliches-Hausman, a static linear model with one regressor measured with error:

$$\begin{aligned} y_{it} &= \xi_{it}\beta + \alpha_i + \eta_{it}, \\ x_{it} &= \xi_{it} + v_{it}, \end{aligned} \qquad t = 1, \ldots, T; i = 1, \ldots, N, \tag{2.1}$$

where β is a parameter of interest, y_{it} is the tth observed response from the ith individual, ξ_{it} is an unobserved covariate associated with y_{it}, η_{it} is a regression error, x_{it} is an observed surrogate for ξ_{it} with unobserved measurement error v_{it}, and α_i is an unobserved individual fixed effect or random effect that may be correlated with ξ_i. The estimation of the effect α_i is not considered until Section 5.3. For each i, let $y_i = (y_{i1}, \ldots, y_{iT})'$, $x_i = (x_{i1}, \ldots, x_{iT})'$, $\xi_i = (\xi_{i1}, \ldots, \xi_{iT})'$, $\eta_i = (\eta_{i1}, \ldots, \eta_{iT})'$, and $v_i = (v_{i1}, \ldots, v_{iT})'$.

Assumption A. The random $3T$-vectors $(\xi_i', \eta_i', v_i')'$, $i = 1, \ldots, N$, are i.i.d. with $E\eta_i = 0$, $Ev_i = 0$, and a finite positive definite covariance matrix. The random T-vectors ξ_i, η_i, and v_i are independent.

Under this assumption, homoskedasticity and independence across i of data is assumed but heteroskedasticity and correlation across t (within i) is allowed for. The covariance matrix of $(\xi_i', \eta_i', v_i')'$ is a block diagonal matrix with three $T \times T$ diagonal blocks, which are the covariance matrices of ξ_i, η_i, and v_i.

Replacing ξ_i by $x_i - v_i$ in (2.1), we obtain

$$y_i = x_i\beta + l_T\alpha_i + \varepsilon_i, \tag{2.2}$$

where $\varepsilon_i = \eta_i - v_i\beta$ and l_T is the $T \times 1$ vector of ones. To obtain consistent estimators, Griliches-Hausman suggested that we first find a set of instrumental variable estimators of the form

$$\widehat{\beta}_P = \left(w'x\right)^{-1} w'y, \quad w = \left(I_N \otimes P\right) x,$$

where I_N is the identity matrix of order N, x is the stacked column vector of all x_i's, $\otimes$ is the Kronecker product (see Abadir and Magnus (2005), Chapter 10 for properties of the Kronecker product and the *vec* operator below), and P is a $T \times T$ deterministic known matrix satisfying

(IV.1) $P' l_T = 0$,

(IV.2) $E[x_i' P' \varepsilon_i] = 0$,

(IV.3) $E[x_i' P' x_i] \neq 0$.

Condition (IV.1) ensures that multiplying P' to both sides of (2.2) eliminates the unobserved α_i. Conditions (IV.2)$-$(IV.3) state that the instrument Px_i must be uncorrelated with the error ε_i but correlated with x_i. Let

$$\mathcal{P} = \{\text{all } T \times T \text{ matrices } P\text{'s satisfying (IV.1)}-\text{(IV.3)}\}. \tag{2.3}$$

To use Griliches-Hausman's method we need to assume that $\mathcal{P} \neq \emptyset$, i.e., there exists at least one P satisfying (IV.1)$-$(IV.3). A further discussion about the condition $\mathcal{P} \neq \emptyset$ is given in Section 4.

For each $P \in \mathcal{P}$, $\widehat{\beta}_P$ is a consistent but possibly inefficient estimator of β, since it is a moment estimator. Griliches-Hausman suggested two ways to combine moment estimators to obtain an efficient estimator of β:

(**W1**) ''*...to combine such estimators optimally we use the Generalized Method of Moments estimator, developed by* Hansen (1982)*; and* White (1982)*, ...''* – Griliches and Hausman (1986, p.104, lines 30-33),

(**W2**) *''After obtaining a complete set of instruments, the efficient β is calculated as a weighted average of the β's from each w. The inverse of the variance-covariance matrix of the β's is the appropriate weighting matrix.''* – Griliches and Hausman (1986, p.115, lines 26-28).

There are infinitely many P's in $\mathcal{P}$ when $\mathcal{P}$ is not empty. Thus, prior to implementing (W1) or (W2), as argued by Griliches-Hausman, we need to find a complete set of $P_1, \cdots, P_K$ ($K \leq T^2$) in the sense that (1) $P_k \in \mathcal{P}$, $k = 1, \ldots, K$; (2) $P_1, \ldots, P_K$ are linearly independent; (3) every $P \in \mathcal{P}$ is a linear combination of $P_1, \ldots, P_K$. A complete set $\{P_1, \ldots, P_K\}$ is referred to as a basis of $\mathcal{P}$.

To apply this method we first need to answer the following questions that are not fully addressed in Griliches-Hausman.

1. Does $\mathcal{P}$ have a basis?
2. If a basis exists, how do we find it?
3. After we find a basis $P_1, \ldots, P_K$, how do we implement Griliches-Hausman's method by (W1) or (W2)? Furthermore, are the estimators resulting from (W1) and (W2) the same?

Note that $\mathcal{P}$ is not a linear subspace of $\mathbb{R}^{T^2}$ and, thus, it is not obvious whether $\mathcal{P}$ has a basis. Let

$$\mathcal{P}_0 = \{P : P \text{ is a } T \times T \text{ matrix satisfying (IV.1) and (IV.2)}\},$$
$$\mathcal{P}_1 = \{P : P \text{ is a } T \times T \text{ matrix satisfying (IV.1) and (IV.2), but not (IV.3)}\}.$$

Here $\mathcal{P}_0$ and $\mathcal{P}_1$ are linear subspaces of $\mathbb{R}^{T^2}$, $\mathcal{P}_1 \subseteq \mathcal{P}_0$, and $\mathcal{P} = \mathcal{P}_0 \backslash \mathcal{P}_1$. With this understanding, we can answer Question 1 with the following lemma.

Lemma 1. *Let V_0 and V_1 be two linear subspaces of $\mathbb{R}^p$ with $V_1 \subsetneq V_0$. Then the maximum number of linearly independent vectors in $S = V_0 \backslash V_1$ is $K =$ the dimension of V_0. Furthermore, if $s_1, \ldots, s_K$ are linearly independent vectors in S, then any vector in S is a linear combination of $s_1, \ldots, s_K$.*

To answer the second question, we need the following lemma that characterizes (IV.1)−(IV.3) in unified algebraic forms. Observe that for every measurement error covariance matrix $E[v_i v_i']$, there exists a known matrix R_0 such that $vec(E[v_i v_i']) = R_0 \varphi$, where the *vec* operator stacks the columns of a matrix into a column vector, and φ is the vector of free parameters in $E[v_i v_i']$. The matrix R_0 represents the structure of $E[v_i v_i']$. If φ is an $m \times 1$ vector, then R_0 is a $T^2 \times m$ matrix of full rank.

Lemma 2. (1) *$P' l_T = 0$ if and only if $vec(P)'(I_T \otimes l_T) = 0$;*
(2) *$E[x_i' P' \varepsilon_i] = 0$ if and only if $vec(P)' R_0 = 0$;*
(3) *$E[x_i' P' x_i] \neq 0$ if and only if $vec(P)'[vec(E[v_i v_i']) + vec(E[\xi_i \xi_i'])] \neq 0$.*

It follows from Lemma 2 that

$$\mathcal{P}_0 = \left\{P : P' l_T = 0, vec(P)' R_0 = 0\right\} = \left\{P : vec(P) \in [\mathcal{C}(L)]^{\perp}\right\}, \tag{2.4}$$

where the matrix $L = [R_0 \vdots I_T \otimes l_T]$ is $T^2 \times (m+T)$, $\mathcal{C}(L)$ denotes the column space of L, and $[\mathcal{C}(L)]^{\perp}$ denotes the orthogonal complement of $\mathcal{C}(L)$. Hence

$$K = \dim(\mathcal{P}_0) = T^2 - \dim(\mathcal{C}(L)).$$

To obtain a basis of $\mathcal{P}$, we first find a basis of $[\mathcal{C}(L)]^{\perp}$, say $\{\nu_1, \ldots, \nu_K\}$. For example, a set of eigenvectors of matrix LL' corresponding to the zero eigenvalue of LL' can serve this purpose. By the assumption that $\mathcal{P} \neq \emptyset$, at least one of them, say ν_1, satisfies condition (IV.3). Now let

$$\widetilde{\nu}_i = \begin{cases} \nu_1, & i = 1; \\ \nu_i + \nu_1, & K \geq i \geq 2. \end{cases}$$

Then $\{P_1, \ldots, P_K\}$ is a basis of $\mathcal{P}$, where $\widetilde{\nu}_i = vec\,(P_i)\,, i = 1, \ldots, K$.

