Assessing the Performance of Different Estimators that Deal with Measurement Error in Linear Models*

Heitor Almeida
University of Illinois
& NBER
halmeida@illinois.edu

Murillo Campello
Cornell University
& NBER
campello@cornell.edu

Antonio F. Galvao Jr.
University of Wisconsin
agalvao@uwm.edu

This Draft: April 26, 2012

Abstract

We describe different procedures to deal with measurement error in linear models, and assess their performance in finite samples using Monte Carlo simulations, and data on corporate investment. We consider the standard instrumental variables approach proposed by Griliches and Hausman (1986) as extended by Biorn (2000) [OLS-IV], the Arellano and Bond (1991) instrumental variable estimator, and the higher-order moment estimator proposed by Erickson and Whited (2000, 2002). Our analysis focuses on characterizing the conditions under which each of these estimators produces unbiased and efficient estimates in a standard "errors in variables" setting. In the presence of fixed effects, under heteroscedasticity, or in the absence of a very high degree of skewness in the data, the EW estimator is inefficient and returns biased estimates for mismeasured and perfectly-measured regressors. In contrast to the EW estimator, IV-type estimators (OLS-IV and AB-GMM) easily handle individual effects, heteroscedastic errors, and different degrees of data skewness. The IV approach, however, requires assumptions about the autocorrelation structure of the mismeasured regressor and the measurement error. We illustrate the application of the different estimators using empirical investment models. Our results show that the EW estimator produces inconsistent results when applied to real-world investment data, while the IV estimators tend to return results that are consistent with theoretical priors.

JEL Classification Numbers: G31, C23.
Keywords: Investment equations, measurement error, Monte Carlo simulations, instrumental variables, GMM.

**This article is based on Almeida, H., M. Campello, and A. Galvao, 2010, Measurement Errors in Investment Equations, Review of Financial Studies, 23, pages 3279-3382.

1 Introduction

OLS estimators are the work-horse of empirical research in many fields in applied economics. Researchers see a number of advantages in these estimators. Most notably, they are easy to implement and the results they generate are easy to replicate. Another appealing feature of OLS estimators is that they easily accommodate the inclusion of individual (e.g., firm and time) idiosyncratic effects. Despite their popularity, however, OLS estimators are weak in dealing with the problem of errors-in-variables. When the independent (right-hand side) variables of an empirical model are mismeasured, coefficients estimated via standard OLS are inconsistent (*attenuation bias*). This poses a problem since, in practice, it is hard to think of any empirical proxies in applied research whose measurement is not a concern.

In this paper, we describe three estimators that deal with the problem of mismeasurement; namely, the standard instrumental variables approach extended by Biorn (2000) [OLS-IV], the Arellano and Bond (1991) instrumental variable estimator [AB-GMM], and the higher-order moment estimator proposed by Erickson and Whited (2000, 2002) [henceforth, EW]. We also assess the performance of these estimators in finite samples using Monte Carlo simulations, and illustrate their application using data on corporate investment.

While we provide a formal presentation in the next section, it is useful to discuss the intuition behind the estimation approaches we analyze. All approaches share the attractive property that they do not require the researcher to look for instruments outside the model being considered.[1] They differ, however, on how identification is achieved.

Both instrumental variable approaches rely on assumptions about the serial correlation of the latent variable and the innovations of the model (the model's disturbances and the measurement error). There are two conditions that must hold to ensure identification. First, the true value of the mismeasured regressor must have some degree of autocorrelation. In this case, lags of the mismeasured regressor are natural candidates for the instrumental set since they contain information about the current value of the mismeasured regressor.[2] This condition is akin to the standard requirement that the instrument be correlated with the variable of interest. The other necessary condition is associated with the *exclusion restriction* that is standard in IV methods and relates to the degree of serial correlation of the innovations. A standard assumption guaranteeing identification is that the

[1] Naturally, if extraneous instruments are available they can help solve the identification problem. See Rauh (2006) for the use of discontinuities in pension contributions as a source of variation in cash flows in an investment model. Bond and Cummins (2000) use information contained in financial analysts' forecasts to instrument for investment demand.

[2] Lags of the well-measured variable may also be included in the instrument set if they are believed to also contain information about the mismeasured one.

measurement error process is independently and identically distributed. This condition ensures that past values of the observed variables are uncorrelated with the current value of the measurement error, validating the use of lags of observed variables as instruments. Under certain conditions, one can also allow for autocorrelation in the measurement error. Examples in which identification works are when autocorrelation is constant over time, or when it evolves according to a moving average process.[3] The first assumption ensures identification because it means that while lagged values of the measurement error are correlated with its current value, any past shocks to the measurement error process do not persist over time. The moving average assumption allows for shocks to persist over time, but it imposes restrictions on the instrumental set. In particular, as we show below, it precludes the use of shorter lags of observed variables in the instrument set, as the information contained in short lags may be correlated with the current value of the measurement error.

The EW estimator is based on high-order moments of residuals obtained by "partialling out" perfectly measured regressors from the dependent, observed mismeasured, and latent variables, as well as high-order moments of the innovations of the model. The key idea is to create a set of auxiliary equations as a function of these moments and cross-moments. Implementation then requires a high degree of skewness in the distribution of the partialled out latent variable. Our analysis also shows that the presence of individual fixed effects and heteroskedasticity also impact the performance of the EW estimator, particularly so if they are both present in the data process.

We perform a series of Monte Carlo simulations to compare the performance of the EW and IV estimators in finite samples. Emulating the types of environments commonly found by empirical researchers, we set up a panel data model with individual-fixed effects and potential heteroscedasticity in the errors. Monte Carlo experiments enable us to study those estimators in a "controlled environment," where we can investigate the role played by each element (or assumption) of an estimator in evaluating its performance. Our simulations compare the EW and IV (OLS-IV and AB-GMM) estimators in terms of bias and root mean squared error (RMSE), a standard measure of efficiency.

We consider several distributional assumptions to generate observations and errors. Experimenting with multiple distributions is important because researchers often find a variety of distributions in real-world applications, and because one ultimately does not observe the distribution of the mismeasurement term. Since the EW estimator is built around the notion of skewness of the relevant distributions, we experiment with three skewed distributions (lognormal, chi-square, and F-distribution), using the standard normal (non-skewed) as a benchmark. The simulations also allow for significant correlation between mismeasured and well-measured regressors (as in Erickson and Whited

[3] See, among others, Biorn (2000), Wansbeek (2001), and Xiao, Shao, and Palta (2008).

(2000, 2002)), so that the attenuation bias of the mismeasured regressor affects the coefficient of the well-measured regressor.

Our simulation results can be summarized as follows. First, we examine the identification test proposed by Erickson and Whited (2002). This is a test that the data contain a sufficiently high degree of skewness to allow for the identification of their model. We study the *power* of the EW identification test by generating data that *do not* satisfy its null hypothesis of non-skewness. In this case, even for the most skewed distribution (lognormal), the test rejects the null hypothesis only 47% of the time — this is far less than desirable, given that the null is false. The power of the test becomes even weaker after we treat the data for the presence of fixed effects in the true model ("within transformation"). In this case, the rejection rate under the lognormal distribution drops to 43%. The test's power declines even further when we consider alternative skewed distributions (chi-square and F distributions). The upshot of this first set of experiments is that the EW model too often rejects data that are generated to fit its identifying assumptions. These findings may help explain some of the difficulties previous researchers have reported when attempting to implement the EW estimator.

We then study the bias and efficiency of the EW and IV estimators. Given that the true model contains fixed effects, it is appropriate to apply the within transformation to the data. However, because most empirical implementations of the EW estimator have used data in "level form" (that is, not treated for the presence of fixed effects),[4] we also experiment with cases in which we do not apply the within transformation. EW propose three estimators that differ according to the number of moments used: GMM3, GMM4, and GMM5. We consider all of them in our experiments.

In a first round of simulations, we impose error homoscedasticity. When we implement the EW estimator with the data in level form, we find that the coefficients returned are significantly biased even when the data have a high degree of skewness (i.e., under the lognormal case, which is EW's preferred case). Indeed, for the mismeasured regressors the EW biases are in excess of 100% of their true value. As should be expected, the performance of the EW estimator improves once the within transformation is used. In the case of the lognormal distribution, the EW estimator bias is relatively small. In addition, deviations from the lognormal assumption tend to generate significant biases for the EW estimator.

In a second round of simulations, we allow for heteroscedasticity in the data. We focus our attention on simulations that use data that are generated using a lognormal distribution after applying the within transformation (the best case scenario for the EW estimator). Heteroscedasticity introduces heterogeneity to the model and consequently to the distribution of the partialled out

[4] Examples are Whited (2001, 2006), Hennessy (2004), and Colak and Whited (2007).

dependent variable, compromising identification in the EW framework. The simulations show that the EW estimator is biased and inefficient for both the mismeasured and well-measured regressors. In fact, biases emerge even for very small amounts of heteroscedasticity, where we find biases of approximately 40% for the mismeasured regressor. Paradoxically, biases "switch signs" depending on the degree of heteroscedasticity that is allowed for in the model. For instance, for small amounts of heteroscedasticity, the bias of the mismeasured regressor is negative (i.e., the coefficient is biased downwards). However, the bias turns positive for a higher degree of heteroscedasticity. Since heteroscedasticity is a naturally occurring phenomenon in corporate data, our simulations imply that empirical researchers might face serious drawbacks when using the EW estimator.

Our simulations also show that, in contrast to the EW estimator, the bias in the IV estimates is small and insensitive to the degree of skewness and heteroscedasticity in the data. Focusing on the OLS-IV estimator, we consider the case of time-invariant correlation in the error structure and use the second lag of the observed mismeasured variable as an instrument for its current (differenced) value.[5] We also allow the true value of the mismeasured regressor to have a moderate degree of autocorrelation. Our results suggest that the OLS-IV estimator renders fairly unbiased estimates. In general, that estimator is also distinctly more efficient than the EW estimator.

We also examine the OLS-IV estimator's sensitivity to the autocorrelation structures of the mismeasured regressor and the measurement error. First, we consider variations in the degree of autocorrelation in the process for the true value of the mismeasured regressor. Our simulations show that the IV bias is largely insensitive to variations in the autoregressive (AR) coefficient (except for extreme values of the AR coefficient). Second, we replace the assumption of time-invariant autocorrelation in the measurement error with a moving average (MA) structure. Our simulations show that the OLS-IV bias remains small if one uses long enough lags of the observable variables as instruments. In addition, provided that the instrument set contains suitably long lags, the results are robust to variations in the degree of correlation in the MA process. As we discuss below, understanding these features (and limitations) of the IV approach is important given that the researcher will be unable to pin down the process followed by the measurement error process.

To illustrate the performance of these alternative estimators on real data, in the final part of our analysis we estimate empirical investment models under the EW and IV frameworks. Concerns about measurement errors have been emphasized in the context of the empirical investment model introduced by Fazzari, Hubbard, and Petersen (1988), where a firm's investment is regressed on a

[5] The results for the Arellano-Bond GMM estimator are similar to those of the OLS-IV estimator. To save space and because the OLS-IV estimator is easier to implement, we focus on this estimator.

proxy for investment demand (Tobin's q) and cash flows. Theory suggests that the correct proxy for a firm's investment demand is captured by *marginal* q, but this quantity is unobservable and researchers use instead its measurable proxy, *average* q. Since the two variables are not the same, a measurement problem naturally arises (Hayashi, 1982; and Poterba, 1988). Following Fazzari, Hubbard, and Petersen (1988), investment–cash flow sensitivities became a standard metric in the literature that examines the impact of financing imperfections on corporate investment (Stein, 2003). These empirical sensitivities are also used for drawing inferences about efficiency in internal capital markets (Lamont, 1997; and Shin and Stulz, 1998), the effect of agency on corporate spending (Hadlock, 1998; and Bertrand and Mullainathan, 2005), the role of business groups in capital allocation (Hoshi, Kashyap, and Scharfstein, 1991), and the effect of managerial characteristics on corporate policies (Bertrand and Schoar, 2003; and Malmendier and Tate, 2005).

Theory does not pin down exact values for the expected coefficients on q and cash flow in an investment model. However, two conditions would seem reasonable in practice. First, given that the estimator is addressing measurement error in q that may be "picked up" by cash flow (joint effects of attenuation bias and regressor covariance), we should expect the q coefficient to go up and the cash flow coefficient to go down, when compared to standard (likely biased) OLS estimates. Second, we would expect the q and cash flow coefficients to be non-negative after addressing the problem of mismeasurement. If the original q-theory of investment holds and the estimator does a good job of addressing mismeasurement, then the cash flow coefficient would be zero. Alternatively, the cash flow coefficient could be positive because of financing frictions.[6]

Using data from Compustat from 1970 to 2005, we estimate investment equations in which investment is regressed on proxies for q and cash flow. Before doing so, we perform standard tests to check for the presence of individual-fixed effects and heteroscedasticity in the data. In addition, we perform the EW identification test to check whether the data contain a sufficiently high degree of skewness.

Our results are as follows. First, our tests reject the hypotheses that the data do not contain firm-fixed effects and that errors are homoscedastic. Second, the EW identification tests indicate that the data fail to display sufficiently high skewness. These initial tests suggest that the EW estimator is not suitable for standard investment equation applications. In fact, we find that, when applied to the data, the EW estimator returns coefficients for q and cash flow that are highly unstable across different years. Moreover, following the EW procedure for panel models (which comprises combining

[6]See Hubbard (1998) and Stein (2003) for comprehensive reviews. We note that the presence of financing frictions does not necessarily imply that the cash flow coefficient should be positive. See Chrinko (1993) and Gomes (2001) for arguments suggesting that financing frictions are not sufficient to generate positive cash flow coefficients.

yearly cross-sectional coefficients into single estimates), we obtain estimates for q and cash flow that do not satisfy the conditions discussed above. In particular, EW estimators do not reduce the cash flow coefficient relative to that obtained by standard OLS, while the q coefficient is never statistically significant. In addition, those estimates are not robust with respect to the number of moments used: EW's GMM3, GMM4, and GMM5 models procedure results that are inconsistent with one another. These results suggest that the presense of heteroskedasticity and fixed effects in real world investment data hampers identification when using the EW estimator.

In contrast to EW, the OLS-IV procedure yields estimates that are fairly sensible. The q coefficient goes up by a factor of 3 to 5, depending on the set of instruments used. At the same time, the cash flow coefficient goes down by about two-thirds of the standard OLS value. Similar conclusions apply to the AB-GMM estimator. We also examine the robustness of the OLS-IV to variations in the set of instruments used in the estimation, including sets that feature only longer lags of the variables in the model. The OLS-IV coefficients remain fairly stable after such changes. These results suggest that real world investment data likely satisfies the assumptions that are required for identification of IV estimators.

The remainder of the paper is structured as follows. We start the next section discussing in detail the EW estimator, clarifying the assumptions that are needed for its implementation. Subsequently, we show how alternative IV models deal with the errors-in-variables problem. In Section 3, we use Monte Carlo simulations to examine the performance of alternative estimators in small samples and when we relax the assumptions that are required for identification. In Section 4, we take our investigation to actual data, estimating investment regressions under the EW and IV frameworks. Section 5 concludes the paper.

2 Dealing with Mismeasurement: Alternative Estimators

2.1 The Erickson-Whited estimator

In this section, we discuss the estimator proposed in companion papers by Erickson and Whited (2000, 2002). Those authors present a two-step generalized method of moments (GMM) estimator that exploits information contained in the high-order moments of residuals obtained from perfectly-measured regressors (similar to Cragg, 1997). We follow EW and present the estimator using notation of cross-section estimation. Let (y_i, z_i, x_i), $i = 1, ..., n$, be a sequence of observable vectors, where $x_i \equiv (x_{i1}, ..., x_{iJ})$ and $z_i \equiv (1, z_{i1}, ..., z_{iL})$. Let $(u_i, \chi_i, \varepsilon_i)$ be a sequence of unobservable vectors, where $\chi_i \equiv (\chi_{i1}, ..., \chi_{iJ})$ and $\varepsilon_i \equiv (\varepsilon_{i1}, ..., \varepsilon_{iJ})$. Consider the following model:

$$y_i = z_i\alpha + \chi_i\beta + u_i, \tag{1}$$

where y_i is the dependent variable, z_i is a perfectly-measured regressor, χ_i is a mismeasured regressor, u_i is the innovation of the model, and $\alpha \equiv (\alpha_0, \alpha_1, ..., \alpha_L)'$ and $\beta \equiv (\beta_1, ..., \beta_J)'$. The measurement error is assumed to be additive such that:

$$x_i = \chi_i + \varepsilon_i, \tag{2}$$

where x_i is the observed variable and ε_i is the measurement error. The observed variables are y_i, z_i, and x_i; and by substituting (2) in (1), we have:

$$y_i = z_i\alpha + x_i\beta + \nu_i,$$

where $\nu_i = u_i - \varepsilon_i\beta$. In the new regression, the observable variable x_i is correlated with the innovation term ν_i, causing the coefficient of interest, β, to be biased.

To compute the EW estimator it is necessary to first partial out the effect of the well-measured variable, z_i, in (1) and (2) and rewrite the resulting expressions in terms of residual populations:

$$y_i - z_i\mu_y = \eta_i\beta + u_i \tag{3}$$

$$x_i - z_i\mu_x = \eta_i + \varepsilon_i, \tag{4}$$

where $(\mu_y, \mu_x, \mu_\chi) = [E(z_i'z_i)]^{-1}E[z_i'(y_i, x_i, \chi_i)]$ and $\eta_i \equiv \chi_i - z_i\mu_\chi$. One can then consider a two-step estimation approach, where the first step is to substitute least squares estimates $(\hat{\mu}_y, \hat{\mu}_x) \equiv \left(\sum_{i=1}^n z_i'z_i\right)^{-1} \sum_{i=1}^n z_i'(y_i, x_i)$ into (3) and (4) to obtain a lower dimensional errors-in-variables model. The second step consists of estimating β using GMM using high-order sample moments of $y_i - z_i\hat{\mu}_y$ and $x_i - z_i\hat{\mu}_x$. Estimates of α are then recovered via $\mu_y = \alpha + \mu_x\beta$. Thus, the estimators are based on equations giving the moments of $y_i - z_i\mu_y$ and $x_i - z_i\mu_x$ as functions of β and the moments of $(u_i, \varepsilon_i, \eta_i)$.

