Treating Measurement Error in Tobin's q

Author(s): Timothy Erickson and Toni M. Whited

Source: *The Review of Financial Studies*, April 2012, Vol. 25, No. 4 (April 2012), pp. 1286-1329

Published by: Oxford University Press. Sponsor: The Society for Financial Studies.

Stable URL: https://www.jstor.org/stable/41407861

Treating Measurement Error in Tobin's q

Timothy Erickson
Bureau of Labor Statistics

Toni M. Whited
University of Rochester

We compare the ability of three measurement error remedies to deliver unbiased estimates of coefficients in investment regressions. We examine high-order moment estimators, dynamic panel estimators, and simple instrumental variables estimators that use lagged mismeasured regressors as instruments. We show that recent investigations of this question are largely uninformative. We find that all estimators can perform well under correct specification, all can be biased under misspecification, and misspecification is easiest to detect in the case of high-order moment estimators. We develop and demonstrate a minimum distance technique that extends the high-order moment estimators to be used on unbalanced panel data. (*JEL* C15, C26, E22, G31)

Tobin's q is arguably the most common regressor in corporate finance. However, in its usual role as a proxy for unobservable investment opportunities, it likely contains a great deal of measurement error because of a conceptual gap between true investment opportunities and observable measures of Tobin's q. We compare three estimators that remedy the ensuing biased regression coefficients: instrumental variables (IV), the dynamic panel estimators from Arellano and Bond (1991, AB hereafter), and the high-order moment estimators in Erickson and Whited (2000, 2002, EW hereafter). The first two employ lagged mismeasured regressors as instruments, but the EW estimators do not employ conventional instruments. Instead, they obtain identification from the third- and higher-order moments of the regression variables. As such, identification requires nonnormally distributed mismeasured regressors. Although this assumption limits the applicability of these estimators, this assumption is ideally suited to Tobin's q because this variable, like many other valuation ratios in finance, has a highly skewed distribution.

We thank Stephen Barr, Yunjeen Kim, Egor Matveyev, and Thu Vo for excellent research assistance. For helpful comments and discussions we thank the editor, Geert Bekaert, two anonymous referees, Tor Bakke, Sreedhar Bharath, Jason Chen, Harry DeAngelo, Jed Frees, Tom George, Arthur Korteweg, Chris Parsons, Leigh Riddick, Jerry Warner, Missaka Warusawitharana, and Mark Westerfield. We thank Heitor Almeida, Murillo Campello, and Antonio Galvao for providing us with their Monte Carlo programs. All of the analysis, views, and conclusions expressed in this article are those of the authors and do not necessarily reflect the views or policies of the Bureau of Labor Statistics. Send correspondence to Toni M. Whited, William E. Simon Graduate School of Business Administration, University of Rochester, Rochester, NY 14627; telephone: (585) 275-3916. E-mail: toni.whited@simon.rochester.edu.

doi:10.1093/rfs/hhr120 Advance Access publication November 16, 2011

One goal of this article is to evaluate the finite sample properties of these three estimators using Monte Carlo experiments. We find that when the assumptions of all three estimators are satisfied, the EW estimators often outperform the IV and AB estimators in terms of bias and dispersion. The EW estimators sometimes underperform on data that have undergone a within transformation to eliminate fixed effects. They always underperform when a mismeasured regressor has a nearly symmetrical distribution, but they perform well with modestly skewed distributions.

We also perform Monte Carlo experiments to examine the performance of the EW, IV, and AB estimators under the types of misspecification that might afflict investment regressions—our application of interest. We find striking evidence that IV and AB can produce the same biased results as does OLS if the measurement error is serially correlated. Yet it is almost impossible to detect this misspecification using conventional specification tests. The EW estimators, in contrast, are unaffected by a serially correlated measurement error. We find that other misspecifications, in particular fixed effects and heteroskedasticity, often produce bias in the EW estimators but not in the IV and AB estimators. However, this misspecification is often easier to detect in the case of the EW estimators. Finally, we find that the *t*-statistics on slope coefficients from the EW estimators over reject null hypotheses. However, we also show that using bootstrapped critical values for these statistics ameliorates but does not eliminate this problem.

This comparison is of interest for two reasons. First, these techniques have produced different results on the same problem. Since Fazzari, Hubbard, and Petersen (1988), researchers have tested for the presence of finance constraints via regressions of investment on cash flow and a mismeasured proxy for q, with the size of the cash flow coefficient measuring the degree of the constraints. Cash flow coefficient estimates from the EW estimators are typically near zero, whereas IV estimates based on lagged instruments are often positive and significant (Almeida and Campello 2007). Second, all three estimators are purely statistical remedies, and economic theory sometimes provides little guidance in choosing one of them. It is therefore useful to understand the statistical properties of these estimators in situations resembling real research questions.

A recent paper by Almeida, Campello, and Galvao (2010, ACG hereafter) also conducts a serious comparison between the EW estimators and traditional instrumental variables estimators. However, our results differ sharply from those in ACG. They argue that the EW estimators are biased and inefficient in finite samples, even though they are asymptotically consistent and efficient (Erickson and Whited 2002), and even when all estimator assumptions are satisfied. They conclude that the AB and IV estimators perform better. Therefore, the second goal of this article is to understand the differences between ACG's results and ours.

ACG reach different conclusions primarily because they inaccurately compute the overidentified EW estimators, which require numerical minimization of an objective function. Established practice is to check the resulting estimates to ensure that they correspond to a global, rather than a local, minimum (e.g., Heckman and Singer 1984). However, most of the poor Monte Carlo performance of the EW estimators in ACG can be traced to a tiny number of extreme estimates, each of which is from a local minimum. This issue affects performance because a Monte Carlo evaluation of an estimator's bias is itself a statistical estimate, computed by taking the difference between the true coefficient and the average of the coefficient estimates over thousands of simulated data sets. Averaging in these few extreme estimates yields a spurious evaluation of bias.

This starting value issue implies that different several-thousand-trial runs of many of the ACG Monte Carlos yield estimates of coefficient bias that differ both in magnitude and in sign, and the range of bias estimates can be many times the value of the true regression coefficient. Estimates of bias from a Monte Carlo should be nearly identical when one runs the Monte Carlo repeatedly, allowing the simulated samples to consist of different random draws in each run. However, in several of ACG's Monte Carlos, different sets of random draws in different runs yield different collections of extreme estimates. Different runs then yield different evaluations of estimator bias.

We remedy this problem by using just two different starting values for the overidentified EW estimators and then choosing the one that produces the lowest GMM objective function. We thereby eliminate local minima, and we find bias estimates of near zero from the ACG Monte Carlos that satisfy all of the estimator assumptions.

Our third goal is to extend the EW estimators by developing two new econometric contributions. The first is a minimum distance technique that extends the EW estimators to be used on unbalanced panel data. The second is a new identification diagnostic test. Both improve the ease or confidence with which the EW estimators can be implemented in practice.[1]

To demonstrate these new procedures and to illustrate the practical issues in implementing the EW estimators, we analyze a standard investment data set from Compustat. We examine the two most widely used specifications of investment regressions. In the first, used for example by Fazzari, Hubbard, and Petersen (1988), all variables are deflated by a measure of the capital stock, and the proxy for Tobin's q attempts to isolate investment opportunities in fixed assets. In the second, used for example by Rauh (2006) and Chava and Roberts (2008), all variables are deflated by a measure of total assets, and the proxy for Tobin's q captures investment opportunities in all assets. We find economically plausible parameter estimates. Although we do find limited evidence of positive significant cash flow coefficients using the EW estimators, those significant estimates are much smaller in magnitude than the corresponding OLS estimates.

---

1 We emphasize that neither contribution addresses any discussion in ACG.

The rest of the article is organized as follows. Section 1 reviews the assumptions and intuition for the estimators we consider. Section 2 presents our Monte Carlo experiments, and Section 3 discusses ACG's Monte Carlo experiments. Section 4 presents our data analysis and provides practical advice, and Section 5 concludes. The Appendix derives our econometric contributions.

1. Estimator Assumptions and Intuition

This section details the assumptions that the EW, IV, and AB estimators have in common and those that they do not. It also provides a brief exposition of these estimators and explains the statistical and economic intuition behind them.

1.1 Assumptions

Table 1 provides an informal schematic layout of the key assumptions that these estimators require. Equations (1)–(4) in Panel A of Table 1 demonstrate that all three estimators have in common the basic structure of the classical linear errors-in-variables model. In our application, y_i is the ratio of investment to the capital stock, χ_i is the true q of firm i (marginal q from classic q theory), x_i is an estimate of its true q, and z_i is a row vector of perfectly measured regressors that contains 1 and the ratio of cash flow to the capital stock. For the IV and AB estimators, all of these variables also have a time subscript, $t = 1, \ldots, T$, because these two estimators require panel data. The EW estimators are, in contrast, cross-sectional estimators. We discuss in the Appendix how coefficient estimates from different cross-sections in a panel

Table 1
Measurement error estimator assumptions

	Assumption Type	Erickson and Whited	IV and Arellano and Bond
Panel A			
Assumptions in Common	Model Structure	$y_i = z_i\alpha + \chi_i\beta + u_i$ (1) $x_i = \gamma + \chi_i + \varepsilon_i$ (2)	$y_{it} = z_{it}\alpha + \chi_{it}\beta + u_{it}$ (3) $x_{it} = \gamma + \chi_{it} + \varepsilon_{it}$ (4)
		$(\varepsilon_i, u_i, z_i, \chi_i)$ are *i.i.d.*	$(\varepsilon_{it}, u_{it}, z_{it}, \chi_{it})$ are independent across i.
Panel B			
Differing Assumptions	Error-Regressor Dependence	(u_i, ε_i) are independent of each other and of (z_i, χ_i).	$(u_{it}, \varepsilon_{it})$ are uncorrelated with each other and with (z_{it}, χ_{it}).
	Distributional Assumptions	χ_i is not normally distributed.	Free distributions.
	Serial Correlation Assumptions	Free correlations.	$E(\varepsilon_{it}\varepsilon_{is}) = 0$, $s < t + j,\ j > 0$

can be combined optimally via a minimum distance procedure to produce a single coefficient estimate for the entire panel. This procedure allows for arbitrary serial correlation of the regression and measurement errors.

Although the three estimators share this basic structure, they differ in their identifying assumptions. Panel B shows that the IV and AB estimators require standard assumptions about a zero error-regressor correlation. The EW estimators require the stronger assumption of independence, which is common in nonlinear estimators (e.g., Schennach 2004). The assumption that u_i is independent of (uncorrelated with) (χ_i, z_i) can be violated if the regression error in (1) or (3) contains a fixed effect that is correlated with the regressors. In the case of the EW estimators, this problem can be dealt with in panel data by removing the firm-specific means from the observable regression variables (a within transformation). For the AB and IV estimators, fixed effects need to be removed by first differencing. In both cases, transforming the data allows the error-regressor assumptions to hold.

The next two sets of assumptions sharply differentiate the two types of estimators. Identification of the EW estimators requires that χ_i not be normally distributed. The IV and AB estimators require no such distributional assumptions. In contrast, the IV and AB estimators require that the degree of serial correlation in the measurement error be limited. This assumption identifies the model by allowing the use of $j+1$ times lagged values of x_{it} to be used as instruments for the first difference of x_{it}. In contrast, the EW estimators impose no restrictions on the serial correlation of either the regression error or measurement error.

1.2 Development

We first review the IV and AB estimators. For both, one takes first differences of the regression to eliminate a possible fixed effect. Next, one uses at least twice-lagged values of Tobin's q as an instrument for current Tobin's q. The lag length must be longer than the degree of the moving-average autocorrelation of the measurement error. Therefore, measurement error that follows an autoregressive process invalidates identification of these estimators. The simple IV estimator pools all of the cross-sections in the panel in one estimation. The AB estimator differs from the simple IV estimator in that it treats each cross-section in the panel as a separate regression. Therefore, the number of moment conditions used to define a GMM system grows as the sample period grows. The AB estimator then pools the cross-sectional estimates using the GMM weighting matrix.

The EW estimators require more explanation. To derive tractable moment conditions for GMM estimation, we partial out of Equations (1) and (2) the vector of perfectly measured variables, z_i; that is, we reexpress (1) and (2) in terms of the residuals from the population linear projections of (y_i, x_i, χ_i) on

z_i. Let $(\dot{y}_i, \dot{x}_i, \dot{\chi}_i)$ be these residuals. Then (1) and (2) can be written as

$$\dot{y}_i = \beta \dot{\chi}_i + u_i \tag{5}$$

$$\dot{x}_i = \dot{\chi}_i + \varepsilon_i. \tag{6}$$

If we square (5), multiply the result by (6), and take expectations of both sides, we obtain

$$E\left(\dot{y}_i^2 \dot{x}_i\right) = \beta^2 E\left(\dot{\chi}_i^3\right). \tag{7}$$

Analogously, if we square (6), multiply the result by (5), and take expectations, we obtain

$$E\left(\dot{y}_i \dot{x}_i^2\right) = \beta E\left(\dot{\chi}_i^3\right). \tag{8}$$

If $\beta \neq 0$ and $E\left(\dot{\chi}_i^3\right) \neq 0$, dividing (7) by (8) produces a consistent estimator for β:

$$\beta = E\left(\dot{y}_i^2 \dot{x}_i\right) / E\left(\dot{y}_i \dot{x}_i^2\right). \tag{9}$$

An estimator can be derived from (9) by replacing the population moments by sample moments. Doing so produces a third-order moment estimator. Inspection of (9) shows that the assumptions $\beta \neq 0$ and $E\left(\dot{\chi}_i^3\right) \neq 0$ are necessary for identification. Erickson and Whited (2000, 2002) show that these assumptions can be tested via the null hypothesis that $E\left(\dot{y}_i^2 \dot{x}_i\right) = 0$ and $E\left(\dot{y}_i \dot{x}_i^2\right) = 0$. We refer to this test hereafter as an identification diagnostic test, and we explore below its finite sample properties. We also examine the properties of a related test outlined in the Appendix.

The innovation in Erickson and Whited (2000, 2002) consists of moving beyond the estimator given by (9), which is from Geary (1942). They combine the information in the moment conditions (7) and (8) with information in moment conditions of orders four and higher via GMM to obtain more efficient overidentified estimators for β. We refer to the third- through fifth-order moment estimators as GEARY, GMM4, and GMM5, and we refer to this group collectively as EW estimators.

It is possible to estimate many interesting quantities besides β because the EW estimators also deliver consistent estimates of the moments $E\left(\dot{\chi}_i^2\right)$, $E\left(\varepsilon_i^2\right)$, $E\left(u_i^2\right)$, $E\left(\dot{\chi}_i^3\right)$, etc. For example, one can estimate the coefficients of determination (R^2) for (1) and (2), which we denote as ρ^2 and τ^2. The latter is an index of measurement quality that ranges between 0 and 1. Because econometric measurement error remedies do not allow one to recover the actual individual measurement errors, a gauge of their severity is useful. Finally, one can also estimate the coefficient vector α, which can be recovered by the identity

$$\alpha \equiv \mu_y - \beta \mu_x, \tag{10}$$

in which (μ_y, μ_x) are the vectors of coefficients in the population projections of (y_i, x_i) on z_i. This identity is a straightforward byproduct of partialing and depends in no way on high-order moments.

The identity (10) reveals why measurement error in Tobin's q biases the cash-flow coefficient. Let $(\alpha_1, \mu_{1y}, \mu_{1x})$ denote the elements of (α, μ_y, μ_x) corresponding to cash flow, so that $\alpha_1 \equiv \mu_{1y} - \beta\mu_{1x}$ relates β to the cash-flow coefficient. Because μ_{1y} and μ_{1x} can be consistently estimated by OLS even in our errors-in-variables model, we need only look at the term $\beta\mu_{1x}$ to see the effects of measurement error on the estimate of α_1. Measurement error biases the OLS estimate of β downward. If $\mu_{1x} = 0$, then this downward bias does not affect the cash flow coefficient, α_1. However, because Tobin's q and cash flow are positively correlated, $\mu_{1x} > 0$. Also, because the variance of Tobin's q is much greater than the variance of cash flow, μ_{1x} is large. Therefore, a small downward bias in β causes a large upward bias in the OLS estimate of the cash flow coefficient.

