Higher Order Moments for Differential Measurement Error, with Application to Tobin’s q and Corporate Investment

Karim Chalak*  
University of Manchester  
Daniel Kim  
University of Waterloo

May 3, 2024

Abstract

We extend the classical measurement error model to allow the proxy for the latent vector to directly affect the response of interest, thereby violating the proxy exclusion restriction. We discuss several economic settings in which this type of differential measurement error occurs. We show that higher order moments can partially identify the model parameters. In the leading case of a scalar latent variable, the identification set consists of two points which we characterize in closed-form. However, signing the effects of the latent variables or distinguishing between certain moments of the latent variables and proxy errors can secure point identification. We propose a plug-in estimator as well as a generalized method of moments estimator. After conducting simulations, we apply our framework to estimate the firm investment equation using Tobin’s q as a proxy for marginal q. We document evidence of the influence of the financial market on the firms’ investment decisions.

JEL codes: C21, C26, G11, G30.
Keywords: differential measurement error, exclusion restriction, feedback, higher order moments, investment, q theory, Tobin’s q.

*Karim Chalak, Department of Economics, University of Manchester, karim.chalak@manchester.ac.uk; Daniel Kim, University of Waterloo, daniel.kim2@uwaterloo.ca.
†Acknowledgments: We thank the participants in the University of Manchester Econometrics Lunch and Brown Bag Economics seminars, and the 93th Annual Meeting of the Southern Economic Association. Any errors are the authors’ responsibility.
1 Introduction

This paper extends the classical measurement error model, in which a vector $W$ serves as a proxy for a key latent vector $U$, to allow $W$ to directly affect the outcome variable $Y$. Specifically, consider the following equations for the outcome $Y$ and the proxy $W$:

$$Y = X'\beta + W'\phi + U'\delta + \eta \quad \text{where} \quad W = \gamma X + U + \varepsilon.$$

The econometrician observes data on the outcome $Y$ and, if present, any covariates $X$. She also observes the error-laden proxy $W$ for $U$ but she does not observe realizations of $U$, the disturbance $\eta$, or the measurement error $\varepsilon$. The “error” $\varepsilon$ is construed broadly as the discrepancy between $W$ and $U$ (after projecting on any covariates $X$) and may very well encode economic quantities of interest. A key point of departure from the standard assumptions in the literature is that the proxy $W$ can directly affect $Y$. That is, $\phi$ need not be zero. As demonstrated in Chalak and Kim (2021), under the classical uncorrelation restrictions, relaxing the proxy exclusion restriction $\phi = 0$ renders the coefficients in the $Y$ equation unidentified (i.e., each coefficient’s identification region spans the real line). We demonstrate that restricting the dependence among the higher order moments of $(U, \varepsilon, \eta)$ partially identifies $\phi$ and $\delta$ as well as moments of $U$, $\varepsilon$, and $\eta$. When $U$ and $W$ are scalars, the identification region consists of a set of two points, which we characterize analytically in closed-form. Moreover, signing $\delta$ or distinguishing between the $U$ and $\varepsilon$ moments secures point identification.

The results of this paper are useful in several economics settings. As discussed further below, Section 6 applies the paper’s econometric method to estimate the firm investment equation (see, e.g., Erickson and Whited, 2000). In the standard specification, investment depends on the latent “marginal q” (the firm’s expected marginal return of capital). “Tobin’s q” (the ratio of the firm’s market value to its assets’ replacement value) is then used as an excluded error-laden proxy for “marginal q.” The analysis in Section 6 relaxes the standard specification to allow Tobin’s q to directly affect investment, reflecting the influence that the financial market exerts on the firms’ management decisions. In addition, as the examples in Section 2.1 demonstrate, the paper’s framework applies in several other contexts. First, in information transmission models (see, e.g., Sobel, 2020), an “informed sender” who observes the state of
nature $U$ may send a noisy message $W$ to an “uninformed receiver” whose payoff may depend on both $U$ and $W$ when talk is not “cheap,” as well as on the players’ actions $X$. Second, in signaling extensions of human capital models (see, e.g., Weiss, 1995), years of education $W$ can measure learning as well as convey a signal about the “type” $U$ of workers, and both $W$ and $U$ can affect wage. Third, in production function estimation, intermediate inputs $W$ (e.g., energy consumption) are often used as proxies for firm productivity $U$ and can directly enter the value-added production function (see, e.g., Levinsohn and Petrin, 2003). Whereas the literature maintains that the intermediate input is a deterministic error-free proxy for productivity, this paper’s setting allows for error-laden proxies. Other examples arise in settings where externalities and reputation effects matter. For instance, production using a polluting technology can be less costly than alternative green technologies, but public information on its use may be detrimental to the firm’s profitability. Similarly, a brand may serve as a proxy for the unobserved quality of a product and command a price premium. For instance, the journal in which a paper is published may serve as an imperfect proxy for the quality of the paper while also having a direct effect on career trajectories (see, e.g., Heckman and Moktan, 2020).

The leading assumption in the econometrics literature is that the measurement error is classical. In its weakest form, this assumes that $(i)$ $\varepsilon$ is uncorrelated with $(U, \eta)$ (and $X$) and $(ii)$ $\phi = 0$ so that the proxy $W$ is excluded from the outcome equation. This model presumes that agents act based on $U$ but that the econometrician observes only a noisy measure $W$ for $U$. It is well known that $\delta$ (and $\beta$) are not point identified in this case. Nevertheless, the Gini-Frisch bounds can partially identify these coefficients (see, e.g., Gini, 1921; Frisch, 1934; Klepper and Leamer, 1984; Bollinger, 2003; Tamer, 2010; and Chalak, 2024). At the other end of the spectrum is the assumption that the error is of the Berkson type (see, e.g., Berkson (1950) and Schennach (2022)). In this case, the roles of $U$ and $W$ are interchanged and $W$ is now assumed to be uncorrelated with $\varepsilon$. This occurs, for example, when agents do not observe $U$ and act instead based on an optimal predictor $W$ for $U$ (see, e.g., Hyslop and Imbens, 2001). It is easy to see that, when $\phi = 0$, a regression of $Y$ on $W$ consistently estimates $\delta$ and $\beta$ provided $(\varepsilon', \eta)$ and $(X', W')'$ are uncorrelated. By considering a setting in which agents act based on their privately observed $U$ as
well as on the commonly observed \( W \), this paper reconciles these polar cases.