To give a concrete example of how the EW estimator works, we explore the case of $J = 1$. The more general case is discussed below. By substituting

$$(\hat{\mu}_y, \hat{\mu}_x) \equiv \left(\sum_{i=1}^n z_i'z_i\right)^{-1} \sum_{i=1}^n z_i'(y_i, x_i)$$

into (1) and (2) one can estimate β, $E(u_i^2)$, $E(\varepsilon_i^2)$, and $E(\eta_i^2)$ via GMM. Estimates of the l^{th} element of α are obtained by substituting the estimate of β and the l^{th} elements of $\hat{\mu}_y$ and $\hat{\mu}_x$ into:

$$\alpha_l = \mu_{yl} - \mu_{xl}\beta, \ l \neq 0.$$

There are three second-order moment equations:

$$E[(y_i - z_i\mu_y)^2] = \beta^2 E(\eta_i^2) + E(u_i^2), \tag{5}$$

$$E[(y_i - z_i\mu_y)(x_i - z_i\mu_x)] = \beta E(\eta_i^2), \tag{6}$$

$$E[(x_i - z_i\mu_x)^2] = E(\eta_i^2) + E(\varepsilon_i^2). \tag{7}$$

The left-hand side quantities are consistently estimable, but there are only three equations with which to estimate four unknown parameters on the right-hand side. The third-order product moment equations, however, consist of two equations in two unknowns:

$$E[(y_i - z_i\mu_y)^2(x_i - z_i\mu_x)] = \beta^2 E(\eta_i^3), \tag{8}$$

$$E[(y_i - z_i\mu_y)(x_i - z_i\mu_x)^2] = \beta E(\eta_i^3). \tag{9}$$

It is possible to solve these two equations for β. Crucially, a solution exists if the identifying assumptions $\beta \neq 0$ and $E(\eta_i^3) \neq 0$ are true, and one can test the contrary hypothesis (that is, $\beta = 0$ and/or $E(\eta_i^3) = 0$) by testing whether their sample counterparts are significative different from zero.

Given β, equations (5)–(7) and (25) can be solved for the remaining right-hand side quantities. One obtains an overidentified equation system by combining (5)–(25) with the fourth-order product moment equations, which introduce only one new quantity, $E(\eta_i^4)$:

$$E[(y_i - z_i\mu_y)^3(x_i - z_i\mu_x)] = \beta^3 E(\eta_i^4) + 3E(\eta_i^2)E(u_i^2), \tag{10}$$

$$E[(y_i - z_i\mu_y)^2(x_i - z_i\mu_x)^2] = \beta^2[E(\eta_i^4) + E(\eta_i^2)E(u_i^2)] + E(u_i^2)[E(\eta_i^2) + E(\varepsilon_i^2)], \tag{11}$$

$$E[(y_i - z_i\mu_y)(x_i - z_i\mu_x)^3] = \beta[E(\eta_i^4) + 3E(\eta_i^2)E(\varepsilon_i^2)]. \tag{12}$$

The resulting eight-equation system (5)–(12) contains the six unknowns (β, $E(u_i^2)$, $E(\varepsilon_i^2)$, $E(\eta_i^2)$, $E(\eta_i^3)$, $E(\eta_i^4)$). It is possible to estimate this vector by numerically minimizing a quadratic form that minimizes asymptotic variance.

The conditions imposed by EW imply restrictions on the residual moments of the observable variables. Such restrictions can be tested using the corresponding sample moments. EW also propose a test for residual moments that is based on several assumptions.[7] These assumptions imply testable restrictions on the residuals from the population regression of the dependent and proxy variables on

[7] First, the measurement errors, the equation error, and all regressors have finite moments of sufficiently high order. Second, the regression error and the measurement error must be independent of each other and of all regressors. Third, the residuals from the population regression of the unobservable regressors on the perfectly-measured regressors must have a nonnormal distribution.

the perfectly-measured regressors. Accordingly, one can develop Wald-type partially-adjusted statistics and asymptotic null distributions for the test. Empirically, one can use the Wald test statistic and critical values from a chi-square distribution to test whether the last moments are equal to zero. This is an *identification test*, and if in a particular application one cannot reject the null hypothesis, then the model is unidentified and the EW estimator may not be used. We study the finite sample performance of this test and its sensitivity to different data-generating processes in the next section.

It is possible to derive more general forms of the EW estimator. In particular, the EW estimators are based on the equations for the moments of $y_i - z_i\mu_y$ and $x_i - z_i\mu_x$ as functions of β and the moments u_i, ε_i, and η_i. To derive these equations, write (3) as $y_i - z_i\mu_y = \sum_{j=1}^{J} \eta_{ij}\beta_j + u_i$, where J is the number of well-measured regressors, and the j^{th} equation in (4) as $x_{ij} - z_i\mu_{xj} = \eta_{ij} + \varepsilon_{ij}$, where μ_{xj} is the j^{th} column of μ_x and $(\eta_{ij}, \varepsilon_{ij})$ is the j^{th} row of $(\eta'_{ij}, \varepsilon'_{ij})$. Next write

$$E\left[(y_i - z_i\mu_y)^{r_0} \prod_{j=1}^{J}(x_i - z_i\mu_x)^{r_j}\right] = E\left[\left(\sum_{j=1}^{J} \eta_i\beta + u_i\right)^{r_0} \prod_{j=1}^{J}(\eta_i + \varepsilon_i)^{r_j}\right], \tag{13}$$

where (r_0, r_1,..., r_J) are nonnegative integers. After expanding the right hand-side of (13) using the multinomial theorem, it is possible to write the above moment condition as:

$$E\left[g_i(\mu)\right] = c(\theta),$$

where $\mu = vec(\mu_y, \mu_x)$, $g_i(\mu)$ is a vector of distinct elements of the form $(y_i - z_i\mu_y)^{r_0} \prod_{j=1}^{J}(x_i - z_i\mu_x)^{r_j}$, $c(\theta)$ contains the corresponding expanded version of the right-hand side of (13), and θ is a vector containing the elements of β and the moments of (u_i, ε_i, η_i). The GMM estimator θ is defined as:

$$\hat{\theta} = \arg\min_{t \in \Theta} (\bar{g}_i(\hat{\mu}) - c(t))' \hat{W} (\bar{g}_i(\hat{\mu}) - c(t)),$$

where $\bar{g}_i(s) \equiv \sum_{t=1}^{n} g_i(s)$ for all s, and $\hat{W}$ is a positive definite matrix. Assuming a number of regularity conditions,[8] the estimator is consistent and asymptotically normal.

It is important to notice that the estimator proposed by Erickson and Whited (2002) was originally designed for cross-sectional data. To accommodate a panel-like structure, Erickson and Whited (2000) propose transforming the data before the estimation using the within transformation or differencing. To mimic a panel structure, the authors propose the idea of combining different cross-sectional GMM estimates using a Minimum Distance estimator (MDE).

[8] More specifically, these conditions are as follows: (z_i, χ_i, u_i, ε_i), is an independent and identically distributed sequence; u_i and the elements of z_i, χ_i, and ε_i have finite moments of every order; (u_i, ε_i) is independent of (z_i, χ_i), and the individual elements in (u_i, ε_i) are independent of each other; E(u_i)=0 and E(ε_i)=0; E[(z_i, χ_i)'(z_i, χ_i)] is positive definite; every element of β is nonzero; and the distribution of η satisfies E[$(\eta_i c)^3$]$\neq$0 for every vector of constants $c = (c_1, ..., c_J)$ having at least one nonzero element.

The MDE estimator is derived by minimizing the distance between the auxiliary parameter vectors under the following restrictions:

$$f(\beta, \hat{\theta}) = H\beta - \hat{\theta} = 0,$$

where the $R \cdot K \times K$ matrix H imposes $(R-1) \cdot K$ restrictions on θ. The $R \cdot K \times 1$ vector $\hat{\theta}$ contains the R stacked auxiliary parameter vectors, and β is the parameter of interest. Moreover, H is defined by a $R \cdot K \times K$–dimensional stacked identity matrix.

The MDE is given by the minimization of:

$$D(\beta) = f(\beta, \hat{\theta})' \hat{V}[\hat{\theta}]^{-1} f(\beta, \hat{\theta}), \tag{14}$$

where $\hat{V}[\hat{\theta}]$ is the common estimated variance–covariance matrix of the auxiliary parameter vectors.

In order to implement the MDE, it is necessary to determine the covariances between the cross-sections being pooled. EW propose to estimate the covariance by using the covariance between the estimators' respective influence functions.[9] The procedure requires that each cross-section have the same sample size, that is, the panel needs to be balanced.

Thus, minimization of D in equation (14) leads to:

$$\hat{\beta} = (H' \hat{V}[\hat{\theta}]^{-1} H)^{-1} H' \hat{V}[\hat{\theta}]^{-1} \hat{\theta},$$

with variance–covariance matrix:

$$\hat{V}[\hat{\beta}] = (H' \hat{V}[\hat{\theta}]^{-1} H)^{-1}.$$

H is a vector in which R is the number of GMM estimates available (for each time period) and $K = 1$, $\hat{\theta}$ is a vector containing all the EW estimates for each period, and β is the MDE of interest. In addition, $\hat{V}[\hat{\theta}]$ is a matrix carrying the estimated variance–covariance matrices of the GMM parameter vectors. In order to implement the MDE it is necessary to determine the covariances between the cross-sections being pooled. EW propose to estimate the covariance by using the covariance between the estimators' respective influence functions.[10] The procedure requires that each cross-section have the same sample size, that is, the panel needs to be balanced.

2.2 An OLS-IV Framework

In this section, we revisit the work of Griliches and Hausman (1986) and Biorn (2000) to discuss a class of OLS-IV estimators that can help address the errors-in-variables problem.

[9] See Erickson and Whited (2002) Lemma 1 for the definition of their proposed influence function.

[10] See Erickson and Whited (2002) Lemma 1 for the definition of their proposed influence function.

Consider the following single-equation model:

$$y_{it} = \gamma_i + \chi_{it}\beta + u_{it,} \quad i = 1, ..., N,\ t = 1, ..., T, \tag{15}$$

where u_{it} is independently and identically distributed, with mean zero and variance σ_u^2, and Cov(χ_{it}, u_{is}) = Cov$(\gamma_i, u_{it}) = 0$ for any t and s, but Cov$(\gamma_i, \chi_{it}) \neq 0$, y is an observable scalar, χ is a $1 \times K$ vector, and β is $K \times 1$ vector. Suppose we do not observe χ_{it} itself, but rather the error-ridden measure:

$$x_{it} = \chi_{it} + \varepsilon_{it}, \tag{16}$$

where Cov$(\chi_{it}, \varepsilon_{it}) =$ Cov$(\gamma_i, \varepsilon_{it}) =$ Cov$(u_{is}, \varepsilon_{it}) = 0$, Var$(\varepsilon_{it}) = \sigma_\varepsilon^2$, Cov$(\varepsilon_{it}, \varepsilon_{it-1}) = \gamma_\varepsilon \sigma_\varepsilon^2$, and ε is a $1 \times K$ vector. If we have a panel data with $T > 3$, by substituting (16) in (15), we can take first differences of the data to eliminate the individual effects γ_i and obtain:

$$y_{it} - y_{it-1} = (x_{it} - x_{it-1})\beta + [(u_{it} - \varepsilon_{it}\beta) - (u_{it-1} - \varepsilon_{it-1}\beta)]. \tag{17}$$

Because of the correlation between the mismeasured variable and the innovations, the coefficient of interest is known the be biased.

Griliches and Hausman (1986) propose an instrumental variables approach to reduce the bias. If the measurement error ε_{it} is i.i.d. across i and t, and x is serially correlated, then, for example, x_{it-2}, x_{it-3}, or $(x_{it-2} - x_{it-3})$ are valid as instruments for $(x_{it} - x_{it-1})$. The resulting instrumental variables estimator is consistent even though T is finite and N might tend to infinity.

As emphasized by Erickson and Whited (2000), for some applications the assumption of i.i.d. measurement error can be seen as too strong. Nonetheless, it is possible to relax this assumption to allow for autocorrelation in the measurement errors. While other alternatives are available, here we follow the approach suggested by Biorn (2000).[11]

Biorn (2000) relaxes the i.i.d. condition for innovations in the mismeasured equation and proposes alternative assumptions under which consistent IV estimators of the coefficient of the mismeasured regressor exists. Under those assumptions, as we will show, one can use the lags of the variables already included in the model as instruments. A notable point is that the consistency of these estimators is robust to potential correlation between individual heterogeneity and the latent regressor.

Formally, consider the model described in equations (15)–(16) and assume that (χ_{it}, u_{it}, ε_{it}, γ_i) are independent across individuals. For the necessary orthogonality assumptions, we refer the reader to Biorn (2000), since these are quite standard. More interesting are the assumptions about the measurement errors and disturbances. The standard Griliches-Hausman's assumptions are:

[11] A more recent paper by Xiao, Shao, and Palta (2008) also shows how to relax the classical Griliches-Hausman assumptions for measurement error models.

(A1) $E(\varepsilon_{it}'\varepsilon_{i\theta}) = 0_{KK}$, $t \neq \theta$,

(A2) $E(u_{it}u_{i\theta}) = 0$, $t \neq \theta$,

which impose nonautocorrelation on innovations. It is possible to relax these assumptions in different ways. For example, we can replace (A1) and (A2) with:

(B1) $E(\varepsilon_{it}'\varepsilon_{i\theta}) = 0_{KK}$, $|t-\theta| > \tau$,

(B2) $E(u_{it}u_{i\theta}) = 0$, $|t-\theta| > \tau$.

This set of assumptions is weaker since (B1) and (B2) allow for a vector moving average (MA) structure up to order τ ($\geq$1) for the innovations. Alternatively, one can use the following assumptions:

(C1) $E(\varepsilon_{it}'\varepsilon_{i\theta})$ is invariant to t, θ, $t \neq \theta$,

(C2) $E(u_{it}u_{i\theta})$ is invariant to t, θ, $t \neq \theta$.

Assumptions (C1) and (C2) allow for a different type of autocorrelation, more specifically they allow for any amount of autocorrelation that is time invariant. Assumptions (C1) and (C2) will be satisfied if the measurement errors and the disturbances have individual components, say $\varepsilon_{it} = \varepsilon_{1i} + \varepsilon_{2it}$, $u_{it} = u_{1i} + u_{2it}$, where ε_{1i}, ε_{2it}, u_{1i}, u_{2it} are i.i.d. Homoscedasticity of ε_{it} and/or u_{it} across i and t need not be assumed; the model accommodates various forms of heteroscedasticity.

Biorn also considers assumptions related to the distribution of the latent regressor vector χ_{it}:

(D1) $E(\chi_{it})$ is invariant to t,

(D2) $E(\gamma_i\chi_{it})$ is invariant to t.

Assumptions (D1) and (D2) hold when χ_{it} is stationary. Note that χ_{it} and i need not be uncorrelated.

To ensure identification of the slope coefficient vector when panel data are available, it is necessary to impose restrictions on the second-order moments of the variables (χ_{it}, u_{it}, ε_{it}, γ_i). For simplicity, Biorn assumes that this distribution is the same across individuals and that the moments are finite. More specifically, $C(\chi_{it}, \chi_{it}) = \sum_{t\theta}^{\chi\chi}$, $E(\chi_{it}\eta_i) = \sum_t^{\chi\eta}$, $E(\varepsilon_{it}'\varepsilon_{it}) = \sum_{t\theta}^{\varepsilon\varepsilon}$, $E(u_{it}u_{it}) = \sigma_{t\theta}^{uu}$, $E(\eta_i^2) = \sigma^{\eta\eta}$, where C denotes the covariance matrix operator. Then, it is possible to derive the second-order moments of the observable variables and show that they only depend on these matrices and the coefficient β.[12] In this framework, there is no need to use assumptions based on higher-order moments.

Biorn proposes several strategies to estimate the slope parameter of interest. Under the OLS-IV framework, he proposes estimation procedures of two kinds:

[12] Formally, one can show that: $C(x_{it}, x_{i\theta}) = \sum_{t\theta}^{\chi\chi} + \sum_{t\theta}^{\varepsilon\varepsilon}$, $E(x_{it}, y_{i\theta}) = \sum_{t\theta}^{\chi\chi}\beta + \sum_t^{\chi\eta}$, and $E(y_{it}, y_{i\theta}) = \beta'\sum_{t\theta}^{\chi\chi}\beta + \sum_t^{\chi\eta}\beta + \beta'(\sum_\theta^{\chi\eta})' + \sigma_{t\theta}^{uu} + \sigma^{\eta\eta}$.

- *OLS-IV A*: The equation is transformed to differences to remove individual heterogeneity and is estimated by OLS-IV. Admissible instruments for this case are the level values of the regressors and/or regressands for other periods.

- *OLS-IV B*: The equation is kept in level form and is estimated by OLS-IV. Admissible instruments for this case are differenced values of the regressors and/or regressands for other periods.

Using moment conditions from the OLS-IV framework, one can define the estimators just described. In particular, using the mean counterpart and the moment conditions, one can formally define the *OLS-IV A* and *OLS-IV B* estimators.

In particular, the estimator for *OLS-IV A* can be defined as:

$$\hat{\beta}_{xp(t\theta)} = \left[\sum_{i=1}^{N} x'_{ip}\left(\Delta x_{it\theta}\right)\right]^{-1}\left[\sum_{i=1}^{N} x'_{ip}\left(\Delta y_{it\theta}\right)\right],$$

where $(t,\, \theta,\, p)$ are indices. Let the dimension of β be defined by K. If $K = 1$, it is possible to define the following estimator for a given $(t,\, \theta,\, p)$:

$$\hat{\beta}_{yp(t\theta)} = \left[\sum_{i=1}^{N} y_{ip}\left(\Delta x_{it\theta}\right)\right]^{-1}\left[\sum_{i=1}^{N} y_{ip}\left(\Delta y_{it\theta}\right)\right].$$

If $K > 1$, the latter estimator is infeasible, but it is possible to modify the former estimator by replacing one element in x'_{ip} by y_{ip}.