Finally, these estimators can be applied in principle only to *i.i.d.* samples. For panel data, which usually exhibit time dependence, we suggest estimating the model on each cross-section of data, and then pooling the cross-sectional estimates. One option for pooling is the method in Fama and MacBeth (1973). However, this technique does not pool the cross-sections efficiently. It gives equal weight to estimates that are estimated with differing levels of precision, and it ignores any temporal dependence between the different cross-sections. The other option is a minimum distance technique that we outline in the Appendix. This technique does put more weight on more precise yearly estimates and optimally adjusts for arbitrary temporal dependence. Efficient pooling is important because the EW estimators have more sampling error than those based on second moments. They therefore require more data, which in turn implies that EW estimates obtained in individual time periods can exhibit more time variation than do OLS estimates.

1.3 Intuition

The economic intuition behind the EW estimators is easiest to see by considering the simplest, that is, the one given by (9). Recall that the problem for OLS in the classical errors-in-variable model can be shown by using (5)–(6) to write the relationship between the observable variables as

$$\dot{y}_i = \beta\dot{x}_i + (u_i - \beta\varepsilon_i), \tag{11}$$

and then noting that $\dot{x}_i$ and the composite error $u_i - \beta\varepsilon_i$ are correlated because they both depend on ε_i. In such situations, economists have been taught to look for an instrumental variable: an observable variable w_i that is correlated with $\dot{x}_i$ but *not* correlated with $u_i - \beta\varepsilon_i$. The economic intuition behind IV estimation comes from using economic reasoning to find observable variables for which these two correlation requirements can be verified. It is unlikely that anyone's economic intuition would ever suggest the instrument $w_i = \dot{y}_i\dot{x}_i$, which leads to exactly the same estimator as (9), but economic reasoning can be used as it would for any proposed instrument to verify whether it satisfies the two IV correlation conditions.

First, consider the zero-correlation requirement:

$$E[\dot{y}_i\dot{x}_i(u_i - \beta\varepsilon_i)] = 0. \tag{12}$$

Substituting (5)–(6) into (12) lets one re-express the left-hand side of (12) as a linear combination of the third-order product moments $E\left(\dot{\chi}_i^2 u_i\right)$, $E\left(\dot{\chi}_i u_i^2\right)$, $E\left(\dot{\chi}_i^2 \varepsilon_i\right)$, $E\left(\dot{\chi}_i \varepsilon_i^2\right)$, $E\left(u_i^2 \varepsilon_i\right)$, $E\left(u_i \varepsilon_i^2\right)$, and $E\left(\dot{\chi}_i u_i \varepsilon_i\right)$. The assumptions of the classical model in Table 1 and the process of partialing ensure that all three variables, $\dot{\chi}_i$, u_i, and ε_i, have zero means and are independent of each other. Hence, all the product moments equal zero, verifying (12). Economic intuition or reasoning can be applied to understand the zero-correlation requirement by verifying whether the assumptions of the classical model are plausible. In the case of investment regressions, this intuition is carefully spelled out in Erickson and Whited (2000, pp. 1036–37) in the context of the q theory of investment.

Now consider the second IV requirement, that $\dot{y}_i\dot{x}_i$ and $\dot{x}_i$ be correlated:

$$E[\dot{y}_i\dot{x}_i^2] \neq 0. \tag{13}$$

Rewriting the left-hand side of this equation yields a linear combination of third-order product moments *and* the third moment $E\left(\dot{\chi}_i^3\right)$. The product moments vanish as before, so (13) is verified if $\dot{\chi}_i$ has a skewed distribution, a property that can also be evaluated using economic reasoning. At a basic level, these estimators are ideally suited to investment regressions because marginal q is technically a shadow value. Therefore, it cannot be negative and cannot be normally distributed. However, marginal q might be approximately normally distributed, so one must also use economic theory to understand the degree of skewness of marginal q. For example, in Abel (1983), the firm chooses optimal investment to maximize the expected discounted value of cash flows and the source of uncertainty is the output price, which follows a geometric Brownian motion. In this model, even though stock returns are approximately normally distributed, marginal q is a linear function of the output price, so that marginal q is approximately lognormally distributed, and hence highly skewed.

In sum, the intuition behind the simplest high-order moment estimator is analogous to intuition behind an IV estimator, in the sense that one must use economic theory to verify instrument validity and relevance. We also note that third-order moments are ubiquitous in applied work, since both OLS and IV estimators are functions of such moments whenever regressions contain interaction terms or squares of a regressor. The EW estimators simply use interactions as instruments.

Economic intuition is harder to find for the IV and AB estimators, because q theory says nothing about whether the measurement error in q follows an autoregressive or a moving average process. Identification of these estimators hinges on this distinction because the use of lagged regressors as instruments requires limited serial correlation of the measurement error process. If, as argued in Erickson and Whited (2000), the measurement error is highly persistent, then the use of lagged values of observable Tobin's q as instruments is invalid. In general, although it is easy to find instrumental variables that satisfy the condition that the instrument be correlated with the mismeasured regressor (e.g., the use of lagged returns in Lewellen and Lewellen 2010), and although it is easy to find variables that satisfy the usual exclusion restriction (any irrelevant variable will do), it is very difficult to find variables that satisfy both.

2. Monte Carlos

In this section, we run four sets of Monte Carlo experiments. We examine the performance of the IV, AB, and EW estimators under correct specification, their performance under misspecification, the reaction of the EW estimators to skewness of the mismeasured regressor, and the performance of identification diagnostics for the EW estimators.

2.1 Calibration and specification

To make our Monte Carlo experiments relevant for investment regressions, we calibrate them so that the distribution of the simulated data closely resembles the distribution of actual investment data. This issue is important because one weakness of Monte Carlo experiments is that they are only informative about an estimator at one point in a parameter space and for one joint distribution of the data. The performance of *any* estimator can depend strongly on the parameter values and distributions chosen. Therefore, for the purposes of providing guidance for empirical researchers, it is important to pick a point in the parameter space and a data distribution that are relevant to the empirical problem at hand. For further discussion of these points in the asset-pricing econometrics literature, see, for example, Kocherlakota (1990), Hansen, Heaton, and Yaron (1996), and Shanken and Zhou (2007).

We pattern the first after the investment regressions in Fazzari, Hubbard, and Petersen (1988) and Erickson and Whited (2000). In this specification, investment and cash flow are deflated by a measure of the capital stock, as is the proxy for true q, which can be expressed as

$$q^{FHP} \equiv \frac{D_i + E_i - C_i}{K_i}. \tag{14}$$

Here, D_i is a measure of the book value of total debt, E_i is the market value of equity, C_i is the replacement value of the firm's current assets, and K_i is the replacement value of the firm's capital stock. If capital and current assets are the firm's only two assets, the numerator of (14) represents the market value of the capital stock. This proxy has both good and bad features. First, if the firm has intangible assets that cannot be subtracted from the numerator of (14), then q^{FHP} is biased. Fortunately, the intercept, γ, in (4) allows for biased measurement. Second, q^{FHP} can occasionally be negative, which is not a desirable feature of a proxy for a quantity that cannot by definition be negative. Third, q^{FHP} does have the redeeming feature that, from an intuitive standpoint, it does a better job of isolating investment opportunities in capital goods than does the other widely used proxy—the market-to-book ratio.

The second experimental design is based on specifications from Chava and Roberts (2008) and Rauh (2006). In this design, investment and cash flow are deflated by a measure of total assets, as is the proxy for true q, the market-to-book ratio. Although this proxy has the advantage that it is never negative, as discussed in Erickson and Whited (2006), it has the disadvantage that it is a proxy for investment opportunities in all assets—not just capital goods.

Table 2 presents summary statistics from our sample of Compustat firms. Details concerning variable construction are in the Appendix.

Panel A summarizes variables deflated by the capital stock, and Panel B presents statistics describing variables deflated by total assets. Three important features stand out. First, q^{FHP} has a higher mean than the market-to-book

Table 2
Summary statistics

	Mean	Variance	Third Standardized Moment	Fourth Standardized Moment	Fifth Standardized Moment	Serial Correlation
Panel A Capital Stock Deflator						
Investment	0.129	0.014	3.152	18.524	119.993	0.499
q	2.279	17.588	6.253	59.475	639.942	0.721
Cash Flow	0.176	0.113	−1.467	30.122	−143.878	0.461
Panel B Total Assets Deflator						
Investment	0.068	0.003	2.622	14.088	81.082	0.525
Market-to-book	1.492	1.119	3.897	26.460	203.984	0.665
Cash Flow	0.087	0.011	−2.064	14.801	−81.044	0.403

Calculations are based on a sample of manufacturing firms from the 2009 Compustat Industrial Files. The sample period is from 1967 to 2008, and the sample contains 57,583 observations. In Panel A, investment and cash flow are deflated by the gross capital stock. In Panel B, investment and cash flow are deflated by total book assets. The variable q is a proxy for Tobin's q intended to isolate variation in investment opportunities for property, plant, and equipment. "Market-to-book" is a proxy for Tobin's q intended to isolate variation in investment opportunities for total assets. $AR(1)$ denotes a first-order autoregressive coefficient, calculated via the technique in Han and Phillips (2010), using firms that have at least five years of data. This technique produces consistent estimates of autoregressive coefficient in panel data by double-differencing the data. Third through fifth moments are scaled by the standard deviation raised to the corresponding power.

ratio, which could be the result of rents to capital or biased measurement. Second, q^{FHP} has higher variance, skewness, and kurtosis than does the market-to-book ratio. Third, q^{FHP} is more highly autocorrelated than the market-to-book ratio.[2] This feature is important because IV estimators based on lagged regressors perform poorly in finite samples when the regressors are highly autocorrelated (Blundell and Bond 1998).

We now turn to a description of the data-generating process. Our simulated data consist of a panel of length 10 with a cross-sectional size of 1,500. These dimensions are roughly the average time-series and cross-sectional dimensions of our real data set. We create our simulated samples as follows. First, we generate time-zero $i.i.d.$ cross-sections of the four variables $(\chi_{i0}, z_{i0}, u_{i0}, \varepsilon_{i0})$, in which each variable has a gamma distribution with a mean of zero, a variance of one, and shape parameters described below.[3] We then generate the entire panel $(\chi_{it}, z_{it}, u_{it}, \varepsilon_{it})$ by updating $AR\,(1)$ processes for 20 periods and keeping the last 10 time periods, which removes the effects of initial conditions.

The specific $AR\,(1)$ processes are

$$\chi_{it} = \delta_\chi + \phi_\chi \chi_{i,t-1} + v_{it}^{\chi} \tag{15}$$

$$z_{it} = \delta_z + \phi_z z_{i,t-1} + v_{it}^{z} \tag{16}$$

$$u_{it} = \phi_u u_{i,t-1} + v_{it}^{u} \tag{17}$$

$$\varepsilon_{it} = \phi_\varepsilon \varepsilon_{i,t-1} + v_{it}^{\varepsilon}. \tag{18}$$

Here, ϕ_j, $j = (\chi, z, u, \varepsilon)$ are the autocorrelation coefficients of the $AR\,(1)$ processes governing $(\chi_{it}, z_{it}, u_{it}, \varepsilon_{it})$, and $\left(v_{it}^{\chi}, v_{it}^{z}, v_{it}^{u}, v_{it}^{\varepsilon}\right)$ are the $i.i.d.$ innovations to these processes, again with zero-mean, unit-variance gamma distributions. We set the parameters $\left(\delta_\chi, \delta_z\right)$ so that the means of (χ_{it}, z_{it}) equal the means of (x_{it}, z_{it}) in our data. We initially set $\phi_u = \phi_\varepsilon = 0$. From the autocorrelation estimates in Table 2, we set $\phi_\chi = 0.72$ and $\phi_z = 0.46$ for the design based on q^{FHP}. For the design based on the market-to-book ratio, we set $\phi_\chi = 0.67$ and $\phi_z = 0.40$.[4] We then multiply the vector (χ_{it}, z_{it}) by our data estimate of the square root of cov (x_{it}, z_{it}), which we calculate via an

---

2 We estimate these first-order autoregressive coefficients using the method in Han and Phillips (2010), which delivers consistent estimates of autoregressive coefficients in panel data by double-differencing the data.

3 Gamma distributions are characterized by two parameters: a scale parameter that we set equal to one and a shape parameter, k. The excess skewness and kurtosis of this distribution are $2/\sqrt{k}$ and $6/k$. See Bekaert and Engstrom (2011) for another use of gamma distributions to capture nonnormal data features.

4 It is also possible to use $ARMA\,(1,1)$ processes here by adding terms in $v_{i,t-1}^{\chi}$ and $v_{i,t-1}^{z}$ to Equations (15) and (16) above. This alternative specification may be warranted because the decay in our actual regression variables may be slower than that implied by our $AR(1)$ coefficient estimates, as is the case in many economic time series. When we have used $ARMA(1,1)$ processes, we have found that the performance of the AB and IV estimators deteriorates.

eigenvalue decomposition. This last procedure has a negligible effect on the persistence of (χ_{it}, z_{it}).

With the variables $(\chi_{it}, z_{it}, u_{it}, \varepsilon_{it})$, we then construct the observable variables (x_{it}, y_{it}) from (3) and (4), with the intercept of the measurement equation set to zero, which has no effect on the Monte Carlo results. We set the intercept in (3) to match the mean of y_{it} in our data. To generate these observable variables, we also need to choose values for the following three key parameters: β, α_1, and τ^2. For the q^{FHP} design, these three parameters equal $(0.02, 0.05, 0.5)$, and for the market-to-book design they equal $(0.03, 0.08, 0.25)$. These settings approximately equal the average estimates from our data analysis in the next section, and they imply regression R^2s that are approximately equal to those from our data analysis. The one difference is that we set α_1 higher than the average estimate we obtain. The intent is to determine whether the EW estimators can erroneously produce estimated coefficients that are near zero, even when the true values are positive. We do not consider the case in which $\alpha_1 = 0$ because it is treated thoroughly in Erickson and Whited (2000).

Finally, we choose distributions for $\left(v_{it}^{\chi}, v_{it}^{z}, v_{it}^{u}, v_{it}^{\varepsilon}\right)$ and $(\chi_{i0}, z_{i0}, u_{i0}, \varepsilon_{i0})$ so that our simulated data vector (x_{it}, y_{it}, z_{it}) has first and second moments identically equal, and higher moments approximately equal to those from our data. For the q^{FHP} design, in each time period we use the following cross-sectional gamma distributions. The innovations $\left(v_{it}^{\chi}, v_{it}^{z}, v_{it}^{u}, v_{it}^{\varepsilon}\right)$ and initial variables $(\chi_{i0}, z_{i0}, u_{i0}, \varepsilon_{i0})$ have shape parameters of 0.027, 0.8, 0.25, and 0.023, respectively. For the market-to-book design, we use shape parameters of 0.02, 0.45, 0.45, and 0.2. It is important to note that because $(\chi_{it}, z_{it}, u_{it}, \varepsilon_{it})$ are the weighted sums of gamma distributions, their skewness and kurtosis are less than those of any of the individual gamma distributions. We report the actual skewness and kurtosis of our simulated variables with our simulation results.