When \( \phi = 0 \) and \( W \) is contaminated with classical error, a regression fails to point-identify \( \delta \) (and \( \beta \)). However, assuming that \( \eta, \varepsilon, \) and \( U \) are jointly independent, rather than merely uncorrelated, generally implies restrictions not only on the second moments of \( (W,Y) \) but also on the moments of order higher than two. Whereas a regression exploits the second order moments of the observables, it does not make use of the distribution of \( (W,Y) \) more generally. Reiersøl (1941, 1950) shows that higher order moments, such as the skewness of the distribution, can be used to point-identify the equation coefficients in this case. This identification approach requires that higher order moments carry useful information about the parameters of interest. For instance, it rules out that \( U \) follows a symmetric distribution such as the normal distribution. However, this approach does not require extraneous information such as the availability of standard instrumental variables or double measurements. This insight has been harnessed and extended to identify and estimate linear, parametric, and nonparametric models with measurement error (see e.g. Geary, 1942; Cragg, 1997; Dagenais and Dagenais, 1997; Erickson and Whited, 2002; Schennach and Hu, 2013; Erickson, Jiang, and Whited, 2014).

These identification results assume that \( \phi = 0 \). However, this assumption may not be always suitable, as the examples discussed above demonstrate. When \( \phi \neq 0 \), the measurement error is “differential” since the distribution of \( Y|(W,U,X) \) differs from that of \( Y|(U,X) \), and the standard econometric results are not directly applicable. The literature recognizes “the potential importance of differential measurement error” (Bound, Brown, and Mathiowetz, 2001), and recent work sheds light on its consequences. Indeed, Chalak and Kim (2021) demonstrate that in the classical uncorrelated measurement error model, relaxing the proxy exclusion restriction \( \phi = 0 \) invalidates the classical Gini-Frisch bounds, and the equation coefficients are no longer identified. Chalak and Kim (2021) demonstrate how using auxiliary assumptions, e.g. on the extent of the measurement error, can nevertheless bound the parameters. See also the analysis in Imai and Yamamoto (2010) and Hu and Lewbel (2012) for the case of a binary treatment. Importantly, general identification results using higher order moments in the differential measurement error case have been lacking so far.

This paper fills this gap in the literature by putting forward constructive identi-
fication results in the relatively less explored yet general setting of differential measurement error. It considers a leading setting for differential measurement in which “W is not merely a mismeasured version of [U], but is a separate variable acting as a type of proxy for [U]” (Carroll, Ruppert, Stefanski, and Crainiceanu, 2006, p. 36). We demonstrate that higher order moments partially identify the model parameters, including the effects δ and φ of the latent U and its proxy W on the outcome Y. When U is a scalar, the identification set consists of two points. Moreover, simply signing the effect of U or ordering certain moments of U and ε point identifies the model parameters. In the course of our analysis, we relax certain assumptions that are sometimes maintained in literature on nondifferential measurement error. In particular, we relax the assumption that the measurement errors are jointly independent across proxies (see, e.g., Erickson and Whited (2002) and Erickson, Jiang, and Whited (2014)) to allow for arbitrary dependence. Nor do we impose distributional assumptions such as requiring that the disturbance and measurement errors are normally distributed (see, e.g., Dagenais and Dagenais, 1997) or that the measurement error distribution is known as in the nonparametric separable model deconvolution results (see, e.g., Fan and Truong, 1993). Table 1 situates the contribution of this paper relative to the literature.

<table>
<thead>
<tr>
<th></th>
<th>δ = 0</th>
<th>δ ≠ 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>φ = 0</td>
<td>Point identification using higher order moments</td>
<td>Partial identification using higher order moments; sign restrictions on e.g. δ can secure point identification</td>
</tr>
<tr>
<td>φ ≠ 0</td>
<td>Point identification using moments of order 2 (regression)</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Higher Order Moments for Differential Measurement Error

This table summarizes various identification results using higher order moments.

The results in the paper illustrate an instance in which a parametric model identifies the parameters partially (see, e.g., Chernozhukov, Hong, and Tamer, 2007). Akin to “label switching” (see, e.g., Stephens, 2000), the multiplicity of solutions here reflects the fact that interchanging the roles of the latent U and ε can generate observationally equivalent models. In this case, a cursory argument that asserts point
identification merely because the number of moments exceeds that of the unknowns leads to erroneous inference. Instead, a careful identification analysis enables characterizing, in closed-form for the scalar case, the number of possible solutions implied by the model and to select the correct solution based on reasonable economic assumptions, such as restricting the sign of $\delta$ or distinguishing between certain moments of $U$ and $\varepsilon$. Note that, without the proxy exclusion restriction, traditional instrumental variables methods fail to recover $\phi$ and $\delta$ since a relevant instrument that is correlated with $W$ is likely correlated with the latent term $U$ and therefore invalid. If available, standard instruments may nevertheless help recover the full effect $\phi + \delta$ of $U$ on $Y$. Intersecting this information with the paper’s identification region can also point identify the model. An added benefit of this paper’s identification analysis is that it can lead not only to a generalized method of moment (GMM) estimator but also to a closed-form plug-in estimator which does not rely on numerical optimization algorithms, as the paper’s simulations illustrate.

Section 6 applies the paper’s framework to estimate the firm investment equation. The theory of investment (see, e.g., Abel and Eberly (1994)) postulates that investment is determined by the firm’s expected marginal return of capital, “marginal q.” Tobin (1969) and Tobin and Brainard (1977) suggest that the financial sector plays a more prominent role in the investment decision, and argue that investment is an increasing function of the ratio of the firm’s market value to its assets’ replacement value, “Tobin’s q.” Although marginal q and Tobin’s average q generally differ, Hayashi (1982) gives conditions under which these coincide.

Since researchers do not observe marginal q, it is common to use Tobin’s q as an error-laden proxy for marginal q. The literature considers various methods to account for the measurement error in Tobin’s q and reports mixed empirical conclusions (see e.g., Erickson and Whited (2000, 2012), Almeida, Campello, and Galvao (2010), and Chalak and Kim (2020)). A common assumption underlying these various estimates is that $\phi = 0$. In this case, managers base their investment decisions on their expectations of the marginal return of capital. Tobin and Brainard (1990), argue however that “one might expect financial markets to influence managerial decisions.” Several papers document empirical evidence of such influences (e.g., Chen, Goldstein, and Jiang (2007), Foucault and Fresard (2012), and Bond, Edmans, and Goldstein
(2012)). Whereas this literature allows for a nonzero $\phi$, it does not explicitly account for measurement error.

We apply this paper’s framework to allow the standard firm investment equation to depend on both the manager’s private information, encoded in marginal $q$, as well as on the additional information that the manager extracts from the public financial market proxy, encoded in the error-laden Tobin’s $q$. Our framework therefore considers a setting in which managers who possess private information may find it advantageous to partly conform to the market. Indeed, this proposition dates back at least to Keynes (1936, p. 287) who states:

“although the private investor is seldom himself directly responsible for new investment, nevertheless the entrepreneurs, who are directly responsible, will find it financially advantageous, and often unavoidable, to fall in with the ideas of the market, even though they themselves are better instructed.”