Now we answer the third question. Let $\{P_1, \ldots, P_K\}$ be a basis of $\mathcal{P}$. According to Griliches-Hausman, as cited above in (W1) and (W2), we have two ways to obtain an efficient estimator. We first study the GMM approach (W1). The moment conditions associated with $P_1, \ldots, P_K$ are

$$E\left[x_i' P_k' \left(y_i - x_i\beta\right)\right] = 0, \quad k = 1, \ldots, K.$$

Because $x_i' P_k' \left(y_i - x_i\beta\right) = vec(P_k)' \left[x_i \otimes \left(y_i - x_i\beta\right)\right]$ for each k, these equations can be written as

$$HE\left[x_i \otimes \left(y_i - x_i\beta\right)\right] = 0, \tag{2.5}$$

where $H = \left[vec\,(P_1)\,, \ldots, vec\,(P_K)\right]'$. Since $P_1, \ldots, P_K$ are linearly independent, H is a matrix of full rank and the system of orthogonality conditions in (2.5) contains no redundant equations. The (asymptotically) efficient two-step GMM estimator can be derived from (2.5), using the technique described in Hall (2005). Specifically, let

$$A_i = H\left[x_i \otimes y_i\right], \quad B_i = H\left[x_i \otimes x_i\right], \quad \overline{A} = \frac{1}{N}\sum_{i=1}^{N} A_i, \quad \overline{B} = \frac{1}{N}\sum_{i=1}^{N} B_i. \tag{2.6}$$

Then the GMM estimator using weighting matrix W is the minimizer of $(\overline{A} - \overline{B}\beta)' W (\overline{A} - \overline{B}\beta)$ over β, which is given by

$$\widehat{\beta}_{GMM} = \left(\overline{B}' W \overline{B}\right)^{-1} \overline{B}' W \overline{A}.$$

A typical consistent first-step GMM estimator is the unweighted GMM estimator where W is taken as the identity matrix:

$$\widehat{\beta}_{GMM1} = \left(\overline{B}' \overline{B}\right)^{-1} \overline{B}' \overline{A}.$$

Under Assumption A, the matrix

$$\Gamma = E\left[\left(A_i - B_i\beta\right)\left(A_i - B_i\beta\right)'\right] \tag{2.7}$$

is well-defined and does not depend on i. In terms of the asymptotic variance, $W = \Gamma^{-1}$, if it is known, is the optimal weighting matrix. Since Γ^{-1} is unknown, we replace it by a consistent estimator

$$\widehat{W} = \left[\frac{1}{N}\sum_{i=1}^{N} \left(A_i - B_i\widehat{\beta}_{GMM1}\right)\left(A_i - B_i\widehat{\beta}_{GMM1}\right)'\right]^{-1}. \tag{2.8}$$

A two-step GMM estimator is then

$$\widehat{\beta}_{GMM2} = \left(\overline{B}'\widehat{W}\overline{B}\right)^{-1}\overline{B}'\widehat{W}\overline{A}. \tag{2.9}$$

Griliches-Hausman provided a GMM estimator at (21) in their paper. To compute their estimator one needs to find a basis $P_1, \ldots, P_K$ using difference transformations, where some P_k's have to be obtained manually. To compute our GMM estimator $\widehat{\beta}_{GMM2}$, we only need to determine the matrix H in (2.5), as discussed previously.

From standard statistical theory, $\widehat{\beta}_{GMM2}$ in (2.9) is consistent for β and asymptotically normal, as $N \to \infty$. It is asymptotically efficient (optimal) in the sense that the asymptotic variance of $\widehat{\beta}_{GMM2}$ is no larger than that of $\widehat{\beta}_{GMM}$ with any W, and is smaller than that of $\widehat{\beta}_{GMM}$ unless W converges to Γ^{-1}.

One natural question is whether the asymptotic variances of $\widehat{\beta}_{GMM2}$ in (2.9) are the same for different bases of $\mathcal{P}$. If $\{P_1, \ldots, P_K\}$ and $\{\widetilde{P}_1, \ldots, \widetilde{P}_J\}$ are two bases of $\mathcal{P}$, then $J = K$. Let $\widetilde{H}$ be the same as H in (2.5) with P_k replaced by $\widetilde{P}_k$. Then, there exists a $K \times K$ nonsingular matrix Π such that $\widetilde{H} = \Pi H$. The set of moment conditions in (2.5) and

$$\widetilde{H}E\left[x_i \otimes (y_i - x_i\beta)\right] = 0$$

are **equivalent variants** of each other in the sense that any moment condition in one of them is a linear combination of the moment conditions in the other. The following result provides sufficient conditions for two sets of moment conditions to be equivalent variants of each other, and shows that equivalent variants produce two-step GMM estimators with the same efficiency.

Lemma 3. *Let Q_1 and Q_2 be two matrices with the same number of columns.*

(i) *Suppose that $Q_1\zeta = 0$ if and only if $Q_2\zeta = 0$, for any vector ζ. Then the set of moment conditions $Q_1E\left[g(\beta)\right] = 0$ is an equivalent variant of the set of moment conditions $Q_2E\left[g(\beta)\right] = 0$, where g is a given function of β. Furthermore, two-step GMM estimators based on $Q_1E\left[g(\beta)\right] = 0$ and $Q_2E\left[g(\beta)\right] = 0$ have the same asymptotically normal distribution.*

(ii) *Suppose that $Q_1\zeta = 0$ implies $Q_2\zeta = 0$ for any vector ζ. Then the two-step GMM estimator based on $Q_1E\left[g(\beta)\right] = 0$ is at least as efficient as the two-step GMM estimator based on $Q_2E\left[g(\beta)\right] = 0$.*

As a direct consequence of Lemma 3 and the previous discussion, we have the following.

Theorem 4. *Suppose Assumption A holds and $\mathcal{P} \neq \emptyset$. For any basis $P_1, \ldots, P_K$ of $\mathcal{P}$, the GMM estimator defined by (2.9) is asymptotically optimal among all such GMM estimators.*

Now, we discuss the second way of constructing an efficient estimator. As cited in (W2), Griliches-Hausman mentioned optimally weighting a complete set of IV estimators, but did not provide an explicit weighting formula.

Let $P_1, \ldots, P_K$ be a basis of $\mathcal{P}$ and $\widehat{\beta}_{P_k} = \left(w_k' x\right)^{-1} w_k' y$, with $w_k = (I_N \otimes P_k)\, x$ for $k = 1, \ldots, K$. For each K-vector $\lambda = (\lambda_1, \ldots, \lambda_K)$ satisfying $\lambda_1 + \cdots + \lambda_K = 1$, we define the weighted IV estimator

$$\hat{\beta}(\lambda) = \lambda_1 \hat{\beta}_{P_1} + \cdots + \lambda_K \hat{\beta}_{P_K}. \tag{2.10}$$

Since $\widehat{\beta}_{P_1}, \ldots, \widehat{\beta}_{P_K}$ are consistent for β, $\widehat{\beta}\,(\lambda) = \lambda_1 \widehat{\beta}_{P_1} + \cdots + \lambda_K \widehat{\beta}_{P_K}$ is consistent for β. The following result shows how to find an optimally weighted IV estimator.