The estimator for *OLS-IV B* (equation in level and instruments in difference) can be defined as

$$\hat{\beta}_{x(pq)t} = \left[\sum_{i=1}^{N} \left(\Delta x_{ipq}\right)' x_{it}\right]^{-1}\left[\sum_{i=1}^{N} \left(\Delta x_{ipq}\right)' y_{it}\right].$$

As in the previous case, if the dimension of β, K is equal to 1, it is possible to define the following estimator for $(t,\, p,\, q)$:

$$\hat{\beta}_{y(pq)t} = \left[\sum_{i=1}^{N} \left(\Delta y_{ipq}\right) x_{it}\right]^{-1}\left[\sum_{i=1}^{N} \left(\Delta y_{ipq}\right) y_{it}\right].$$

If $K > 1$, the latter estimator is infeasible, but it is possible to modify the former estimator by replacing one element in Δx_{ip} by Δy_{ip}.

For some applications, it might be useful to impose weaker conditions on the autocorrelation of measurement errors and disturbances. In this case, it is necessary to restrict slightly further the conditions on the instrumental variables. More formally, if one replaces assumptions (A1) and (A2), or (C1) and (C2), by the weaker assumptions (B1) and (B2), then it is necessary to ensure that the IV set has a lag of at least $\tau-2$, and/or lead of at least $\tau+1$ periods of the regressor in order to "clear" the τ period memory of the MA process. Consistency of these estimators is discussed in Biorn (2000).[13]

To sum up, there are two simple ways to relax the standard assumption of i.i.d. measurement errors. Under the assumption of time-invariant autocorrelation, the set of instruments can contain the same variables used under Griliches-Hausman. For example, if one uses the *OLS-IV A* estimator (equation in differences and instruments in levels), then twice-lagged levels of the observable variables can be used as instruments. Under a moving average structure for the innovations in the measurement error [assumptions (B1) and (B2)], identification requires the researcher to use longer lags of the observable variables as instruments. For example, if the innovations follow a MA(1) structure, then consistency of the *OLS-IV A* estimator requires the use of instruments that are lagged three periods and longer. Finally, identification requires the latent regressor to have some degree of autocorrelation (since lagged values are used as instruments). Our Monte Carlo simulations will illustrate the importance of these assumptions, and will evaluate the performance of the OLS-IV estimator under different sets of assumptions about the structure of the errors.

2.3 GMM estimator

Within the broader instrumental variable approach, we also consider an entire class of GMM estimators that deal with mismeasurement. These GMM estimators are close to the OLS-IV estimator discussed above, but may attain appreciable gains in efficiency by combining numerous orthogonality conditions [see Biorn (2000) for a detailed discussion]. GMM estimators that use all the available lags at each period as instruments for equations in first-differences were proposed by Holtz-Eakin, Newey, and Rosen (1988) and Arellano and Bond (1991). We provide a brief discussion in turn.

In the context of a standard investment model, Blundell, Bond, Devereux, and Schiantarelli (1992) use GMM allowing for correlated firm-specific effects, as well as endogeneity (mismeasurement) of q. The authors use an instrumental variables approach on a first-differenced model in which the instruments are weighted optimally so as to form the GMM estimator. In particular, they use q_{it-2} and twice-lagged investments as instruments for the first-differenced equation for firm i in period t. The Blundell, Bond, Devereux, and Schiantarelli estimator can be seen as an application of the GMM instrumental approach proposed by Arellano and Bond (1991), which was originally applied to a dynamic panel.

[13] In particular, if $|t-p|$, $|\theta-p|>\tau$, then (B1) and $\operatorname{rank}\big(E\left[\chi'_{ip}(\Delta\chi_{it\theta})\right]\big)=K$ for some $p\neq t\neq\theta$ ensure consistency of *OLS-IV B*, $\hat{\beta}_{xp(t\theta)}$, and (B2) and the same rank condition ensure consistency of $\hat{\beta}_{yp(t\theta)}$. In the same way, if $|p-t|$, $|q-t|>\tau$, (B1), (D1), (D2) and $\operatorname{rank}\Big(E\left[\left(\Delta\chi_{ipq}\right)'\chi_{it}\right]\Big)=K$ for some $p\neq q\neq t$ ensure consistency of *OLS-IV B*, $\hat{\beta}_{x(pq)t}$, and (B2), (D1), (D2) and the same rank condition ensure consistency of $\hat{\beta}_{y(pq)t}$.

A GMM estimator for the errors-in-variables model of equation (17) based on IV moment conditions takes the form:

$$\hat{\beta} = \left[\left(\Delta x' Z \right) V_N^{-1} \left(Z' \Delta x \right) \right]^{-1} \left(\Delta x' Z \right) V_N^{-1} \left(Z' \Delta y \right),$$

where Δx is the stacked vector of observations on the first difference of the mismeasured variable and Δy is the stacked vector of observations on the first difference of the dependent variable. As in Blundell, Bond, Devereux, and Schiantarelli (1992), the instrument matrix Z has the following form[14]

$$Z_i = \begin{pmatrix} x_1 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & x_1 & x_2 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & x_1 & \cdots & x_{T-2} \end{pmatrix}.$$

According to standard GMM theory, an optimal choice of the inverse weight matrix V_N is a consistent estimate of the covariance matrix of the orthogonality conditions $E(Z_i' \Delta v_i \Delta v_i' Z_i)$, where Δv_i are the first-differenced residuals of each individual. Accordingly, a one-step GMM estimator uses $\hat{V} = \sum_{i=1}^{N} Z_i' D D' Z_i$, where D is the first-difference matrix operator. A two-step GMM estimator uses a robust choice $\tilde{V} = \sum_{i=1}^{N} Z \Delta \hat{v}_i \Delta \hat{v}_i' Z_i$, where $\Delta \hat{v}_i$ are one-step GMM residuals.

Biorn (2000) proposes estimation of linear, static regression equations from panel data models with measurement errors in the regressors, showing that if the latent regressor is autocorrelated or non-stationary, several consistent OLS-IV and GMM estimators exist, provided some structure is imposed on the disturbances and measurement errors. He considers alternative GMM estimations that combine all essential orthogonality conditions. The procedures are very similar to the one described just above under non-autocorrelation in the disturbances. In particular, the required assumptions when allowing autocorrelation in the errors are very similar to those discussed in the previous section. For instance, when one allows for an MA(τ) structure in the measurement error, for instance, one must ensure that the variables in the IV matrix have a lead or lag of at least $\tau + 1$ periods to the regressor.

We briefly discuss the GMM estimators proposed by Biorn (2000). First, consider estimation using the equation in differences and instrumental variables in levels. After taking differences of the

[14] In models with exogenous explanatory variables, Z_i may consist of sub-matrices with the block diagonal (exploiting all or part of the moment restrictions), concatenated to straightforward one-column instruments.

model, there are $(T-1)+(T+1)$ equations that can be stacked for individual i as:

$$\begin{bmatrix} \Delta y_{i21} \\ \Delta y_{i32} \\ \vdots \\ \Delta y_{i,T,T-1} \\ \Delta y_{i31} \\ \Delta y_{i42} \\ \vdots \\ \Delta y_{i,T,T-2} \end{bmatrix} = \begin{bmatrix} \Delta x_{i21} \\ \Delta x_{i32} \\ \vdots \\ \Delta x_{i,T,T-1} \\ \Delta x_{i31} \\ \Delta x_{i42} \\ \vdots \\ \Delta x_{i,T,T-2} \end{bmatrix} \beta + \begin{bmatrix} \Delta \epsilon_{i21} \\ \Delta \epsilon_{i32} \\ \vdots \\ \Delta \epsilon_{i,T,T-1} \\ \Delta \epsilon_{i31} \\ \Delta \epsilon_{i42} \\ \vdots \\ \Delta \epsilon_{i,T,T-2} \end{bmatrix},$$

or compactly:

$$\Delta y_i = \Delta X_i \beta + \Delta\epsilon_i.$$

The IV matrix is the $((2T-3) \times KT(T-2))$ diagonal matrix with the instruments in the diagonal defined by Z. Let:

$$\begin{aligned} \Delta y &= [(\Delta y_1)', \ldots, (\Delta y_N)']', \quad \Delta\epsilon = [(\Delta\epsilon_1)', \ldots, (\Delta\epsilon_N)']' \\ \Delta X &= [(\Delta X_1)', \ldots, (\Delta X_N)']', \quad Z = [Z_1', \ldots, Z_N']'. \end{aligned}$$

The GMM estimator that minimizes $[N^{-1}(\Delta\epsilon)'Z](N^{-2}V)^{-1}[N^{-1}Z'(\Delta\epsilon)]$ for $V = Z'Z$ can be written as:

$$\begin{aligned} \hat{\beta}_{Dx} &= \left[\left[\sum_i (\Delta X_i)' Z_i\right]\left[\sum_i Z_i' Z_i\right]^{-1}\left[\sum_i Z_i'(\Delta X_i)\right]\right]^{-1} \\ &\quad \times \left[\left[\sum_i (\Delta X_i)' Z_i\right]\left[\sum_i Z_i' Z_i\right]^{-1}\left[\sum_i Z_i'(\Delta y_i)\right]\right]. \end{aligned}$$

If $\Delta\epsilon$ has a non-scalar covariance matrix, a more efficient GMM estimator, $\tilde{\beta}_{Dx}$, can be obtained setting $V = V_{Z(\Delta\epsilon)} = E\left[Z'(\Delta\epsilon)(\Delta\epsilon)'Z\right]$, and estimating $\hat{V}_{Z(\Delta\epsilon)}$ by:

$$\frac{\hat{V}_{Z(\Delta\epsilon)}}{N} = \frac{1}{N}\sum_i Z'(\widehat{\Delta\epsilon})(\widehat{\Delta\epsilon})'Z,$$

where $\widehat{\Delta\epsilon}_i = \Delta y_i - (\Delta X_i)\hat{\beta}_{Dx}$. This procedure assumes that (A1) and (A2) are satisfied. However, as Biorn (2000) argues, one can replace them by (B1) or (B2), and then ensure that the variables in the IV matrix have a lead or lag of at least $\tau+1$ periods to the regressor, to "get clear of" the τ period memory of the MA(τ) process. The procedure described below is also based on the same set of assumptions and can be extend similarly.[15]

[15] See Propositions 1* and 2* in Biorn (2000) for a formal treatment of the conditions.

The procedure for estimation using equation in levels and IV's in difference is similar. Consider the T stacked level equations for individual i:

$$\begin{bmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{bmatrix} = \begin{bmatrix} c \\ \vdots \\ c \end{bmatrix} + \begin{bmatrix} x_{i1} \\ \vdots \\ x_{iT} \end{bmatrix} \beta + \begin{bmatrix} \epsilon_{i1} \\ \vdots \\ \epsilon_{iT} \end{bmatrix},$$

or more compactly,

$$y_i = e_T c + X_i \beta + \epsilon,$$

where e_T denotes a $(T \times 1)$ vector of ones. Let the $(T \times T(T-2)K)$ diagonal matrix of instrument be denoted by ΔZ_i. This matrix has the instruments in difference in the main diagonal. In addition, define:

$$\begin{aligned} y &= [y_1', \ldots, y_N']', \quad \epsilon = [\epsilon_1', \ldots, \epsilon_N']' \\ X &= [X_1', \ldots, X_N']', \quad \Delta\ Z = [(\Delta Z_1)', \ldots, (\Delta Z_N)']'. \end{aligned}$$

The GMM estimator that minimizes $[N^{-1}\epsilon'(\Delta Z)'](N^{-2}V_\Delta)^{-1}[N^{-1}(\Delta Z)'\epsilon]$ for $V_\Delta = (\Delta Z)'(\Delta Z)$ is:

$$\begin{aligned} \hat{\beta}_{Lx} &= \left[\left[\sum_i X_i'(\Delta Z_i) \right] \left[\sum_i (\Delta Z_i)'(\Delta Z_i) \right]^{-1} \left[\sum_i (\Delta Z_i)' X_i \right] \right]^{-1} \\ &\quad \times \left[\left[\sum_i X_i(\Delta Z_i) \right] \left[\sum_i (\Delta Z_i)'(\Delta Z_i) \right]^{-1} \left[\sum_i (\Delta Z_i)' y_i \right] \right]. \end{aligned}$$

If ϵ has a non-scalar covariance matrix, a more efficient GMM estimator, $\tilde{\beta}_{Lx}$, can be obtained setting $V_\Delta = V_{(\Delta Z)\epsilon} = E\left[(\Delta Z)'\epsilon\epsilon'(\Delta Z)\right]$, and estimating $\hat{V}_{(\Delta Z)\epsilon}$ by:

$$\frac{\hat{V}_{(\Delta Z)\epsilon}}{N} = \frac{1}{N} \sum_i (\Delta Z)'\hat{\epsilon}\hat{\epsilon}'(\Delta Z),$$

where $\hat{\epsilon} = y_i - X_i \hat{\beta}_{Lx}$.

Finally, let us briefly contrast the OLS-IV and AB-GMM estimators. The advantages of GMM over IV are clear: if heteroskedasticity is present, the GMM estimator is more efficient than the IV estimator; while if heteroskedasticity is not present, the GMM estimator is no worse asymptotically than the IV. Implementing the GMM estimator, however, usually comes with a high price. The main problem, as Hayashi (2000, p. 215) points out, concerns the estimation of the optimal weighting matrix that is at the core of the GMM approach. This matrix is a function of fourth moments, and obtaining reasonable estimates of fourth moments requires very large sample sizes. Problems also arise when the number of moment conditions is high; that is, when there are "too many instruments." This latter problem affects squarely the implementation of the AB-GMM, since it relies on large numbers of lags (especially in long panels). The upshot is that the efficient GMM estimator can have poor small sample properties [see Baum, Schaffer and Stillman (2003) for a discussion]. These problems are well documented and remedies have been proposed by, among others, Altonji and Segal (1996) and Doran and Schmidt (2006).

3 Monte Carlo Analysis

We use Monte Carlo simulations to assess the finite sample performance of the EW and IV estimators discussed in Section 2. Monte Carlo simulations are an ideal experimental tool because they enable us to study those two estimators in a controlled setting, where we can assess and compare the importance of elements that are key to estimation performance. Our simulations use several distributions to generate observations. This is important because researchers will often find a variety of distributions in real-world applications, and because one ultimately does not see the distribution of the mismeasurement term. Our Monte Carlos compare the EW, OLS-IV, and AB-GMM estimators presented in Section 2 in terms of bias and RMSE.[16] We also investigate the properties of the EW identification test, focusing on the empirical size and power of this test.

3.1 Monte Carlo design

A critical feature of panel data models is the observation of multiple data points from the same individuals over time. It is natural to consider that repeat samples are particularly useful in that individual idiosyncrasies are likely to contain information that might influence the error structure of the data-generating process.

We consider a simple data-generating process to study the finite sample performance of the EW and OLS-IV estimators. The response variable y_{it} is generated according to the following model:

$$y_{it} = \gamma_i + \beta\chi_{it} + z_{it}'\alpha + (1 + \rho w_{it})u_{it}, \tag{18}$$

where γ_i captures the individual-specific intercepts, β is a scalar coefficient associated with the mismeasured variable χ_{it}, $\alpha = (\alpha_1, \alpha_2, \alpha_3)'$ is 3×1 vector of coefficients associated with the 3×1 vector

[16] The mean squared error (MSE) of an estimator $\hat{\theta}$ incorporates a component measuring the variability of the estimator (precision) and another measuring its bias (accuracy). An estimator with good MSE properties has small combined variance and bias. The MSE of $\hat{\theta}$ can be defined as: $\mathrm{Var}(\hat{\theta}) + [\mathrm{Bias}(\hat{\theta})]^2$. The root mean squared error (RMSE) is simply the square root of the MSE. This is an easily interpretable statistic, since it has the same unit as the estimator $\hat{\theta}$. For an approximately unbiased estimator, the RMSE is just the square root of the variance, that is, the standard error.

of perfectly-measured variables $z_{it} = (z_{it1}, z_{it2}, z_{it3})$, u_{it} is the error in the model, and ρ modulates the amount of heteroscedasticity in the model. When $\rho = 0$, the innovations are homoscedastic. When $\rho > 0$, there is heteroscedasticity associated with the variable w_{it}, and this correlation is stronger as the coefficient gets larger. The model in (18) is flexible enough to allow us to consider two different variables as w_{it}: (1) the individual specific intercept γ_i, and (2) the well-measured regressor z_{it}.

We consider a standard additive measurement error:

$$x_{it} = \chi_{it} + v_{it}, \tag{19}$$

where χ_{it} follows an AR(1) process:

$$(1 - \phi L)\chi_{it} = \epsilon_{it}. \tag{20}$$

In all simulations, we set $\chi^*_{i,-50} = 0$ and generate χ_{it} for $t = -49, -48, ..., T$, such that we drop the first 50 observations. This ensures that the results are not unduly influenced by the initial values of the χ_{it} process.

Following Biorn (2000), we relax the assumption of i.i.d. measurement error. Our benchmark simulations will use the assumption of time-invariant autocorrelation [(C1) and (C2)]. In particular, we assume that $u_{it} = u_{1i} + u_{2it}$ and $v_{it} = v_{1i} + v_{2it}$. We draw all the innovations $(u_{1i}, u_{2it}, v_{1i}, v_{2it})$ from a lognormal distribution; that is, we exponentiate two normal distributions and standardize the resulting variables to have unit variances and zero means (this follows the approach used by EW). In Section 3.6, we analyze the alternative case in which the innovations follow a MA structure.

The perfectly-measured regressor is generated according to:

$$z_{it} = \mu_i + \varepsilon_{it}. \tag{21}$$

And the fixed effects, μ_i and γ_i, are generated as:

$$\begin{aligned} \mu_i &= e_{1i} \\ \gamma_i &= e_{2i} + \frac{1}{\sqrt{T}} \sum_{t=1}^{T} W_{it}, \end{aligned} \tag{22}$$

where W_{it} is the sum of the explanatory variables. Our method of generating μ_i and γ_i ensures that the usual random effects estimators are inconsistent because of the correlation that exists between the individual effects and the error term or the explanatory variables. The variables (e_{1i}, e_{2i}) are fixed as standard normal distributions.[17]

[17] Robustness checks show that the choice of a standard normal does not influence our results.