Our results help reconcile the previous contradictory findings in the literature. We demonstrate that, without the proxy exclusion restriction, the investment equation parameters are partially identified in a set of two points. The first corresponds to the case in which Tobin’s $q$ is a noisy proxy for marginal $q$. In this case, we find that managers attach more weight to their private signal $U$ and less weight to the additional information $\varepsilon$ extracted from the financial market. The other point corresponds to the case in which Tobin’s $q$ is a relatively accurate proxy for marginal $q$. In this case, the managers investment decisions attach less weight to their private information and more weight to the information from the financial market. In both cases, investment depends on marginal $q$ and Tobin’s $q$, and our estimates demonstrate that the more precise the signal from the financial market is, the more sensitive the management investment decision to it will be.

The remainder of the paper is organized as follows. Section 2 introduces the assumptions and discusses economics examples. Section 3 studies identification and comments on special cases of interest. Section 4 studies estimation and inference. Section 5 reports simulation results. Section 6 discusses the empirical application. Section 7 concludes. Mathematical proofs are gathered in the Appendix.
2 Assumptions and Examples

We consider the following assumptions.

**Assumption A**

1. **Data Generation:** (i) Let the random vector \((W', Y')'\), with \(J \geq 1\), have sufficiently many finite moments. (ii) Let the random variables \(\eta_{1 \times 1}, U_{J \times 1}, \varepsilon_{J \times 1}, X_{k \times 1}\) with \(k \geq 0\), \(W\), and \(Y\) satisfy
   \[
   Y = X'\beta + W'\phi + U'\delta + \eta \quad \text{and} \quad W = \gamma X + U + \varepsilon
   \]
   with constant slope coefficients. The researcher observes realizations of \((W', X', Y')'\) but not of \((U', \eta, \varepsilon')\).

   A\(_1\) decomposes the proxy \(W\) into the signal \(U\) and the noise \(\varepsilon\) (after projecting on the covariates \(X\)).\(^1\) A key feature of A\(_1\) is that the coefficient \(\phi\) on \(W\) in the equation for the outcome \(Y\) is not assumed to be zero. As such the distribution of \(Y|\(U, W, X\)) generally differs from that of \(Y|\(U, X\)), and the measurement error is therefore “differential.” A\(_1\) allows but does not require the presence of perfectly measured covariates \(X\). If these are absent, we drop \(X\) from the analysis \((k = 0)\) and subsume all the explanatory variables in the latent vector \(U\), with \(W\) serving as its error-laden proxy.

2. **Joint Independence:** \((X, U) \perp (\eta, \varepsilon)\) and \(\eta \perp \varepsilon\).

   A\(_2\) assumes that \((X, U), \eta,\) and \(\varepsilon\) are jointly independent. Naturally, \(U\) and \(X\) may be dependent. The components of \(\varepsilon\) need not be uncorrelated or independent. Similarly, A\(_2\) does not restrict the dependence among the residuals \(\eta\). In particular, \(Var(\varepsilon)\) and \(Var(\eta)\) need not be diagonal. Last, it is possible to relax the assumption \(X \perp (\eta, \varepsilon)\) in A\(_2\) to render the analysis conditional on the covariates \(X\) instead.\(^2\) We forgo this here to simplify the notation. Nevertheless, we emphasize that our identification analysis does not rely on the availability of correctly measured covariates \(X\).

   Indeed, these can be dropped from the analysis, in which case \(U\) subsumes all the

---

\(^1\) A\(_1\) is observationally equivalent to \(Y = X'\beta + W'\phi + V'\gamma + \eta\) and \(W = \gamma X + \psi V + \varepsilon\), with \(V\) unobserved. Since only \(\delta = \psi^{-1} \gamma\) may be (partially) identified, we impose the scale normalization \(U \equiv \psi V\).

\(^2\) See for example Chalak (2019, Section A.1 of the Online Supplemental Material) who considers a linear random coefficients model where the coefficients can depend on the covariates.
explanatory variables proxied by \( W \). Section 3.3.3 comments further on the special case in which the researcher assumes that a subset \( X \) of the explanatory variables is perfectly measured.

### 2.1 Examples

Section 6 applies the results of this paper to estimate the firm investment equation using Tobin’s \( q \) as a proxy for marginal \( q \). It relaxes the standard specification to allow Tobin’s \( q \) to directly affect investment. In this setting, managers base their investment decisions on their expectation of the marginal return on capital and on the additional information they extract from the financial market. As the next examples illustrate, the results of this paper can prove useful in several other settings.

**Information Transmission** Consider an “informed sender” who observes a latent state of nature \( U \). She then sends a message \( W \) to an “uninformed receiver” and takes an action \( X_s \). The receiver observes \( W \) (but not \( U \)) and makes a decision \( X_r \). Our framework accommodates two features of this setting. First, \( W \) need not equal \( U \) since the sender may introduce “noise into his signal” (Crawford and Sobel, 1982). Second, a player’s payoff \( Y(U, W, X) \) may depend in general on the latent variable \( U \), the message \( W \), and the actions \( X \equiv (X_r, X_s) \). See e.g. Frankel and Kartik (2019) and Sobel (2020, Section VII) for several examples, including costly lying and feints.

**Signaling Extensions of Human Capital Models** The literature considers two leading mechanisms that explain the positive association between educational attainment and wage. The human capital model (Becker, 1962; Mincer, 1974) views education as an investment in valuable skills. The signaling model (Spence, 1973) argues that education signals worker “types.” As Weiss (1995, p. 134) explains, both mechanism may operate simultaneously: “an accurate measure of the change in wages for a person who goes to school for 12 years instead of 11 would not measure the effect of that year of education on his productivity, but rather the combined effect of one additional year of learning and the effect of being identified as the type of person who has 12 rather than 11 years of schooling.” Nevertheless, disentangling these effects empirically is challenging (for recent progress in this direction, see, e.g., Chalak and
Kim (2021) and Aryal, Bhuller, and Lange (2022)). This paper’s framework offers a way forward that does not rely on the availability of proper instruments. It uses years of education \( W \) as a proxy for a worker’s latent type \( U \), and allows both variables to directly affect wage,

**Intermediate inputs in production functions** Consider estimating a production function where the output \( Y \) depends on capital \( X_k \), labor \( X_l \), and productivity \( U \). To identify the elasticities of capital and labor, one must account for the possibility that productivity \( U \) is correlated with \( X = (X_k, X_l) \). To resolve this challenge, Olley and Pakes (1996) propose modeling investment as a deterministic function \( I = g(U, X_k) \) that is strictly monotonic in productivity given capital. Since \( U = g^{-1}(I, X_k) \), one can then condition on \( I \) in the \( Y \) equation to help resolve the endogeneity issue. Levinsohn and Petrin (2003) point out that data on investment tends to be “lumpy,” and that this makes the monotonicity assumption less likely to hold for investment. Instead, they propose using an intermediate input \( W \) (e.g. energy cost) as proxy for \( U \). Specifically, they assume that \( W = g(U, X_k) \) and that \( W \) is strictly monotonic in \( U \) given \( X_k \) so that

\[
Y = X'\beta + W\phi + U + \eta \quad \text{where} \quad W = g(U, X_k).
\]