Theorem 5. *uppose Assumption A holds and $\mathcal{P} \neq \emptyset$.*

(i) *For each fixed λ, $(\hat{\beta}(\lambda) - \beta)/\sqrt{V(\lambda)}$ converges in distribution to standard normal.*

(ii) *$V(\lambda)$ is minimized at $\lambda^* = \left(B'\Gamma^{-1}B\right)^{-1} diag\,(B)\,\Gamma^{-1}B$ and $V(\lambda^*) = (B'\Gamma^{-1}B)^{-1}$, where Γ is defined in (2.7), $B = (B_1, \ldots, B_K)'$ is a K-vector with $B_k = E(x_i' P_k' x_i)$, and $diag(B)$ is the $K \times K$ diagonal matrix whose kth diagonal element is B_k.*

(iii) *If we estimate λ^* by*

$$\widehat{\lambda} = \left(\overline{B}'\widehat{W}\overline{B}\right)^{-1} diag\left(\overline{B}\right)\widehat{W}\overline{B},$$

where $\widehat{W}$ is given by (2.8) and $\overline{B}$ is given by (2.6), then the weighted IV estimator $\hat{\beta}(\widehat{\lambda})$ is identical to $\hat{\beta}_{GMM2}$ defined at (2.9) and is optimal in the sense that $\hat{\beta}(\widehat{\lambda})$ is asymptotically normal with mean β and asymptotic variance $V(\lambda^*)$.

Theorem 5 indicates that an optimally weighted IV estimator is identical to an efficient two-step GMM estimator. Although optimally weighted IV estimators and efficient two-step GMM estimators are not unique (we may use different consistent estimators of Γ^{-1} and/or different consistent estimators of λ^*), Theorem 5 shows that all of them are asymptotically equivalent in the sense that they are asymptotically normal with mean β and the same asymptotic variance.

Since Griliches-Hausman's approaches in (W1) and (W2) produce the same or asymptotically equivalent estimators, we refer to Griliches-Hausman's GMM method, or simply Griliches-Hausman's approach.

3. Wansbeek's GMM Estimator

Instead of constructing orthogonality conditions by means of instrumental variables, Wansbeek constructed moment conditions directly. Let $A_T =$

$I_T - (1/T)l_T l_T'$ denote the within transformation, $R = (I_T \otimes A_T)\, R_0$, and $M_R = I_{T^2} - R\,(R'R)^- R'$, where $(R'R)^-$ is a generalized inverse of $R'R$ and R_0 is the correlation structure matrix in $vec(E[v_i v_i']) = R_0 \varphi$. Under (2.1) with Assumption A, Wansbeek derived

$$M_R\,(I_T \otimes A_T)\, E\,[x_i \otimes (y_i - x_i \beta)] = 0, \tag{3.1}$$

and suggested constructing GMM estimators based on (3.1). However, some equations in (3.1) are linear combinations of others. Hence, the covariance matrix of $M_R\,(I_T \otimes A_T)\,[x_i \otimes (y_i - x_i\beta)]$ is not invertible and the efficient GMM estimator in the form of (2.9) is not directly obtainable based on (3.1). The redundancy in (3.1) comes from the singularity of $M_R\,(I_T \otimes A_T)$, which is caused by the singularity of A_T and M_R.

To continue we need the following concept. A set of moment conditions is **essential** if, in this set of moment conditions, no single moment condition can be written as a linear combination of the rest. If a set of moment conditions is not essential, then we should find its essential equivalent variant, from which we can easily obtain GMM estimators in the form of (2.9). The question is, given a set of moment conditions that is not essential, how do we find its essential equivalent variant?

Biørn and Klette (1998) and Biørn (2000) provided solutions for identifying essential conditions in some special cases related to Griliches-Hausman's GMM. For the moment conditions in (3.1), we suggest using singular value decomposition to obtain essential equivalent variant. Specifically, suppose the singular value decomposition of $M_R\,(I_T \otimes A_T)$ is $M_R\,(I_T \otimes A_T) = U \Lambda V'$, where U and V are $T^2 \times r$ matrices whose columns are orthogonal to each other, r is the rank of $M_R\,(I_T \otimes A_T)$, and Λ is a $r \times r$ nonsingular diagonal matrix. By Lemma 3, an essential equivalent variant of (3.1) is

$$V' E\,[x_i \otimes (y_i - x_i \beta)] = 0. \tag{3.2}$$

Now, the asymptotically efficient GMM estimator based on (3.2) is given by (2.9) with $A_i = V'\,[x_i \otimes y_i]$ and $B_i = V'\,[x_i \otimes x_i]$.

Wansbeek's choice of (A_T, M_R) is not the only way of constructing moment conditions. Consider an arbitrary $l \times T$ matrix A satisfying $A l_T = 0$ and a corresponding conformable matrix M satisfying $M\,(I_T \otimes A)\, R_0 = 0$. Following the procedures above, we can obtain the GMM estimator (2.9) based on (A, M) and the essential equivalent variant of

$$M\,(I_T \otimes A)\, E\,[x_i \otimes (y_i - x_i \beta)] = 0. \tag{3.3}$$

The set of moment conditions constructed by Wansbeek in (3.1) is a special case of (3.3). Let

$$\mathcal{S} = \{(A, M) : Al_T = 0 \text{ and } M\,(I_T \otimes A)\,R_0 = 0\}$$

be the collection of all pairs (A, M) that can be used for constructing moment conditions. Besides Wansbeek's (A_T, M_R), another popular pair in $\mathcal{S}$ is (D_1, M_{D_1}), where D_1 is the $(T-1) \times T$ matrix representing the first order difference transformation, i.e., in the tth row of D_1, the tth element is -1, the $(t+1)$th element is 1, and the rest of the elements are 0, $t = 1, \ldots, T-1$, and

$$M_{D_1} = I_{T(T-1)} - R_{D_1}\left(R'_{D_1} R_{D_1}\right)^{-} R'_{D_1}, \qquad R_{D_1} = (I_T \otimes D_1)\,R_0.$$

Now, two questions arise naturally.

- Among all $(A, M) \in \mathcal{S}$ leading to the GMM estimator (2.9), is there an optimal choice in the sense that the GMM estimator (2.9) has the smallest asymptotic variance?
- Is Wansbeek's choice (A_T, M_R) optimal?

To answer the above questions we establish the following result.

Theorem 6.(i) *The two-step GMM estimator* (2.9) *based on a pair* $(A, M) \in \mathcal{S}$ *is optimal if* (A, M) *satisfies* (1) *rank* $(A) = T - 1$ *and* (2) M *has the maximum possible rank, rank* $(M) = T^2 -$ *rank* (R_A), *where* $R_A = (I_T \otimes A)\,R_0$.
(ii) *The sets of moment conditions based on any pairs* (A_1, M_1) *and* (A_2, M_2) *in* $\mathcal{S}$ *satisfying conditions in part* (i) *are equivalent variants of each other.*

It is easy to check that both (A_T, M_R) and (D_1, M_{D_1}) satisfy the conditions in Theorem 6(i). Hence, they both lead to GMM estimators that are optimal among GMM estimators (2.9) based on pairs in $\mathcal{S}$.

As the last result of this section, we study the relationship between Griliches-Hausman's and Wansbeek's approaches. Griliches-Hausman's set of moment conditions (2.5) bears a striking resemblance to Wansbeek's set of conditions (3.1). This hints at an intimate relationship between them. The following result tells us that any moment condition in (2.5) is a linear combination of moment conditions in (3.1), and vice versa. Therefore, the two methods lead to asymptotically equivalent GMM estimators.

Theorem 7. *Consider model* (2.1) *with Assumption A. Then the set of moment conditions in Griliches-Hausman, as specified in* (2.5), *is an equivalent variant of the set of moment conditions in Wansbeek, as specified in* (3.1).

4. Estimability of β

In this section we elaborate on the estimability of β under Griliches-Hausman's or Wansbeek's approach.