Four shape parameters give more than enough degrees of freedom to match the skewness of the three observable variables (x_{it}, y_{it}, z_{it}).[5] This issue is important because the EW estimators require that the unobservable χ_{it} be nonnormal. Because the observable variable, x_{it}, equals the sum, $\chi_{it} + \varepsilon_{it}$, one can mimic high skewness in the observable variable x_{it} with either a highly skewed χ_{it} or a highly skewed ε_{it}. To pin down the skewness of χ_{it}, we also match the following higher-order cross moments: $E\left(x_{it}^2 y_{it}\right)$ and $E\left(x_{it} y_{it}^2\right)$.

The final design issue is starting values for the two overidentified estimators, GMM4 and GMM5. We use two different starting values for β: the Geary estimator and the OLS estimator. We then use the resulting estimate that corresponds to the lower of the GMM objective functions. We use the OLS estimate as an alternative starting value because it is simple and because this practice is standard in the estimation of many nonlinear models, such as logits

[5] See the discussion in Almeida, Campello, and Galvao (2010).

and probits. Further, although the OLS estimate is biased, it has a low variance and therefore does not produce any extreme starting values. Ideally, one should try many different starting values along a grid, but in a Monte Carlo with thousands of trials, computational constraints limit us to a grid with only two points.

In our Monte Carlos, we present results for the EW estimators based on moments of order three through five, in which the yearly estimates have been pooled with the minimum distance method described in the Appendix. We also report results for three other estimators. The first is a cross-sectional OLS estimator, in which the yearly estimates have also been pooled via minimum distance. The second is a simple instrumental variables estimator in which twice- and three-times-lagged q and cash flow are used as instruments for the first difference of q. The IV estimator is therefore overidentified, as the number of instruments exceeds the number of mismeasured regressors. Third, we present the results from the AB estimator in which twice-lagged levels of q and cash flow are used as instrumental variables for the differences of q. This estimator is also overidentified, but by many more degrees of freedom because the estimator uses each cross-section of data to construct separate moment conditions, which results in two moment conditions for each year of the panel.

2.2 Results from correct specification

This section reports results from Monte Carlos in which all estimators are simulated under assumptions guaranteeing zero asymptotic bias. The results therefore refer to finite sample bias.

2.2.1 Main results. Table 3 reports the results from our Monte Carlo based on the q^{FHP} data-generating process (DGP). We present the mean and median bias as a fraction of the true coefficient for β, α_1, ρ^2, and τ^2, which are, respectively, the coefficient on the mismeasured regressor, the coefficient on the perfectly measured regressor; the true regression R^2 (the actual R^2 in the case of OLS); and the index of measurement quality (the R^2 of the measurement equation). Also, for each parameter we report the mean absolute deviation of the estimates from their true value (MAD) and the probability the estimate is within 20% of the true value (probability concentration). Panel A contains the results for simulated data in levels, that is, that have not undergone either first differencing or a within transformation, both of which remove potential fixed effects. Panel B contains results for transformed data. For the AB and IV estimators, the transformation is first differencing, and for OLS and the EW estimators, it is the within transformation. Because the IV and AB estimators are defined *only* in first-differenced form, we report no levels results for these two estimators in Panel A.

First, we compare the performance of the OLS and EW estimators in levels (Panel A of Table 3) with the IV and AB estimators in differenced form

Table 3
Monte Carlo performance of EW and IV estimators: q^{FHP} DGP

Panel A Levels	OLS	GEARY	GMM4	GMM5
Mean Bias ($\hat{\beta}$)	−0.531	−0.027	−0.028	−0.036
Median Bias ($\hat{\beta}$)	−0.531	−0.027	−0.030	−0.038
MAD($\hat{\beta}$)	0.011	0.001	0.001	0.001
P($\mid \hat{\beta} - \beta \mid \leq 0.2\beta$)	0.000	0.999	0.996	0.996
P(t_{β}) (asymptotic)		0.183	0.349	0.541
P(t_{β}) (bootstrap)		0.066	0.054	0.022
Mean Bias ($\hat{\alpha_1}$)	0.756	−0.023	0.006	0.023
Median Bias ($\hat{\alpha_1}$)	0.757	−0.006	0.017	0.032
MAD($\hat{\alpha_1}$)	0.038	0.013	0.008	0.007
P($\mid \hat{\alpha_1} - \alpha_1 \mid \leq 0.2\alpha_1$)	0.000	0.592	0.720	0.774
P(t_{α}) (asymptotic)		0.542	0.510	0.500
P(t_{α}) (bootstrap)		0.068	0.072	0.086
Mean Bias ($\hat{\rho}^2$)	−0.390	−0.005	−0.002	−0.018
Median Bias ($\hat{\rho}^2$)	−0.393	−0.005	−0.003	−0.019
MAD($\hat{\rho}^2$)	0.185	0.021	0.019	0.020
P($\mid \hat{\rho}^2 - \rho^2 \mid \leq 0.2\rho^2$)	0.000	0.990	0.995	0.996
Mean Bias ($\hat{\tau}^2$)		0.098	0.132	0.201
Median Bias ($\hat{\tau}^2$)		0.098	0.133	0.203
MAD($\hat{\tau}^2$)		0.061	0.070	0.101
P($\mid \hat{\tau}^2 - \tau^2 \mid \leq 0.2\tau^2$)		0.815	0.762	0.489
Nominal 5% J-Test Rejection Rate			0.091	0.167

Panel B		Within			First Differenced	
Fixed Effects Transformation	OLS	GEARY	GMM4	GMM5	IV	AB
Mean Bias ($\hat{\beta}$)	−0.659	−0.112	−0.168	−0.140	−0.028	−0.170
Median Bias ($\hat{\beta}$)	−0.660	−0.096	−0.106	−0.069	−0.030	−0.169
MAD($\hat{\beta}$)	0.013	0.003	0.004	0.003	0.002	0.004
P($\mid \hat{\beta} - \beta \mid \leq 0.2\beta$)	0.000	0.799	0.693	0.765	0.870	0.581
P(t_{β}) (asymptotic)		0.323	0.683	0.645	0.205	0.248
P(t_{β}) (bootstrap)		0.052	0.088	0.137	0.038	0.121
Mean Bias ($\hat{\alpha_1}$)	0.810	0.575	0.361	0.228	0.030	0.188
Median Bias ($\hat{\alpha_1}$)	0.811	0.272	0.112	0.061	0.031	0.190
MAD($\hat{\alpha_1}$)	0.040	0.159	0.041	0.038	0.007	0.011
P($\mid \hat{\alpha_1} - \alpha_1 \mid \leq 0.2\alpha_1$)	0.000	0.158	0.428	0.564	0.737	0.506
P(t_{α}) (asymptotic)		0.772	0.613	0.457	0.110	0.217
P(t_{α}) (bootstrap)		0.116	0.173	0.191	0.054	0.117
Mean Bias ($\hat{\rho}^2$)	−0.450	−0.055	−0.141	−0.120		
Median Bias ($\hat{\rho}^2$)	−0.454	−0.054	−0.141	−0.121		
MAD($\hat{\rho}^2$)	0.118	0.027	0.042	0.035		
P($\mid \hat{\rho}^2 - \rho^2 \mid \leq 0.2\rho^2$)	0.000	0.856	0.639	0.733		
Mean Bias ($\hat{\tau}^2$)		0.070	0.140	0.166		
Median Bias ($\hat{\tau}^2$)		0.069	0.148	0.175		
MAD($\hat{\tau}^2$)		0.059	0.075	0.077		
P($\mid \hat{\tau}^2 - \tau^2 \mid \leq 0.2\tau^2$)		0.706	0.573	0.535		
Nominal 5% J-Test Rejection Rate			0.036	0.125	0.065	0.105

(*continued*)

Table 3
Continued

Panel C Summary Statistics	Mean	Variance	Third Standardized Moment	Fourth Standardized Moment	Fifth Standardized Moment
Investment (y_{it})	0.129	0.014	3.018	18.035	128.398
Observable q (x_{it})	2.280	17.588	6.261	66.325	896.228
Cash flow (z_{it})	0.176	0.113	−1.549	7.467	−30.866
True q (χ_{it})	2.280	8.800	5.657	55.587	691.654
Measurement error (ε_{it})	0.000	8.787	11.546	181.357	3387.482

Indicated expectations and probabilities are estimates based on 10,000 Monte Carlo samples of size 1,500. The samples are generated by

$$y_{it} = z_i\alpha + \chi_{it}\beta + u_{it}$$
$$x_{it} = \chi_{it} + \varepsilon_{it},$$

in which χ_{it} is the unobservable true regressor, x_{it} is its observable counterpart, ε_{it} is the measurement error, u_{it} is the regression error, z_{it} is a vector containing 1 and a single perfectly measured regressor, and y_{it} is the dependent variable, which is analogous to investment in an investment regression. The coefficient α_1 on the perfectly measured regressor is analogous to the cash flow coefficient in an investment regression, and the coefficient β is analogous to the coefficient on observable q. The variables χ_{it}, ε_{it}, u_{it}, and z_{it} are pseudo-gamma distributed variables. ρ^2 is the population R^2 of the first equation, and τ^2 is the population R^2 of the second. GMMn denotes the EW estimator based on moments up to order $M = n$. GEARY denotes the third-order moment estimator from Geary (1942). OLS denotes estimates obtained by regressing y_{it} on x_{it} and z_{it}. IV denotes estimates obtained by a regression containing first differences of these variables with twice- and three-times-lagged values of x_{it} and z_{it} as instruments. AB indicates the Arellano and Bond (1991) estimator, which uses twice-lagged values of x_{it} and z_{it} as instruments. Bias is expressed as a fraction of the true coefficient value. MAD indicates "Mean Absolute Deviation." "Fixed Effects Transformation" refers to within-transformed data for the OLS and EW estimators and to first-differenced data for the IV and AB estimators. "P(t) (asymptotic)" and "P(t) (bootstrap)" are the actual sizes of two-sided t-tests with nominal significance levels of 5%. The first test is calculated with asymptotic critical values, and the second is calculated with a bootstrapped critical value from 1,000 bootstrap trials. t_α refers to a test of the null that the estimate of α_1 equals its true value, and t_β refers to a test of the null that the estimate of β equals its true value. The J-test is a test of the overidentifying restrictions of the model. Indicated third through fifth moments are scaled by the standard deviation raised to the corresponding power.
True Values: $\beta = 0.02$, $\alpha_1 = 0.05$, $\rho^2 = 0.341$, and $\tau^2 = 0.5$.

(Panel B of Table 3). The first column of Panel A shows the bias induced by measurement error. The R^2 and β are biased downward, and α_1 is biased upward because of the positive correlation between the mismeasured and perfectly measured regressors. In contrast, the simple IV estimator shown in Panel B performs well. Coefficient bias and MAD are small, and the probability concentration is over 0.85. We also find that the AB estimator does not perform as well. It produces downward-biased estimates of β and upward-biased estimates of α_1. As in Blundell and Bond (1998), the bias stems from high regressor serial correlation. Finally, a comparison of Panel A with the last two columns of Panel B shows that all three EW estimators outperform the simple IV and the AB estimators. The bias for all parameters is near zero, the MADs are low, and the probability concentrations are near 1. The EW estimators also produce highly accurate estimates of ρ^2, and the third- and fourth-order moment estimators deliver accurate estimates of the index of measurement quality, τ^2.

Table 3 also reports the actual size of three sets of tests. The first is a nominal 5% test of the overidentifying restrictions of the IV, AB, GMM4, and GMM5 estimators. The second set of tests are nominal 5% two-sided t-tests that the slope coefficients, β and α_1, equal their true values, in which we use test-critical values from the asymptotic distribution. The third set of tests is identical to the second, except that we use the critical values from the block bootstrap in Hall and Horowitz (1996). To compute the critical values, we resample at the firm level and use 1,000 bootstrap trials.[6]

The overidentification test (J-test) is approximately correctly sized for the IV estimator, but the test overrejects slightly for GMM4, GMM5, and the AB estimator. Nominal asymptotic 5% t-tests for β have actual sizes of 0.21 and 0.25 for the IV and AB estimators. The overrejection rate is usually higher for the EW estimators, with the actual sizes ranging between 0.18 and 0.54. The situation is worse for the t-tests on α_1. The tests overreject slightly for the IV and AB estimators, but the actual test size ranges between 0.50 and 0.54 for the EW estimators. This problem is due to standard errors that are too small—a result also in Erickson and Whited (2000). In terms of investment regressions, the result implies that one is likely to reject incorrectly the null hypothesis that the cash flow coefficient is zero.[7] Using bootstrapped critical values ameliorates this problem. We find that for both β and α_1, the actual and nominal bootstrapped test sizes are close for the EW, IV, and AB estimators.

Panel B of Table 3 compares the performance of the IV and AB estimators with that of the EW estimators when the latter are applied to within-transformed data. We find that the performance of the EW estimators falls because doing a within transformation increases the regression error variance relative to the regressor variance. This issue affects the high-order moment estimators more than traditional estimators for two related reasons. First, regressions with high error variances generally require more data for *any* estimator to be able to obtain precise coefficient estimates. Second, at *any* error variance more data are required for precise estimation of high-order moments than for the second-order moments on which traditional regression analysis is based.

The EW estimators for β and α_1 perform better than the AB estimator in terms of median bias and probability concentrations, but unambiguously worse than the simple IV estimator. The GEARY estimator for α_1 performs much worse than GMM4 or GMM5. The EW estimators produce a positive bias on α_1, but the bias for GMM5 is small, and the positive sign of the bias implies that spurious findings of zero cash flow coefficients are unlikely. The GEARY estimator is the best for ρ^2 and τ^2.[8] Finally, the rejection rates for the

6 These particular calculations are based on 1,000 Monte Carlo trials.

7 This result is DGP dependent and thus relevant only to investment regressions. For example, Riddick and Whited (2009) report under rejection for a different DGP that resembles a savings regression.

8 To gauge the performance of EW estimators for ρ^2 and τ^2, we need to calculate the true values of these coefficients after we perform a within transformation. To do so, we estimate the true within ρ^2 and τ^2 by

asymptotic t-tests are largely similar to those for the data in levels. However, the bootstrap test rejection rates for both β and α_1 are too high for GMM4 and GMM5, ranging between 0.17 and 0.19. Thus, although using a bootstrap ameliorates the problem of small standard errors in some cases, it does not always fix it completely. Nonetheless, this result leads us to use the more conservative bootstrapped critical values in the data analysis that follows.

Panel C of Table 3 presents summary statistics for the simulations. A comparison with Table 2 shows that we exactly match the first- and second-order moments of the observable data and approximately match the higher moments, with the one exception being the kurtosis of cash flow, where the simulated value is one-fourth the actual value. Not shown are the moments $E\left(y_{it}^2 x_{it}\right)$ and $E\left(y_{it} x_{it}^2\right)$. Their values in the Monte Carlo of 0.048 and 2.41 are quite close to the corresponding data moments of 0.044 and 2.5. We also present the moments of the unobservable mismeasured regressor and the measurement error. The calibration of this Monte Carlo leads to skewness of the measurement error that is twice as large as the skewness of the unobservable mismeasured regressor.

Table 4 reports results from a Monte Carlo based on the market-to-book DGP. The organization of this table is identical to that of Table 3. Panel A shows that when used on data in levels, the EW estimators of β markedly outperform both the IV and AB estimators according to all metrics. For α_1, the EW estimators outperform in terms of bias and MAD but underperform in terms of probability concentrations. This performance ranking reverses when we use the EW estimators on within-transformed data. Panel B shows that the IV estimator outperforms the EW estimators, but the GMM5 estimator has performance comparable to the AB estimator in terms of all three metrics. The median biases of α_1 for the IV and AB estimators are 7.9% and 33.2%, and the median biases of α_1 for the EW estimators range from 13% to 19%, even though all assumptions assuring consistency are satisfied. It is important to interpret these results in the context of the tiny true coefficients in these regressions. For example, for the EW estimators, this bias at worst translates into a coefficient estimate of 0.095, when the truth is 0.08. Also, the median bias of these estimates is positive, which assuages concerns that the EW estimators deliver estimates of α_1 (the cash flow coefficient) near zero when the true value is positive. Finally, for this data-generating process, the t-test that α_1 equals its true value overrejects strongly with the use of an asymptotic critical value. As in Table 3, we find that using bootstrapped critical values lowers the test rejection rates, but the improvement is not as strong as in the case of the q^{FHP} DGP.