We employ four different schemes to generate the disturbances $(\epsilon_{it}, \varepsilon_{it})$. Under Scheme 1, we generate them under a normal distribution, $N(0, \sigma_u^2)$. Under Scheme 2, we generate them from a lognormal distribution, $LN(0, \sigma_u^2)$. Under Scheme 3, we use a chi-square with 5 degrees of freedom, χ_5^2. Under Scheme 4, we generate the innovations from a $F_{m,n}$-distribution with $m = 10$ and $n = 40$. The latter three distributions are right-skewed so as to capture the key distributional assumptions behind the EW estimator. We use the normal (non-skewed) distribution as a benchmark.

Naturally, in practice, one cannot determine how skewed — if at all — is the distribution of the partially out latent variable. One of our goals is to check how this assumption affects the properties of the estimators we consider. Figure 1 provides a visual illustration of the distributions we employ. By inspection, at least, the three skewed distributions we study appear to be plausible candidates for the distribution governing mismeasurement, assuming EW's prior that measurement error must be markedly rightly-skewed.

– Figure 1 about here –

As in Erickson and Whited (2002), our simulations allow for cross-sectional correlation among the variables in the model. We do so because this correlation may aggravate the consequences of mismeasurement of one regressor on the estimated slope coefficients of the well-measured regressors. Notably, this source of correlation is emphasized by EW in their argument that the inferences of Fazzari, Hubbard, and Petersen (1988) are flawed in part due to the correlation between q and cash flows. To introduce this correlation in our application, for each period in the panel we generate $(\chi_i, z_{i1}, z_{i2}, z_{i3})$ using the correspondent error distribution and then multiply the resulting vector by $[\text{var}(\chi_i, z_{i1}, z_{i2}, z_{i3})]^{1/2}$ with diagonal elements equal to 1 and off-diagonal elements equal to 0.5.

In the simulations, we experiment with $T = 10$ and $N = 1{,}000$. We set the number of replications to 5,000 and consider the following values for the remaining parameters:

$$(\beta, \alpha_1, \alpha_2, \alpha_3) = (1, -1, 1, -1)$$

$$\phi = 0.6, \ \sigma_u^2 = \sigma_{e1}^2 = \sigma_{e2}^2 = 1,$$

where the set of slope coefficients β, α_i is set similarly to EW.

Notice that the parameter ϕ controls the amount of autocorrelation of the latent regressor. As explained above, this autocorrelation is an important requirement for the identification of the IV estimator. While we set $\phi = 0.6$ in the following experiments, we also conduct simulations in which we check the robustness of the results with respect to variations in ϕ between 0 and 1 (see Section 3.6).

3.2 The EW Identification Test

We study the EW identification test in a simple panel data set up. In the panel context, it is important to consider individual-fixed effects. If the data contain fixed effects, according to Erickson and Whited (2000), a possible strategy is to transform the data first and then apply their high-order GMM estimator. Accordingly, throughout this section our estimations consider data presented in two forms: "level" and "within." The first refers to data in their original format, without the use of any transformation; estimations in level-form ignore the presence of fixed effects.[18] The second applies the within transformation to the data — eliminating fixed effects — before the model estimation.

We first compute the empirical size and power of the test. Note that the null hypothesis is that the model is incorrectly specified, such that $\beta = 0$ and/or $E[\eta_i^3] = 0$. The *empirical size* is defined as the number of rejections of the null hypothesis when the null is true — ideally, this should hover around 5%. In our case, the empirical size is given when we draw the innovations $(\epsilon_{it}, \varepsilon_{it})$ from a non-skewed distribution, which is the normal distribution since it generates $E[\eta_i^3] = 0$. The *empirical power* is the number of rejections when the null hypothesis is false — ideally, this should happen with very high probability. In the present case, the empirical power is given when we use skewed distributions: lognormal, chi-square, and F-distribution.

Our purpose is to investigate the validity of the skewness assumption once we are setting $\beta \neq 0$. Erickson and Whited (2002) also restrict every element of β to be nonzero. We conduct a Monte Carlo experiment to quantify the second part of this assumption. It is important to note that we can compute the $E[(\eta_i)^3]$ since, in our controlled experiment, we generate χ_i, and therefore observe it.

Since the EW test is originally designed for cross-sectional data, the first difficulty the researcher faces when implementing a panel test is aggregation. Following EW, our test is computed for each year separately. We report the average of empirical rejections over the years.[19] To illustrate the size and power of the test for the panel data case, we set the time series dimension of the panel to $T = 10$. Our tests are performed over $5,000$ samples of cross-sectional size equal to $1,000$. We use a simple homoscedastic model with $\rho = 0$, with the other model parameters given as above.

Table 1 reports the empirical size and power of the statistic proposed by EW for testing the null hypothesis $H_0 : E(\dot{y}_i^2 \dot{x}_i) = E(\dot{y}_i \dot{x}_i^2) = 0$. This hypothesis is equivalent to testing $H_0 : \beta = 0$ and/or $E(\eta_i^3) = 0$. Table 1 reports the frequencies at which the statistic of test is rejected at the 5% level of significance, for respectively, the normal, lognormal, chi-square, and F-distributions of the

[18] To our knowledge, all but one of the empirical applications of the EW model use the data in level form. In other words, firm-fixed effects are ignored outright in panel setting estimations of parameters influencing firm behavior.

[19] The results using the median are similar.

data-generating process. Recall, when the null hypothesis is true we have the size of the test, and when the null is false we have the power of the test.

The results reported in Table 1 imply an average size of approximately 5% for the test. In particular, the first two rows in the table show the results in the case of a normal distribution for the residuals (implying that we are operating under the null hypothesis). For both the level and within cases, the empirical sizes match the target significance level of 5%.

When we move to the case of skewed distributions (lognormal, chi-square, and F), the null hypothesis is not satisfied by design, and the number of rejections delivers the empirical power of the test. In the case when the data is presented in levels and innovations are drawn from a lognormal distribution (see row 2), the test rejects about 47% of the time the null hypothesis of no skewness. Using within data, the test rejects the null hypothesis 43% of the time. Not only are these frequencies low, but comparing these results one can see that the within transformation slightly reduces the power of the test.

The results associated with the identification test are more disappointing when we consider other skewed distributions. For example, for the F-distribution, we obtain only 17% of rejections of the null hypothesis in the level case, and only 28% for the within case. Similarly, poor statistical properties for the model identification test are observed in the chi-square case.

– Table 1 about here –

3.3 Bias and Efficiency of the EW, OLS-IV, and AB-GMM Estimators

In this section we present simulation results that assess the finite sample performance of the estimators discussed in Section 2. The simulations compare the estimators in terms of bias and efficiency under several distributional assumptions. In the next subsection we consider the cross-sectional setting, focusing on the properties of the EW estimator. Subsequently, we examine the panel case in detail, comparing the performance of the EW, OLS-IV, and AB-GMM estimators in terms of bias and efficiency.

3.3.1 The cross-sectional case

We generate data using a simple model as in (18)–(19) with $T = 1$, such that there are no fixed effects, no autocorrelation ($\phi = 0$), and no heteroscedasticity ($\rho = 0$). The other parameters are $(\beta, \alpha_1, \alpha_2, \alpha_3) = (1, -1, 1, -1)$. Table 2 shows the results for bias and RMSE for four different distributions: lognormal, Chi-Square, F-distribution, and standard normal. For each distribution we estimate the model using three different EW estimators: EW-GMM3, EW-GMM4, and EW-GMM5.

These estimators are based on the respective third, fourth, and fifth moment conditions. By combining the estimation of 4 parameters, under 4 different distributions, for all 3 EW estimators — a total of 48 estimates — we aim at establishing robust conclusions about the bias and efficiency of the EW approach.

Panel A of Table 2 presents the results for bias and RMSE when we use the lognormal distribution to generate innovations $(\epsilon_i, \varepsilon_i)$ that produce χ_i and z_i. Under this particular scenario, point estimates are approximately unbiased, and the small RMSEs indicate that coefficients are relatively efficiently estimated.

Panels B and C of Table 2 present the results for the chi-square and F-distribution, respectively. The experiments show that coefficient estimates produced by the EW approach are generally very biased. For example, Panel B shows that the β coefficient returned for the EW-GMM4 and EW-GMM5 estimators is biased downwards by approximately 35%. Panel C shows that for EW-GMM3, the β coefficient is biased upwards about 35%. Paradoxically, for EW-GMM4 and EW-GMM5, the coefficients are biased downwards by approximately 25%. The coefficients returned for the perfectly-measured regressors are also noticeably biased. And they, too, switch bias signs in several cases. Panels B and C show that the EW RMSEs are very high. Notably, the RMSE for EW-GMM4 under the chi-square distribution is 12.23, and under F-distribution it is 90.91. These RMSE results highlight the lack of efficiency of the EW estimator. Finally, Panel D presents the results for the normal distribution case, which has zero skewness. In this case, the EW estimates are severely biased and the RMSEs are extremely high. The estimated coefficient for the mismeasured variable using EW-GMM3 has a bias of 1.91 (about 3 times larger than its true value) and a RMSE of 2305.

These results reveal that the EW estimators only have acceptable performance in the case of very strong skewness (lognormal distribution). They relate to the last section in highlighting the poor identification of the EW framework, even in the most basic cross-sectional set up. Crucially, for the other skewed distributions we study, the EW estimator is significantly biased for *both* the mismeasured and the well-measured variables. In addition, the RMSEs are quite high, indicating low efficiency.

– Table 2 about here –

3.3.2 The panel case

We argue that a major drawback of the EW estimator is its limited ability to handle individual heterogeneity — fixed effects and error heteroscedasticity — in panel data. This section compares the impact of individual heterogeneity on the EW, OLS-IV, and AB-GMM estimators in a panel setting.

In the first round of experiments, we assume error homoscedasticity by setting the parameter ρ in (18) equal to zero. We shall later allow for changes in this parameter.

Although the EW estimations are performed on a period-by-period basis, one generally wants a single coefficient for each of the variables in an empirical model. To combine the various (time-specific) estimates, EW suggest the minimum distance estimator (MDE) described below. Accordingly, the results presented in this section are for the MDE that combines the estimates obtained from each of the 10 time periods considered. For example, EW-GMM3 is the MDE that combines the 10 different cross-sectional EW-GMM3 estimates in our panel.

The OLS-IV is computed after differencing the model and using the second lag of the observed mismeasured variable x, x_{t-2}, as an instrument for Δx_t. The AB-GMM estimates (Arellano and Bond, 1991) use all the orthogonality conditions, with all available lags of x's as instrumental variables. We also concatenate the well-measured variables z's in the instruments' matrix. The AB-GMM estimator is also computed after differencing equation (18). To highlight the gains of these various estimators vis-à-vis the standard (biased) OLS estimator, we also report the results of simulations for OLS models using equation in first-difference without instruments.

We first estimate the model using data in level form. While the true model contains fixed effects (and thus it is appropriate to use the within transformation), it is interesting to see what happens in this case since most applications of the EW estimator use data in level form, and as shown previously, the EW identification test performs slightly better using data in this form.

The results are presented in Table 3. The table makes it clear that the EW method delivers remarkably biased results when ignoring the presence of fixed effects. Panel A of Table 3 reports the results for the model estimated with the data under strong skewness (lognormal). In this case, the coefficients for the mismeasured regressor are very biased, with biases well in excess of 100% of the true coefficient for the EW-GMM3, EW-GMM4, and EW-GMM5 estimators. The biases for the well-measured regressors are also very strong, all exceeding 200% of the estimates' true value. Panels B and C report results for models under chi-square and F distributions, respectively. The EW method continues to deliver very biased results for all of the estimates considered. For example, the EW-GMM3 estimates that are returned for the mismeasured regressors are biased downwardly by about 100% of their true values — those regressors are deemed irrelevant when they are not. Estimates for the well-measured regressors are positively biased by approximately 200% — they are inflated by a factor of 3. The RMSEs reported in Panels A, B, and C show that the EW methodology produces very inefficient estimates even when one assumes pronounced skewness in the data.

Finally, Panel D reports the results for the normal distribution. For the non-skewed data case, the EW framework can produce estimates for the mismeasured regressor that are downwardly biased by about 90% of their true parameter values for all models. At the same time, that estimator induces an upward bias of larger than 200% for the well-measured regressors.

– Table 3 about here –

Table 4 reports results for the case in which we apply the within transformation to the data. Here, we introduce the OLS, OLS-IV, and AB-GMM estimators. We first present the results associated with the set up that is *most favorable* for the EW estimations, which is the lognormal case in Panel A. The EW estimates for the lognormal case are relatively unbiased for the well-measured regressors (between 4% and 7% deviation from true parameter values). The same applies for the mismeasured regressors. Regarding the OLS-IV, Panel A shows that coefficient estimates are unbiased in all models considered. AB-GMM estimates are also approximately unbiased, while standard OLS estimates are very biased. In terms of efficiency, the RMSEs of the EW-GMM3 are somewhat smaller than those of the OLS-IV and AB-GMM for the well-measured and mismeasured regressors. However, for the mismeasured regressor, both OLS-IV and AB-GMM have smaller RMSEs than EW-GMM4 and EW-GMM5.

Panel B of Table 4 presents the results for the chi-square distribution. One can see that the EW yields markedly biased estimates in this case. The bias in the mismeasured regressor is approximately 38% (downwards), and the coefficients for the well-measured variable are also biased (upwards). In contrast, the OLS-IV and AB-GMM estimates for both well-measured and mismeasured regressors are approximately unbiased. In terms of efficiency, as expected, the AB-GMM presents slightly smaller RMSEs than the OLS-IV estimator. These IV estimators' RMSEs are much smaller than those associated with the EW estimators.

Panels C and D of Table 4 show the results for the F and standard normal distributions, respectively. The results for the F-distribution in Panel C are essentially similar to those in Panel B: the instrumental variables estimators are approximately unbiased while the EW estimators are very biased. Finally, Panel D shows that deviations from a strongly skewed distribution are very costly in terms of bias for the EW estimator, since the bias for the mismeasured regressor is larger than 70%, while for the well-measured it is around 20%. A comparison of RMSEs shows that the IV estimators are more efficient in both the F and normal cases. In all, our simulations show that standard IV methods almost universally dominate the EW estimator in terms of bias and efficiency.

We reiterate that the bias and RMSE of the IV estimators in Table 4 are all relatively invariant to the distributional assumptions, while the EW estimators are all very sensitive to those assumptions.

In short, this happens because the EW relies on the high-order moment conditions as opposed to the OLS and IV estimators.

– Table 4 about here –

3.4 Heteroscedasticity

One way in which individual heterogeneity may manifest itself in the data is via error heteroscedasticity. Up to this point, we have disregarded the case in which the data has a heteroscedastic error structure. However, most empirical applications in corporate finance entail the use of data for which heteroscedasticity might be relevant. It is important that we examine how the EW and the IV estimators are affected by heteroscedasticity.[20]

The presence of heteroscedasticity introduces heterogeneity in the model and consequently in the distribution of the partialled out dependent variable. This compromises identification in the EW framework. Since the EW estimator is based on equations giving the moments of $(y_i - z_i\mu_y)$ and $(x_i - z_i\mu_\chi)$ as functions of β and moments of $(u_i, \varepsilon_i, \eta_i)$, the heteroscedasticity associated with the fixed effects (α_i) or with the perfectly measured regressor (z_{it}) distorts the required moment conditions associated with $(y_i - z_i\mu_y)$, yielding biased estimates. These inaccurate estimates enter the minimum distance estimator equation and consequently produce incorrect weights for each estimate along the time dimension. As our simulations of this section demonstrate, this leads to biased MDE estimates, where the bias is a function of the amount of heteroscedasticity.

We examine the biases imputed by heteroscedasticity by way of graphical analysis. The graphs we present below are useful in that they synthesize the outputs of numerous tables and provide a fuller visualization of the contrasts we draw between the EW and OLS-IV estimators. The graphs depict the sensitivity of those two estimators with respect to heteroscedasticity as we perturb the coefficient ρ in equation (18).

In our simulations, we alternatively set $w_{it} = \gamma_i$ or $w_{it} = z_{it}$. In the first case, heteroscedasticity is associated with the individual effects. In the second, heteroscedasticity is associated with the the well-measured regressor. Each of our figures describes the biases associated with the mismeasured and the well-measured regressors for each of the OLS-IV, EW-GMM3, EW-GMM4, and EW-GMM5 estimators.[21] In order to narrow our discussion, we only present results for the highly skewed distribution case (lognormal distribution) and for data that is treated for fixed effects using the within transformation. As Section 3.3.2 shows, this is the *only case* in which the EW estimator returns

[20] We focus on the OLS-IV estimator hereinafter for the purpose of comparison with the EW estimator.

[21] Since estimation biases have the same features across all well-measured regressors of a model, we restrict attention to the first well-measured regressor of each of the estimated models.

relatively unbiased estimators for the parameters of interest. In all the other cases (data in levels, and for data generated by chi-square, F, and normal distributions), the estimates are strongly biased even under the assumption of homoscedasticity.[22]

Figure 2 presents the simulation results under the assumption that $w_{it} = \gamma_i$, as we vary the amount of heteroscedasticity by changing the parameter ρ.[23] The results for the mismeasured coefficients show that biases in the EW estimators are generally small for ρ equal to zero (this is the result reported in Section 3.3.2). However, as this coefficient increases, the bias quickly becomes large. For example, for $\rho = 0.45$, the biases in the coefficient of the mismeasured variable are, respectively, –11%, –20%, and –43%, for the EW-GMM3, EW-GMM4, and EW-GMM5 estimators. Notably, those biases, which are initially negative, turn positive for moderate values of ρ. As heteroscedasticity increases, some of the biases diverge to positive infinite. The variance of the biases of the EW estimators is also large. The results regarding the well-measured variables using EW estimators are analogous to those for the mismeasured one. Biases are substantial even for small amounts of heteroscedasticity, they switch signs for some level of heteroscedasticity, and their variances are large. In sharp contrast, the same simulation exercises show that the OLS-IV estimates are approximately unbiased even under heteroscedasticity. While the EW estimator may potentially allow for some forms of heteroskedasticity, it is clear that it is not well equipped to deal with this problem in more general settings.