Under Griliches-Hausman's approach, β is estimable if and only if there exists at least one matrix P satisfying (IV.1)−(IV.3). Let ϑ be the vector of free parameters in $E[\xi_i\xi_i']$ and $vec(E[\xi_i\xi_i']) = R_\xi\vartheta$, where R_ξ is a known matrix representing the correlation structure of ξ_i. From Lemma 2, $\mathcal{P} \neq \emptyset$ if and only if

$$vec(P)' R_0 = 0 \text{ and } vec(P)'(I_T \otimes l_T) = 0 \nRightarrow vec(P)' R_\xi = 0. \tag{4.1}$$

For Wansbeek's approach, β is estimable if and only if $M_R(I_T \otimes A_T) E[x_i \otimes x_i] \neq 0$. Since $M_R(I_T \otimes A_T) E[v_i \otimes v_i] = 0$, β is estimable if and only if

$$M_R(I_T \otimes A_T) R_0 = 0 \nRightarrow M_R(I_T \otimes A_T) R_\xi = 0. \tag{4.2}$$

Condition (4.2) is equivalent to condition (4.1), as we have shown in (A.1) that $Hu = 0$ if and only if $M_R(I_T \otimes A_T)u = 0$ for any vector u. This means that when Griliches-Hausman's method fails, so does Wansbeek's method, and vice versa. Therefore, the two approaches are not only equivalent in producing asymptotically equivalent efficient GMM estimators, but also equivalent in terms of when they fail.

Obviously, (4.1) does not hold if R_ξ is a submatrix of R_0, or if any column of R_ξ is a linear combination of the columns of R_0. An easier-to-check characterization of (4.1) is given as follows. Its proof is omitted.

Lemma 8. *Let $\Pi = [\nu_1, \ldots, \nu_K]$, where $\{\nu_1, \ldots, \nu_K\}$ is a basis of $[\mathcal{C}(L)]^\perp$ given in* (2.4). *Then* (4.1) *holds if and only* $\Pi' R_\xi \neq 0$.

Note that the conclusion in Lemma 8 is invariant to the choice of Π. We illustrate the application of Lemma 8 by two examples with $T = 3$.

Suppose first that $E[v_iv_i'] = \sigma_v^2 I_3$ and $E[\xi_i\xi_i'] = (\sigma_\xi^2 - c)I_3 + cl_3l_3'$, where I_3 is the identity matrix of order 3, and l_3 is the 3 dimensional vector of ones. Then $R_0 = vec(I_3)$, $\varphi = \sigma_v^2$,

$$R_\xi = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}',$$

and $\vartheta = [\sigma_\xi^2, c]'$. Using $L = [R_0 \vdots I_3 \otimes l_3]$, we obtain the matrix Π in Lemma 8 as

$$\Pi = \begin{bmatrix} -1 & 1 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 1 & -1 \end{bmatrix}'.$$

It is easy to verify that $\Pi' R_\xi = 0$. Therefore, β is not estimable.

Suppose next that

$$E\left[v_i v_i'\right] = \begin{bmatrix} \sigma_{v1}^2 & 0 & 0 \\ 0 & \sigma_{v2}^2 & 0 \\ 0 & 0 & \sigma_{v3}^2 \end{bmatrix} \quad \text{and} \quad E\left[\xi_i \xi_i'\right] = \begin{bmatrix} \sigma_\xi^2 & \lambda & 0 \\ \lambda & \sigma_\xi^2 & \lambda \\ 0 & \lambda & \sigma_\xi^2 \end{bmatrix}.$$

Then

$$R_0 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}', \quad R_\xi = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}',$$

$\varphi = [\sigma_{v1}^2, \sigma_{v2}^2, \sigma_{v3}^2]'$, and $\vartheta = [\sigma_\xi^2, \lambda]'$. Using $L = [R_0 \vdots I_3 \otimes l_3]$, we obtain

$$\Pi = \begin{bmatrix} 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \end{bmatrix}' \quad \text{and} \quad \Pi' R_\xi = \begin{bmatrix} 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}' \neq 0.$$

Hence, β is estimable.

5. Extensions

5.1. Models with strictly exogenous variables

A covariate without measurement error is said to be *strictly exogenous* if it is uncorrelated with both the regression disturbance and the measurement error. With strictly exogenous variables in the model we can construct additional orthogonality conditions.

Consider the model

$$\begin{aligned} y_i &= \xi_i \beta + Z_i \gamma + l_T \alpha_i + \eta_i, \\ x_i &= \xi_i + v_i, \end{aligned} \qquad t = 1, \ldots, T; i = 1, \ldots, N, \tag{5.1}$$

where Z_i with dimension $T \times q$ is a set of strictly exogenous covariates. By the strict exogeneity of Z_i we have moment conditions

$$E\left[vec\left(Z_i\right) \otimes \left(A_T \epsilon_i\right)\right] = 0. \tag{5.2}$$

Now we can give the moment conditions of Griliches-Hausman and Wansbeek for the model specified by (5.1). Since (5.2) can be expressed as

$$\left(I_{qT} \otimes A_T\right) E\left[vec\left(Z_i\right) \otimes \left(y_i - x_i \beta - Z_i \gamma\right)\right] = 0,$$

an extension of (2.5) under Griliches-Hausman's approach is

$$\begin{bmatrix} H & \\ & \left(I_{qT} \otimes A_T\right) \end{bmatrix} E\left\{ \begin{bmatrix} x_i \\ vec\left(Z_i\right) \end{bmatrix} \otimes \left(y_i - x_i \beta - Z_i \gamma\right) \right\} = 0, \tag{5.3}$$

and an extension of (3.1) under Wansbeek's approach is

$$\begin{bmatrix} M_R\left(I_T \otimes A_T\right) & \\ & \left(I_{qT} \otimes A_T\right) \end{bmatrix} E\left\{ \begin{bmatrix} x_i \\ vec\left(Z_i\right) \end{bmatrix} \otimes \left(y_i - x_i\beta - Z_i\gamma\right) \right\} = 0. \qquad (5.4)$$

As acknowledged by Wansbeek himself, he did not use the full information from exogeneity to construct moment conditions. The moment conditions used by Wansbeek to represent exogeneity of Z_i are

$$E\left[Z_i'\left(A_T\epsilon_i\right)\right] = 0, \qquad (5.5)$$

which are some linear combinations of the moment conditions in (5.2). By Lemma 3, the GMM estimator based on (5.2) is more efficient than the GMM estimator based on (5.5).

Note also that both (5.3) and (5.4) contain redundant conditions. We can apply singular value decomposition (as we did in Section 3) to these equations to derive GMM estimators based on the essential variants of (5.3) and (5.4).

The equivalence between (5.3) and (5.4) follows immediately from Theorem 7, i.e., Griliches-Hausman's and Wansbeek's methods are equivalent in the case where exogenous covariates enter into the regression.

5.2. Extension to nonclassical assumptions

The independence of the true covariate ξ_i and the measurement error v_i is a "classical" measurement error assumption (Hausman (2001), Bound, Brown and Mathiowetz (2001)). Violations of this classical measurement error assumption have appeared in a number of economic applications, such as in Bound and Krueger (1991) where a negative correlation between the true covariate (true earnings) and measurement error was found to be significant. We now relax this assumption to allow the measurement error and the true covariate to be correlated. We illustrate the extension of the two GMM methods using the model specified by (2.1). The results for more general situations are similar. Let $\Delta = E\left[v_i\xi_i'\right]$ be the correlation matrix between the measurement error and the true covariate. Assume $vec\left(\Delta\right) = \Lambda_0\eta$, where Λ_0 is a known matrix and η is a vector of free parameters in Δ. Both Griliches-Hausman's and Wansbeek's methods can be easily adapted to this situation.

Consider the extension of (2.5). The conditions (IV.1)$-$(IV.3) in Section 2 should be replaced by the following new requirements for matrix P.