---

using OLS on (1) and (2) in which x_i has *not* been substituted in for χ_i, that is, we do OLS on the true regression and the true measurement equation. To ensure that the estimates are precise, we use 10 million simulated cross-sectional observations over 10 simulated years.

Table 4
Monte Carlo performance of EW and IV estimators: Market-to-book DGP

Panel A Levels	OLS	GEARY	GMM4	GMM5
Mean Bias ($\hat{\beta}$)	−0.799	−0.095	−0.055	−0.165
Median Bias ($\hat{\beta}$)	−0.799	−0.096	−0.058	−0.154
MAD($\hat{\beta}$)	0.024	0.004	0.003	0.006
P($\mid \hat{\beta} - \beta \mid \leq 0.2\beta$)	0.000	0.801	0.879	0.649
P(t_β) (asymptotic)		0.184	0.300	0.680
P(t_β) (bootstrap)		0.133	0.081	0.057
Mean Bias ($\hat{\alpha_1}$)	0.622	0.054	−0.075	0.028
Median Bias ($\hat{\alpha_1}$)	0.622	0.039	−0.013	0.079
MAD($\hat{\alpha_1}$)	0.050	0.045	0.024	0.022
P($\mid \hat{\alpha_1} - \alpha_1 \mid \leq 0.2\alpha_1$)	0.000	0.399	0.524	0.490
P(t_α) (asymptotic)		0.575	0.548	0.621
P(t_α) (bootstrap)		0.102	0.097	0.128
Mean Bias ($\hat{\rho}^2$)	−0.415	−0.025	−0.008	−0.081
Median Bias ($\hat{\rho}^2$)	−0.417	−0.027	−0.007	−0.082
MAD($\hat{\rho}^2$)	0.044	0.013	0.012	0.013
P($\mid \hat{\rho}^2 - \rho^2 \mid \leq 0.2\rho^2$)	1.000	0.885	0.913	0.872
Mean Bias ($\hat{\tau}^2$)		−0.073	−0.120	−0.183
Median Bias ($\hat{\tau}^2$)		−0.069	−0.115	−0.179
MAD($\hat{\tau}^2$)		0.041	0.040	0.054
P($\mid \hat{\tau}^2 - \tau^2 \mid \leq 0.2\tau^2$)		0.681	0.700	0.525
Nominal 5% J-Test Rejection Rate			0.075	0.098

Panel B Fixed Effects Transformation	OLS	GEARY	Within GMM4	GMM5	First Differenced IV	AB
Mean Bias ($\hat{\beta}$)	−0.851	−0.156	−0.274	−0.387	−0.113	−0.545
Median Bias ($\hat{\beta}$)	−0.852	−0.154	−0.275	−0.383	−0.128	−0.545
MAD($\hat{\beta}$)	0.026	0.006	0.009	0.012	0.009	0.017
P($\mid \hat{\beta} - \beta \mid \leq 0.2\beta$)	0.000	0.537	0.337	0.205	0.402	0.112
P(t_β) (asymptotic)		0.135	0.670	0.854	0.000	0.524
P(t_β) (bootstrap)		0.134	0.190	0.205	0.039	0.368
Mean Bias ($\hat{\alpha_1}$)	0.597	0.605	−0.530	0.028	0.067	0.331
Median Bias ($\hat{\alpha_1}$)	0.597	0.129	0.198	0.134	0.079	0.332
MAD($\hat{\alpha_1}$)	0.048	0.337	0.144	0.059	0.017	0.027
P($\mid \hat{\alpha_1} - \alpha_1 \mid \leq 0.2\alpha_1$)	0.000	0.187	0.223	0.363	0.552	0.244
P(t_α) (asymptotic)		0.678	0.749	0.653	0.000	0.451
P(t_α) (bootstrap)		0.129	0.290	0.304	0.054	0.340
Mean Bias ($\hat{\rho}^2$)	−0.411	−0.075	−0.154	−0.167		
Median Bias ($\hat{\rho}^2$)	−0.413	−0.078	−0.165	−0.174		
MAD($\hat{\rho}^2$)	0.027	0.016	0.017	0.017		
P($\mid \hat{\rho}^2 - \rho^2 \mid \leq 0.2\rho^2$)	0.013	0.665	0.586	0.586		
Mean Bias ($\hat{\tau}^2$)		−0.361	−0.387	−0.387		
Median Bias ($\hat{\tau}^2$)		−0.380	−0.433	−0.441		
MAD($\hat{\tau}^2$)		0.072	0.079	0.078		
P($\mid \hat{\tau}^2 - \tau^2 \mid \leq 0.2\tau^2$)		0.237	0.198	0.197		
Nominal 5% J-Test Rejection Rate			0.048	0.072	0.067	0.078

(continued)

Table 4
Continued

Panel C Summary Statistics	Mean	Variance	Third Standardized Moment	Fourth Standardized Moment	Fifth Standardized Moment
Investment (y_{it})	0.068	0.003	2.571	13.426	80.549
Observable q (x_{it})	1.492	1.120	3.688	23.640	187.489
Cash flow (z_{it})	0.087	0.011	−2.193	11.642	−63.775
True q (χ_{it})	1.492	0.281	6.365	68.799	934.743
Measurement error (ε_{it})	0.000	0.839	4.365	30.763	270.426

Indicated expectations and probabilities are estimates based on 10,000 Monte Carlo samples of size 1,500. The samples are generated by

$$\begin{aligned} y_{it} &= z_i\alpha + \chi_{it}\beta + u_{it} \\ x_{it} &= \chi_{it} + \varepsilon_{it}, \end{aligned}$$

in which χ_{it} is the unobservable true regressor, x_{it} is its observable counterpart, ε_{it} is the measurement error, u_{it} is the regression error, z_{it} is a vector containing 1 and a single perfectly measured regressor, and y_{it} is the dependent variable, which is analogous to investment in an investment regression. The coefficient α_1 on the perfectly measured regressor is analogous to the cash flow coefficient in an investment regression, and the coefficient β is analogous to the coefficient on observable q. The variables χ_{it}, ε_{it}, u_{it}, and z_{it} are pseudo-gamma distributed variables. ρ^2 is the population R^2 of the first equation, and τ^2 is the population R^2 of the second. GMMn denotes the EW estimator based on moments up to order $M = n$. GEARY denotes the third-order moment estimator from Geary (1942). OLS denotes estimates obtained by regressing y_{it} on x_{it} and z_{it}. IV denotes estimates obtained by a regression containing first differences of these variables with twice- and three-times-lagged values of x_{it} and z_{it} as instruments. AB indicates the Arellano and Bond (1991) estimator, which uses twice-lagged values of x_{it} and z_{it} as instruments. Bias is expressed as a fraction of the true coefficient value. MAD indicates "Mean Absolute Deviation." "Fixed Effects Transformation" refers to within-transformed data for the OLS and EW estimators and to first-differenced data for the IV and AB estimators. "P(t) (asymptotic)" and "P(t) (bootstrap)" are the actual sizes of two-sided t-tests with nominal significance levels of 5%. The first test is calculated with asymptotic critical values, and the second is calculated with a bootstrapped critical value from 1,000 bootstrap trials. t_α refers to a test of the null that the estimate of α_1 equals its true value, and t_β refers to a test of the null that the estimate of β equals its true value. The J-test is a test of the overidentifying restrictions of the model. Indicated third through fifth moments are scaled by the standard deviation raised to the corresponding power.
True Values: $\beta = 0.03$, $\alpha_1 = 0.08$, $\rho^2 = 0.126$, and $\tau^2 = 0.25$.

Panel C of Table 4 again shows that we match the first- and second-order moments of the data exactly and the higher moments approximately. We also approximately match the two cross moments $E\left(y_{it}^2 x_{it}\right)$ and $E\left(y_{it} x_{it}^2\right)$, which are 0.024 and 0.001 in the Monte Carlo and 0.017 and 0.001 in the data. Interestingly, our calibration of the market-to-book DGP leaves us with skewness of χ_{it} that is slightly larger than the skewness of the measurement error.

Three general conclusions can be extracted from Tables 3 and 4. First, GMM4 and GMM5 appear to offer in most cases a large boost in performance relative to the GEARY estimator. Second, doing a within transformation changes regression error variance relative to the regressor variance and adversely affects the EW estimators. This conclusion implies that one should take care when using the EW estimators on regressions that are well known to contain fixed effects, such as leverage regressions (Lemmon, Roberts, and Zender 2008). Third, all estimators perform much worse on the market-to-book DGP than on the q^{FHP} DGP. The lower skewness and kurtosis of the market-to-book ratio affects the EW estimators. More importantly, both ρ^2 and τ^2 are much lower for this DGP, as is the case in our data. In general, most estimators perform worse when the data cloud is more dispersed and the measurement quality is worse.

2.2.2 Results from a larger sample size. We consider two further issues. The first is the effect of sample size on the performance of the EW estimators on within-transformed data. This exercise is of interest because the cross-sectional sample size of 1,500 used in Tables 3 and 4 is the approximate size of a typical cross-section of *manufacturing* firms in Compustat. However, many empirical corporate finance studies that include Tobin's q use all of the unregulated, nonfinancial firms in Compustat, and the typical size of a cross-section in this case is approximately 4,000.

Table 5 presents the results from this simulation. For brevity, we omit results for the OLS estimator, ρ^2, and τ^2. We also include results for the EW estimator based on moments up to order six (GMM6) because we see performance improvements here for this estimator that are not present for the smaller sample size of 1,500.

Table 5 shows that increasing the sample size to 4,000 markedly improves the performance of the estimators on within-transformed data. Panel A presents results for the q^{FHP} DGP. Here, the performance of all of the estimators for β is strong and varies little across estimators. In contrast, the IV and AB estimators of α_1 dominate the GEARY and GMM4 estimators, but are only slightly better than the GMM5 and GMM6 estimators in terms of bias, MAD, and probability concentrations. We also find that increasing the sample size does nothing to ameliorate the result of an oversized nominal 5% t-test that uses asymptotic critical values. Once again, the actual sizes of the tests that use bootstrapped critical values are closer to the nominal sizes, and we still observe slight overrejection.

Panel B of Table 5 shows the results for the market-to-book DGP. As in the case of a smaller sample size, the performance of all estimators deteriorates. Here, the GEARY estimator is the best estimator for β, the AB estimator is the worst, and the IV estimator is the second best. The best estimator for α_1 is the IV estimator, the worst is the GEARY estimator, and the second best is GMM6. The test statistic results are similar to those in the top panel. In general, GMM5 and GMM6 perform well, with little bias and probability concentrations between 0.59 and 0.66. In sum, we learn that although a within transformation can hurt the performance of the EW estimators for a sample size of 1,500, a within transformation only produces a slight deterioration of performance for a larger—and empirically relevant—sample size of 4,000.

2.2.3 Results from varying skewness. The second issue is the effect of the skewness of the mismeasured regressor on estimator performance. To this end, we return to a sample size of 1,500 and alter the q^{FHP} data-generating

Table 5
Monte Carlo performance of EW and IV estimators: Large sample size

	First Differenced IV	First Differenced AB	GEARY	Within GMM4	Within GMM5	Within GMM6
Panel A q^{FHP} DGP						
Mean Bias ($\hat{\beta}$)	−0.009	−0.084	−0.050	−0.068	−0.033	−0.053
Median Bias ($\hat{\beta}$)	−0.010	−0.085	−0.046	−0.049	−0.013	−0.032
MAD($\hat{\beta}$)	0.001	0.002	0.001	0.002	0.001	0.001
P($\mid \hat{\beta} - \beta \mid \leq 0.2\beta$)	0.988	0.901	0.979	0.933	0.958	0.948
P(t_β) (asymptotic)	0.473	0.123	0.194	0.422	0.296	0.535
P(t_β) (bootstrap)	0.031	0.088	0.045	0.029	0.153	0.135
Mean Bias ($\hat{\alpha_1}$)	0.010	0.094	−0.128	0.048	0.041	0.069
Median Bias ($\hat{\alpha_1}$)	0.012	0.094	0.053	0.027	−0.005	0.013
MAD($\hat{\alpha_1}$)	0.004	0.006	0.082	0.015	0.008	0.008
P($\mid \hat{\alpha_1} - \alpha_1 \mid \leq 0.2\alpha_1$)	0.936	0.813	0.281	0.577	0.875	0.892
P(t_α) (asymptotic)	0.350	0.119	0.767	0.609	0.380	0.374
P(t_α) (bootstrap)	0.049	0.081	0.072	0.078	0.191	0.224
Nominal 5% J-Test Rejection Rate	0.065	0.090		0.035	0.186	0.347
Panel B Market-to-book DGP						
Mean Bias ($\hat{\beta}$)	−0.046	−0.345	−0.104	−0.190	−0.208	−0.276
Median Bias ($\hat{\beta}$)	−0.052	−0.345	−0.104	−0.189	−0.191	−0.241
MAD($\hat{\beta}$)	0.005	0.011	0.004	0.006	0.007	0.009
P($\mid \hat{\beta} - \beta \mid \leq 0.2\beta$)	0.624	0.252	0.725	0.518	0.514	0.413
P(t_β) (asymptotic)	0.000	0.354	0.107	0.546	0.674	0.809
P(t_β) (bootstrap)	0.047	0.226	0.096	0.121	0.094	0.093
Mean Bias ($\hat{\alpha_1}$)	0.027	0.210	−0.060	−0.279	0.013	0.048
Median Bias ($\hat{\alpha_1}$)	0.033	0.212	0.061	0.123	0.038	0.054
MAD($\hat{\alpha_1}$)	0.010	0.018	0.149	0.085	0.025	0.032
P($\mid \hat{\alpha_1} - \alpha_1 \mid \leq 0.2\alpha_1$)	0.808	0.464	0.269	0.315	0.591	0.655
P(t_α) (asymptotic)	0.000	0.324	0.689	0.739	0.570	0.563
P(t_α) (bootstrap)	0.065	0.211	0.099	0.169	0.170	0.200
Nominal 5% J-Test Rejection Rate	0.066	0.094		0.051	0.098	0.182

Indicated expectations and probabilities are estimates based on 10,000 Monte Carlo samples of size 4,000. Panel A presents results from the data-generating process in Table 3, and Panel B presents results from the data-generating process in Table 4. In both cases, α_1 is the coefficient on the perfectly measured regressor, and β is the coefficient on the mismeasured regressor. GMMn denotes the EW estimator based on moments up to order $M = n$. GEARY denotes the third-order moment estimator from Geary (1942). OLS denotes estimates obtained by regressing y_i on x_i and z_i. IV denotes estimates obtained by using lagged mismeasured and perfectly measured variables as instruments. AB indicates the Arellano and Bond (1991) estimator. Bias is expressed as a fraction of the true coefficient value. MAD indicates "Mean Absolute Deviation." "P(t) (asymptotic)" and "P(t) (bootstrap)" are the actual sizes of two-sided t-tests with nominal significance levels of 5%. The first test is calculated with asymptotic critical values, and the second is calculated with a bootstrapped critical value from 1,000 bootstrap trials. t_α refers to a test of the null that the estimate of α equals its true value, and t_β refers to a test of the null that the estimate of β equals its true value. The J-test is a test of the overidentifying restrictions of the model.

process by allowing the gamma distribution that generates the mismeasured regressor to have a shape parameter that varies between 0.0006 and 840 so that the skewness of the mismeasured regressor, χ_{it}, ranges from 19 to 0.2.[9] The results from this experiment are in Figure 1, which contains three panels. Panel A reports the bias in the coefficient on the perfectly measured regressor as a function of skewness of the mismeasured regressor. Panel B reports the MAD, and Panel C reports probability concentrations.