As pointed out in Ackerberg, Caves, and Frazer (2015), these assumptions can entail functional dependencies among the variables that render identification challenging. Our framework relaxes the assumption that \( W \) is a deterministic function of \( U \) and \( X_k \) and allows \( g \) to admit an error term. Assuming a linear specification, we then have\(^3\)

\[
Y = X'\beta + W\phi + U\delta + \eta \quad \text{where} \quad W = U + \gamma_k X_k + \varepsilon.
\]

**Additional examples** The paper’s framework also applies in contexts where reputation and externalities matter. For instance, it may be cheaper for a firm to produce using a polluting technology, but public information on this usage may be detrimental to the firm’s ability to borrow to raise capital. Similarly, a brand may proxy a product’s quality and command a price premium. For instance, the journal where a

\(^3\)Rather than normalizing \( \delta = 1 \), we normalize the coefficient on \( U \) in the \( W \) equation to 1.
paper is published may serve as an imperfect proxy for the quality of the paper and can also have a direct effect on career trajectories (see, e.g., Heckman and Moktan, 2020).

2.2 Projecting on the Covariates

It is convenient to project all the variables on the covariates \( X \) and to express them as deviations from their projected means. Assume \( Var(X) \) is nonsingular and let

\[
\hat{A}' \equiv e_{A,X}' \equiv [A - E(A)]' - [X - E(X)]'Var(X)^{-1}Cov(X, A)
\]

denote the residual from the linear regression of a random vector \( A \) on \( X \). If there are no covariates, set \( \hat{A} \equiv [A - E(A)] \). By construction, \( E(e_{A,X}) = 0 \) and \( Cov(X, e_{A,X}) = 0 \). Given \( A_2 \), we obtain the following system of projected equations

\[
\begin{align*}
\tilde{Y} &= \tilde{W}'\phi + \tilde{U}'\delta + \tilde{\eta} \\
\tilde{W} &= \tilde{U} + \tilde{\varepsilon}
\end{align*}
\]

where \( \tilde{U}, \tilde{\varepsilon}, \) and \( \tilde{\eta} \) have mean zero and are jointly independent, \( \tilde{U} \perp (\tilde{\eta}, \tilde{\varepsilon}) \) and \( \tilde{\eta} \perp \tilde{\varepsilon} \).

We study the identification of \( \phi \) and \( \delta \) using this projected system. This encompasses the case in which \( X \) is absent and all the variables are measured with error. While we focus on the identification of \( \phi, \delta, \) and e.g. moments of \( \tilde{U} \) and \( \tilde{\varepsilon} = \varepsilon - E(\varepsilon) \), we note that

\[
\beta - \gamma'\delta = Var(X)^{-1}Cov(X, Y) - Var(X)^{-1}Cov(X, W)(\phi + \delta).
\]

Therefore, if \( \gamma = 0 \) (so that \( W = U + \varepsilon \)) then the identification region for \( \beta \) can be inferred from that of \( \phi + \delta \).

Except when otherwise noted, we leave the covariates \( X \) implicit in what follows and we drop the tilde accents to simplify the notation.

3 Identification

3.1 A Scalar Latent Variable: Closed-Form Solution

We begin the analysis by focusing on the leading case in which \( J = 1 \), with \( U \) and \( W \) scalars. For non-negative integers \( l_1, \ldots, l_d \), and distinct random variables \( V_1, \ldots, V_d \),
we denote the central moment of order \( \sum_{k=1}^{d} l_h \) by:

\[
\mu_{V_1, \ldots, V_d}^{l_1, \ldots, l_d} \equiv E[(V_1 - E(V_1))^{l_1} \cdots (V_d - E(V_d))^{l_d}].
\]

Leaving \( X \) implicit, together, \( A_1 \) and \( A_2 \) allow expressing the moments of \((W, Y)\) as a function of a vector \( \theta^* \) of unknown parameters. For example, using the moments of order at most 4, we can generate a system \( S(\theta^*) = 0 \) composed of the following 9 moment equations

\begin{align*}
\mu_{W,Y}^{1,1} &= (\phi + \delta) \mu_U^2 + \phi \mu_\varepsilon^2 \\
\mu_{W}^{2} &= \mu_U^2 + \mu_\varepsilon^2 \\
\mu_{Y}^{2} &= (\phi + \delta)^2 \mu_U^2 + \phi^2 \mu_\varepsilon^2 + \mu_\eta^2 \\
\mu_{W,Y}^{2,1} &= (\phi + \delta) \mu_U^3 + \phi \mu_\varepsilon^3 \\
\mu_{W,Y}^{1,2} &= (\phi + \delta)^2 \mu_U^3 + \phi^2 \mu_\varepsilon^3 \\
\mu_{W}^{3} &= \mu_U^3 + \mu_\varepsilon^3 \\
\mu_{W,Y}^{3,1} &= (\phi + \delta) \mu_U^4 + \phi \mu_\varepsilon^4 + 3(\phi + \delta) \mu_U^2 \mu_\varepsilon^2 + 3 \phi \mu_U \mu_\varepsilon^2 \\
\mu_{W,Y}^{2,2} &= (\phi + \delta)^2 \mu_U^4 + \phi^2 \mu_U^2 \mu_\varepsilon^2 + \mu_U^2 \mu_\eta^2 + 4(\phi + \delta) \phi \mu_U^2 \mu_\varepsilon^2 + (\phi + \delta)^2 \mu_U^2 \mu_\varepsilon^2 + \phi^2 \mu_\varepsilon^4 + \mu_\varepsilon^2 \mu_\eta^2, \text{ and} \\
\mu_{W,Y}^{1,3} &= (\phi + \delta)^3 \mu_U^4 + 3(\phi + \delta) \mu_U^2 \mu_\varepsilon^2 + 3(\phi + \delta) \phi^2 \mu_U^2 \mu_\varepsilon^2 + \phi^3 \mu_\varepsilon^4 + 3(\phi + \delta)^2 \phi \mu_U \mu_\varepsilon^2 + 3 \phi \mu_\varepsilon^4 \mu_\eta^2,
\end{align*}

with the 9 unknowns\(^4\) \( \theta^* \) appearing on the right hand side:

\[
\theta^* \equiv (\phi, \delta, \mu_U^2, \mu_\varepsilon^2, \mu_\eta^2, \mu_U^3, \mu_\varepsilon^3, \mu_\varepsilon^4, \mu_\eta^4).
\]

Note that the moment \( \mu_Y^3 = (\phi + \delta)^3 \mu_U^3 + \phi^3 \mu_\varepsilon^3 + \mu_\eta^3 \) involves the unknown \( \mu_\eta^3 \) which does not appear in the above equations. If \( \mu_\eta^3 \) is of direct interest, adding \( \mu_Y^3 \) results in a larger system of 10 equations and 10 unknowns. A similar comment applies to \( \mu_Y^4 \). Conversely, the moment \( \mu_W^4 = \mu_U^4 + 6 \mu_U^2 \mu_\varepsilon^2 + \mu_\varepsilon^4 \) supplies an additional equation involving the components of \( \theta^* \). Thus, we sometimes augment system (3) with the \( \mu_W^4 \) moment to conduct overidentification tests.