– Figure 2 about here –

3.5 Identification of the EW estimator in panel data

Our Monte Carlo experiments show that the EW estimator has a poor handle of individual-fixed effects and that biases arise for deviations from the assumption of strict lognormality. Biases in the EW framework are further magnified if one allows for heteroscedasticity in the data (even under lognormality). The biases arising from the EW framework are hard to measure and sign, ultimately implying that it can be very difficult to replicate the results one obtains under that framework.

To better understand these results, we now discuss in more mathematical details the identification of the EW estimator for the panel data case for both the model in level and after the within transformation. Extending the EW estimator to panel data seems to be a nontrivial task. EW have proposed to break the problem for each time-series, and estimate a cross-section model for each t,

[22] Our simulation results (available upon request) suggest that introducing heteroscedasticy makes the performance of the EW estimator even worse in these cases.

[23] The results for $w_{it} = z_{it}$ are quite similar to those we get from setting $w_{it} = \gamma_i$. We report only one set of graphs to save space.

and after that combine the estimates using a Minimum Distance estimator. In what follows we show that this procedure might affect the identification condition.

Consider the following model:

$$y_{it} = \alpha_i + \beta\chi_{it} + u_{it}, \quad i = 1, ..., N; t = 1, ..., T, \tag{23}$$

where u_{it} is independent and identically distributed, with mean zero and variance σ_u^2. Assume that the $Var(x_{it}) = \sigma_\chi^2$. The independent variable and unobserved effects are exogenous, that is, $Cov(\chi_{it}, u_{is}) = Cov(\alpha_i, u_{it}) = 0$ for any t and s. However, $Cov(\alpha_i, \chi_{it}) \neq 0$. Now, assume that we do not observe the true variable χ_{it}^*, but rather a mismeasured variable, that is, you observe the following variable with an error

$$x_{it} = \chi_{it} + e_{it}, \quad i = 1, ..., N; t = 1, ..., T, \tag{24}$$

where $Cov(x_{it}, e_{is}) = Cov(\alpha_i, e_{is}) = Cov(u_{it}, e_{is}) = 0$, and $Var(e_{it}) = \sigma_e^2$, $Cov(e_{it}, e_{it-1}) = \gamma\sigma_e^2$.

In addition, assume here that there is no variable z_{it} ($\alpha = 0$ in equation 1) to simplify the argument.

3.5.1 Model in level

As mentioned before, EW propose to fix a particular time series and estimate the model using the cross-section data. Without loss for generality, fix $t = 1$. Thus, equations (23) and (24) become

$$y_{i1} = \alpha_i + \beta\chi_{i1} + u_{i1}, \quad i = 1, ..., N, \tag{25}$$

and

$$x_{i1} = \chi_{i1} + e_{i1}, \quad i = 1, ..., N. \tag{26}$$

However, the unobserved individual specific intercepts, α_i, is still present in equation (25) and in addition $Cov(\alpha_i, \chi_{i1}) \neq 0$. Therefore, one can see that it is impossible to estimate β consistently since α_i's are unobserved. This argument is easily extended for every $t = 1, ..., T$. Thus, the estimator for each fixed t is inconsistent, and consequently the Minimum Distance estimator is inconsistent by construction. Therefore, we conclude that the EW Minimum Distance estimator produces inconsistent estimates for panel data model with fixed effects.

3.5.2 Model after within transformation

Given the inconsistency of the model in levels presented in the last section, one strategy is to previously transform the data to eliminate the fixed effects. One suggestion is to use the within transformation in the data before estimation.

In order to analyze the model after the transformation, let's assume that $T = 2$ for simplification. Using the within transformation in (23) and (24) we obtain

$$y_{it} - \bar{y}_i = \beta(\chi_{it} - \bar{\chi}_i) + (u_{it} - \bar{u}_i),$$

and

$$x_{it} - \bar{x}_i = (\chi_{it} - \bar{\chi}_i) + (e_{it} - \bar{e}_i),$$

where $\bar{y}_i = \frac{1}{2}(y_{i1} + y_{i2})$, $\bar{\chi}_i = \frac{1}{2}(\chi_{i1} + \chi_{i2})$, and so on.

Now, again EW propose to use a particular time-series and estimate the model using the cross-section data. Without loss of generality let's set $t = 1$. The model can be written as

$$y_{i1} - \bar{y}_i = \beta(\chi_{i1} - \bar{\chi}_i) + (u_{i1} - \bar{u}_i),$$

and

$$x_{i1} - \bar{x}_i = (\chi_{i1} - \bar{\chi}_i) + (e_{it} - \bar{e}_i),$$

Now substituting the definition of the deviations and rearranging we have

$$y_{i1} - \frac{1}{2}(y_{i1} + y_{i2}) = \beta\big(\chi_{i1} - \frac{1}{2}(\chi_{i1} + \chi_{i2})\big) + \big(u_{i1} - \frac{1}{2}(u_{i1} + u_{i2})\big),$$
$$y_{i1} + y_{i2} = \beta\big(\chi_{i1} + \chi_{i2}\big) + \big(u_{i1} + u_{i2}\big),$$

or:

$$x_{i1} - \frac{1}{2}(x_{i1} + x_{i2}) = \big(\chi_{i1} - \frac{1}{2}(\chi_{i1} + \chi_{i2})\big) + \big(e_{i1} - \frac{1}{2}(e_{i1} + e_{i2})\big),$$
$$x_{i1} + x_{i2} = \big(\chi_{i1} + \chi_{i2}\big) + \big(e_{i1} + e_{i2}\big).$$

Finally, our model can be described as

$$y_{i1} + y_{i2} = \beta\big(\chi_{i1} + \chi_{i2}\big) + \big(u_{i1} + u_{i2}\big),$$

and

$$x_{i1} + x_{i2} = \big(\chi_{i1} + \chi_{i2}\big) + \big(e_{i1} + e_{i2}\big),$$

Let's now define $Y_i = y_{i1} + y_{i2}$, $X_i = x_{i1} + x_{i2}$, $U_i = u_{i1} + u_{i2}$, $\nu_i = \chi_{i1} + \chi_{i2}$ and $E_i = e_{i1} + e_{i2}$.

So, the model could be rewrite the model as

$$Y_i = \beta\nu_i + U_i$$

and

$$X_i = \nu_i + E_i.$$

Finally, note that the requirements for identification now are on the high order moments of (ν, U, E). However, note that $\nu_i = \chi_{i1} + \chi_{i2}$, which is a sum of two random variables. As it is well know from the econometrics literature convolution of random variables are in general a nontrivial object.

One example of why the identification condition worsens considerably is the following. Consider a model where χ_{i1} and χ_{i2} are independent Chi-square distributions with 2 degrees of freedom. The skewness of the Chi-square with k degrees of freedom is $\sqrt{8/k}$. Note that the sum of two independent Chi-squares with k degrees of freedom is a Chi-square with $2k$ degrees of freedom. Therefore, the skewness of the $\nu_i = \chi_{i1} + \chi_{i2}$ drops from 2 for the model using one distribution to 1.41 for the model using the summation of both χ_{i1} and χ_{i2}.

From this simple analysis one could conclude that the identification conditions required for EW estimator deteriorates considerably when using the within transformation to eliminate the fixed effects in panel data. Thus, the required conditions to achieve unbiased estimates with EW are very strong.

3.6 Revisiting the OLS-IV Assumptions

Our Monte Carlo simulations show that the OLS-IV estimator is consistent even when one allows for autocorrelation in the measurement error structure. We have assumed, however, some structure on the processes governing innovations. In this section, we examine the sensitivity of the OLS-IV results with respect to our assumptions about measurement error correlation and the amount of autocorrelation in the latent regressor. These assumptions can affect the quality of the instruments and therefore should be examined in some detail.

We first examine conditions regarding the correlation of the measurement errors and disturbances. The assumption of time-invariant autocorrelation for measurement errors and disturbances implies that past shocks to measurement errors do not affect the current level of the measurement error. One way to relax this assumption is to allow for the measurement error process to have a moving average structure. This structure satisfies Biorn's assumptions (B1) and (B2). In this case, Proposition 1 in Biorn (2000) shows that for a MA(τ) the instruments should be of order of at most $t - \tau - 2$. Intuitively, the set of instruments must be "older" than the memory of the measurement error process. For example, if the measurement error is MA(1), then one must use third- and longer-lagged instruments to identify the model.

To analyze this case, we conduct Monte Carlo simulations in which we replace the time invariant assumption for innovation u_{it} and v_{it} with a MA(1) structure for the measurement error process. The degree of correlation in the MA process is set to $\theta = 0.4$. Thus, the innovation in equations (18) and (19) have the following structure:

$$u_{it} = u_{1it} - \theta u_{1it-1} \text{ and } v_{it} = v_{1it} - \theta v_{1it-1},$$

with $|\theta| \leq 1$, and u_{1it} and v_{1it} are i.i.d. lognormal distributions. The other parameters in the simulation remain the same.

The results are presented in Table 5. Using MA(1) in the innovations and the third lag of the latent regressor as an instrument (either on its own or in combination with the fourth lag), the bias of the OLS estimator is very small (approximately 2% to 3%). The bias increases somewhat when we use only the fourth lag. While the fourth is an admissible instrument in this case, using longer lags decreases the implied autocorrelation in the latent regressor [which follows an AR(1) process by equation (20)]. This effect decreases somewhat the quality of the instruments. Notice also that when we do not eliminate short lags from the instrument set, the identification fails. For example, the bias is 60% when we use the second lag as an instrument. These results thus underscore the importance of using long enough lags in this MA case. Table 5 also reports results based on a MA(2) structure. The results are qualitatively identical to those shown in the MA(1) case. Once again, the important condition for identification is to use long enough lags (no less than four lags in this case).[24]

– Table 5 about here –

The second condition underlying the use of the OLS-IV is that the latent regressor is not time invariant. Accordingly, the degree of autocorrelation in the process for the latent regressor is an important element of the identification strategy. We assess the sensitivity of the OLS-IV results to this condition by varying the degree of autocorrelation through the autoregressive coefficient in the AR(1) process for the latent regressor. In these simulations, we use a time-invariant autocorrelation condition for the measurement error, but the results are very similar for the MA case.

Figure 3 shows the results for the bias in the coefficients of interest for the well-measured and mismeasured variables, using the second lag of the mismeasured variable as an instrument. The results show that the OLS-IV estimator performs well for a large range of the autoregressive coefficient. How-

[24] We note that if the instrument set uses suitably long lags, then the OLS-IV results are robust to variations in the degree of correlation in the MA process. In unreported simulations under MA(1), we show that the OLS bias is nearly invariant to the parameter θ.

ever, as expected, when the ϕ coefficient is very close to zero or one, we have evidence of a weak instruments problem. For example, when $\phi = 1$, then $\Delta\chi_{it}$ is uncorrelated with any variable dated at time $t-2$ or earlier. These simulations show that, provided that one uses adequately lagged instruments, the exact amount of autocorrelation in the latent variable is not a critical aspect of the estimation.

– Figure 3 about here –

The simulations of this section show how the performance of the OLS-IV estimator is affected by changes in assumptions concerning measurement errors and latent regressors. In practical applications, it is important to verify whether the results obtained with OLS-IV estimators are robust to the elimination of short lags from the instrumental set. This robustness check is particularly important given that the researcher will be unable to pin down the process followed by the measurement error. Our empirical application below incorporates this suggestion. In addition, identification relies on some degree of autocorrelation in the process for the latent regressor. While this condition cannot be directly verified, we can perform standard tests of instrument adequacy that rely on "first-stage" test statistics calculated from the processes for the observable variables in the model.

Another important assumption in the OLS is non-autocorrelation in both u_{it} and v_{it}. For example, these innovations cannot follow an autoregressive process. When, this is the case, the IV strategy of using lags of mismeasured variable as valid instruments is invalid (see Biorn, 2000).

3.7 Distributional Properties of the EW and OLS-IV Estimators

A natural question is whether our simulation results are rooted in the lack of accuracy of the asymptotic approximation of the EW method. Inference in models with mismeasured regressors are based on asymptotic approximations, hence inference based on estimators with poor approximations might lead to wrong inference procedures. For instance, we might select wrong critical values for a test under poor asymptotic approximations, and make inaccurate statements under such circumstances. In this section, we use the panel data simulation procedure of Section 3.3.2 to study and compare the accuracy of the asymptotic approximation of the EW and IV methods. To save space, we restrict our attention to the mismeasured regressor coefficient for the EW-GMM5 and OLS-IV cases. We present results where we draw the data from the lognormal, chi-square, and F distributions. The EW-GMM5 estimator is computed after the within transformation and the OLS-IV uses second lags as instruments.

One should expect both the IV and EW estimators to have asymptotically normal representations, such that when we normalize the estimator by subtracting the true parameter and divide by the standard deviation, this quantity behaves asymptotically as a normal distribution. Accordingly, we compute the empirical density and the distribution functions of the normalized sample estimators and their normal approximations. These functions are plotted in Figure 4. The true normal density and distribution functions (drawn in red) serve as benchmarks. The graphs in Figure 4 depict the accuracy of the approximation. We calculate the density of the estimators using a simple Gaussian Kernel estimator, and also estimate the empirical cumulative distribution function.[25]

Consider the lognormal distribution (first panel). In that case, the OLS-IV (black line) displays a very precise approximation to the normal curve in terms of both density and distribution. The result for the OLS-IV is robust across all of the distributions considered (lognormal, chi-square, and F). These results are in sharp contrast to those associated with the EW-GMM5 estimator. This estimator presents a poor asymptotic approximation for all distributions examined. For the lognormal case, the density is not quite centered at zero, and its shape does not fit the normal distribution. For the chi-square and F-distributions, Figure 4 shows that the shapes of the density and distribution functions are very unlike the normal case, with the center of the distribution located far away from zero. These results imply that inference procedures using the EW estimator might be asymptotically invalid in simple panel data with fixed effects, even when the relevant distributions present high skewness.

– Figure 4 about here –

4 Empirical Application

We apply the EW and OLS-IV estimators to Fazzari, Hubbard, and Petersen's (1988) investment equation. This is the most well-known model in the corporate investment literature and we use this application as a way to illustrate our Monte Carlo-based results. In the Fazzari, Hubbard, and Petersen model, a firm's investment spending is regressed on a proxy for investment demand (Tobin's q) and the firm's cash flow. Theory suggests that the correct proxy for the firm's investment demand is *marginal* q, but this quantity is unobservable and researchers use instead its measurable proxy, *average* q. Because average q measures marginal q imperfectly, a measurement problem naturally arises. Erickson and Whited (2002) uses the Fazzari, Hubbard, and Petersen model to motivate the adoption of their estimator in applied work in panel data.

A review of the corporate investment literature shows that virtually all empirical work in the area considers panel data models with firm-fixed effects (Kaplan and Zingales, 1997; Rauh, 2006; and

[25] The empirical cumulative distribution function F_n is a step function with jumps i/n at observation values, where i is the number of tied observations at that value.

Almeida and Campello, 2007). From an estimation point of view, there are distinct advantages in exploiting repeated observations from individuals to identify the model (Blundell, Bond, Devereux, and Schiantarelli, 1992). In an investment model setting, exploiting firm effects contributes to estimation precision and allows for model consistency in the presence of unobserved idiosyncrasies that may be simultaneously correlated with investment and q. The baseline model in this literature has the form:

$$I_{it}/K_{it} = \eta_i + \beta q^*_{it} + \alpha CF_{it}/K_{it} + u_{it}, \tag{27}$$

where I denotes investment, K capital stock, q^* is marginal q, CF cash flow, η is the firm-specific effect, and u is the innovation term.

As mentioned earlier, if q^* is measured with error, OLS estimates of β will be biased downwards. In addition, given that q and cash flow are likely to be positively correlated, the coefficient α is likely to be biased upwards in OLS estimations. In expectation, these biases should be reduced by the use of estimators like the ones discussed in the previous section.

4.1 Theoretical Expectations

In order to better evaluate the performance of the two alternative estimators, we develop some hypotheses about the effects of measurement-error correction on the estimated coefficients β and α from equation (27). Theory does not pin down the exact values that these coefficients should take. Nevertheless, one could argue that the two following conditions should be reasonable.

First, an estimator that addresses measurement error in q in a standard investment equation should return a higher estimate for β and a lower estimate for α when compared with standard OLS estimates. Recall, measurement error causes an attenuation bias on the estimate for the coefficient β. In addition, since q and cash flow are likely to be positively correlated, measurement error should cause an upward bias on the empirical estimate returned for α under the standard OLS estimation. Accordingly, if one denotes the OLS and the measurement-error consistent estimates, respectively, by (β^{OLS}, α^{OLS}) and (β^{MEC}, α^{MEC}), one should expect:

Condition 1 *.$\beta^{OLS} < \beta^{MEC}$ and $\alpha^{OLS} > \alpha^{MEC}$.*

Second, one would expect the coefficients for q and the cash flow to be non-negative after treating the data for measurement error. The q-theory of investment predicts a positive correlation between investment and q (e.g., Hayashi, 1982). If the theory holds and the estimator does a good job of adjusting for measurement error, then the cash flow coefficient should be zero ("neoclassical view").

However, the cash flow coefficient could be positive either because of the presence of financing frictions (as posited by Fazzari, Hubbard, and Petersen, 1988),[26] or due to fact that cash flow picks up variation in investment opportunities even after we apply a correction for mismeasurement in q. Accordingly, one should observe:

Condition 2 *.$\beta^{MEC} \geq 0$ and $\alpha^{MEC} \geq 0$.*

Notice that these conditions are fairly weak. If a particular measurement-error consistent estimator does not deliver these basic results, one should have reasons to question the usefulness of that estimator in applied work.

4.2 Data Description

Our data collection process follows that of Almeida and Campello (2007). We consider a sample of manufacturing firms over the 1970 to 2005 period with data available from Compustat. Following those authors, we eliminate firm-years displaying asset or sales growth exceeding 100%, or for which the stock of fixed capital (the denominator of the investment and cash flow variables) is less than \$5 million (in 1976 dollars). Our raw sample consists of 31,278 observations from 3,084 individual firms. Summary statistics for investment, q, and cash flow are presented in Table 6. These statistics are similar to those reported by Almeida and Campello, among other papers. To save space we omit the discussion of these descriptive statistics.