Three important results stand out in Figure 1. First, although the GEARY estimator is the best in terms of all metrics for very high levels of skewness, its performance falls rapidly once the skewness of χ_{it} falls below approximately 3. Second, the IV and AB estimators markedly outperform the higher-order moment estimators for nearly symmetric distributions with skewness less than 0.5. Finally, for modest levels of skewness greater than approximately 1, the GMM4 and GMM5 estimators continue to outperform the IV and AB estimators. In general, the overidentified high-order moment estimators are surprisingly robust to decreasing the skewness of the mismeasured regressor. However, their performance falls dramatically for almost symmetric distributions.

2.3 Results from misspecification

In Tables 3 through 5, we identify instances in which the EW estimators outperform the AB and IV estimators and vice versa. However, in those Monte Carlos, the assumptions for consistency of all three estimators are satisfied. It is likely, however, that real investment data do not satisfy these assumptions. We therefore study the performance and biases of these different estimators under misspecification. We consider three types: a fixed effect that is correlated with the regressors, heteroskedasticity, and a serially correlated measurement error. The first two types affect the EW estimators, and the third type affects the two IV estimators.[10] ACG examine the effects of the first two sources on the EW estimator, but only address the effects of limited moving average serial correlation on the IV and AB estimators. Instead, we examine autoregressive serial correlation, which renders these estimators inconsistent. We consider not only the bias these problems induce, but also the power of specification tests to detect them. For brevity, we present the results from the q^{FHP} DGP because the market-to-book DGP yields similar results.

We start with heteroskedasticity, which we model by multiplying u_{it} by $1 + h\,(\chi_{it} + z_{it})$, $h = 1, \ldots, 10$. This specification is similar to that in ACG.

---

[9] It is also possible to add offsetting skewness to ε_{it} so that the skewness of x_{it} remains unchanged. These results are almost identical to those reported below.

[10] Of course, these three sources of misspecification are not the only ones that might afflict investment regressions. For example, Erickson and Whited (2000) find useful power of the EW J-test to detect other types of misspecification such as a mismeasured capital stock or a nonlinear regression.

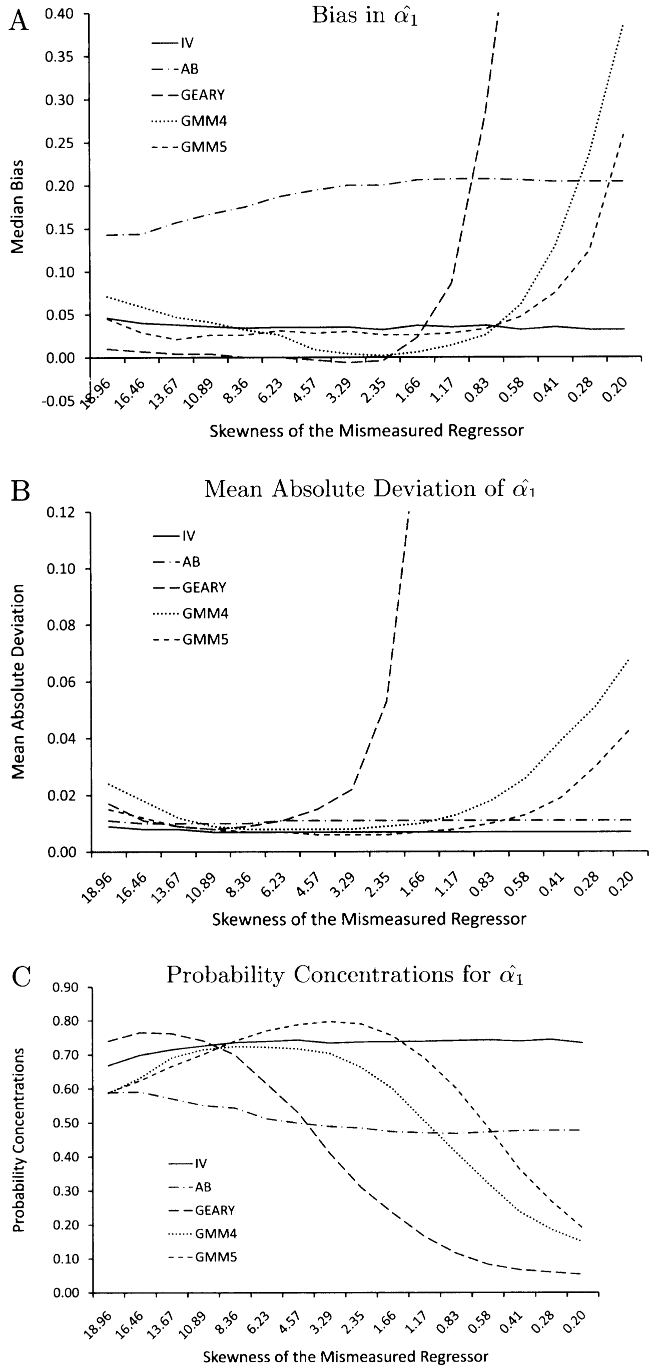

Figure 1
(Continued)

←

Figure 1
Monte Carlo performance of EW estimators: Varying degrees of skewness
Indicated median biases, mean absolute deviations, and probability concentrations are from Monte Carlo experiments based on the design used to generate Table 3. All of these figures are for the coefficient on the perfectly measured regressor, α_1, which is the coefficient on cash flow in an investment regression. The probability concentrations are computed as the fraction of Monte Carlo trials in which an estimate is within 20% of its true value. The horizontal axis on each graph is the skewness of the true unobserved regressor, χ_i.

However, to avoid changing the variance of the regression error, as do ACG, we standardize this new error term to have the same variance as the original error term. This way of modeling heteroskedasticity allows us to continue to approximate the third and higher moments of (x_{it}, y_{it}, z_{it}). We consider simulations in which the data are in levels and are within transformed.

The results are in Panels A and B of Figure 2, each of which contains two side-by-side plots. We report the bias in the coefficient on the perfectly measured regressor (α_1) in the left plot. In the right plot, we report the fraction of Monte Carlo samples in which the J-test rejects the model overidentifying restrictions.[11] The most important result here is the large bias in the GEARY estimator relative to GMM4 or GMM5. The former is sufficiently biased to produce the wrong sign, both for data in levels and within-transformed data. In contrast, for data in levels, GMM4 and GMM5 are not biased downward enough to drive a positive cash flow coefficient to zero. For the within-transformed data, these two estimators produce negligible bias. For example, the true value of α_1 is 0.05, and at most we observe that the GMM5 yields an average coefficient of 0.058. For the data in levels, the J-test accompanying GMM5 has the useful power (approximately 60%) to detect heteroskedasticity of this form. Not depicted in the figure are two further results that are useful for interpreting our data analysis. We find that heteroskedasticity increases the MAD, and we find large bias in the estimate of ρ^2, which at least doubles relative to its true value.

We next consider a fixed effect that is correlated with the regressors, which we model as

$$f_i \equiv T^{-1} \sum_{t=1}^{T} (\chi_{it} + z_{it}) .$$

Here, $T = 10$ is the length of the simulated panel. We then define the regression error as $\sigma_u (u_{it} + \pi f_i) \Big/ \sqrt{\sigma_u^2 + \pi^2 \sigma_f^2}$, the variance of which equals the variance of the original u_{it}. The parameter π adjusts the correlation between the fixed effect and the regressors. Panel C of Figure 2 contains the results from Monte Carlos run on simulated data that have not undergone a

[11] We focus on α_1 because it is the central coefficient of interest in investment regressions.

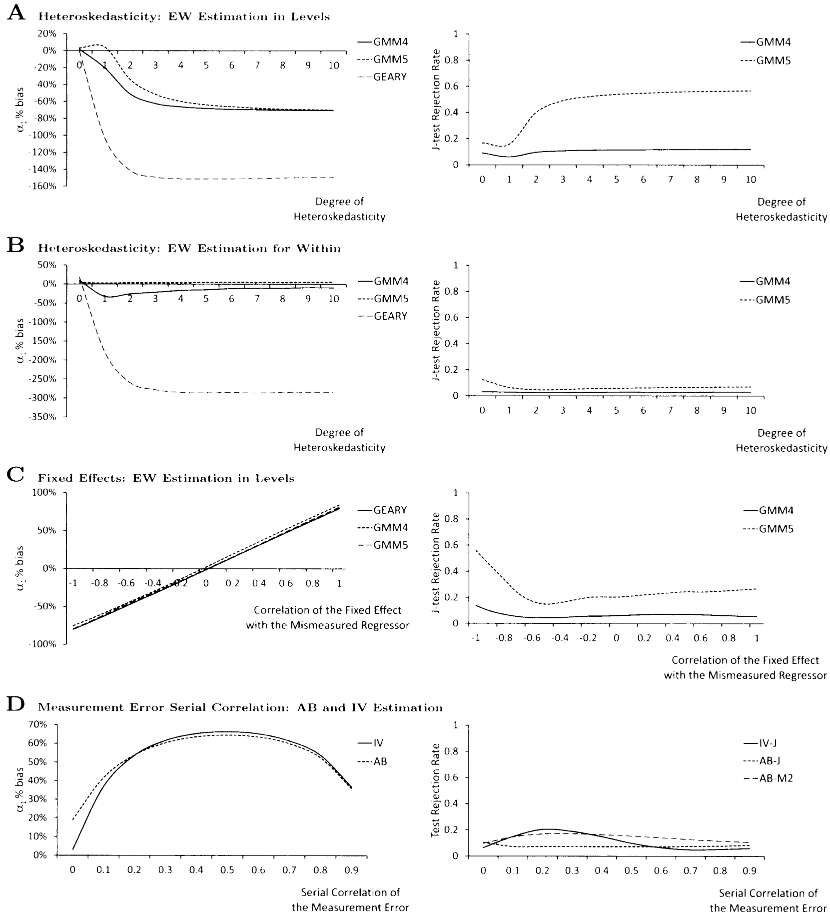

Figure 2
Bias and tests of overidentifying restrictions in misspecified Monte Carlos
Indicated biases and rejection rates are from Monte Carlo experiments based on the design used to generate Table 3. For the top four graphs, this design has been altered to include a heteroskedastic regression error. For the next two graphs, this design has been altered to include a fixed effect that is correlated with both the mismeasured and the perfectly measured regressors. For the bottom two graphs, this design has been altered to include a serially correlated measurement error. The graphs on the left side of the figure depict the bias in the coefficient on the perfectly measured regressor, and the graphs on the right side depict the fraction of Monte Carlo trials in which the overidentifying restrictions are rejected via the *J*-test. For the AB estimator, we also record the rejection rate for their second-order residual serial correlation test, which we denote "M2."

within transformation. In the upper left panel, we see substantial bias in the EW estimators for either a high negative or a positive correlation between f_i and (χ_{it}, z_{it}). This result is not surprising, given that the EW estimators are asymptotically biased in the presence of a regressor-error correlation, which is induced here by the fixed effect. Unfortunately, as seen in the right panel in the third row, this bias is for the most part difficult to detect. It is accompanied by a high J-test rejection rate only for a high negative correlation between the fixed effect and the regressors.

Nonetheless, these Monte Carlo experiments provide useful information for empirical researchers. Although the EW estimators are biased on untransformed data in the presence of a fixed effect that is correlated with the regressors, they are largely unbiased on data that has undergone a within transformation. Therefore, a data analyst who wants to use the EW estimators on panel data should calculate estimates both on data in levels and on within-transformed data. This informal check is probably the best fixed effects diagnostic available,[12] given the strong assumptions about serial correlation that are necessary for formal testing for fixed effects in the presence of measurement error (Holtz-Eakin, Newey, and Rosen 1988).

Panel D presents the IV and AB results in the case of a serially correlated measurement error, in which we let ϕ_u in (17) range between 0 and 1. Here, we also report the rejection rate of the AB $m2$ test, which detects serially correlated regression residuals. The bottom left figure shows that even small amounts of serial correlation result in enough bias to make the results from either of these estimators resemble the biased OLS results from Table 3. Further, neither the test of overidentifying restrictions nor the $m2$ test has useful power to detect misspecification. The power curves depicted in this figure are low even for extreme amounts of serial correlation. This result is important because it means that these two instrumental variable estimators are not robust to this source of misspecification and that it is difficult to tell that anything is wrong via traditional specification tests.[13] In contrast, in these Monte Carlos, the EW estimators remain unbiased in the presence of a serially correlated measurement error.

2.4 Identification diagnostics

Next, we use Monte Carlos to examine the power and size of the identification diagnostic test in Erickson and Whited (2000, 2002). This is a test of the joint null hypothesis that $\beta = 0$ and $E\left(\dot{\chi}^3\right) = 0$. These two conditions are necessary to obtain an estimator from (9). However, this test detects only skewness. Because the EW estimators also obtain identification from kurtosis, we also examine a new test, derived in the Appendix, that detects both skewness and kurtosis. Both tests are directly analogous to a weak instrument

[12] One can conduct a Wald test of the null hypothesis that the EW estimators produce identical estimates on level data and within-transformed data. Unfortunately, this test overrejects strongly in samples of the sizes we consider.

[13] This problem also affects the IV estimators proposed by Ağca and Mozumdar (2010), which use lagged values of Tobin's q as instruments for an analyst-based measure of q from Cummins, Hassett, and Oliner (2006), and vice versa. This method only works if the measurement errors have limited serial correlation, and if these two measures of q are independent. The latter case is unlikely. The Cummins et al. measure is based on analysts' expectations of future profits, and Tobin's q is based on stock market expectations of future profits. To the extent that the market capitalizes analysts' opinions, the two measures cannot be independent.

diagnostic, given the interpretation, discussed above, of the GEARY estimator as an IV estimator.[14]

Our Monte Carlos are based on the q^{FHP} DGP. To consider varying degrees of skewness and kurtosis of the unobserved mismeasured regressor, χ_{it}, we let the shape parameter vary as in Figure 1 so that skewness of χ_{it} ranges from 19 to 0.2, but we hold the variance of χ_{it} constant. We also consider our two sample sizes of 1,500 and 4,000. These design features are important because we are interested in test power, and power is not a single number, but a function of sample size and the distance of the true parameter from its hypothesized null value.

Before conducting this experiment, we first change the cross-sectional distribution of χ_{it} to a normal. We find that a nominal 5% test rejects the null between 5.8% and 7.2% of the time for sample sizes of 1,500 and 4,000, both with and without a within transformation. We conclude that the tests are approximately correctly sized.

Figure 3 depicts the power of the tests. It contains four plots arranged in two panels, with Panel A corresponding to a sample size of 1,500 and Panel B corresponding to a sample size of 4,000. The plots on the left side of each panel are for data in levels, and those on the right are for within-transformed data.

Two patterns in Figure 3 are of note. First, we find that the new test is usually slightly more powerful than the old test. Second, both tests have excellent power for intermediate levels of skewness when the data are in levels, but this power drops when the data are in deviations from individual means, and the drop is more pronounced for a sample size of 1,500.

The most striking feature of Figure 3 is the hump-shaped power curves. All four panels depict low power for both extremely high and extremely low skewness and higher power for intermediate levels of skewness. The low power for low skewness distributions makes immediate sense, and the low power for high skewness can be understood as follows. Increasing the skewness of a gamma distribution also increases the kurtosis. The data therefore contain many more "outliers," and this increased noise in the data lowers the power of the tests to detect nonnormalities.

A comparison of Figures 1 and 3 also shows that the low power of the identification diagnostics for high levels of skewness can be accompanied by low bias and low dispersion of the EW estimators themselves. This finding means that these tests are conservative in that rejection of the null is sufficient but by no means necessary for an expectation of good estimator performance. Finally, and most importantly, if a researcher has a strong prior from economic theory, then that information may dominate the test results in a Bayesian sense, especially given our result that the estimators can perform well even when the

[14] Both ACG and Ağca and Mozumdar (2010) misinterpret this test as a formal specification test.