When \( \phi = 0 \), the measurement error is nondifferential and the parameters are generally point identified (see e.g. Reiersøl, 1950; Erickson and Whited, 2002; Erickson, Jiang, and Whited, 2014). In what follows, we demonstrate that when \( \phi = 0 \) system

\(^4\)Recall that by \( A_1 \), the means of \( \varepsilon \) and \( \eta \) are zero.
$S(\theta^*) = 0$ admits two roots despite the presence of differential measurement error. Moreover, we show that simply imposing a sign restriction on $\delta$ (the coefficient on $U$ in the $Y$ equation) or on $\mu_r^U - \mu_r^\epsilon$ for $r \in \{2, 3, 4\}$ (for instance, determining whether the signal to total variance $\frac{\mu_2^U}{\mu_2^W}$ exceeds $\frac{1}{2}$) point identifies the system parameters $\theta^*$.

In the special case when $\phi = 0$, it is well known that there are instances in which the parameters are not identified. In particular, this occurs when $\delta = 0$ or when $U$ has a symmetric distribution (see, e.g., Erickson and Whited, 2002). Such instances can also occur in the case of differential measurement error considered here. For example, suppose that the distributions of $U$ and $\epsilon$ are symmetric, $\mu_3^U = 0$ and $\mu_3^\epsilon = 0$. Then the moments of order three ($\mu_3^{U,Y}$, $\mu_3^{W,Y}$, and $\mu_3^W$) are uninformative about the other components of $\theta^*$ and system $S(\theta^*) = 0$ is underdetermined. In this case, the system of equations can have infinitely many roots and $\theta^*$ is not point-identified. We comment on these special cases in what follows.

To derive the result, Theorem 3.1 makes use of the moment equations (3) to express all the components of $\theta^*$ as a function $M(\cdot)$ of $\phi$ (where the notation for $M(\cdot)$ leaves implicit the estimable moments of $(Y,W)$). Specifically, for $j \in \{U,\epsilon,\eta\}$ and $l \in \{U,\epsilon\}$, this expresses $\delta$, the variances $\mu_2^j$, the third moments $\mu_3^l$, and the fourth moments $\mu_4^l$ as mappings $D(\cdot)$, $V_j(\cdot)$, $S_l(\cdot)$, and $K_l(\cdot)$ of $\phi$ respectively. We then use these mappings to reduce the problem of solving the system $S(\theta^*) = 0$ to that of solving a nonlinear equation in $\phi$. We label this nonlinear equation by

$$N(\phi) = 0$$

and relegate an explicit representation for it, along with the rather lengthy mathematical derivations, to the Appendix.

**Theorem 3.1** Assume $A_1$ and $A_2$ (with $X$ implicit) and let $\phi$, $\delta$, $\phi + \delta$, and $\mu_3^U$ be nonzero. Then $\theta^*$ can be expressed as a function $M(\cdot)$ of $\phi$, with the other components
of $M(\cdot)$ defined recursively by:

$$\phi = F(\phi) \equiv \phi$$

$$\mu_\varepsilon^3 = S_\varepsilon(\phi) \equiv \frac{\mu_{W,Y}^{3,1} - (\mu_{W,Y}^{1,2})^2}{\phi(\mu_{W,Y}^1 - \mu_{W,Y}^{1,2}) - (\phi\mu_{W,Y}^1 - \mu_{W,Y}^{1,2})}$$

$$\delta = D(\phi) \equiv T(\phi) - \phi \equiv \frac{\mu_{W,Y}^{1,2} - \phi^2 S_\varepsilon(\phi)}{\mu_{W,Y}^{1,2} - \phi S_\varepsilon(\phi) - \phi}$$

$$\mu_\varepsilon^3 = S_U(\phi) \equiv \frac{\mu_{W,Y}^{2,1} - \phi S_\varepsilon(\phi)}{T(\phi)}$$

$$\mu_\varepsilon^2 = V_\varepsilon(\phi) \equiv \frac{\mu_{W,Y}^{1,1} - T(\phi)\mu_{W}^2}{-D(\phi)}$$

$$\mu_\varepsilon^2 = V_U(\phi) \equiv \mu_{W}^2 - V_\varepsilon(\phi)$$

$$\mu_\eta^2 = V_\eta(\phi) \equiv \mu_{W}^2 - T(\phi)^2 V_U(\phi) - \phi^2 V_\varepsilon(\phi)$$

$$\mu_\varepsilon^4 = K_\varepsilon(\phi) \equiv \frac{\mu_{W,Y}^{2,2} - T(\phi)\mu_{W,Y}^{3,1} + D(\phi)\phi + 2T(\phi)\mu_{W,Y}^{1,1} + 2T(\phi)\mu_{W,Y}^{1,1} + 2T(\phi)\mu_{W,Y}^{1,1} + 2T(\phi)\mu_{W,Y}^{1,1}}{-\phi D(\phi)}$$

$$\mu_U^4 = K_U(\phi) \equiv \frac{\mu_{W,Y}^{3,1} - \phi K_\varepsilon(\phi) - 3T(\phi)\mu_{W,Y}^{1,1} + 3T(\phi)\mu_{W,Y}^{1,1} + 3T(\phi)\mu_{W,Y}^{1,1} + 3T(\phi)\mu_{W,Y}^{1,1}}{T(\phi)}$$

Lemma A.1 in the Appendix gives reduced-form expressions for certain components of $M(\phi)$. While these more primitive expressions can be informative, we express the components of $M(\phi)$ in Theorem 3.1 in their recursive form as this reveals more clearly the substitution order in which system $S(\theta^*) = 0$ is solved.

Next, Theorem 3.2 shows that equation $N(\phi) = 0$ admits two roots $f^-$ and $f^+$. To see this, note that $S(\theta^*) = 0$ is based on the data generating equations

$$Y = U(\delta + \phi) + \varepsilon\phi + \eta \quad \text{and} \quad W = U + \varepsilon.$$  

However, swapping $\varepsilon$ and $U$ while simultaneously exchanging their respective coefficients $\phi$ and $\phi + \delta$ leads to an observationally equivalent data generating system. The fact that $N(\phi) = 0$ admits two roots reflects this fact. Nevertheless, this observation also suggests a way to distinguish between these two roots. Specifically, since one cannot distinguish between $(\phi + \delta)$ and $\phi$, the roots $f^-$ and $f^+$ imply two possible solutions for $\delta = (\phi + \delta) - \phi$ that are of equal magnitude but of opposite signs, $D(f^-) = -D(f^+)$. Correspondingly, the roots $f^-$ and $f^+$ also interchange the
possible solutions for the moments of $U$ and $\varepsilon$. For example, $V_U(f^-) = V_\varepsilon(f^+)$ and $V_U(f^+) = V_\varepsilon(f^-)$. It follows, that correctly restricting the sign of $\delta$ or $\mu_\varepsilon^r - \mu_U^r$ for $r = 2, 3, 4$ identifies the unique root $f^* \in \{f^-, f^+\}$ for $\phi$ and, therefore, point identifies $\theta^*$ via the mapping $M(f^*)$.