– Table 6 about here –

4.3 Testing for the Presence of Fixed Effects and Heteroscedasticity

Before estimating our investment models, we conduct a series of tests for the presence of firm-fixed effects and heteroscedasticity in our data. As a general rule, these phenomena might arise naturally in panel data applications and should not be ignored. Importantly, whether they appear in the data can have concrete implications for the results generated by different estimators.

We first perform a couple of tests for the presence of firm-fixed effects. We allow for individual firm intercepts in equation (27) and test the null hypothesis that the coefficients associated with those firm effects are jointly equal to zero (Baltagi, 2005). Table 7 shows that the F-statistic for this test is 4.4 (the associated p-value is 0.000). Next, we contrast the random effects OLS and the fixed effects OLS

[26] However, financial constraints are not sufficient to generate a strictly positive cash flow coefficient because the effect of financial constraints is capitalized in stock prices and may thus be captured by variations in q (Chirinko, 1993; and Gomes, 2001).

estimators to test again for the presence of fixed effects. The Hausman test statistic reported in Table 7 rejects the null hypothesis that the random effects model is appropriate with a test statistic of 8.2 (p-value of 0.017). In sum, standard tests strongly reject the hypothesis that fixed effects can be ignored.

We test for homoscedasticity using two different panel data-based methods. First, we compute the residuals from the least squares dummy variables estimator and regress the squared residuals on a function of the independent variables [see Frees (2004) for additional details]. We use two different combinations of independent regressors — (q_{it}, CF_{it}) and $(q_{it}, q_{it}^2, CF_{it}, CF_{it}^2)$ — and both of them robustly reject the null hypothesis of homoscedasticity. We report the results for the first combination in Table 7, which yields a test statistic of 55.2 (p-value of 0.000). Our second approach for testing the null of homoscedasticity is the standard random effects Breusch-Pagan test. Table 7 shows that the Breusch-Pagan test yields a statistic of 7,396.2 (p-value of 0.000). Our tests hence show that the data strongly reject the hypothesis of error homoscedasticity.

– Table 7 about here –

4.4 Implementing the EW Identification Test

Our preliminary tests show that one should control for fixed effects when estimating investment models using real data. In the context of the EW estimator, it is thus appropriate to apply the within transformation before the estimation. However, in this section we also present results for the data in level form to illustrate the point made in Section 3.2 that applying the within transformation compromises identification in the EW context. Prior papers adopting the EW estimator have ignored (or simply dismissed) the importance of fixed effects (e.g., Whited, 2001, 2006).

We present the results for EW's identification test in Table 8. Using the data in level form, we reject the hypothesis of no identification in 12 out of 30 years (or 36% rejection). For data that is transformed to accommodate fixed effects (within transformation), we find that in only 7 out of 33 (or 21%) of the years between 1973 and 2005 one can reject the null hypothesis that the model is not identified at the usual 5% level of significance. These results suggest that the power of the test is low and decreases further after applying the within transformation to the data. These results are consistent with Almeida and Campello's (2007) use of the EW estimator. Working with a 15-year Compustat panel, those authors report that they could only find a maximum of three years of data passing the EW identification test.

– Table 8 about here –

The results in Table 8 reinforce the notion that it is quite difficult to operationalize the EW estimator in real-world applications; particularly in situations in which the within transformation is appropriate due to the presence of fixed effects. We recognize that the EW identification test rejects the model for most of the data at hand. However, recall from Section 3.2 that the test itself is likely to be misleading ("over-rejecting" the data). In the next section, we take the EW estimator to the data (a standard Compustat sample extract) to illustrate the issues applied researchers face when using that estimator, contrasting it to an easy-to-implement alternative.

4.5 Estimation Results

We estimate equation (27) using the EW, OLS-IV, and AB-GMM estimators. For comparison purposes, we also estimate that investment equation using standard OLS and OLS with fixed effects (OLS-FE). The estimates for the standard OLS are likely to be biased, providing a benchmark to evaluate the performance of the other estimators. As discussed in Section 4.1, we expect estimators that improve upon the problem of mismeasurement to deliver results that satisfy Conditions 1 and 2 above.

As is standard in the empirical literature, we use an unbalanced panel in our estimations. Erickson and Whited (2000) propose a minimum distance estimator (MDE) to aggregate the cross-sectional estimates obtained for each sample year, but their proposed MDE is designed for balanced panel data. Following Riddick and Whited (2008), we use a Fama-MacBeth procedure to aggregate the yearly EW estimations.[27]

To implement our OLS-IV estimators, we first take differences of the model in equation (27). We then employ the estimator denoted by *OLS-IV A* from Section 2.2, using lagged levels of q and cash flow as instruments for (differenced) q_{it}. Our Monte Carlos suggest that identification in this context may require the use of longer lags of the model variables. Accordingly, we experiment with specifications that use progressively longer lags of q and cash flow to verify the robustness of our results.

Table 9 reports our findings. The OLS and OLS-FE estimates, reported in columns (1) and (2) respectively, disregard the presence of measurement error in q. The EW-GMM3, EW-GMM4, and EW-GMM5 estimates are reported in columns (3), (4), and (5). For the OLS-IV estimates reported in column (6), we use q_{t-2} as an instrument.[28] The AB-GMM estimator, reported in column (7), uses lags of q as instruments. Given our data structure, this implies using a total of 465 instruments. We account for firm-fixed effects by transforming the data.

– Table 9 about here –

[27] Fama-McBeth estimates are computed as a simple standard errors for yearly estimates. An alternative approach could use the Hall-Horowitz bootstrap. For completeness, we present in the appendix the actual yearly EW estimates.

[28] In the next section, we examine the robustness of the results with respect to variation in the instrument set.

When using OLS and OLS-FE, we obtain the standard result in the literature that both q and cash flow attract positive coefficients [see columns (1) and (2)]. In the OLS-FE specification, for example, we obtain a q coefficient of 0.025, and a cash flow coefficient of 0.121. Columns (3), (4), and (5) show that the EW estimator does not deliver robust inferences about the correlations between investment, cash flow, and q. The q-coefficient estimate varies significantly with the set of moment conditions used, even flipping signs. In addition, *none* of the q coefficients is statistically significant. The cash flow coefficient is highly inflated under EW, and in the case of the EW-GMM4 estimator it is more than three times *larger* than the (supposedly biased) OLS coefficient. These results are inconsistent with Conditions 1 and 2 above. These findings agree with the Monte Carlo simulations of Section 3.3.2, which also point to a very poor performance of the EW estimator in cases in which fixed effects and heteroscedasticity are present.

By comparison, the OLS-IV delivers results that are consistent with Conditions 1 and 2. In particular, the q coefficient increases from 0.025 to 0.063, while the cash flow coefficient drops from 0.131 to 0.043. These results suggest that the proposed OLS-IV estimator does a fairly reasonable job at addressing the measurement error problem. This conclusion is consistent with the Monte Carlo simulations reported above, which show that the OLS-IV procedure is robust to the presence of fixed effects and heteroscedasticity in simulated data. The AB-GMM results also generally satisfy Conditions 1 and 2. Notice, however, that the observed changes in the q and *cash flow* coefficients ("corrections" relative to the simple, biased OLS estimator) are less significant than those obtained under the OLS-IV estimation.

4.6 Robustness of the Empirical OLS-IV Estimator

It is worth demonstrating that the OLS-IV we consider is robust to variations in the set of instruments that is used for identification. While the OLS-IV delivered results that are consistent with our priors, note that we examined a just-identified model, for which tests of instrument quality are not available. As we have discussed previously, OLS-IV estimators should be used with care in this setting, since the underlying structure of the error in the latent variable is unknown. In particular, the Monte Carlo simulations suggest that it is important to show that the results remain when we use longer lags to identify the model.

We present the results from our robustness checks in Table 10. We start by adding one more lag of q (i.e., q_{t-3}) to the instrumental set. The associated estimates are in the first column of Table 10. One can observe that the slope coefficient associated with q increases even more with the new instrument (up to 0.090), while that of the cash flow variable declines further (down to 0.038). One problem with this estimation, however, is the associated J-statistic. If we consider a 5% hurdle rule, the J-statistic of 4.92 implies that, with this particular instrumental set, we reject the null hypothesis that the identification restrictions are met (p-value of 3%). As we have discussed, this could be expected if, for example, the measurement error process has a MA structure. This suggests that the researcher should look for longer lagging schemes, lags that "erase" the MA memory of the error structure.

– Table 10 about here –

Our next set of estimations use longer lagging structures for our proposed instruments and even a instrumental set with only lags of cash flow, the exogenous regressors in the model. We use combinations of longer lags of q (such as the fourth and fifth lags) and longer lags of cash flow (fourth and fifth lags). This set of tests yield estimates that more clearly meet standard tests for instrument validity.[29] Specifically, the J-statistics now indicate we do not reject the hypothesis that the exclusion restrictions are met. The results reported in columns (2) through (7) of Table 10 also remain consistent with Conditions 1 and 2. In particular, the q coefficient varies from approximately 0.040 to 0.091, while the cash flow coefficient varies roughly from 0.044 to 0.046. These results are consistent with our simulations, which suggest that these longer lag structures should deliver relatively consistent, stable estimates of the coefficients for q and cash flow in standard investment regressions.

5 Concluding Remarks

OLS estimators have been used as a reference in empirical work in financial economics. Despite their popularity, those estimators perform poorly when dealing with the problem of errors-in-variables. This is a serious problem since in most empirical applications one might raise concerns about issues such as data quality and measurement errors.

This paper uses Monte Carlo simulations and real data to assess the performance of different estimators that deal with measurement error, including EW's higher-order moment estimator and alternative instrumental variable-type approaches. We show that in the presence of individual-fixed effects, under heteroscedasticity, or in the absence of high degree of skewness in the data, the EW estimator returns biased coefficients for both mismeasured and perfectly-measured regressors. The IV estimator requires assumptions about the autocorrelation structure of the measurement error, which we characterize and discuss in the paper.

[29] All of the F-statistics associated with the first-stage regressions have p-values that are close to zero. These statistics (reported in Table 10) suggest that we do not incur a weak instrument problem when we use longer lags in our instrumental set.

We also estimate empirical investment models using the two methods. Because real-world investment data contain firm-fixed effects and heteroscedasticity, the EW estimator delivers coefficients that are unstable across different specifications and not economically meaningful. In contrast, a simple OLS-IV estimator yields results that conform to theoretical expectations. We conclude that real world investment data is likely to satisfy the assumptions that are required for identification of OLS-IV, but that the presence of heteroskedasticy and fixed effects causes the EW estimator to return biased coefficients.

References

[1] Agca, S., and A. Mozumdar. 2007. Investment-Cash Flow Sensitivity: Myth or Reality? *mimeo*, George Washington University.

[2] Almeida, H., and M. Campello. 2007. Financial Constraints, Asset Tangibility and Corporate Investment. *Review of Financial Studies* 20, 1429-1460.

[3] Altonji, J., and L. Segal. 1996. Small-Sample Bias in GMM Estimation of Covariance Structures. *Journal of Business & Economic Statistics* 14, 353-366.

[4] 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, 277-297.

[5] Bakke, T., and T. Whited. 2009. What Gives? A Study of Firms' Reactions to Cash Shortfalls. Mimeo, University of Rochester.

[6] Baltagi, B. 2005. *Econometric Analysis of Panel Data*, John Wiley and Sons Press, 3rd Edition.

[7] Baum, C., M. Schaffer, and S. Stillman. 2003. Instrumental Variables and GMM: Estimation and Testing. *Stata Journal* 3, 1-31.

[8] Bertrand, M., E. Duflo, and S. Mullainathan. 2004. How Much Should We Trust Differences-in-Differences Estimates? *Quarterly Journal of Economics* 119, 249-275.

[9] Bertrand, M., and S. Mullainathan. 2005. Bidding for Oil and Gas Leases in the Gulf of Mexico: A Test of the Free Cash Flow Model? *mimeo*, University of Chicago and MIT.

[10] Bertrand, M., and A. Schoar. 2003. Managing with Style: The Effect of Managers on Firm Policies. *Quarterly Journal of Economics* 118, 1169-1208.

[11] Biorn, E. 2000. Panel Data with Measurement Errors: Instrumental Variables and GMM Procedures Combining Levels and Differences. *Econometric Reviews* 19, 391-424.

[12] Blanchard, O., F. Lopez-de-Silanes, and A. Shleifer. 1994. What Do Firms Do with Cash Windfalls? *Journal of Financial Economics* 36, 337-360.

[13] Blundell, R., S. Bond, M. Devereux, and F. Schiantarelli. 1992. Investment and Tobin's Q: Evidence from Company Panel Data. *Journal of Econometrics* 51, 233-257.

[14] Bond, S., and J. Cummins. 2000. The Stock Market and Investment in the New Economy. Brookings Papers on Economic Activity 13, 61-124.

[15] Chirinko, R. 1993. Business Fixed Investment Spending: Modeling Strategies, Empirical Results, and Policy Implications. *Journal of Economic Literature* 31, 1875-1911.

[16] Colak, G., and T. Whited. 2007. Spin-offs, Divestitures, and Conglomerate Investment. *Review of Financial Studies* 20, 557-595.

[17] Cragg, J. 1997. Using Higher Moments to Estimate the Simple Errors-In-Variables Model. *RAND Journal of Economics* 28, 71-91.

[18] Doran, H., and P. Schmidt. 2006. GMM Estimators with Improved Finite Sample Properties Using Principal Components of the Weighting Matrix, with an Application to the Dynamic Panel Data Model. *Journal of Econometrics*133, 387-409.

[19] Erickson, T., and T. Whited. 2000. Measurement Error and the Relationship between Investment and *Q*. *Journal of Political Economy* 108, 1027-1057.

[20] Erickson, T., and T. Whited. 2002. Two-Step GMM Estimation of the Errors-In-Variables Model Using High-Order Moments. *Econometric Theory* 18, 776-799.

[21] Fazzari S., R. G. Hubbard, and B. Petersen. 1988. Financing Constraints and Corporate Investment. *Brooking Papers on Economic Activity* 1, 141-195.

[22] Frees, E. 2004. *Longitudinal and Panel Data Analysis and Applications in the Social Sciences*, Cambridge University Press, New York, NY.

[23] Gan, J. 2007. Financial Constraints and Corporate Investment: Evidence from an Exogenous Shock to Collateral. *Journal of Financial Economics* 85, 709-734.

[24] Gomes, J. 2001. Financing Investment. *American Economic Review* 91, 1263-85.

[25] Griliches, Z., and J. A. Hausman. 1986. Errors in Variables in Panel Data. *Journal of Econometrics* 31, 93-118.

[26] Hadlock, C. 1998. Ownership, Liquidity, and Investment. *RAND Journal of Economics* 29, 487-508.

[27] Hayashi, F. 1982. Tobin's Marginal q and Average q: A Neoclassical Interpretation. *Econometrica* 50, 213-24.

[28] Hayashi, F. 2000. *Econometrics.* Princeton, NJ: Princeton University Press.

[29] Hennessy, C. 2004. Tobin's Q, Debt Overhang, and Investment. *Journal of Finance* 59, 1717-1742.

[30] Holtz-Eakin, D., W. Newey, and H. S. Rosen. 1988. Estimating Vector Autoregressions with Panel Data. *Econometrica* 56, 1371-1395.

[31] Hoshi, T., A. Kashyap, and D. Scharfstein. 1991. Corporate Structure, Liquidity, and Investment: Evidence from Japanese Industrial Groups. *Quarterly Journal of Economics* 106, 33-60.

[32] Hubbard, R. G. 1998. Capital Market Imperfections and Investment. *Journal of Economic Literature* 36, 193-227.

[33] Kaplan, S., and L. Zingales. 1997. Do Financing Constraints Explain Why Investment Is Correlated with Cash Flow? *Quarterly Journal of Economics* 112, 169-215.

[34] Lamont, O. 1997. Cash Flow and Investment: Evidence from Internal Capital Markets. *Journal of Finance* 52, 83-110.

[35] Lyandres, E. 2007. External Financing Costs, Investment Timing, and Investment-Cash Flow Sensitivity. *Journal of Corporate Finance* 13, 959-980.

[36] Malmendier, U., and G. Tate. 2005. CEO Overconfidence and Corporate Investment. *Journal of Finance* 60, 2661-2700

[37] Petersen, M. 2009. Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches. *Review of Financial Studies* 22, 435-480.

[38] Poterba, J. 1988. Financing Constraints and Corporate Investment: Comment. *Brookings Papers on Economic Activity* 1, 200-204.

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

[40] Riddick, L., and T. Whited. 2008. The Corporate Propensity to Save. Forthcoming, *Journal of Finance.*

[41] Shin, H., and R. Stulz. 1998. Are Internal Capital Markets Efficient? *Quarterly Journal of Economics* 113, 531-552.

[42] Stein, J. 2003. Agency Information and Corporate Investment, in G. Constantinides, M. Harris, and R. Stulz (eds.), *Handbook of the Economics of Finance*, Elsevier/North-Holland, Amsterdam.

[43] Xiao, Z., J. Shao, and M. Palta. 2008. A Unified Theory for GMM Estimation in Panel Data Models with Measurement Error. Mimeo, University of Wisconsin Madison.

[44] Wansbeek, T. J. 2001. GMM Estimation in Panel Data Models with Measurement Error. *Journal of Econometrics* 104, 259-268.

[45] Whited, T. 2001. Is It Inefficient Investment that Causes the Diversification Discount? *Journal of Finance* 56, 1667-1691.

[46] Whited, T. 2006. External Finance Constraints and the Intertemporal Pattern of Intermittent Investment. *Journal of Financial Economics* 81, 467-502.