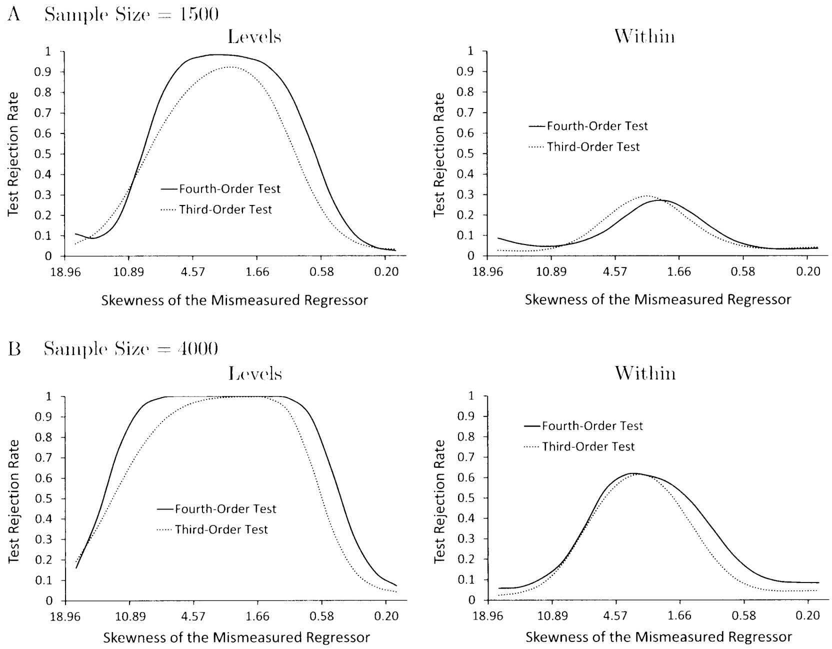

Figure 3
Power of identification diagnostics
Indicated test rejection rates are from Monte Carlo experiments based on the design used to generate Tables 3 and 5. The figures plot the probability that a nominal 5% identification diagnostic rejects the null of an unidentified model. "Third-Order Test" refers to the test from Erickson and Whited (2000, 2002) that detects skewness. "Fourth-Order Test" refers to the test described in the Appendix that detects both skewness and kurtosis.

test fails to produce a rejection. For example, as explained above, economic theory implies that marginal q is likely to be highly skewed.

3. The ACG Monte Carlos

ACG also use Monte Carlo experiments to assess the finite-sample performance of the EW estimators. They find, as we do, that the estimators are biased if one violates the assumptions that ensure consistency. However, this general problem is true of *all* estimators. More striking are their claims of bias and large estimator dispersion when no estimator assumptions are violated.

One immediate difference between the two studies is the general design. In particular, the simulated data in ACG do not approximate actual data on investment, q, and cash flow. Also, the regression coefficients they consider have values of either 1 or -1, which are much larger in absolute value than the small coefficients one generally obtains in investment regressions. Although these design features render the ACG Monte Carlos uninformative about investment regressions, these features are not the main reason behind why their results differ from ours. Instead, their conclusions are inaccurate because of three further features of their Monte Carlo design.

First, ACG's Monte Carlos produce a spurious evaluation of bias because they use the GEARY estimator as a starting value to compute the GMM4 and GMM5 estimators, without determining whether the resulting estimates are from a global minimum. The GEARY estimator has a large variance in some of the ACG Monte Carlos, so it occasionally produces extreme estimates in some of the individual Monte Carlo trials. In turn, these extreme starting values then cause GMM4 and GMM5 to produce occasional extreme estimates, each of which corresponds to a local minimum. The overall Monte Carlo estimate of bias can then be traced to these few extreme estimates, because the rest of the individual Monte Carlo trials yield nearly unbiased coefficient estimates.

We show that trying just two alternative starting values eliminates bias by eliminating the local minima that correspond to extreme estimates. As in our own Monte Carlos, we use both the Geary estimator and the OLS estimator, and we then pick the estimate corresponding to the lower of the two GMM objective functions.

The starting value problem implies that different runs of the same Monte Carlo program can produce average biases that differ both in magnitude and in sign. Evaluation of bias from a Monte Carlo should be approximately the same when one runs the Monte Carlo repeatedly, allowing the simulated samples to consist of different sets of random draws in each run. However, running the ACG Monte Carlos repeatedly yields different sets of random draws, which yield different collections of extreme estimates and therefore different evaluations of bias.

Exacerbating this problem is a second design issue that makes many of ACG's results difficult to replicate, even approximately. ACG do not use a random number generator in which one can set the seed. Computer-generated random numbers are deterministic sequences, in which an arbitrary number—the seed—is used to calculate the first number in the sequence. In the absence of a set seed, the computer software chooses an unrecoverable seed, and it is therefore impossible to find the run of the Monte Carlo corresponding to the ACG results, and it is extremely difficult to approximate many of their results.

Third, ACG calculate root mean squared error (RMSE) incorrectly. This calculation can grossly overstate true RMSE, thereby making the EW estimators appear very inaccurate. As explained in Kennedy (2003) or Shanken and Zhou (2007), the Monte Carlo estimate of RMSE is the square root of the average over the Monte Carlo trials of the squared deviations of the estimates from the true parameter value. Instead, ACG calculate the average over the Monte Carlo trials of the asymptotic standard errors, square that average, add it to the Monte Carlo estimate of bias, and then take the square root of the sum. This calculation is incorrect in part because they swap the order of averaging and squaring. More important ly, this calculation incorrectly equates an asymptotic standard error with a finite-sample estimator standard deviation.

To establish the issues with starting values and RMSE, we reconsider the results from Panel C of their Table 2, in which they examine the performance of the EW estimators when χ_i has an F distribution with 10 and 40 degrees of freedom, which is close to a normal distribution, only having a coefficient of skewness of 1.3. We choose these results because they are the worst results ACG report under correct specification.

Table 6 depicts the starting value and RMSE issues. We generate all results using the original ACG Monte Carlo programs, which we augment to include a seed for the random number generator. The table contains four panels. For Panels A and B, we use a random number generator seed of 1, and for Panels C and D we use a seed of 2. Panels A and C on the left report results in which we use the GEARY estimate as a starting value for GMM4 and GMM5, as do ACG. Panels B and D on the right report results in which we use both the Geary and the OLS estimate as starting values and then choose the resulting GMM estimate corresponding to the smaller minimized objective function.

The first and most important result can be seen by comparing Panels A and C of Table 6, which reveals that different runs of the program can produce large differences in the Monte Carlo estimates of coefficient bias. For example, with a true coefficient value of 1, the bias for the GMM4 estimate of β switches from -0.121 to -0.932. Indeed, 200 runs of their program with seeds numbered 1 through 200 produce biases for the GMM4 estimator of β that range between -35.3 and 0.795. Further, in only one out of these 200 runs are the estimates of bias in β for all three estimators within 20% of the estimates they report, and no run produces corresponding estimates of RMSE (their calculation) within 20% of their results.

In contrast, Panels B and D of Table 6 show that using two different starting values for GMM4 and GMM5 results in almost no bias. Also, the seed of the random number generator no longer matters materially because we eliminate the extreme estimates that come from local minima. Therefore, we conclude that if ACG had computed the overidentified EW estimators correctly, they would have found that they perform well.

The second notable result in Panels A and C of Table 6 is the large difference between the correct calculation and ACG's calculation of RMSE. The latter is usually much larger than the former, sometimes up to 26 times as large. In contrast, in Panels B and D, we find that both the correct and incorrect way of calculating RMSE produce low estimates that are close to each other. It is therefore the interaction of the starting value issue and the calculation error that results in extremely large estimates of RMSE. This result also means that the asymptotic standard error is not a good approximation of the estimator standard deviation when the estimate comes from a local minima, but that it is a good approximation when the estimate comes from a global minimum.

Finally, the Monte Carlo estimates of bias for the GEARY estimator in Panels B and D of Table 6 are identical to those in Panels A and C. The

Table 6
Reexamining the ACG Monte Carlo experiments

		Panel A Seed = 1 GEARY Starting Value				Panel B Seed = 1 GEARY and OLS Starting Values			
		β	α_1	α_2	α_3	β	α_1	α_2	α_3
GEARY	Bias	0.331	−0.084	−0.090	−0.078	0.331	−0.084	−0.090	−0.078
	ACG RMSE	130.888	27.938	37.201	29.905	130.888	27.938	37.201	29.905
	RMSE	10.059	2.202	2.804	2.490	10.059	2.202	2.804	2.490
GMM4	Bias	−0.121	0.029	0.031	0.028	0.020	−0.007	−0.004	−0.005
	ACG RMSE	4.384	0.973	1.328	0.962	0.224	0.081	0.079	0.079
	RMSE	2.385	0.566	0.690	0.530	0.299	0.105	0.093	0.096
GMM5	Bias	−0.105	0.023	0.028	0.024	0.013	−0.004	−0.003	−0.003
	ACG RMSE	2.789	0.728	0.707	0.699	0.157	0.067	0.067	0.067
	RMSE	3.120	0.903	0.803	0.732	0.188	0.076	0.075	0.077
		Panel C Seed = 2 GEARY Starting Value				Panel D Seed = 2 GEARY and OLS Starting Values			
		β	α_1	α_2	α_3	β	α_1	α_2	α_3
GEARY	Bias	−0.541	0.199	0.111	0.076	−0.541	0.199	0.111	0.076
	ACG RMSE	1286.938	406.270	239.596	220.821	1286.938	406.270	239.596	220.821
	RMSE	42.690	14.730	8.362	7.800	42.690	14.730	8.362	7.800
GMM4	Bias	−0.932	0.301	0.195	0.185	0.025	−0.007	−0.005	−0.007
	ACG RMSE	154.723	53.217	29.568	28.368	0.241	0.085	0.082	0.084
	RMSE	42.876	15.063	8.023	7.724	0.369	0.115	0.094	0.111
GMM5	Bias	−0.552	0.144	0.149	0.115	0.016	−0.005	−0.003	−0.005
	ACG RMSE	22.495	7.384	4.815	4.179	0.158	0.067	0.067	0.067
	RMSE	23.765	6.404	6.872	4.129	0.190	0.076	0.075	0.075

Indicated expectations and probabilities are estimates based on 5,000 Monte Carlo samples of size 1,000. The samples are generated by

$$y_i = \alpha_0 + z_{i1}\alpha_1 + z_{i2}\alpha_2 + z_{i3}\alpha_3 + \chi_i\beta + u_i$$
$$x_i = \gamma + \chi_i + \varepsilon_i,$$

in which $(\chi_i, z_{i1}, z_{i2}, z_{i3})$ are each distributed as an F with 10 and 40 degrees of freedom. ε_i and u_i are lognormal variables. GMM*n* denotes the GMM estimator based on moments up to order $M = n$. RMSE indicates "Root Mean Squared Error," and is the Monte Carlo estimate of this quantity. ACG RMSE is the square root of the following. The asymptotic standard error is averaged over Monte Carlo trials, squared, and added to the square of the Monte Carlo estimate of bias. "GEARY and OLS Starting Values" indicates that both the OLS estimate and the GEARY estimate are used as starting values for GMM4 and GMM5. The resulting coefficient estimates correspond to the estimation with the lower objective function value. "Seed = 1" and "Seed = 2" refer to the seeds of the Matlab 7.9.0.529 random number generator. **True Values:** $\beta = 1$, $\alpha_1 = -1$, $\alpha_2 = 1$, and $\alpha_3 = -1$.

reason is that this estimator is exactly identified and does not depend on starting values. To investigate this issue further, we run two simulations of the GEARY estimator with 1,000,000 trials. We still find different biases, again because this estimator occasionally produces extreme estimates. Interestingly, this result does not mean that the estimator is necessarily biased. Instead, in this Monte Carlo, either this estimator has no moments, or it has such a large variance that more than 1,000,000 trials are needed to estimate bias accurately. What can be said is that the GEARY estimator is therefore unreliable in the case of the specific distributional assumptions in this particular Monte Carlo and for a sample size of 1,000 (the ACG choice).

This result does not mean that the GEARY estimator is universally unreliable. For example, 5,000 trials is sufficient to estimate median (as opposed to mean) bias near zero. Also, increasing the sample size from 1,000 to 5,000 results in a Monte Carlo estimate of mean bias near zero. Nonetheless, the comparison of the GEARY estimator with the GMM4 and GMM5 estimators shows that the idea from Erickson and Whited (2002) of combining the information in many moment conditions can greatly improve finite sample performance.

We address one more result in ACG's Monte Carlos, which is serious bias in the EW estimators when they model a heteroskedastic regression error. These conclusions are from an experiment in which the *variance* of the regression error increases at the same time as does heteroskedasticity. It is therefore impossible to isolate heteroskedasticity. However, when we rerun their program but hold the variance constant while increasing heteroskedasticity, we find small coefficient biases that range from -0.12 to -0.17 for β and from 0.027 to 0.045 for α. For this particular simulation, heteroskedasticity has little effect on the EW estimators. The erratic behavior of these estimators observed by ACG arises in part from starting value problems and in part from the increased error variance. As explained above, the second issue affects the EW estimators more than traditional estimators because a great deal of data is required to estimate high-order moments precisely. Thus, the heteroskedasticity Monte Carlos in ACG confirm the fact that high-order moment estimators require more data than estimators based on first- and second-order moments.

In general, starting value problems afflict all of the ACG Monte Carlos of the overidentified EW estimators. The problem is especially severe for their Monte Carlos of the EW estimators under correct specification, as the reported bias is virtually all due to including extreme estimates from non-global minima in the averaging. In contrast, the ACG Monte Carlo result that EW estimators are biased under the misspecification of correlated regressor and fixed effects is sustained, although the purging of local-minima estimates changes the results somewhat. However, as explained above in Section 2.3, this result is not surprising because it comes from violation of the identifying assumption that the error and regressor are uncorrelated. The result is also not new. It is discussed in Erickson and Whited (2000), and Riddick and Whited

(2009) conduct Monte Carlos of the EW estimators with correlated errors and regressors.

4. Data Analysis

In this section, we analyze a sample of firms from Compustat. We also explain the differences between our real-data results and those in ACG. This discussion leads to the formulation of practical advice in implementing the EW estimators.

4.1 Our data analysis

As described in Section 2.1, in our data analysis we consider two different regression specifications: one in which all regressors are deflated by the capital stock and in which we use q^{FHP}, and another in which all variables are deflated by total assets and in which we use market-to-book. For both cases, we consider data in levels and within-transformed data.

First, we discuss the results from the two identification diagnostic tests. If we deflate the data by the capital stock, for the data in levels, the old test rejects in 71% of the 42 years in our panel, and the new test rejects in 74%. For the data in deviations from firm means, both tests reject in 29% of the years. For the data deflated by total assets, both tests reject between 29% and 31% of the time for both data in levels and in deviations from firm means.

Our estimation results are in Table 7. We obtain these results by using a fixed-effects OLS estimator with clustered standard errors, a simple IV estimator in differenced form with twice- and three-times-lagged levels of q and cash flow for instruments, and the AB estimator with twice-lagged levels of q and cash flow as instruments. (The IV and AB estimates are largely unaffected by including more or fewer lags of q and cash flow as instruments.) The EW estimates are computed by first searching over a grid of starting values to obtain global minima in each year of data, and then by pooling the yearly estimates via the minimum distance procedure described in the Appendix.

Table 7 contains four panels. Panel A presents results from the capital-stock specification in which the variables are in levels, and Panel B presents results from this specification, except that the variables have undergone a fixed-effects transformation. Panels C and D are analogous to Panels A and B, except that they contain results for the total-assets specification.