**Theorem 3.2** Assume $A_1$ and $A_2$ (with $X$ implicit) and let $\phi$, $\delta$, $\phi + \delta$, and $\mu_U^3$ be nonzero. Let

$$
A \equiv \mu_{W,Y}^{2,1}(3\mu_{W,Y}^{1,1}\mu_W^2 - \mu_{W,Y}^{3,1}) + \mu_W^3(\mu_{W,Y}^{2,2} - 2\mu_{W,Y}^{1,1}\mu_{W,Y} - \mu_W^2\mu_Y^2),
$$

$$
B \equiv \mu_W^3(-\mu_{W,Y}^{1,3} + 3\mu_{W,Y}^{1,1}\mu_Y^2) + \mu_{W,Y}^{1,2}(\mu_{W,Y}^{3,1} - 3\mu_W^2\mu_{W,Y}),
$$

and

$$
C \equiv \mu_{W,Y}^{2,1}(\mu_{W,Y}^{1,3} - 3\mu_{W,Y}^{1,1}\mu_Y^2) - \mu_{W,Y}^{1,2}(\mu_{W,Y}^{2,2} - 2\mu_{W,Y}^{1,1}\mu_{W,Y} - \mu_W^2\mu_Y^2),
$$

and

$$
\Delta \equiv B^2 - 4AC.
$$

(i) $\Delta \geq 0$ and, provided $A \neq 0$, the system of equations (3) admits two roots for $\theta^*$ given by $M(f^-)$ and $M(f^+)$ where

$$
f^- = \frac{-B - \sqrt{\Delta}}{2A} \quad \text{and} \quad f^+ = \frac{-B + \sqrt{\Delta}}{2A}. \quad (5)
$$

(ii) The components of $M(\cdot)$ corresponding to $\delta$ and to the moments of $U$ and $\varepsilon$ obey

$$
D(f^-) = -D(f^+),
$$

$$
V_U(f^-) = V_\varepsilon(f^+) \quad \text{and} \quad V_U(f^+) = V_\varepsilon(f^-),
$$

$$
S_U(f^-) = S_\varepsilon(f^+) \quad \text{and} \quad S_U(f^+) = S_\varepsilon(f^-),
$$

$$
K_U(f^-) = K_\varepsilon(f^+) \quad \text{and} \quad K_U(f^+) = K_\varepsilon(f^-).
$$

It follows that if the sign of $\delta$, $\mu_\varepsilon^2 - \mu_U^2$, $\mu_\varepsilon^3 - \mu_U^3$, or $\mu_\varepsilon^4 - \mu_U^4$ is known then the correct root $f^* \in \{f^-, f^+\}$ is point identified and $M(f^*)$ point identifies $\theta^*$.

Condition (i) of Theorem 3.2 assumes that $A \neq 0$. From Lemma A.2 in the Appendix, we have that $A = 0$ if and only if

$$
\delta = 0 \quad \text{or} \quad (\phi + \delta)\mu_\varepsilon^3(\mu_U^4 - 3\mu_U^2\mu_U^2) = \phi\mu_U^3(\mu_\varepsilon^4 - 3\mu_\varepsilon^2\mu_\varepsilon^2).
$$

This fails, for example, when $U$ or $\varepsilon$ is symmetric and mesokurtic, as is the case for the normal distribution. This also fails if $\varepsilon$ is degenerate, although in this case a regression of $Y$ on $W$ consistently estimates $\phi + \delta$. 

15
3.2 A Vector of Latent Variables

This section generalizes Section 3.1’s analysis to accommodate a vector of latent variables, with \( J \geq 1 \). As in the scalar case, we express the higher order moments involving \((W,Y)\) as a function of \( \phi, \delta \), and of the higher order moments involving \((U,\varepsilon,\eta)\). By the multinomial theorem,

\[
W^{l_j} = (U_j + \varepsilon_j)^{l_j} = \sum_{d_j=0}^{l_j} \frac{l_j!}{d_j!(l_j-d_j)!} U_j^{d_j} \varepsilon_j^{l_j-d_j} \quad \text{for } j = 1, \ldots, J
\]

and

\[
Y^q = \left[ (\sum_{j=1}^{J} U_j(\phi_j + \delta_j) + \varepsilon_j\phi_j) + \eta \right]^q
\]

where the sum is over all combinations of nonnegative \( c \) \((\text{involving } \phi, \delta)\) as a function of \( W, Y \) \(l\). By the multinomial theorem,

\[
\prod_{s=1}^{J} (\phi_s + \delta_s)^{c_s} U_s^{c_s} \prod_{t=J+1}^{2J} \phi_t^{c_t} \varepsilon_t^{c_t} \eta^{c_2J+1}
\]

where \( q \) is the number of \( \phi, \delta \) unknowns. Setting \( q = 0 \) yields the moment \( \mu_{W_1,\ldots,W_J}^{l_1,\ldots,l_J} \) whereas setting \( l = 0 \) yields the moment \( \mu_{y}^{q} \).

By sequentially stacking moments of the form \( \mu_{W_1,\ldots,W_J,Y}^{l_1,\ldots,l_J,q} \) in increasing order, one can generate a system with at least as many equations than unknowns. Specifically, recall that there are \( \frac{(R+h-1)!}{h!(R-1)!} \) ways to select a combination of \( h \) items from \( R \) possibilities with repetition. Here, \( R = J + 1 \) corresponds to the dimension of \((W',Y)'\) and \( h \equiv l + q \) corresponds to the order of \( \mu_{W_1,\ldots,W_J,Y}^{l_1,\ldots,l_J,q} \). Thus, ignoring the \( (0) \) means when \( h = 1 \), there are \( \sum_{h=2}^{r} \frac{(J+h-1)!}{h!(J-1)!} \) \( (J+h-1)! \) \( \mu_{W_1,\ldots,W_J,Y}^{l_1,\ldots,l_J,q} \) of order at most \( r \). These moments involve the unknowns \( \phi \) and \( \delta \), \( 2 \sum_{s=2}^{r} \frac{(J+h-1)!}{h!(J-1)!} \) \( \mu_{W_1,\ldots,W_J,Y}^{l_1,\ldots,l_J,q} \) of unknowns of the form \( \mu_{U_1,\ldots,U_J}^{l_1,\ldots,l_J} \) or \( \mu_{\varepsilon_1,\varepsilon_J}^{l_1,\ldots,l_J} \), and \( r - 1 \) unknowns of the form \( \mu_{y}^{q} \) for \( 1 < s \leq r \). In total, there
are $\sum_{h=2}^{r} \frac{(J+h)!}{h!J!}$ moment equations involving $2J + 2\sum_{i=2}^{r} \frac{(J+i-1)!}{i!(J-1)!} + r - 1$ unknowns, which we label by the vector $\theta^*$.