- $vec\left(P\right)'\left(I_T \otimes l_T\right) = 0$;
- $vec\left(P\right)'\left[R_0 \vdots \Lambda_0\right] = 0$;
- $vec\left(P\right)'\left[vec\left(E\left[\xi_i\xi_i'\right]\right)\right] \neq 0$.

The resulting moment conditions are still given by (2.5).

To generalize (3.1), we can still use (3.1) with M_R defined by $M_R = I_{T^2} - R\,(R'R)^{-}\,R'$, where $R = (I_T \otimes A_T)\,[R_0 \,\vdots\, \Lambda_0]$.

The equivalence between the two new sets of moment conditions can be established similarly as before.

If the structure of Δ is a special class within the structure class of $E\,[v_i v_i']$, i.e., columns of Λ_0 are linear combinations of the columns of R_0, then

$$vec\,(P)'\,[R_0 \,\vdots\, \Lambda_0] = 0 \Leftrightarrow vec\,(P)'\,R_0 = 0,$$

and we can also take $R = (I_T \otimes A_T)\,R_0$. This means that we can ignore the correlation between measurement error and true covariate if its structure is simpler than the serial correlation structure of measurement errors.

In theory it is also possible to generalize both methods when η_i and $(\xi_i', v_i')'$ are dependent, or when all three vectors η_i, ξ_i, and v_i are dependent. However, one needs to be cautious that more general model assumptions imply more restrictions on the corresponding transformations, and a too general model setup may lead to non-estimability of parameters.

5.3 Estimation of the individual effect α_i

Under Griliches-Hausman's and Wansbeek's approaches, the individual effect α_i under model (2.1) is treated as a nuisance effect. Under some assumptions, some effects related to the α_i's can be estimated. For example, if the α_i's are i.i.d. random effects, then a consistent estimator of $E(\alpha_i)$ (as $N \to \infty$) is $N^{-1}\sum_{i=1}^{N}\widehat{\alpha}_i$, where $\widehat{\alpha}_i = T^{-1}\sum_{t=1}^{T}(y_{it} - x_{it}\widehat{\beta}_{GMM2})$, $i = 1, \cdots, N$. For fixed effect α_i, we need some assumption since α_i is unobserved. For instance, $\widehat{\alpha}_i$ is a consistent estimator of α_i if $T \to \infty$; if $\alpha_i = \psi' t_i$ for a vector of covariate t_i observed without measurement error and an unknown parameter vector ψ, then a consistent estimator of ψ (as $N \to \infty$) is $(\sum_{i=1}^{N} t_i t_i')^{-1}\sum_{i=1}^{N} t_i\widehat{\alpha}_i$, provided that $N^{-1}\sum_{i=1}^{N} t_i t_i'$ converges to a positive definite matrix. Combining the previous arguments, we can also handle the situation where α_i follows a mixed-effect model.

6. Application to the q Theory of Investment

Economic theory on firm investment suggests that a firm's optimal investment decision is solely determined by its marginal q, the ratio of the market value of an additional unit of capital to its replacement cost. According to this so called q theory of investment, a firm's financing structure is irrelevant to its investment decision after controlling for q. Therefore, both financially unconstrained firms and financially constrained firms should be insensitive to their internal financing capabilities (such as the amount of cash flow). However, empirical studies (Fazzari, Hubbard and Petersen (1988), Barnett and Sakellaris (1998), and others) found that financial factors such as cash flow and sales enter into the investment-q regression significantly as additional regressors. Moreover, financially constrained firms were found to be more sensitive to cash flow than unconstrained firms. Erickson and Whited (2000) argued that the previous controversial and disappointing empirical results were caused by the analysts neglecting the mismeasurement of marginal q, as marginal q is unobservable and empirical proxies such as *average* q, the ratio of the market value of existing capital to its replacement cost (Hayashi (1982)), could contain considerable error. Using a GMM method involving higher order moments to adjust for measurement error in q, Erickson and Whited (2000) found that cash flow did not have effect on investment; they concluded that the empirical failure of the q theory of investment was due to inappropriate treatment of the measurement error of marginal q (neglecting either the measurement error of q or the serial correlation of the measurement error). Erickson and Whited (2000) showed that the measurement error of marginal q is serially correlated. Hence, estimation methods neglecting this correlation structure may produce misleading results.

We revisit the investment-cash flow sensitivity controversy by applying Wansbeek's GMM method described in the previous sections. Although Griliches-Hausman's approach produces asymptotically equivalent GMM estimators to Wansbeek's approach, and can be programmed using standard software package, we prefer Wansbeek's approach since it is computationally simpler. We consider the model

$$\frac{I_{i,t}}{K_{i,t-1}} = \mu_0 + \beta q_{i,t} + \gamma \frac{CF_{i,t}}{K_{i,t-1}} + \alpha_i + \eta_{it}, \tag{6.1}$$

where $I_{i,t}$ represents firm i's investment in period t, $K_{i,t-1}$ is its capital stock at the beginning of period t, $CF_{i,t}$ is its cash flow in period t, and $q_{i,t}$ is the firm's marginal q value at the beginning of period t. The observable q value is the average q and denoted by $Q_{i,t}$. We assume that

$$Q_{i,t} = q_{i,t} + \lambda_0 + v_{it}. \tag{6.2}$$

Compared with the model defined by (5.1), (6.1)−(6.2) have the additional quantities μ_0 and λ_0. These do not have any effect on our GMM estimator of β and γ since they will be swept out by the transformation we use to sweep out the unobserved individual effect α_i. We assume that the measurement error is uncorrelated with the true covariate marginal q, and that cash flow is strictly exogenous.

Table 1. Descriptive statistics for separate samples of firms

	Variable	Q1	Median	Q3	Mean	s.d.
FC firms	$\frac{I}{K}$	0.0902	0.161	0.295	0.234	0.234
(445 firms)	$\frac{CF}{K}$	0.1130	0.188	0.289	0.229	0.171
	Average q	0.9810	1.147	1.488	1.340	0.666
FUC firms	$\frac{I}{K}$	0.1510	0.238	0.417	0.346	0.346
(494 firms)	$\frac{CF}{K}$	0.4600	0.709	1.400	1.066	0.904
	Average q	1.2260	1.711	2.474	2.027	1.198

FC = Financially constrained; FUC = Financially unconstrained.

Q1 denotes the first quartile and Q3 denotes the third quartile.

Our data come from COMPUSTAT database (years 1992-1995), the same data source used by Erickson and Whited (2000). We adopt the accounting definitions of $I_{i,t}, K_{i,t}, CF_{i,t}$ and $Q_{i,t}$ as in Kaplan and Zingales (1997). We follow Lamont, Polk and Saa-Requejo (2001) to classify firms into financially constrained and financially unconstrained categories. Specifically, we construct a composite index of financial constraints called the KZ index for each firm in each year, and each year we rank all firms accordingly. A firm is classified as "financially constrained for year t" if its KZ index is among the top 33% of all firms at year t; it is classified as "financially unconstrained for year t" if its KZ index is among the bottom 33% of all firms in year t. We refer to a firm as "financially constrained" if it is classified as financially constrained for every year from 1992 to 1995. "Financially unconstrained" firms are defined similarly. We follow the conventional practice of removing outlying observations from the data analysis. Specifically, we keep observations with the following characteristics: (1) investment to capital ratio is between 0 and 3; (2) average q is between 0 and 10; (3) cash flow to capital ratio is between 0 and 5. Some descriptive statistics are given in Table 1.

We apply efficient two-step GMM estimation separately to financially constrained firms and unconstrained firms. The results are given in Table 2. The OLS estimates were based on the within transformation suggested by Wallace and Hussain (1969), and the OLS-1 estimates were obtained by applying the OLS method to the first order differences. To consider the serial correlation of the measurement errors, we calculated four GMM estimates based on (2.9) with different serial correlation structures of measurement errors: GMM-1 assumes $v_{i1}, \ldots, v_{iT}$ are i.i.d., where v_{it} is the tth component of v_i; GMM-2 assumes $\{v_{i1}, \ldots, v_{iT}\}$ is stationary with $E[v_{i1}v_{it}] = 0$ if and only if $t > 1$; GMM-3 assumes $\{v_{i1}, \ldots, v_{iT}\}$ is stationary with $E[v_{is}v_{it}] = \rho^{|s-t|}E(v_{i1}^2)$ for some $\rho \in (-1, 1)$; and GMM-4 assumes nonstationary MA(1) measurement errors, which is what GMM-2 assumes except that $\{v_{i1}, \ldots, v_{iT}\}$ is not stationary.