Several general patterns stand out in Table 7. First, all of the slope coefficient estimates are economically reasonable, in the sense that they imply positive average elasticities that are less than one. Second, the EW estimates of the coefficients on q are larger than their OLS counterpart, and the EW estimates of the cash flow coefficient are smaller than their OLS counterparts. The IV and AB estimates lie between the EW and OLS estimates. Third, the EW estimators deliver estimates of the regression R^2 that are larger than the OLS R^2. All three patterns are symptomatic of the attenuation bias on β given by OLS and of the

Table 7
Regressions of investment on Tobin's q and cash flow

Panel A
Levels, Capital Stock Deflator

	OLS	GEARY	GMM4	GMM5	IV	AB
q	0.009**	0.015*	0.010*	0.010		
	(0.000)	(0.001)	(0.000)	(0.000)		
Cash flow	0.068**	0.014	0.018	0.015		
	(0.003)	(0.005)	(0.004)	(0.004)		
ρ^2	0.184	0.318*	0.309*	0.305*		
		(0.011)	(0.010)	(0.009)		
τ^2		0.555*	0.594*	0.564*		
		(0.017)	(0.017)	(0.017)		
Fraction significant cash flow coefficients		0.071	0.119	0.119		
J-test rejections			0.119	0.238		

Panel B
Transformed, Capital Stock Deflator

	OLS	GEARY	GMM4	GMM5	IV	AB
q	0.009**	0.018*	0.016*	0.011	0.013**	0.008**
	(0.000)	(0.001)	(0.001)	(0.000)	(0.002)	(0.001)
Cash flow	0.070**	0.017	0.026	0.027	0.040**	0.051**
	(0.004)	(0.006)	(0.004)	(0.004)	(0.004)	(0.006)
ρ^2	0.170	0.247*	0.213*	0.195*		
		(0.013)	(0.009)	(0.007)		
τ^2		0.332*	0.340*	0.293*		
		(0.019)	(0.019)	(0.017)		
Fraction significant cash flow coefficients		0.024	0.167	0.262		
J-test rejections/(p-value)			0.095	0.048	(0.000)	(0.000)

(continued)

Table 7
Continued

	OLS	GEARY	GMM4	GMM5	IV	AB
Panel C						
Levels, Total Assets Deflator						
Market-to-book	0.006**	0.038*	0.036*	0.033*		
	(0.001)	(0.003)	(0.001)	(0.001)		
Cash flow	0.147**	0.070*	0.079*	0.078*		
	(0.006)	(0.008)	(0.007)	(0.007)		
ρ^2	0.140	0.169*	0.180*	0.177*		
		(0.010)	(0.009)	(0.008)		
τ^2		0.213*	0.099*	0.105*		
		(0.016)	(0.010)	(0.010)		
Fraction of significant cash flow coefficients		0.119	0.333	0.357		
J-test rejections			0.024	0.024		
Panel D						
Transformed, Total Assets Deflator						
Market-to-book	0.010**	0.040*	0.032*	0.019	0.020**	0.010**
	(0.001)	(0.003)	(0.001)	(0.001)	(0.002)	(0.002)
Cash flow	0.115**	0.013	0.046	0.092*	0.056**	0.065**
	(0.005)	(0.010)	(0.007)	(0.009)	(0.004)	(0.007)
ρ^2	0.130	0.152*	0.120*	0.125*		
		(0.011)	(0.008)	(0.008)		
τ^2		0.150*	0.109*	0.092*		
		(0.014)	(0.010)	(0.010)		
Fraction of significant cash flow coefficients		0.071	0.261	0.262		
J-test rejections/(p-value)			0.024	0.143	(0.000)	(0.000)

Calculations are based on a sample of manufacturing firms from the 2009 Compustat Industrial Files. The sample period is from 1967 to 2008. All estimates are of a regression of investment on a proxy for Tobin's q and cash flow. GMMn denotes the GMM estimator based on moments up to order $M = n$, and GEARY denotes the third-order moment estimator from Geary (1942). All estimates are obtained by pooling yearly estimates via minimum distance. OLS denotes ordinary least squares estimation. IV denotes instrumental variable estimation, with two and three lags of cash flow and the proxy for Tobin's q as instruments for the first difference of the proxy for Tobin's q. AB denotes estimation of this same first-differenced regression with the same instruments, but using the GMM technique in Arellano and Bond (1991). In the top two panels, investment and cash flow are deflated by the gross capital stock. In the bottom two panels, investment and cash flow are deflated by total assets. "Transformed" refers to data that have undergone the within transformation in the case of the EW estimators and to data that are first differenced in the case of the IV and AB estimators. The variable q is a proxy for Tobin's q intended to isolate variation in investment opportunities for property, plant, and equipment. "Market-to-book" is a proxy for Tobin's q intended to isolate variation in investment opportunities for total assets. ρ^2 is an estimate of the R^2 of the regression. τ^2 is an index of measurement quality for the proxy for Tobin's q that lies between 0 and 1. "J-test rejections" refers to the fraction of years in which the tests of the overidentifying restrictions of the model produce rejections. For the IV and AB estimators J-test p-values are reported in parentheses. Standard errors are in parentheses under the parameter estimates. An asterisk indicates 5% significance using bootstrapped critical values. Two asterisks indicate 5% significance using asymptotic critical values.

positive correlation between q and cash flow. Fourth, the estimates of τ^2 for q^{FHP} are approximately 0.5 and those for market-to-book are between 0.1 and 0.2, indicating that the former has higher information content than the latter. This difference in information content undoubtedly explains the low rejection rate of the identification diagnostics in the case of the market-to-book ratio. If this proxy has little information content, then these tests have a difficult time detecting any nonnormality in true q. Fifth, the estimates of ρ^2 and τ^2 are lower in the case of the within-transformed data, thus pointing to differences in the cross-sectional and within variation in our data. Finally, all of the tests of overidentifying restrictions for the IV and AB estimators reject.

To allay concerns about heteroskedasticity, we note two results.[15] First, few tests of overidentifying restrictions for the EW estimators reject, even for the data in levels, a result that should be interpreted in light of the Monte Carlo evidence from Section 2.3 showing that the GMM5 J-test has useful power to detect heteroskedasticity. Second, our Monte Carlos indicate that the GMM4 and GMM5 estimators are not affected strongly by heteroskedasticity, and our qualitative conclusions can be established by looking only at the results from these estimators.

One striking result is that most estimates of the slope coefficients from the data in levels are not *economically* different from the results from the within-transformed data. This finding suggests a correlation between fixed effects and the regressors is unlikely to have contaminated our results. If fixed effects were correlated with the regressors, then the EW estimator assumptions would be violated for the estimations from the first panel but not for the estimations from the second panel, and we ought to have seen different results.

The results on the cash flow coefficients are of particular interest. First, the OLS estimates of this coefficient are smaller than those reported by Fazzari, Hubbard, and Petersen (1988) or Erickson and Whited (2000). This result is consistent with the tendency for cash flow coefficients to have decreased over time, documented in Chen and Chen (2011). Also consistent with the results in Chen and Chen (2011) is our finding that only a few of the underlying *yearly* EW estimates are significant. Third, the AB and IV estimates are significantly greater than zero, and they tend to be closer to the OLS estimates than the EW estimates. Fourth, using 5% critical values from the bootstrap in Hall and Horowitz (1996), we find statistically insignificant coefficients when we apply the EW estimators to the data in levels that are constructed with the capital stock deflator. A bootstrap is warranted by the overrejection of the asymptotic tests in our Monte Carlos. This result confirms a similar one in

15 We do not use standard tests for heteroskedasticity because they are based on the assumption of a normal error. They can therefore over reject the null of no heteroskedasticity when the error is homoskedastically but nonnormally distributed.

Erickson and Whited (2000).[16] Fifth, in contrast to the results in Erickson and Whited (2000), for the remaining three data specifications, only five out of nine EW estimates of the cash flow coefficients are insignificantly different from zero. Again, this result comes from using the bootstrapped critical values.

Although we do find some evidence of positive cash flow coefficients, the evidence is not overwhelming. Also, we do not conclude that cash flow sensitivities are good indicators of finance constraints. Recent advances in dynamic theory have demonstrated that these coefficients are hard to interpret because they can stem from many different sources, only one of which is finance constraints (Gomes 2001; Moyen 2004; Hennessy and Whited 2007; Abel and Eberly 2011).

4.2 The ACG data analysis

Our data results differ sharply from the data results in ACG, who report that the EW estimators deliver economically implausible results when applied to data on investment, Tobin's q, and cash flow. In particular, they find that GMM4 and GMM5 produce negative estimates of the coefficients on q and cash flow. These negative values are inconsistent with economic theory, which predicts nonnegative coefficients. These results are puzzling not only because they differ from our results, but because they differ sharply from the plausible data estimates obtained by several empirical researchers, such as Hennessy (2004), Alderson and Betker (2009), Chen and Chen (2010), and Papanikolaou and Panousi (2011), as well as Erickson and Whited (2000), Whited (2001), Riddick and Whited (2009), and Bakke and Whited (2010). We therefore seek to understand these differences and, at the same time, provide practical guidance to applied researchers in implementing the EW estimators—guidance that is missing from Erickson and Whited (2000).

The implausible coefficients reported by ACG are in part due to their use of the GEARY estimates as starting values for GMM4 and GMM5. As is the case in their Monte Carlos, this choice sometimes produces extreme results from local optima that strongly influence the Fama-MacBeth averages of the yearly EW estimates reported by ACG. We therefore recommend using many different starting values for β along a grid to find global minima, where the grid should contain the OLS estimate of β, as well as the GEARY estimate.

The starting value issue arises because the EW estimates are highly sensitive to starting value choices in the ACG data set. We conjecture that this problem stems from ACG's decision to deflate investment and cash flow by the net capital stock, while leaving the denominator of the market-to-book ratio as total assets. This decision is inconsistent with basic q theory, which implies that the same divisor should be used for all regression variables (Hayashi and

[16] This result differs from the finding in Ağca and Mozumdar (2010) that this specification yields significant cash flow coefficients. The difference arises because their results are an artifact of one or two tail observations per year.

Inoue 1991). Further, if firms differ in their capital intensity, this problem introduces unnecessary heteroskedasticity into the regression. As described in Section 2.3, when we simulate a heteroskedastic regression error, the dispersion of the EW estimators rises. Therefore, the different divisors problem likely makes it more difficult than necessary to find global optima.

These data construction choices also mechanically alter the correlation structure of investment, q, and cash flow in ways that do not reflect any economic forces. First, using different divisors for q and investment breaks their natural positive correlation, which leads to lower estimates of β. This issue implies that in these data the identification assumption that $\beta \neq 0$ is close to being violated, and no estimator provides reliable estimates when it is nearly unidentified. At the same time, the common divisor of investment and cash flow, the net capital stock, drifts down over time because of accumulated depreciation. This drift can inflate the correlation between investment and cash flow.[17] Although one can think of examples of individual firms in which one might for various reasons use different denominators, for the analysis of a large cross-section, we therefore advise using the same divisor for all regression variables.

5. Conclusion

We use Monte Carlos to compare the finite sample performance of three econometric measurement error remedies: the high-order moment estimators from Erickson and Whited (2000, 2002), simple IV estimators, and the dynamic panel estimators in Arellano and Bond (1991), where the second two use lagged values of cash flow and Tobin's q as instruments for the current first-difference of Tobin's q.

Our Monte Carlos show that under correct specification, the EW estimators sometimes outperform the simple IV estimators, and vice versa. Both estimators usually dominate the AB estimators. More interesting are our findings on the performance of the three estimators under misspecification. In particular, we find that estimators using lagged instruments are not robust to serial correlation of the measurement error, which can invalidate the instruments when the degree of serial correlation is larger than the number of times one lags the instruments. This problem causes these estimators to deliver the same biased estimates as does OLS, and standard specification tests cannot detect the bias. We also find that untreated fixed effects can seriously bias the EW estimators, but that doing a within transformation on the data removes much of this bias, although sometimes at the cost of increased variance for smaller sample sizes. We therefore recommend using the EW estimators on data both

[17] This data construction choice also drives some of the large cash flow coefficients in Ağca and Mozumdar (2010).

in levels and in deviations from firm means to determine if the two sets of estimates are economically different. Finally, we find that heteroskedasticity can also bias some of the EW estimators but that the bias for the overidentified EW estimators is not large enough to distort qualitative inference. Further, the *J*-test has useful power to detect heteroskedasticity.

This comparison is partly motivated by Almeida, Campello, and Galvao (2010), who claim that the EW estimators have poor finite sample performance relative to the AB and IV estimators. However, we show that their simulation results are mostly due to miscalculating the overidentified EW estimators by using a restrictive starting value procedure that does not include checks to see if the computed minimums of the GMM objective function are merely local rather than global. One consequence of this miscalculation is that different runs of the same Monte Carlo simulation can produce materially different results, and the serendipitous lesson from this is that the EW estimators can be sensitive to starting values, as is often true of nonlinear estimators. We remedy this starting value problem for Monte Carlo experimentation by using *two* starting values for each simulated data set and then choosing which of the two resulting estimates gives the smaller objective function value. When we do this, the appearance of bias found by ACG disappears.

Another contribution is development of a minimum distance technique that extends the EW estimators to unbalanced panels, thereby increasing their practical usefulness as a robustness test for measurement error. We demonstrate this technique on a new data set, illustrating how to be careful with fixed effects and starting values. We find that fixed effects do not matter in an economically material way for these kinds of investment regressions. We also find economically reasonable coefficient estimates, as well as limited evidence of positive cash flow coefficients.

Finally, neither Erickson and Whited (2000, 2002) nor this article recommend that the EW estimators be used whenever a researcher suspects the presence of measurement error. Applying either IV or the EW estimators to any regression is only an imperfect solution to the problem of mismeasured regressors. We therefore recommend careful consideration of the assumptions behind all of these techniques before applying them. As in Erickson and Whited (2000), Whited (2001), and Riddick and Whited (2009), we also recommend examining their finite sample performance in data-relevant Monte Carlos before using them on new data, especially if the sample size differs greatly from the ones we consider here or if the mismeasured regressor appears to be nearly normally distributed. Nonetheless, in the particular case of firm-level investment regressions, the results here mirror those in Erickson and Whited (2000) by showing that the EW estimators are appropriate for handling measurement error in Tobin's q. Because Tobin's q appears in many other corporate finance regressions, the EW estimators might be useful in those applications as well.

Appendix

This Appendix describes our data, presents our new identification diagnostic test, and outlines our method for pooling an unbalanced panel.

Data

Our sample comes from the universe of all manufacturing firms on the Annual 2009 Compustat Files from 1967 to 2008. We screen firm-year observations as in ACG, except that we require the gross (instead of the net) capital stock to be greater than 5 million real 1982 dollars. We screen on the gross (instead of the net) capital stock because we use the gross capital stock as the scaling variable in most of our regressions. We also use two further sensible screens: We require that the variables needed to construct our regression variables not be missing, and we winsorize our regression variables over the entire panel at once at the conservative 0.1% level to remove extreme outliers. We end up with a sample that contains between 631 and 1,587 firms per year. We construct investment as Compustat item CAPX. Cash flow is the sum of items IB and DPR. For our first set of regressions, we construct Tobin's q as a variant of q^{FHP} given by (14). The numerator is DLTT plus DLC plus PRCC_F times CSHO minus AC. The denominator is PPEGT. We deflate investment and cash flow by PPEGT. For our second set of regressions, we deflate investment and cash flow by total assets (AT), and we construct the numerator of the market-to-book ratio as TA+PRCC_F times CSHO minus CEQ minus TXDB, where these last two variables are book common equity and balance sheet deferred taxes.

It is possible to use a perpetual inventory algorithm to estimate the replacement cost of capital or a recursive algorithm to estimate the market value of debt. However, Erickson and Whited (2006) demonstrate that these types of algorithms add little in terms of the measurement quality of various proxies for marginal q, or even more directly, for the replacement value of the capital stock. We therefore stick with the book values.