Providing a closed-form solution for $\theta^*$ is more challenging when $J > 1$. Nevertheless, examining equation (6) reveals the following insights about the identification region for $\theta^*$.

**The Order Condition Holds** First, note that for a fixed $J$, the ratio of equations to unknowns diverges as $r$ increases. For example, for $J = 2$, setting $r = 4$ yields a system of 31 moment equations involving 31 unknowns, whereas setting $r = 5$ yields a system of 52 equations in 44 unknowns. Similarly, setting $J = 3$ and $r = 5$ yields a system of 121 moment equation in 114 unknowns. Thus, the “order condition” for identification holds. As in the scalar case, certain moments (such as $\mu_Y^3$ or $\mu_Y^4$) may involve ancillary unknowns that do not appear elsewhere. If so, removing these moments reduces the dimensions of the equations and of the unknowns.

**The Identification Set Contains at Least $2^J$ Points** In special cases, the moment equations may carry redundant information (as in the joint normality example discussed above) and may have infinitely many solutions. Otherwise, let $\mu_U$ and $\mu_\varepsilon$ collect respectively the $U$ and $\varepsilon$ moments that appear in equation (6). We have that equation (6) is symmetric in $(\phi, \mu_\varepsilon)$ and $(\phi + \delta, \mu_U)$. In particular, for any given solution, there exist other solutions that obtain by swapping some of the components of $\varepsilon$ and $U$ and interchanging their respective coefficients in $\phi$ and $\phi + \delta$. For instance, if $J = 2$ then one can swap the roles of $\varepsilon_1$ and $U_1$, $\varepsilon_2$ and $U_2$, neither, or both, and interchange the components of $\phi$ and $\phi + \delta$ correspondingly. Since $U$ and $\varepsilon$ are of dimension $J \times 1$, the identification set for $\theta^*$ therefore generally contains at least $2^J$ points.

**Simple Restrictions Can Shrink the Identification Set** Equation (6) also reveals that one can generally discern between these $2^J$ points by imposing restrictions on the signs of $\delta_j$ since this would help distinguish between $U_j$ and $\varepsilon_j$. Similarly, distinguishing between the moments $\mu_{U,j}^r$ and $\mu_{\varepsilon,j}^r$ of order at least 2 ($r \geq 2$) can shrink the identification region.
3.3 Special Cases

We comment briefly on two special cases in which the dimension of the unknowns \( \theta^* \) is reduced. We also discuss the special case of perfectly measured covariates.

3.3.1 Multiple Proxies per Latent Variable

Multiple proxies may be available for one or more latent variables. For example, suppose that the first \( m \) components of \( W \) are proxies for the same latent variable \( V_1 \).

Then, for \( j = 1, \ldots, m \), one can set \( U_j = V_1 \) and restrict the corresponding coefficients on \( U_j \) in the \( Y \) equation to \( \delta_j = \frac{1}{m} \gamma_1 \). This reduces the dimension of the unknowns in \( \delta \) to \( J - m + 1 \). Using the observed moments of order at most \( r \), this also reduces the number of unknowns of the form \( \mu_{l_1, \ldots, l_J}^{U_1, \ldots, U_J} \) to \( \sum_{l=2}^{r} \frac{(J-m+l)!}{l!(J-m)!} \).

3.3.2 Independent Measurement Errors

Our analysis allows the components of the measurement error vector \( \varepsilon \) to be correlated or dependent. The literature often assumes that the components of \( \varepsilon \) are jointly independent. This implies that \( \mu_{\varepsilon_{l_1}, \ldots, \varepsilon_{l_J}} = \mu_{\varepsilon_{l_1}} \times \ldots \times \mu_{\varepsilon_{l_J}} \).

Ignoring the means, this reduces the number of unknown \( \varepsilon \) moments of order at most \( r \) to \( J \times (r-1) \).

3.3.3 Perfectly Measured Covariates

If available, perfectly measured covariates \( X \) can be used to construct instrumental variables (IVs) (see, e.g., Lewbel, 1997). For example, suppose that \( W = U + \varepsilon \) and let

\[
Y = X' \beta + W' \phi + U' \delta + \eta = X' \beta + W' (\phi + \delta) - \varepsilon' \delta + \eta.
\]

If the covariates \( X \) are independent of \((\varepsilon, \eta)\) then, provided they are relevant, nonlinear functions of \( X \) may be used as instruments for \( W \) in the above equation. Note that, similar to traditional IVs, this enables identifying \( \beta \) and \( \phi + \delta \) but does not identify \( \phi \) and \( \delta \) separately. In particular, consider estimating \( \phi \) using an IV regression of \( Y \) on \( X \) and \( W \). Then a valid instrument for \( W \) that is uncorrelated with \( U' \delta + \eta \) is likely to be also uncorrelated with \( W \) and therefore irrelevant. Nevertheless, intersecting the \( \phi + \delta \) estimand with this paper’s identification region can point identify the model parameters. Thus, one can consider supplementing the moments...
in Sections 3.1 and 3.2 with moments that involve other instruments. The analysis in Sections 3.1 and 3.2 does not require the availability of perfectly measured covariates - it holds even when only data on the error laden proxies $W$ for $U$ is available.

4 Estimation and Inference

Let $\{(Y_i, W_i)\}_{i=1}^n$ be a sample of $n$ independent and identically distributed observations. We consider closed-form and generalized method of moments (GMM) estimators for the vector of unknowns $\theta^*$. The plug-in estimator is based on the closed-form expressions from Theorem 3.2. For any expression $E$ involving population averages, let $\hat{E}$ denote the sample analogue expression that replaces expectations with sample averages. The plug-in estimators for $\theta^- \equiv M(f^-)$ and $\theta^+ \equiv M(f^+)$ are given by

$$\hat{\theta}^- = \hat{M}(\hat{f}^-) \quad \text{and} \quad \hat{\theta}^+ = \hat{M}(\hat{f}^+)$$

Under regularity conditions sufficient to invoke the law of large numbers, central limit theorem, and the delta method, $\hat{\theta}^-$ and $\hat{\theta}^+$ are $\sqrt{n}$ consistent and asymptotically normal estimators for $\theta^-$ and $\theta^+$ respectively. As shown in Theorem 3.2(iii), $\theta^* \in \{\theta^-, \theta^+\}$ and one can use additional restrictions to point-identify $\theta^*$.