Table 2. Effects of marginal q and cash flow on firm's investment: comparison of various estimation methods.

		Method					
		OLS	OLS-1	GMM-1	GMM-2	GMM-3	GMM-4
FC firms	β	0.013	0.005	0.019	0.013	0.019	0.054
		(0.012)	(0.015)	(0.011)	(0.011)	(0.012)	(0.008)
	γ	0.464	0.433	0.415	0.410	0.369	0.605
		(0.047)	(0.055)	(0.049)	(0.050)	(0.053)	(0.052)
FUC firms	β	-0.017	-0.039	-0.039	-0.059	-0.056	-0.009
		(0.010)	(0.013)	(0.012)	(0.013)	(0.014)	(0.011)
	γ	0.169	0.190	0.200	0.203	0.199	0.266
		(0.011)	(0.014)	(0.021)	(0.021)	(0.021)	(0.023)

The numbers in the parenthesis are standard errors of the parameter estimates.

FC = Financially constrained; FUC = Financially unconstrained.

OLS-1: OLS applying to first order differences.

GMM-1: Two-step GMM estimator with i.i.d. measurement errors

GMM-2: Two-step GMM estimator with stationary MA(1) measurement errors.

GMM-3: Two-step GMM estimator with stationary AR(1) measurement errors.

GMM-4: Two-step GMM estimator with nonstationary MA(1) measurement errors.

For financially constrained firms, all GMM estimates for β, the coefficient of marginal q, are positive and larger than the OLS estimate, which reflects the attenuation by measurement error of the coefficient of a mismeasured covariate. The estimate of β from the first order difference OLS estimate (OLS-1) is smaller than that from the OLS, consistent with the argument in Griliches-Hausman that measurement error biases the first order difference estimator toward zero more than it biases the within estimator. GMM estimates for β are not significantly different from zero except for the GMM estimate under nonstationary MA(1) measurement errors. All GMM estimates of cash flow effect γ are significantly different from zero, indicating a robust effect of cash flow on investment for liquidity constrained firms.

For financially unconstrained firms, all GMM estimates of cash flow effect are significantly greater than zero, and all estimates for the coefficient of q are negative (except for the nonstationary MA(1) measurement error situation, all are significant.), suggesting that cash flow might be playing a more important role in determining corporate investment than marginal q.

Compare the results for the two groups of firms, we find that the GMM estimates of cash flow effect for financially constrained firms are approximately twice that for financially unconstrained firms, which indicates that financially constrained firms respond more sensitively to cash flow than financially unconstrained firms.

Therefore, our empirical study suggests that: (1) cash flow is a significant factor for firms' investment, regardless of whether the firm is financially constrained or not; (2) financially constrained firms respond more sensitively to cash flow than financially unconstrained firms; and (3) adjusting for measurement error in marginal q might still lead to results unexplainable by the q theory of investment. These results confirm some of the previous findings (Fazzari, Hubbard and Petersen (1988), etc.) that internal funds play an important role in corporate investment, even after controlling for Tobin's marginal q, and that there is substantial difference between the two types of firms in the effect of cash flow on investment. Accordingly, our results provide different explantation of the effect of measurement error than Erickson and Whited (2000).

7. Concluding Remarks

In this paper we investigated two previously proposed approaches to the measurement error problem in panel data. The main results of this paper have been to (i) provide an easy-to-compute GMM estimator based on Griliches-Hausman's idea of using instrumental variables; (ii) provide a weighted instrumental variable estimator that is as efficient as the GMM estimator in (i); (iii) remove the redundancy of the moment conditions constructed by Wansbeek and show that Wansbeek's GMM estimator is optimal; (iv) show that Griliches-Hausman's and Wansbeek's methods are asymptotically equivalent and also computationally similar so that applied users can choose one that is easy to implement; and (v) discuss the estimability of the parameters and some extensions to both methods.

A key advantage of Griliches-Hausman's and Wansbeek's methods and our extensions is that measurement error is handled without assuming a known measurement error covariance matrix or requiring additional validation/replication data, because we can use panel data as "partial" replicates under certain assumptions on the serial correlation structure of the measurement error (e.g., condition (4.1)). As seen in the application, different assumptions on the serial correlation structure of the measurement error lead to different moment conditions and different GMM estimators. Empirical studies may be carried out to choose between different measurement error correlation structures.

The methods in this paper work for balanced, linear, static panel data. An interesting future research topic is to investigate whether these methods can be applied to unbalanced panel data, nonlinear panel data, and dynamic panel data.

Acknowledgements

The authors thank two referees and an associate editor for their helpful comments and suggestions.

Appendix: Proofs of results

Proof of Lemma 1. Let $d = \dim(V_0)$, $\widetilde{d} = \dim(V_1)$, with $\widetilde{d} < d$. Let $v_1, \ldots, v_{\widetilde{d}}$ be a basis of V_1. Enlarge $\{v_1, \ldots, v_{\widetilde{d}}\}$ to a basis $v_1, \ldots, v_{\widetilde{d}}, v_{\widetilde{d}+1}, \ldots, v_d$ of V_0. Then $v_{\widetilde{d}+1}, \ldots, v_d \in S$. Now consider vectors $v_1 + v_{\widetilde{d}+1}, \ldots, v_{\widetilde{d}} + v_{\widetilde{d}+1}, v_{\widetilde{d}+1}, \ldots, v_d$. It is easy to show that they are linearly independent. All of them are in S, otherwise there is a contradiction. Thus we have constructed a set of d linearly independent vectors in S. Since $S \subseteq V_0$, the maximum number of linearly independent vectors in S should not exceed d, hence the claim is proved. If $s_1, \ldots, s_d$ are linearly independent vectors in S, then they are a basis of V_0, consequently every vector in S can be expressed as a linear combination of them.

Proof of Lemma 2. For (IV.1), it is easy to verify that $P' l_T = 0$ if and only if $vec(P)'(I_T \otimes l_T) = 0$. For (IV.2), using the properties of the Kronecker product, one can show that $E[x_i' P' \varepsilon_i] = -vec(P)' R_0 \varphi \beta$. Hence $E[x_i' P' \varepsilon_i] = 0 \Leftrightarrow vec(P)' R_0 = 0$. Similarly we can show (IV.3).

Proof of Lemma 3. (i) Since $Q_1 \zeta = 0 \Leftrightarrow Q_1' Q_1 \zeta = 0$ and $Q_2 \zeta = 0 \Leftrightarrow Q_2' Q_2 \zeta = 0$, we have that $Q_1' Q_1 \zeta = 0 \Leftrightarrow Q_2' Q_2 \zeta = 0$. Thus if ζ is an eigenvector of $Q_1' Q_1$ associated with the eigenvalue 0, it is an eigenvector of $Q_2' Q_2$ associated with the eigenvalue 0. Hence, the zero-eigenspace of $Q_1' Q_1$ is the same as the zero-eigenspace of $Q_2' Q_2$. Since $Q_1' Q_1$ and $Q_2' Q_2$ have the same dimension, we know that the nonzero-eigenspace of $Q_1' Q_1$ is the same as the nonzero-eigenspace of $Q_2' Q_2$. Let $Q_1 = U_1 \Lambda_1 V_1'$ and $Q_2 = U_2 \Lambda_2 V_2'$ be the singular value decompositions of Q_1 and Q_2, respectively. Then the rows of V_1' form a basis for the nonzero-eigenspace of $Q_1' Q_1$, and the rows of V_2' form a basis for the nonzero-eigenspace of $Q_2' Q_2$. Hence there exists a nonsingular matrix A such that $V_2' = A V_1'$. This means that the set of moment conditions $Q_1 E[g(\beta)] = 0$ is an equivalent variant of the set of moment conditions $Q_2 E[g(\beta)] = 0$. Applying the arguments in part (ii) below we can see that the two GMM estimators have the same asymptotic variance.