Table 1. The performance of the EW identification test

Distribution	Null is	Data Form	Frequency of Rejection
Normal	True	Level	0.05
		Within	0.05
Lognormal	False	Level	0.47
		Within	0.43
χ_3^2	False	Level	0.14
		Within	0.28
$F_{10,40}$	False	Level	0.17
		Within	0.28

This table shows the performance of the EW identification test for different distributional assumptions, displayed in column 1. The tests are computed for the data in levels, and after applying a within transformation. Column 4 shows the the frequencies at which the null hypothesis that the model is not identified is rejected, at the 5% level of significance.

Table 2. The EW estimator: cross-sectional data

		β	α_1	α_2	α_3
Panel A. Lognormal Distribution					
EW-GMM3	Bias	0.0203	-0.0054	-0.0050	-0.0056
	RMSE	0.1746	0.0705	0.0704	0.0706
EW-GMM4	Bias	0.0130	-0.0034	-0.0033	-0.0040
	RMSE	0.2975	0.1056	0.1047	0.1083
EW-GMM5	Bias	0.0048	-0.0013	-0.0013	-0.0019
	RMSE	0.0968	0.0572	0.0571	0.0571
Panel B. Chi-Square Distribution					
EW-GMM3	Bias	-0.0101	0.0092	0.0101	-0.0060
	RMSE	61.9083	16.9275	16.1725	14.7948
EW-GMM4	Bias	-0.3498	0.0938	0.0884	0.0831
	RMSE	12.2386	3.2536	2.9732	3.1077
EW-GMM5	Bias	-0.3469	0.0854	0.0929	0.0767
	RMSE	7.2121	1.8329	1.8720	1.6577
Panel C. F-Distribution					
EW-GMM3	Bias	0.3663	-0.1058	-0.0938	-0.0868
	RMSE	190.9102	53.5677	52.4094	43.3217
EW-GMM4	Bias	-0.2426	0.0580	0.0649	0.0616
	RMSE	90.9125	24.9612	24.6827	21.1106
EW-GMM5	Bias	-0.2476	0.0709	0.0643	0.0632
	RMSE	210.4784	53.5152	55.8090	52.4596
Panel D. Normal Distribution					
EW-GMM3	Bias	1.9179	-0.6397	-0.5073	-0.3512
	RMSE	2305.0309	596.1859	608.2098	542.2125
EW-GMM4	Bias	-1.0743	0.3012	0.2543	0.2640
	RMSE	425.5931	111.8306	116.2705	101.4492
EW-GMM5	Bias	3.1066	-1.0649	-0.9050	-0.5483
	RMSE	239.0734	60.3093	65.5883	58.3686

This table shows the bias and the RMSE associated with the estimation of the model in equations (17) to (21) using the EW estimator in simulated cross-sectional data. β is the coefficient on the mismeasured regressor, and α_1 to α_3 are the coefficients on the perfectly measured regressors. The table shows the results associated with GMM3, GMM4, and GMM5, for all the alternative distributions. These estimators are based on the respective third, fourth, and fifth moment conditions.

Table 3. The EW estimator: panel data in levels

		β	α_1	α_2	α_3
Panel A. Lognormal Distribution					
EW-GMM3	Bias	-1.6450	2.5148	2.5247	2.5172
	RMSE	1.9144	2.5606	2.5711	2.5640
EW-GMM4	Bias	-1.5329	2.5845	2.5920	2.5826
	RMSE	1.9726	2.6353	2.6443	2.6354
EW-GMM5	Bias	-1.3274	2.5468	2.5568	2.5490
	RMSE	1.6139	2.5944	2.6062	2.5994
Panel B. Chi-Square Distribution					
EW-GMM3	Bias	-1.0051	2.2796	2.2753	2.2778
	RMSE	1.1609	2.2887	2.2841	2.2866
EW-GMM4	Bias	-0.9836	2.2754	2.2714	2.2736
	RMSE	1.0540	2.2817	2.2776	2.2797
EW-GMM5	Bias	-0.9560	2.2661	2.2613	2.2653
	RMSE	1.0536	2.2728	2.2679	2.2719
Panel C. F-Distribution					
EW-GMM3	Bias	-0.9926	2.2794	2.2808	2.2777
	RMSE	1.1610	2.2890	2.2904	2.2870
EW-GMM4	Bias	-0.9633	2.2735	2.2768	2.2720
	RMSE	1.0365	2.2801	2.2836	2.2785
EW-GMM5	Bias	-0.9184	2.2670	2.2687	2.2654
	RMSE	2.0598	2.2742	2.2761	2.2725
Panel D. Normal Distribution					
EW-GMM3	Bias	-0.8144	2.2292	2.228	2.2262
	RMSE	0.9779	2.2363	2.2354	2.2332
EW-GMM4	Bias	-0.9078	2.2392	2.2363	2.2351
	RMSE	0.9863	2.2442	2.2413	2.2400
EW-GMM5	Bias	-0.8773	2.2262	2.2225	2.2217
	RMSE	0.9846	2.2316	2.2279	2.2269

This table shows the bias and the RMSE associated with the estimation of the model in equations (17) to (21) using the EW estimator in simulated panel data. The table reports results from data in levels (that is, without applying the within transformation). β is the coefficient on the mismeasured regressor, and $(\alpha_1, \alpha_2, \alpha_3)$ are the coefficients on the perfectly measured regressors. The table shows the results for the EW estimator associated with EW-GMM3, EW-GMM4, and EW-GMM5, for all the alternative distributions. These estimators are based on the respective third, fourth, and fifth moment conditions.

Table 4. OLS, OLS-IV, AB-GMM, and EW estimators: panel data after within transformation

		β	α_1	α_2	α_3
Panel A. Lognormal Distribution					
OLS	Bias	-0.7126	0.1553	0.1558	0.1556
	RMSE	0.7131	0.1565	0.1570	0.1568
OLS-IV	Bias	0.0065	-0.0019	-0.0014	-0.0015
	RMSE	0.1179	0.0358	0.0357	0.0355
AB-GMM	Bias	-0.0248	0.0080	0.0085	0.0081
	RMSE	0.0983	0.0344	0.0344	0.0340
EW-GMM3	Bias	-0.0459	0.0185	0.0184	0.0183
	RMSE	0.0901	0.0336	0.0335	0.0335
EW-GMM4	Bias	-0.0553	0.0182	0.0182	0.0183
	RMSE	0.1405	0.0320	0.0321	0.0319
EW-GMM5	Bias	-0.0749	0.0161	0.0161	0.0161
	RMSE	0.1823	0.0303	0.0297	0.0297
Panel B. Chi-Square Distribution					
OLS	Bias	-0.7126	0.1555	0.1553	0.1556
	RMSE	0.7132	0.1565	0.1563	0.1567
OLS-IV	Bias	0.0064	-0.0011	-0.0017	-0.001
	RMSE	0.1149	0.0348	0.0348	0.0348
AB-GMM	Bias	-0.0231	0.0083	0.0077	0.0081
	RMSE	0.0976	0.0339	0.0338	0.0342
EW-GMM3	Bias	-0.3811	0.0982	0.0987	0.0982
	RMSE	0.4421	0.1133	0.1136	0.1133
EW-GMM4	Bias	-0.3887	0.0788	0.0786	0.0783
	RMSE	0.4834	0.0927	0.0923	0.0919
EW-GMM5	Bias	-0.4126	0.0799	0.0795	0.0798
	RMSE	0.5093	0.0926	0.0921	0.0923

This table shows the bias and the RMSE associated with the estimation of the model in equations (17) to (21) using the OLS, OLS-IV, AB-GMM, and EW estimators in simulated panel data. The table reports results from the estimators on the data after applying the within transformation. β is the coefficient on the mismeasured regressor, and $(\alpha_1, \alpha_2, \alpha_3)$ are the coefficients on the perfectly measured regressors. The table shows the results for the EW estimator associated with EW-GMM3, EW-GMM4, and EW-GMM5, for all the alternative distributions. These estimators are based on the respective third, fourth, and fifth moment conditions.

Table 4. (continued)

		β	α_1	α_2	α_3
Panel C. F-Distribution					
OLS	Bias	-0.7123	0.1554	0.1549	0.1555
	RMSE	0.7127	0.1565	0.1559	0.1566
OLS-IV	Bias	0.0066	-0.0013	-0.0023	-0.001
	RMSE	0.1212	0.0359	0.0362	0.0361
AB-GMM	Bias	-0.0232	0.0079	0.0072	0.0085
	RMSE	0.0984	0.0343	0.0342	0.0344
EW-GMM3	Bias	-0.3537	0.0928	0.0916	0.0917
	RMSE	0.4239	0.1094	0.1086	0.1095
EW-GMM3	Bias	-0.3906	0.0802	0.0790	0.0791
	RMSE	0.4891	0.0939	0.0930	0.0932
EW-GMM3	Bias	-0.4188	0.0818	0.0808	0.0813
	RMSE	0.5098	0.0939	0.0932	0.0935
Panel D. Normal Distribution					
OLS	Bias	-0.7119	0.1553	0.1554	0.1551
	RMSE	0.7122	0.1563	0.1564	0.1562
OLS-IV	Bias	0.0060	-0.0011	-0.0012	-0.0014
	RMSE	0.1181	0.0353	0.0355	0.0358
AB-GMM	Bias	-0.0252	0.0086	0.0085	0.0084
	RMSE	0.0983	0.0344	0.0339	0.0343
EW-GMM3	Bias	-0.7370	0.1903	0.1904	0.1895
	RMSE	0.7798	0.2020	0.2024	0.2017
EW-GMM4	Bias	-0.8638	0.2141	0.2137	0.2137
	RMSE	0.8847	0.2184	0.218	0.2182
EW-GMM5	Bias	-0.8161	0.1959	0.1955	0.1955
	RMSE	0.8506	0.2021	0.2018	0.2017

Table 5. Moving average structures for the measurement error process

Instrument	MA(1)	MA(2)
X_{it-2}	-0.593	-0.368
	(0.60)	(0.38)
X_{it-3}	0.028	-0.707
	(0.30)	(0.71)
X_{it-3}, X_{it-4}	0.025	0.077
	(0.63)	(1.01)
$X_{it-3}, X_{it-4}, X_{it-5}$	-0.011	-0.759
	(0.30)	(0.76)
X_{it-4}	-0.107	-0.144
	(1.62)	(2.01)
X_{it-4}, X_{it-5}	-0.113	-0.140
	(0.58)	(0.59)
X_{it-5}	-0.076	-0.758
	(0.31)	(0.76)

This table shows the bias in the well measured coefficient for OLS-IV using moving average structure for the measurement error process. Numbers in parentheses are the RMSE.

Table 6. Descriptive statistics

Variable	Obs.	Mean	Std. Dev.	Median	Skewness
Investment	22556	0.2004	0.1311	0.17423	2.6871
q	22556	1.4081	0.9331	1.1453	4.5378
Cash flow	22556	0.3179	0.3252	0.27845	-2.2411

This table shows the basic descriptive statistics for q, cash flow, and investment. The data are taken from the annual Compustat industrial files over the 1970 to 2005 period. See text for details.

Table 7. Diagnosis tests

Test	Test statistic	p-value
Pooling test	4.397	0.0000
Random Effects vs. Fixed Effects	8.17	0.0169
Homocedasticity 1	55.19	0.0000
Homocedasticity 2	7396.21	0.0000

This table reports results for specification tests. Hausman test for fixed effects models considers fixed effects models against the simple pooled OLS and the random effects model. A homocedasticity test for the innovations is also reported. The data are taken from the annual Compustat industrial files over the 1970 to 2005 period. See text for details

Table 8. The EW identification test using real data

Level				Within Transformation			
			#Rejections Null				#Rejections Null
1973	t-statistic	1.961	0	1973	t-statistic	1.349	0
	p-value	0.375			p-value	0.509	
1974	t-statistic	5.052	0	1974	t-statistic	7.334	1
	p-value	0.08			p-value	0.026	
1975	t-statistic	1.335	0	1975	t-statistic	1.316	0
	p-value	0.513			p-value	0.518	
1976	t-statistic	7.161	1	1976	t-statistic	5.146	0
	p-value	0.028			p-value	0.076	
1977	t-statistic	1.968	0	1977	t-statistic	1.566	0
	p-value	0.374			p-value	0.457	
1978	t-statistic	9.884	1	1978	t-statistic	2.946	0
	p-value	0.007			p-value	0.229	
1979	t-statistic	9.065	1	1979	t-statistic	1.042	0
	p-value	0.011			p-value	0.594	
1980	t-statistic	9.769	1	1980	t-statistic	7.031	1
	p-value	0.008			p-value	0.03	
1981	t-statistic	10.174	1	1981	t-statistic	7.164	1
	p-value	0.006			p-value	0.028	
1982	t-statistic	3.304	0	1982	t-statistic	2.991	0
	p-value	0.192			p-value	0.224	
1983	t-statistic	5.724	0	1983	t-statistic	9.924	1
	p-value	0.057			p-value	0.007	
1984	t-statistic	15.645	1	1984	t-statistic	6.907	1
	p-value	0			p-value	0.032	
1985	t-statistic	16.084	1	1985	t-statistic	1.089	0
	p-value	0			p-value	0.58	
1986	t-statistic	4.827	0	1986	t-statistic	5.256	0
	p-value	0.089			p-value	0.072	
1987	t-statistic	19.432	1	1987	t-statistic	13.604	1
	p-value	0			p-value	0.001	
1988	t-statistic	5.152	0	1988	t-statistic	1.846	0
	p-value	0.076			p-value	0.397	
1989	t-statistic	0.295	0	1989	t-statistic	0.687	0
	p-value	0.863			p-value	0.709	
1990	t-statistic	0.923	0	1990	t-statistic	1.3	0
	p-value	0.63			p-value	0.522	
1991	t-statistic	3.281	0	1991	t-statistic	3.17	0
	p-value	0.194			p-value	0.205	
1992	t-statistic	2.31	0	1992	t-statistic	2.573	0
	p-value	0.315			p-value	0.276	
1993	t-statistic	1.517	0	1993	t-statistic	1.514	0
	p-value	0.468			p-value	0.469	
1994	t-statistic	2.873	0	1994	t-statistic	4.197	0
	p-value	0.238			p-value	0.123	
1995	t-statistic	0.969	0	1995	t-statistic	1.682	0
	p-value	0.616			p-value	0.431	
1996	t-statistic	17.845	1	1996	t-statistic	4.711	0
	p-value	0			p-value	0.095	
1997	t-statistic	0.14	0	1997	t-statistic	1.535	0
	p-value	0.933			p-value	0.464	
1998	t-statistic	0.623	0	1998	t-statistic	5.426	0
	p-value	0.732			p-value	0.066	
1999	t-statistic	0.354	0	1999	t-statistic	2.148	0
	p-value	0.838			p-value	0.342	
2000	t-statistic	13.44	1	2000	t-statistic	13.502	1
	p-value	0.001			p-value	0.001	
2001	t-statistic	3.159	0	2001	t-statistic	3.309	0
	p-value	0.206			p-value	0.191	
2002	t-statistic	13.616	1	2002	t-statistic	0.693	0
	p-value	0.001			p-value	0.707	
2003	t-statistic	12.904	1	2003	t-statistic	4.006	0
	p-value	0.002			p-value	0.135	
2004	t-statistic	5.212	0	2004	t-statistic	2.801	0
	p-value	0.074			p-value	0.246	
2005	t-statistic	2.365	0	2005	t-statistic	4.127	0
	p-value	0.306			p-value	0.127	
		Sum	12			Sum	7
		% of Years	0.3636			% of Years	0.2121

This table shows the test-statistic and its p-value for the EW identification test, which tests the null hypothesis that the model is not identified. The tests are performed on a yearly basis. In the last columns, we collect the number of years in which the null hypothesis is rejected (sum), and compute the percentage of years in which the null is rejected. The data are taken from the annual Compustat industrial files over the 1970 to 2005 period. See text for details

Table 9. EW, GMM, and OLS-IV coefficients: real world data

Variables	OLS	OLS-FE	EW-GMM3	EW-GMM4	EW-GMM5	OLS-IV	AB-GMM
q	0.0174***	0.0253***	0.0679	-0.3031	0.0230	0.0627***	0.0453***
	(0.002)	(0.003)	(0.045)	(0.302)	(0.079)	(0.007)	(0.006)
Cash flow	0.1310***	0.1210***	0.1299***	0.3841*	0.1554***	0.0434***	0.0460***
	(0.011)	(0.017)	(0.031)	(0.201)	(0.052)	(0.007)	(0.016)
Observations	22556	22556	22556	22556	22556	17348	19748
F-stat p-value (first-step)	–	–	–	–	–	0.000	–

This table shows the coefficients and standard deviations that we obtain when we use the OLS, EW and the GMM estimators in equation (22). The table also displays the standard OLS-FE coefficients (after applying the differencing transformation to treat the fixed effects) in column (2) and OLS-IV in last column. Robust standard errors in parentheses for OLS and GMM, and clustered in firms for OLS-FE and OLS-IV. Each EW coefficient is an average of the yearly coefficients reported in Table A1 and the standard error for these coefficients is a Fama-McBeth standard error. The table shows the EW coefficients for the data after applying the within transformation. The data are taken from the annual Compustat industrial files over the 1970 to 2005 period. See text for details. *, **, and *** represent statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 10. OLS-IV coefficients: robustness tests

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
q	0.0901***	0.0652***	0.0394***	0.0559***	0.0906***	0.0660***	0.0718***
	(0.014)	(0.014)	(0.015)	(0.012)	(0.033)	(0.024)	(0.026)
Cash flow	0.0383***	0.0455***	0.0434***	0.0449***	0.0421***	0.0450***	0.0444***
	(0.007)	(0.011)	(0.011)	(0.010)	(0.008)	(0.012)	(0.012)
Observations	15264	11890	12000	13448	12000	10524	10524
F-stat p-value (first-step)	0.000	0.000	0.000	0.000	0.001	0.000	0.000
J-stat	4.918	3.119	0.122	0.00698	0.497	8.955	5.271
J-stat p-value	0.0266	0.210	0.727	0.933	0.481	0.111	0.261

This table shows the results of varying the set of instruments that are used when applying the OLS-IV estimator to equation (22). In the first column we use the second and third lags of q as instruments for current (differenced) q, as in Table 6. In column (2), we use third, fourth and fifth lags of q as instruments. In column (3), we use the fourth and fifth lags of q and the first lag of cash flow as instruments. In column (4), we use the third lag of q and fourth lag of cash flow as instruments. In column (5), we use the fourth and fifth lags of cash flow as instruments. In column (6), we use $\{q_{t-4}, q_{t-5}, q_{t-6}, CF_{t-3}, CF_{t-4}, CF_{t-5}\}$ as instruments. Finally, in column (7) $\{q_{t-5}, q_{t-6}, CF_{t-3}, CF_{t-4}, CF_{t-5}\}$ as instruments. The estimations correct the errors for heteroscedasticity and firm-clustering. The data are taken from the annual Compustat industrial files over the 1970 to 2005 period. See text for details. *, **, and *** represent statistical significance at the 10%, 5%, and 1% levels, respectively.