Identification Diagnostic Test

The identification diagnostic test from Erickson and Whited (2000, 2002) is based on third-order moments and therefore picks up skewness in the data. However, the EW estimators also obtain identification from kurtosis. We therefore formulate a more powerful identification test based on fourth-order moments that can pick up this source of identification. Like the original test, this test is also for separate cross-sections in a panel.

A sufficient condition for nonnormality is that at least one cumulant of $\dot{\chi}_i$ of order three or higher be nonzero. To exploit a non-zero fourth-order cumulant, we derive moment conditions analogous to (7) and (8) based on the moments $E\left(\dot{y}_i{}^3\dot{x}_i\right)$, $E\left(\dot{y}_i{}^2\dot{x}_i{}^2\right)$, and $E\left(\dot{y}_i\dot{x}_i{}^3\right)$, which are as follows:

$$E\left(\dot{y}_i{}^3\dot{x}_i\right) - 3E(\dot{y}_i\dot{x}_i)\,E\left(\dot{y}_i{}^2\right) = \beta^3\left[E\left(\dot{\chi}_i^4\right) - 3E\left(\dot{\chi}_i^2\right)^2\right] \tag{A1}$$

$$E\left(\dot{y}_i{}^2\dot{x}_i{}^2\right) - E\left(\dot{y}_i{}^2\right)E\left(\dot{x}_i{}^2\right) - 2E(\dot{y}_i\dot{x}_i)^2 = \beta^2\left[E\left(\dot{\chi}_i^4\right) - 3E\left(\dot{\chi}_i^2\right)^2\right] \tag{A2}$$

$$E\left(\dot{y}_i\dot{x}_i{}^3\right) - 3E(\dot{y}_i\dot{x}_i)\,E\left(\dot{x}_i{}^2\right) = \beta\left[E\left(\dot{\chi}_i^4\right) - 3E\left(\dot{\chi}_i^2\right)^2\right]. \tag{A3}$$

Just as the right-hand sides of (7) and (8) contain a power of β times the third cumulant of $\dot{\chi}_i$, the right-hand sides of (A1)–(A3) contain a power of β times the fourth cumulant of $\dot{\chi}_i$ Therefore, we can test the hypothesis that the fourth cumulant of $\dot{\chi}_i$ is nonzero by testing whether the left-hand sides of the above three equations are nonzero.

Our expanded identification test is then a test of the joint hypothesis that the left-hand sides of (7)–(8) and (A1)–(A3) are zero. We use the Wald test in Newey and McFadden (1994), where we adjust the weighting matrix as described in Erickson and Whited (2002).

Minimum Distance Estimation

The EW estimators can be applied only to samples that are arguably *i.i.d.* For panel data, Erickson and Whited (2000) suggest estimating the model on each cross-section of data, and then pooling the cross-sectional estimates via a minimum distance technique.

To simplify notation, we explain the minimum distance technique for a panel with only two time periods, which yields two estimates of β: $\hat{\beta}_1$ and $\hat{\beta}_2$. A more efficient estimator is the value of β that minimizes the weighted sum of the squared distances between itself and $\hat{\beta}_1$ and $\hat{\beta}_2$. The optimal weighting matrix for this minimum distance estimator is the inverse of the asymptotic covariance matrix of the vector $\left(\hat{\beta}_1, \hat{\beta}_2\right)$.

A nice feature of the estimator is that it does not require assumptions about a lack of serial correlation in the measurement errors or the regression errors because these correlations are accounted for by the covariances between the cross-sectional estimates being pooled. However, obtaining this covariance is not immediately obvious because $\hat{\beta}_1$ and $\hat{\beta}_2$ are not jointly estimated. Erickson and Whited (2000) solve this problem by covarying the influence functions for $\hat{\beta}_1$ and $\hat{\beta}_2$.

The influence function for an estimator is a function of an individual observation that has the useful property that its expected outer product is the asymptotic covariance matrix of the estimator. For example, for a single observation, w_i, let $g\left(w_i, \theta\right) \equiv g_i\left(\theta\right)$ be the moment function for the GMM estimator $\hat{\theta} = \operatorname{argmin}\left\{g_i\left(\theta\right)' W g_i\left(\theta\right)\right\}$, where $W \equiv \left\{E\left(g_i\left(\theta_0\right) g_i\left(\theta_0\right)'\right)\right\}^{-1}$ is the optimal GMM weighting matrix. Let G be the gradient of $g_i\left(\theta\right)$. The asymptotic variance of $\hat{\theta}$ is $\left(G'WG\right)^{-1}$, and the influence function for the GMM estimator is $\psi_i = \left(G'WG\right)^{-1} G'W g_i\left(\theta_0\right)$. It is straightforward to see that $E\left(\psi_i \psi_i'\right)$ equals the asymptotic variance of $\hat{\theta}$.

Let $w_i \equiv \left(x_i, y_i, z_i\right)$ be a single observation from the data from a single cross-section. Let $\psi\left(w_i\right) \equiv \psi_i$ be the influence function for our nonlinear estimator $\hat{\beta}$. Then, by definition, $\sqrt{n}\left(\hat{\beta} - \beta\right)$ converges in probability to the random variable $\sqrt{n}\overline{\psi}$, in which $\overline{\psi}$ is the sample mean of the influence function ψ_i. See Newey and McFadden (1994, eqs. 3.3 and 3.7). This property of influence functions implies that $\left(\hat{\beta} - \beta\right)$ has the same asymptotic distribution as $\overline{\psi}$, which in turn implies that $\hat{\beta}$ has the same asymptotic variance covariance matrix as $\overline{\psi}$.

The original minimum distance application in Erickson and Whited (2000) cannot be used on unbalanced panels, so we now generalize it to this case. Suppose the two-period panel is unbalanced, with $n_1 = n_c + n_o$ observations in period 1 and $n_2 = n_c + n_n$ observations in period 2, in which n_c is the number of continuing observations present in both periods, n_o is the number of old observations present only in period 1, and n_n is the number of new observations present only in period 2. Let $\overline{\psi}_1$ be the sample average influence function from the first period, and $\overline{\psi}_2$ be that from the second. These sample averages can be expressed as weighted averages of the sample mean influence functions for the old, continuing, and new observations as follows:

$$\bar{\psi}_1 = \frac{n_o}{n_1}\bar{\psi}_{1o} + \frac{n_c}{n_1}\bar{\psi}_{1c} \tag{A4}$$

$$\bar{\psi}_2 = \frac{n_n}{n_2}\bar{\psi}_{2n} + \frac{n_c}{n_2}\bar{\psi}_{2c}. \tag{A5}$$

Therefore, one can express the covariance between $\bar{\psi}_1$ and $\bar{\psi}_2$ as

$$\begin{aligned} \operatorname{cov}\left(\bar{\psi}_1, \bar{\psi}_2\right) &= \operatorname{cov}\left(\frac{n_o}{n_1}\bar{\psi}_{1o} + \frac{n_c}{n_1}\bar{\psi}_{1c}, \frac{n_n}{n_2}\bar{\psi}_{2n} + \frac{n_c}{n_2}\bar{\psi}_{2c}\right) \\ &= \frac{n_c^2}{n_1 n_2}\operatorname{cov}\left(\bar{\psi}_{1c}, \bar{\psi}_{2c}\right). \end{aligned} \tag{A6}$$

The second equality holds because the covariance between the sample means of the influence functions for nonoverlapping observations is zero.

Next, because $\hat{\beta}$ has the same asymptotic distribution as $\overline{\psi}$, we can write

$$\begin{aligned} \operatorname{cov}\left(\hat{\beta}_1, \hat{\beta}_2\right) &= \operatorname{cov}\left(\bar{\psi}_1, \bar{\psi}_2\right) \\ &= \frac{n_c^2}{n_1 n_2} \operatorname{cov}\left(\bar{\psi}_{1c}, \bar{\psi}_{2c}\right) \\ &= \frac{n_c^2}{n_1 n_2} E\left(\psi_{i1c}\psi_{i2c}'\right). \end{aligned} \tag{A7}$$

The last equality follows because influence functions by definition have population means of 0 and because the influence function for $\bar{\psi}_{1c}$ is trivially ψ_{i1c}, etc. With this covariance matrix for the individual cross-sectional estimates, the minimum distance estimation proceeds exactly as in Erickson and Whited (2000). For an analogous approach for unbalanced panels, see MaCurdy (2007).

It is also possible to pool the cross-sectional estimates from different time periods in a panel by stacking the moment conditions for each time period in the panel. One can then estimate a large system and constrain the parameters to be the same over time. This is essentially the approach taken by Arellano and Bond (1991) and Holtz-Eakin, Newey, and Rosen (1988). In limited Monte Carlo experiments, we have found that this efficient method of pooling performs approximately as well as the method we have described.

References

Abel, A. 1983. Optimal Investment Under Uncertainty. *American Economic Review* 73:228–33.

Abel, A., and J. Eberly. 2011. How *Q* and Cash Flow Affect Investment Without Frictions: An Analytic Explanation. *Review of Economic Studies* 78:1179–1200.

Alderson, M. J., and B. L. Betker. 2009. Were Internal Capital Markets Affected by the "Perfect" Pension Storm? *Journal of Corporate Finance* 15:257–71.

Ağca, S., and A. Mozumdar. 2010. Investment-cash Flow Sensitivity: Fact or Fiction? Working Paper, George Washington University.

Aigner, D. J., C. Hsiao, A. Kapteyn, and T. Wansbeek. 1984. Latent Variable Models in Econometrics. In Z. Griliches and M. D. Intriligator (eds.), *Handbook of Econometrics*, vol. 2, pp. 1321–93. Amsterdam: Elsevier.

Almeida, H., and M. Campello. 2007. Financial Constraints, Asset Tangibility, and Corporate Investment. *Review of Financial Studies* 20:1429–60.

Almeida, H., M. Campello, and A. F. Galvao, Jr. 2010. Measurement Errors in Investment Equations. *Review of Financial Studies* 23:3279–328.

———. 2011. More on the Performance of Higher-order Moment Estimators in Investment Equations. Working Paper, University of Illinois.

Arellano, M., and S. Bond. 1991. Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies* 58:277–97.

Bakke, T.-E., and T. M. Whited. 2010. Which Firms Follow the Market? An Analysis of Corporate Investment Decisions. *Review of Financial Studies* 23:1941–80.

Bekaert, G., and E. Engstrom. 2011. Asset Return Dynamics Under Bad Environment-Good Environment Fundamentals. Working Paper, Columbia University.

Blundell, R., and S. Bond. 1998. Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. *Journal of Econometrics* 87:115–43.

Chava, S., and M. R. Roberts. 2008. How Does Financing Impact Investment? The Role of Debt Covenant Violations. *Journal of Finance* 63:2085–121.

Chen, H., and S. Chen. Forthcoming 2011. Investment-cash Flow Sensitivity Cannot Be a Good Measure of Financial Constraints: Evidence from the Time Series. *Journal of Financial Economics*.

Cummins, J., K. Hassett, and S. S. Oliner. 2006. Investment Behavior, Observable Expectations, and Internal Funds. *American Economic Review* 96:796–810.

Erickson, T., and T. M. Whited. 2000. Measurement Error and the Relationship Between Investment and q. *Journal of Political Economy* 108:1027–57.

Erickson, T., and T. M. Whited. 2002. Two-step GMM Estimation of the Errors-in-variables Model Using High-order Moments. *Econometric Theory* 18:776–99.

———. 2005. Proxy Quality Thresholds: Theory and Applications. *Finance Research Letters* 2:131–51.

———. 2006. On the Accuracy of Different Measures of Q. *Financial Management* 35:5–33.

Fama, E., and J. MacBeth. 1973. Risk, Return, and Equilibrium: Empirical Tests. *Journal of Political Economy* 81:607–36.

Fazzari, S. M., R. G. Hubbard, and B. C. Petersen. 1988. Financing Constraints and Corporate Investment. *Brookings Papers on Economic Activity* 1:141–95.

Gomes, J. 2001. Financing Investment. *American Economic Review* 91:1263–85.

Geary, R. C. 1942. Inherent Relations Between Random Variables. *Proceedings of the Royal Irish Academy A* 47:63–76.

Hall, P., and J. L. Horowitz. 1996. Bootstrap Critical Values for Tests Based on Generalized-method-of-moment Estimators. *Econometrica* 64:891–916.

Han, C., and P. C. B. Phillips. 2010. GMM Estimator for Dynamic Panels with Fixed Effects and Strong Instruments at Unity. *Econometric Theory* 26:119–51.

Hansen, L. P., J. Heaton, and A. Yaron. 1996. Finite-sample Properties of Some Alternative GMM Estimators. *Journal of Business and Economic Statistics* 14:262–80.

Hayashi, F., and T. Inoue. 1991. The Relation Between Firm Growth and Q with Multiple Capital Goods: Theory and Evidence from Panel Data on Japanese Firms. *Econometrica* 59:731–53.

Heckman, J., and B. Singer. 1984. A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data. *Econometrica* 52:271–320.

Hennessy, C. A. 2004. Tobin's Q, Debt Overhang, and Investment. *Journal of Finance* 59:1717–42.

Hennessy, C. A., and T. M. Whited. 2007. How Costly Is External Financing? Evidence from a Structural Estimation. *Journal of Finance* 62:1705–45.

Holtz-Eakin, D. 1988. Testing for Individual Effects in Autoregressive Models. *Journal of Econometrics* 39:297–307.

Holtz-Eakin, D., W. Newey, and H. S. Rosen. 1988. Estimating Vector Autoregressions with Panel Data. *Econometrica* 56:1371–95.

Kennedy, P. 2003. *A Guide to Econometrics*. Malden, MA: Blackwell.

Kocherlakota, N. 1990. On Tests of Representative Consumer Asset-pricing Models. *Journal of Monetary Economics* 26:285–304.

Lemmon, M. L., M. R. Roberts, and J. F. Zender. 2008. Back to the Beginning: Persistence and the Cross-section of Corporate Capital Structure. *Journal of Finance* 63:1575–608.

Lewellen, J., and K. Lewellen. 2010. Internal Cashflow and Investment. Unpublished manuscript, Dartmouth University.

MaCurdy, T. 2007. A Practitioner's Approach to Estimating Intertemporal Relationships Using Longitudinal

Data: Lessons from Applications in Wage Dynamics. In J. J. Heckman and E. E. Leamer (eds.), *Handbook of Econometrics*, vol. 6, pp. 4057–167. Amsterdam: Elsevier.

Moyen, N. 2004. Investment-cash Flow Sensitivities: Constrained versus Unconstrained Firms. *Journal of Finance* 69:2061–92.

Newey, W., and D. McFadden. 1994. Large Sample Estimation and Hypothesis Testing. In R. Engle and D. McFadden (eds.), *Handbook of Econometrics,* vol. 4, pp. 2111–245. Amsterdam: Elsevier.

Papanikolaou, D., and V. Panousi. Forthcoming 2011. Investment, Idiosyncratic Risk, and Ownership. *Journal of Finance*.

Rauh, J. 2006. Investment and Financing Constraints: Evidence from the Funding of Corporate Pension Plans. *Journal of Finance* 61:33–71.

Riddick, L. A., and T. M. Whited. 2009. The Corporate Propensity to Save. *Journal of Finance* 64:1729–66.

Schennach, S. M. 2004. Estimation of Nonlinear Models with Measurement Error. *Econometrica* 72:33–75.

Shanken, J., and G. Zhou. 2007. Estimating and Testing Beta Pricing Models: Alternative Methods and Their Performance in Simulations. *Journal of Financial Economics* 84:40–86.

Whited, T. M. 2001. Is It Inefficient Investment That Causes the Diversification Discount? *Journal of Finance* 56:1667–92.