(ii) By a similar argument as in (i), there exists a $r_2 \times r_1$ matrix A with rank r_2 such that $V_2' = A V_1'$. While $Q_1 E[g(\beta)] = 0 \Leftrightarrow V_1' E[g(\beta)] = 0$, $Q_2 E[g(\beta)] = 0 \Leftrightarrow A V_1' E[g(\beta)] = 0$. Assuming there is no redundancy in $g(\beta)$, $\Omega = var(g(\beta)) > 0$. The asymptotic variance of the efficient GMM estimator based on $V_1' E[g(\beta)] = 0$ is $\left(C' V_1 (V_1' \Omega V_1)^{-1} V_1' C\right)^{-1}$, while the asymptotic variance of the efficient GMM estimator based on $A V_1' E[g(\beta)] = 0$ is $\left(C' V_1 A' (A V_1' \Omega V_1 A')^{-1} A V_1' C\right)^{-1}$. By the theory of linear models,

$$\left(C' V_1 \left(V_1' \Omega V_1\right)^{-1} V_1' C\right)^{-1} \le \left(C' V_1 A' \left(A V_1' \Omega V_1 A'\right)^{-1} A V_1' C\right)^{-1}.$$

Hence the GMM estimator based on $Q_1E[g(\beta)] = 0$ is at least as efficient as the GMM estimator based on $Q_2E[g(\beta)] = 0$.

Proof of Theorem 5. (i) For $k = 1, \ldots, K$, let $A_k(x_i, y_i) = x_i'P_k'y_i$, $B_k(x_i, y_i) = x_i'P_k'x_i$, $\overline{A_k} = (1/N)\sum_{i=1}^{N} A_k(x_i, y_i)$,$\overline{B_k} = (1/N)\sum_{i=1}^{N} B_k(x_i, y_i)$. Then $\widehat{\beta}_{P_k} = \overline{A_k}/\overline{B_k}$. If $e_k^i = A_k(x_i, y_i) - B_k(x_i, y_i)\beta$, $\overline{e_k} = (1/N)\sum_{i=1}^{N} e_k^i = \overline{A_k} - \overline{B_k}\beta$, $e^i = \left(e_1^i, \cdot, e_K^i\right)'$ and $\overline{e} = \sum_{i=1}^{N} e^i$, then $E\left[e_k^i\right] = 0$. Let $\Gamma = var\left(e^i\right)$. By the Central Limit Theorem, $\sqrt{N}\overline{e} \xrightarrow{d} \mathcal{N}(0, \Gamma)$. Since $\widehat{\beta}_{P_k} - \beta = \overline{A_k}/\overline{B_k} - \beta = \overline{e_k}/\overline{B_k}$,

$$\begin{aligned}
\sqrt{N}\left(\widehat{\beta}(\lambda) - \beta\right) &= \lambda_1\sqrt{N}\left(\widehat{\beta}_{P_1} - \beta\right) + \cdots + \lambda_K\sqrt{N}\left(\widehat{\beta}_{P_K} - \beta\right) \\
&= \left(\frac{\lambda_1}{\overline{B_1}}, \ldots, \frac{\lambda_K}{\overline{B_K}}\right)\sqrt{N}\overline{e} \\
&\xrightarrow{d} \mathcal{N}(0, V(\lambda))
\end{aligned}$$

with

$$V(\lambda) = \left(\frac{\lambda_1}{E(B_1)}, \ldots, \frac{\lambda_K}{E(B_K)}\right)\Gamma\left(\frac{\lambda_1}{E(B_1)}, \ldots, \frac{\lambda_K}{E(B_K)}\right)'.$$

(ii) Consider another approach: pool the K equations and use GMM to solve for an efficient estimator. Denote the efficient GMM estimator by $\widehat{\beta}_{GMM}$. Its asymptotic variance is

$$V^* = \left[(E(B_1), \ldots, E(B_K))\Gamma^{-1}(E(B_1), \ldots, E(B_K))'\right]^{-1}.$$

Let $w = \Gamma^{1/2}\left(\frac{\lambda_1}{E(B_1)}, \ldots, \frac{\lambda_K}{E(B_K)}\right)'$, and $v = \Gamma^{-1/2}(E(B_1), \cdots, E(B_K))'$. Then by Cauchy's inequality,

$$V(\lambda)/V^* = w'wv'v \geq \left(w'v\right)^2 = 1,$$

where the equality holds if and only if $w = cv$ for some scalar c, i.e.,

$$\Gamma^{1/2}\left(\frac{\lambda_1}{E(B_1)}, \ldots, \frac{\lambda_K}{E(B_K)}\right)' = c\Gamma^{-1/2}(E(B_1), \ldots, E(B_K))',$$

which is

$$\left(\frac{\lambda_1}{E(B_1)}, \ldots, \frac{\lambda_K}{E(B_K)}\right)' = c\Gamma^{-1}(E(B_1), \ldots, E(B_K))'.$$

Now combining with the fact that $\sum_{k=1}^{K}\lambda_k = 1$, we can find $c = \left(B'\Gamma^{-1}B\right)^{-1}$. Hence the optimal weight is $\lambda^* = \left(B'\Gamma^{-1}B\right)^{-1} diag(B)\Gamma^{-1}B$, since V^* is the

lower bound of the asymptotic variance of any consistent estimator based on the set of K equations.
(iii) It is easy to verify that

$$\begin{aligned}\hat{\beta}(\widehat{\lambda}) &= \left(\overline{B}'\widehat{W}\overline{B}\right)^{-1}\overline{B}'\widehat{W}' diag\left(\overline{B}\right)\left[\frac{\overline{A_1}}{\overline{\overline{B_1}}},\ldots,\frac{\overline{A_K}}{\overline{\overline{B_K}}}\right]' \\ &= \left(\overline{B}'\widehat{W}\overline{B}\right)^{-1}\overline{B}'\widehat{W}'\overline{A} \\ &= \widehat{\beta}_{GMM2}.\end{aligned}$$

Proof of Theorem 6. It is easy to verify that (D_1, M_{D_1}) satisfies conditions (1) and (2) of part (i). We first show that (D_1, M_{D_1}) is optimal. By Lemma 3(ii), we need only check that $M_{D_1}(I_T \otimes D_1)\, u = 0$ implies $M(I_T \otimes A)\, u = 0$ for any $(A, M) \in \mathcal{S}$ and for any u. Observe that the rows of D_1 form a basis of the space $\left\{x \in \mathbb{R}^T : x' l_T = 0\right\}$. This means that each row of A is a linear combination of the rows of D_1, hence there exists a matrix C such that $A = CD_1$. Now observe that $M_{D_1}(I_T \otimes D_1)\, u = 0$ implies that there exists some fixed vector γ such that $(I_T \otimes D_1)\, u = (I_T \otimes D_1)\, R_0\gamma$. By some algebraic manipulation, we have that $M(I_T \otimes A)\, u = M(I_T \otimes A)\, R_0\gamma = 0$. Hence (D_1, M_{D_1}) is optimal. Now we show that any two pairs (A_1, M_1) and (A_2, M_2) in $\mathcal{S}$ generate moment conditions that are equivalent variants of each other, hence any $(A, M) \in \mathcal{S}$ satisfying the condition (1) and (2) of part (i) is optimal. First we observe that $A_1 y = 0 \Leftrightarrow A_2 y = 0 \Leftrightarrow y = cl_T$, where c is some scalar. We establish (C) for any two conformable matrices P and Q satisfying $rank(PQ) = rank(Q)$, we have that $PQu = 0$ implies $Qu = 0$ for any vector u. To see this, note that $\ker(PQ) = [\mathcal{C}(Q'P')]^{\perp}$ and $\ker(Q) = [\mathcal{C}(Q')]^{\perp}$. Now $rank(PQ) = rank(Q)$ implies $rank(Q'P') = rank(Q')$. Since $\mathcal{C}(Q'P') \subseteq \mathcal{C}(Q')$, we have that $\mathcal{C}(Q'P')$ $= \mathcal{C}(Q')$, hence $\ker(PQ) = \ker(Q)$. This means that $PQu = 0$ if and only if $Qu = 0$ for any vector u. Now let $(A, M) \in \mathcal{S}$ satisfy (1) and (2) of part (i). By definition $MR_A = 0$, hence there exists a matrix U such that $M = U\left[I - R_A\left(R_A' R_A\right)^{-} R_A'\right]$. Since M has maximum possible rank, $rank\,(M) = rank\left(I - R_A\left(R_A' R_A\right)^{-} R_A'\right)$. Hence by (C),

$$\begin{aligned}Mu = 0 &\Leftrightarrow U\left[I - R_A\left(R_A' R_A\right)^{-} R_A'\right] u = 0 \\ &\Leftrightarrow \left[I - R_A\left(R_A' R_A\right)^{-} R_A'\right] u = 0 \\ &\Leftrightarrow u \in \mathcal{C}\left(R_A\right).\end{aligned}$$