Table A.1. EW coefficients for real data (within transformation)

	q coefficient			Cash-Flow coefficient		
Year	GMM3	GMM4	GMM5	GMM3	GMM4	GMM5
1973	-0.029	0.000	0.000	0.347	0.265	0.264
	(0.075)	(0.073)	(4.254)	(0.207)	(0.207)	(11.968)
1974	0.050	0.029	0.019	0.168	0.199	0.214
	(0.037)	(0.012)	(0.016)	(0.073)	(0.043)	(0.043)
1975	0.225	0.001	0.000	0.161	0.292	0.292
	(0.475)	(0.149)	(0.125)	(0.281)	(0.095)	(0.094)
1976	0.137	0.001	0.000	0.156	0.276	0.276
	(0.094)	(0.273)	(0.042)	(0.090)	(0.251)	(0.048)
1977	0.082	0.243	0.000	0.203	0.091	0.261
	(0.263)	(0.109)	(0.108)	(0.179)	(0.090)	(0.083)
1978	0.263	0.514	0.281	0.122	(0.067)	0.108
	(0.282)	(0.927)	(0.146)	(0.224)	(0.689)	(0.125)
1979	0.020	0.001	0.001	0.249	0.266	0.266
	(0.161)	(0.048)	(0.031)	(0.155)	(0.056)	(0.044)
1980	0.349	0.116	0.183	0.021	0.219	0.163
	(0.294)	(0.071)	(0.055)	(0.273)	(0.074)	(0.067)
1981	0.334	0.185	0.324	-0.145	0.061	-0.131
	(0.165)	(0.045)	(0.128)	(0.248)	(0.093)	(0.191)
1982	0.109	0.383	0.238	0.125	-0.206	-0.031
	(0.155)	(0.316)	(0.126)	(0.195)	(0.398)	(0.174)
1983	0.081	0.001	0.001	0.132	0.184	0.184
	(0.037)	(0.041)	(0.059)	(0.033)	(0.034)	(0.040)
1984	0.230	0.210	0.185	0.125	0.138	0.154
	(0.083)	(0.050)	(0.043)	(0.067)	(0.052)	(0.048)
1985	0.198	0.349	0.230	0.050	(0.018)	0.035
	(0.483)	(0.137)	(0.024)	(0.212)	(0.086)	(0.032)
1986	0.672	0.244	0.593	-0.179	0.070	-0.133
	(0.447)	(0.089)	(0.162)	(0.303)	(0.079)	(0.128)
1987	0.102	0.104	0.115	0.078	0.078	0.077
	(0.039)	(0.020)	(0.003)	(0.021)	(0.021)	(0.020)
1988	0.129	0.179	0.148	0.030	0.027	0.029
	(0.051)	(0.029)	(0.014)	(0.011)	(0.007)	(0.007)
1989	-0.365	-0.015	-0.111	0.285	0.162	0.196
	(1.797)	(0.082)	(0.196)	(0.642)	(0.063)	(0.078)
1990	-0.437	-0.419	-0.529	0.395	0.386	0.440
	(0.404)	(0.137)	(0.024)	(0.214)	(0.094)	(0.093)
1991	0.384	0.260	0.240	-0.098	0.007	0.023
	(0.225)	(0.105)	(0.038)	(0.199)	(0.099)	(0.055)
1992	0.105	0.102	0.040	0.086	0.088	0.148
	(0.016)	(0.008)	(0.016)	(0.034)	(0.033)	(0.037)
1993	0.274	0.322	0.452	-0.076	-0.118	-0.232
	(0.394)	(0.352)	(0.273)	(0.360)	(0.297)	(0.276)
1994	-0.110	-4.436	-0.047	0.255	3.550	0.207
	(0.136)	(86.246)	(0.011)	(0.108)	(65.488)	(0.045)
1995	-0.574	-8.847	-2.266	0.537	5.898	1.633
	(1.862)	(145.827)	(5.565)	(1.275)	(94.154)	(3.749)
1996	0.220	0.167	0.196	0.101	0.106	0.103
	(0.068)	(0.022)	(0.013)	(0.036)	(0.033)	(0.030)
1997	0.089	0.177	0.158	0.059	0.020	0.028
	(0.082)	(0.042)	(0.021)	(0.042)	(0.041)	(0.034)
1998	-0.620	-0.245	-0.119	0.688	0.355	0.242
	(1.634)	(0.187)	(0.027)	(1.446)	(0.169)	(0.037)
1999	-0.031	-0.003	0.000	0.160	0.126	0.123
	(0.059)	(0.028)	(0.055)	(0.074)	(0.038)	(0.068)
2000	0.071	0.126	0.118	0.032	-0.029	-0.021
	(0.024)	(0.030)	(0.020)	(0.043)	(0.057)	(0.051)
2001	0.050	0.077	0.055	0.034	0.020	0.031
	(0.021)	(0.020)	(0.013)	(0.016)	(0.016)	(0.012)
2002	0.047	0.048	0.048	0.030	0.030	0.030
	(0.128)	(0.016)	(0.014)	(0.033)	(0.013)	(0.012)
2003	0.131	0.066	0.157	-0.013	0.025	-0.027
	(0.043)	(0.025)	(0.010)	(0.031)	(0.026)	(0.014)
2004	0.005	0.030	0.030	0.092	0.079	0.079
	(0.066)	(0.018)	(0.009)	(0.045)	(0.034)	(0.034)
2005	0.049	0.029	0.026	0.078	0.095	0.098
	(0.025)	(0.009)	(0.011)	(0.040)	(0.032)	(0.034)
Fama-MacBeth	0.0679	-0.3031	0.0232	0.1299	0.3841	0.1554
Standard Error	0.0455	0.3018	0.0787	0.0310	0.2027	0.05222

This table shows the coefficients and standard deviations that we obtain when we use the EW estimator in equation (22), estimated year-by-year. The table also shows the results for the EW estimator associated with GMM3, GMM4 and GMM5. The table also shows the EW coefficients for the data that is treated for fixed effects via the within transformation. The data are taken from the annual Compustat industrial files over the 1970 to 2005 period. See text for details.

Figure 1. Theoretical Distributions Examined

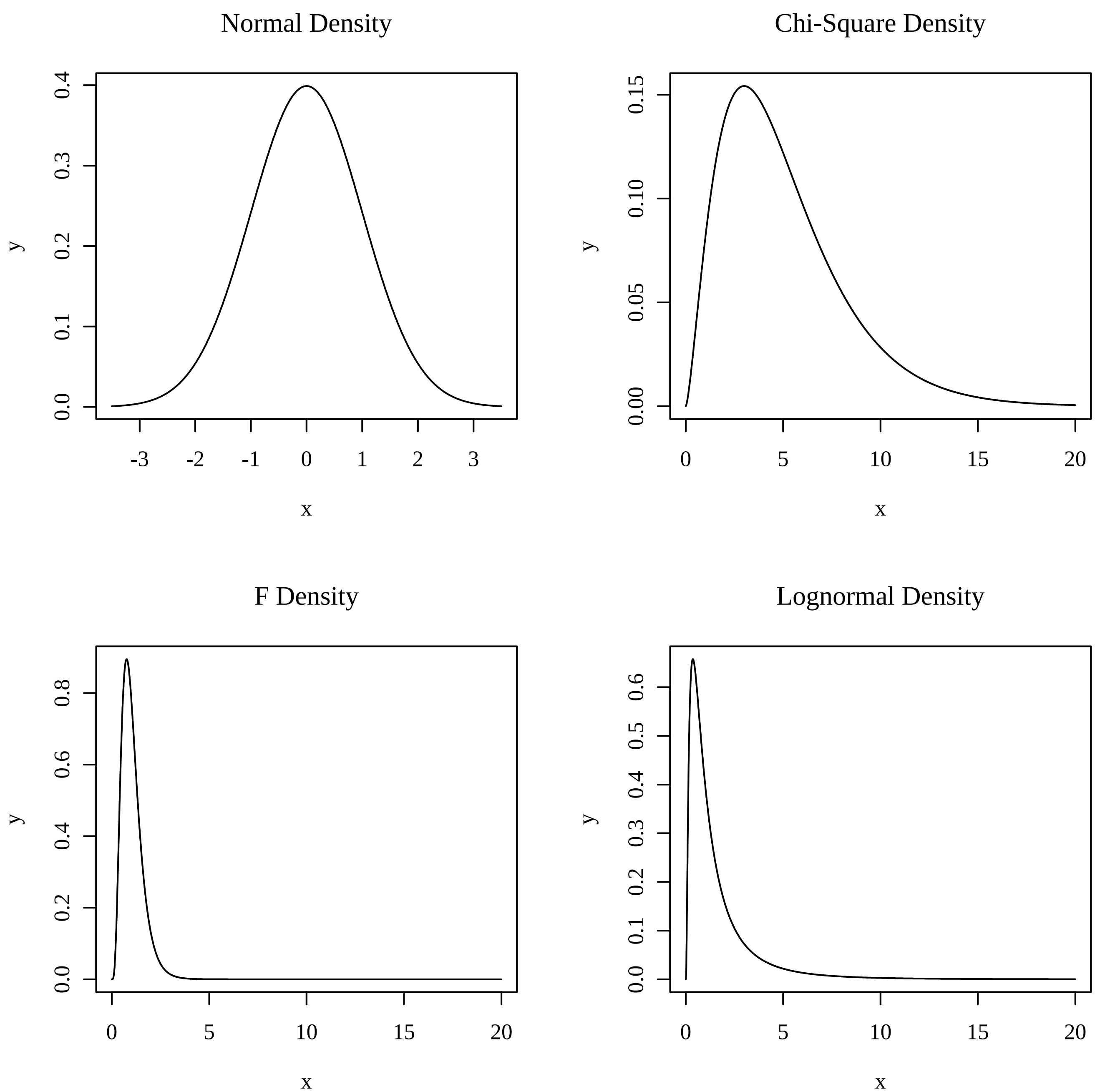
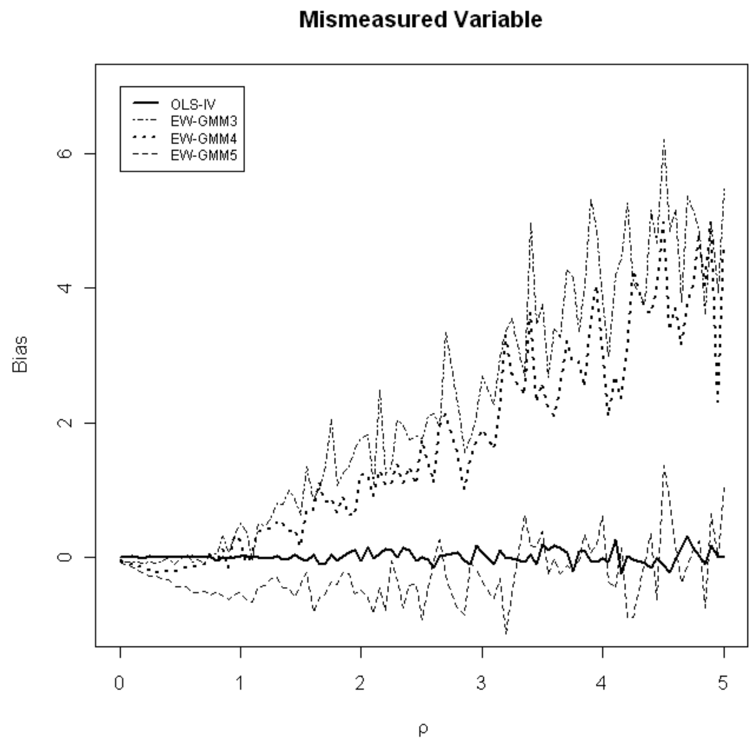


Figure 2. The Effect of Heteroscedasticity on the Bias in EW's GMM and OLS-IV estimators.



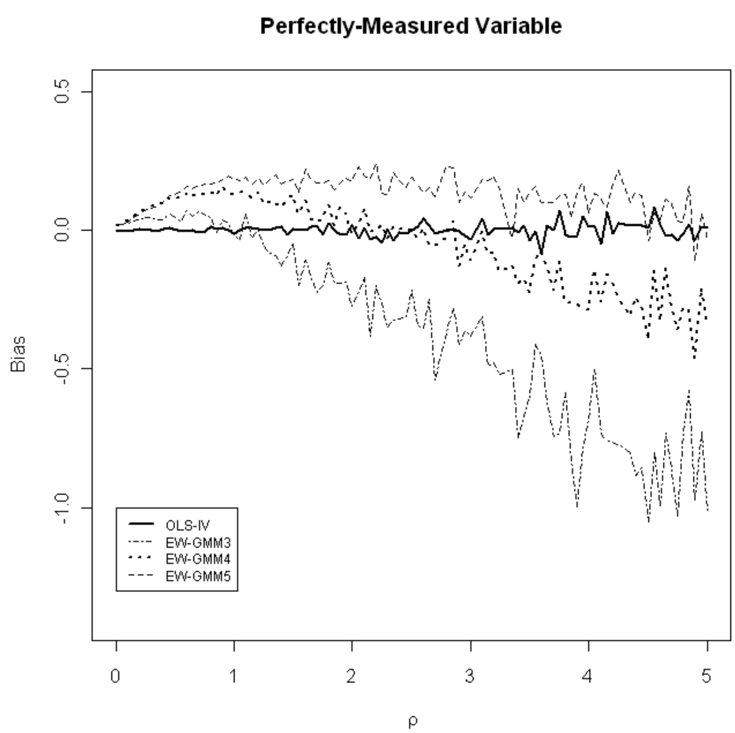


Figure 3. Autocorrelation in the Latent Variable and the OLS-IV Bias

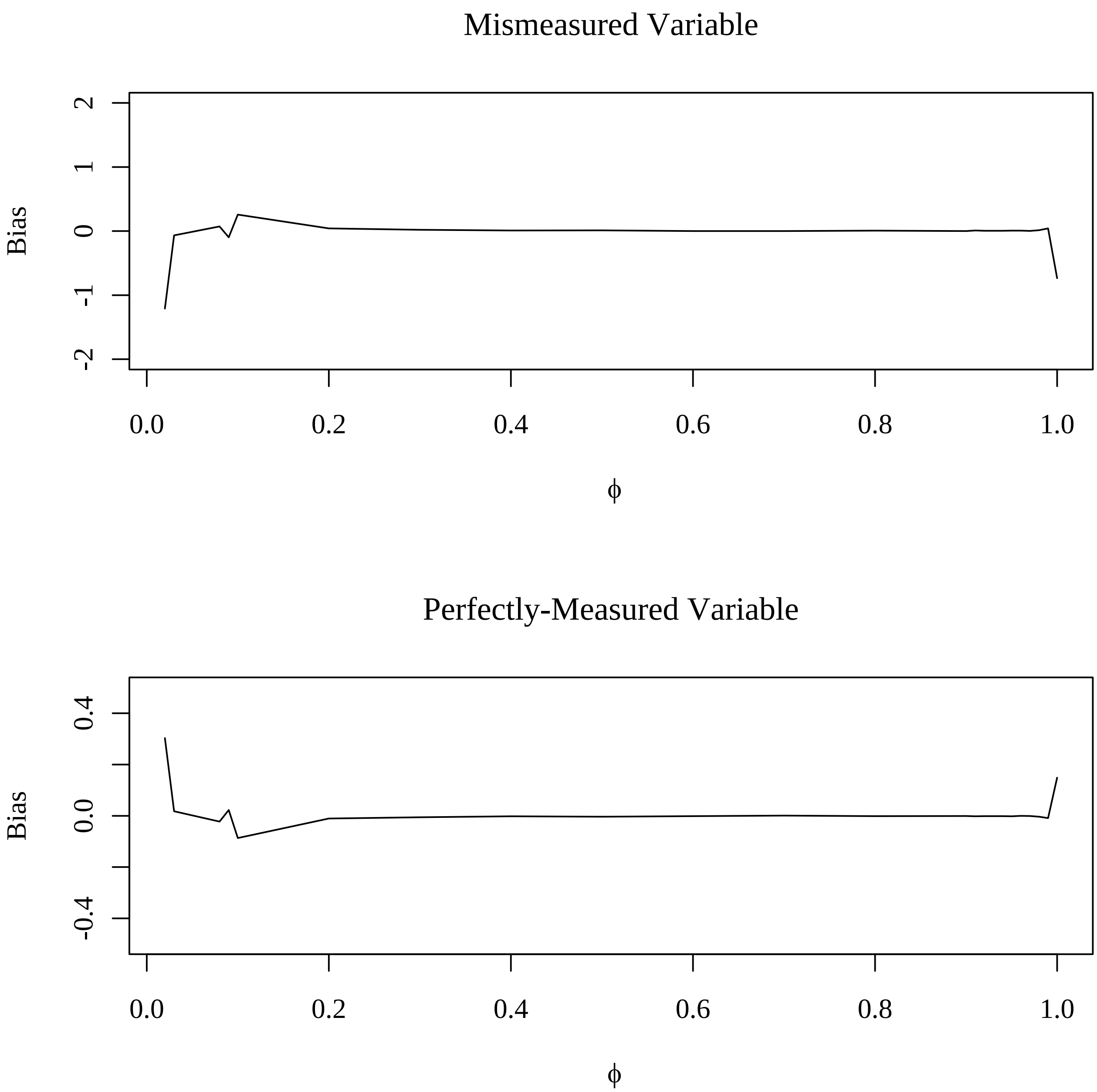


Figure 4. Asymptotic Approximations