Therefore $M_1(I_T \otimes A_1)\, x = 0$ if and only if $(I_T \otimes A_1)\, x \in \mathcal{C}\left((I_T \otimes A_1)\, R_0\right)$, i.e., there exist some vector γ such that $(I_T \otimes A_1)\, x = (I_T \otimes A_1)\, R_0\gamma$. This is again equivalent to $A_1\left(X-\widetilde{\Omega}\right)=0$, where $vec\left(\widetilde{\Omega}\right)=R_0\gamma$ and $vec\,(X)=x$, hence equivalent to $A_2\left(X-\widetilde{\Omega}\right)=0$, and eventually equivalent to $M_2\left(I_T\otimes A_2\right)x=0$.

Proof of Theorem 7. By virtue of Lemma 3(i), we need only check that

$$Hu=0 \text{ if and only if } M_R\left(I_T\otimes A_T\right)u=0, \text{for any vector } u. \tag{A.1}$$

Assume first that $M_R\left(I_T\otimes A_T\right)u=0$. Since M_R is the projection matrix for the column space of R, there exits a vector $\widetilde{\gamma}$ such that $\left(I_T\otimes A_T\right)u=R\widetilde{\gamma}=\left(I_T\otimes A_T\right)R_0\widetilde{\gamma}$. Let $\widetilde{\Omega}$ and U be matrices such that $vec\left(\widetilde{\Omega}\right)=R_0\widetilde{\gamma}$, $vec\,(U)=u$. Then by the properties of the Kronecker product we have $vec\,(A_TU)=vec\left(A_T\widetilde{\Omega}\right)$. This implies $A_T\left(U-\widetilde{\Omega}\right)=0$. Hence there exist scalars $c_1,\ldots$ c_T such that $U=\widetilde{\Omega}+(c_1,\ldots,c_T)\otimes l_T$. Then for any $k=1,\ldots,K$,

$$\begin{aligned}
vec\,(P_k)'\,vec\,(U) &= vec\,(P_k)'\,vec\left(\widetilde{\Omega}\right)+vec\,(P_k)'\,vec\left((c_1,\cdots,c_T)\otimes l_T\right)\\
&= vec\,(P_k)'\,vec\left(\widetilde{\Omega}\right)\\
&= vec\,(P_k)'\,R_0\gamma\\
&= 0.
\end{aligned}$$

Hence, $M_R\left(I_T\otimes A_T\right)u=0$ implies $Hu=0$.

Now suppose $Hu=0$. Then there exist a vector $\theta=[\theta_1\ \theta_2]'$ such that

$$u=L\theta=[R_0\ I_T\otimes l_T]\,[\theta_1\ \theta_2]'=R_0\theta_1+\left(I_T\otimes l_T\right)\theta_2.$$

Hence

$$\begin{aligned}
M_R\left(I_T\otimes A_T\right)u &= M_R\left(I_T\otimes A_T\right)\left[R_0\theta_1+\left(I_T\otimes l_T\right)\theta_2\right]\\
&= M_R\left(I_T\otimes A_T\right)R_0\theta_1+M_R\left(I_T\otimes A_T\right)\left(I_T\otimes l_T\right)\theta_2\\
&= 0+M_R\left(I_T\otimes A_Tl_T\right)\theta_2\\
&= 0.
\end{aligned}$$

This proves (A.1).

References

Abadir, K. and Magnus, J. (2005). *Matrix Algebra*. Cambridge University Press, New York.

Barnett, S. and Sakellaris, P. (1998). Nonlinear response of firm investment to Q: Testing a model of convex and non-convex adjustment costs. *J. Monetary Economics* **42**, 261-288.

Biørn, E. (2000). Panel data with measurement errors: Instrumental variables and GMM procedures combining levels and differences. *Econom. Rev.* **19**, 391-424.

Biørn, E. and Klette, T. (1998). Panel data with errors-in-variables: Essential and redundant orthogonality conditions in GMM-estimation. *Econom. Lett.* **59**, 275-282.

Bound, J., Brown, C. and Mathiowetz, N. (2001). Measurement error in survey data. *Handbook of Econometrics* **5**, 3705-3843.

Bound, J. and Krueger, A. (1991). The extent of measurement error in longitudinal earnings data: Do two wrongs make a right? *J. Labor Economics* **9**, 1-24.

Carroll, R. J., Ruppert, D., Stefanski, L. A. and Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective.* Chapman & Hall/CRC, Boca Raton, FL.

Cui, H., Ng, K. and Zhu, L. (2004). Estimation in mixed effects model with errors in variables. *J. Multivariate Anal.* **91**, 53-73.

Erickson, T. and Whited, T. (2000). Measurement error and the relationship between investment and Q. *J. Political Economy* **108**, 1027-1057.

Fazzari, S., Hubbard, G. R. and Petersen, B. (1988). Financing constraints and corporate investment. *Brookings Paper on Economic Activity* **1**, 141-195.

Fuller, W. (1987). *Measurement Error Models.* John Wiley, New York.

Griliches, Z. and Hausman, J. A. (1986). Errors in variables in panel data. *J. Econometrics* **32**, 93-118.

Hall, A. (2005). *Generalized Method of Moments.* Oxford University Press, New York.

Hansen, L. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* **50**, 1029-1054.

Hausman, J. (2001). Mismeasured variables in econometric analysis: Problems from the right and problems from the left. *J. Economic Perspectives* **15**, 57-67.

Hayashi, F. (1982). Tobin's marginal Q and average Q: A neoclassical interpretation. *Econometrica* **50**, 213-224.

Kaplan, S. and Zingales, L. (1997). Do investment-cash flow sensitivities provide useful measures of financing constraints? *Quarterly J. Economics* **112**, 168-216.

Lamont, O., Polk, C. and Saa-Requejo, J. (2001). Financial constraints and stock returns. *Rev. Financial Studies* **14**, 529-554.

Wallace, T. D. and Hussain, A. (1969). The use of error components models in combining cross-section and time-series data. *Econometrica* **37**, 55-72.

Wansbeek, T. J. (2001). GMM estimation in panel data models with measurement error. *J. Econometrics* **104**, 259-268.

White, H. (1982). Instrumental variables regression with independent observations. *Econometrica* **50**, 483-499.

Zhong, X., Fung, W. and Wei, B. (2002). Estimation in linear models with random effects and errors-in-variables. *Ann. Inst. Statist. Math.* **54**, 595-606.

Department of Statistics, School of Management, Fudan University, Shanghai, China.

E-mail: zhiguo_xiao@fudan.edu.cn

Department of Statistics, University of Wisconsin-Madison, Madison, WI 53706, U.S.A.

E-mail: shao@stat.wisc.edu

Department of Population Health Sciences, University of Wisconsin-Madison, Madison, WI 53726, U.S.A.

E-mail: mpalta@wisc.edu

(Received October 2008; accepted June 2009)