It is also useful to discuss a GMM estimator for $\theta^*$. For $Z = (W, Y)$, let $m(Z)$ stack the moments of the form $\mu_{W_1,\ldots,W_J,Y}^{l_1,\ldots,l_J,q}$ that are generated by equation (6) and let $c(\theta^*)$ stack the corresponding right hand side functions of $\theta^*$. Then $\theta^*$ is a solution to the moment equations

$$E[g(Z, \theta^*)] = E[m(Z) - c(\theta^*)] = 0. \quad (7)$$

Given a square positive definite weighting matrix $\Xi^*$ with conforming dimensions, we obtain

$$\theta^* = \arg\min_{\theta \in \Theta} E[g(Z, \theta)]'\Xi^*E[g(Z, \theta)]. \quad (8)$$

As Theorem 3.2 shows, without further restrictions, this minimization problem generally has multiple solutions and does not point identify $\theta^*$. One can proceed by estimating the set of minimizers of this objective function as in Chernozhukov, Hong, and Tamer (2007). Here, we make use of identification results, such as those in Theorem 3.2, to point identify $\theta^*$ by imposing the correct sign restrictions on $\delta_j$ or $\mu_{\varepsilon_j}^r - \mu_{U_j}^r$. 


for $j = 1, ..., J$ and $r \geq 2$. Let $\Theta^* \subset \Theta$ be the restricted parameter space that reflects these sign restrictions. Replacing $\Theta$ with $\Theta^*$ then guarantees that $\theta^*$ is the unique solution to the population optimization problem:

$$
\theta^* = \arg \min_{\theta \in \Theta^*} E[g(Z, \theta)]\Xi^* E[g(Z, \theta)].
$$

We can then estimate $\theta^*$ using a $\sqrt{n}$ consistent and asymptotically normal GMM estimator $\hat{\theta}$. In the simulations and empirical application, we use a two-step GMM estimator. The properties of GMM estimators are well-established, so we omit a detailed discussion for brevity.

If covariates $X$ are present then one must account for conditioning on these. In this case, the projections of $Y$ and $W$ on $X$ would depend on unknown regression coefficients which satisfy standard regression moment equations. By augmenting the above moments with these regression moment equations, one can redefine $\theta^*$ to include these regression coefficients. The redefined $\theta^*$ is then the solution to the augmented system of moments equations, and can be estimated in one step.

Deriving the asymptotic variances of the closed-form estimators $\hat{\theta}^-$ and $\hat{\theta}^+$ and of the GMM estimator $\tilde{\theta}$ analytically is feasible but somewhat unwieldy. A more convenient path forward is to bootstrap the confidence intervals. For this, we resample, e.g. 1,000 times, with replacement to estimate the empirical distribution for the estimator for $\theta^*$. We then use the 2.5% and 97.5% percentiles of this distribution to construct a 95% confidence interval.

## 5 Monte Carlo Simulations

We report simulation results that illustrate the performance of this paper’s method. We consider Theorem 3.2’s closed-form estimators, the naive unconstrained GMM estimator that places no restriction on the parameter space, and the constrained GMM estimator which imposes point-identifying sign restrictions. We base the GMM estimators on the 9 moments in the system of equations (3) and the equation for $\mu_W^4$. This yields a system of 10 equations in 9 unknowns and enables performing overidentification tests. For the GMM estimators, we set the parameter range in the numerical optimization to be $[0, 1000]$ for the even moments parameters and $[-1000, 1000]$ for the remaining parameters.
We generate $n$ data points on $(U, \epsilon, \eta)$ according to $A_2$, drawn from demeaned jointly independent Gamma distributions. Specifically, $U$ follows a demeaned $\Gamma(2, \sqrt{0.5})$. Similarly, we draw observations on $\epsilon$ and $\eta$ from independent demeaned $\Gamma(1.3, 0.5)$ distributions. We then set $\delta = -1$ and $\phi = 1.2$ and generate data on $Y$ and $W$ according to $A_1$:

$$Y = W\phi + U\delta + \eta \quad \text{and} \quad W = U + \epsilon.$$  

Table 2 reports the parameter means and standard deviations across the simulation iterations when the sample size $n = 100,000$ is large. The first column reports the results for the (inconsistent) regression estimator. For the unconstrained GMM estimator in column 2, we choose the initial point for numerical optimization randomly in $[0, 10]$ ($[-10, 10]$) for every even moment parameter (for every other parameter). Without any further constraints, there is no guarantee that the naive unconstrained GMM estimator will converge to the data-generating parameters. This is reflected in column 2 which reports the mean and standard deviation across the 621 simulation iterations in which the numerical optimization algorithm converged.\(^5\) Further inspection of the simulation results reveals that these GMM estimates are scattered across the parameter space, with some estimates concentrated around the two roots from Theorem 3.2. These results illustrate the challenge involved with using GMM naively based on informal identification arguments.

The next 3 columns report the results for the closed-form estimators. The first estimator incorrectly assumes the classical measurement error proxy exclusion restriction, $\phi = 0$, whereas the second and third estimators correspond to the two roots from Theorem 3.2. As expected, the classical measurement error estimator is inconsistent whereas Theorem 3.2’s closed-form differential measurement error estimators perform well. As shown in Table 2, correctly restricting the sign of $\delta$ or distinguishing between the $U$ and $\epsilon$ moments point-identifies the parameters.

\(^5\)The reported number of convergence is based on the exit condition ($exitflag$) of Matlab’s $fmincon$ nonlinear programming solver.
Table 2: Large Sample ($n = 100,000$) Simulation Results

This table reports parameter means and standard deviations (below each mean) across 1,000 simulation iterations. Columns 2 and 3 report the regression and naive unconstrained GMM estimates. Columns 4 and 5-6 correspond to the classical and differential measurement error closed-form estimators respectively. Columns 7-9 correspond to the analogous constrained GMM estimators. The last 2 rows report the frequency with which the Sargen-Hansen J test exceeds the Chi-squared critical value and the GMM numerical optimization algorithm converges.

<table>
<thead>
<tr>
<th>True value</th>
<th>Regression</th>
<th>GMM Unconstrained</th>
<th>Closed form</th>
<th>GMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi + \delta$</td>
<td>0.2</td>
<td>0.445</td>
<td>0.484</td>
<td>0.779</td>
</tr>
<tr>
<td></td>
<td>0.003</td>
<td></td>
<td></td>
<td>0.011</td>
</tr>
<tr>
<td>$\phi$</td>
<td>1.200</td>
<td></td>
<td>0.420</td>
<td>0.200</td>
</tr>
<tr>
<td></td>
<td>3.507</td>
<td></td>
<td></td>
<td>0.000</td>
</tr>
<tr>
<td>$\delta$</td>
<td>-1.000</td>
<td></td>
<td>0.064</td>
<td>1.001</td>
</tr>
<tr>
<td></td>
<td>5.294</td>
<td></td>
<td></td>
<td>0.011</td>
</tr>
<tr>
<td>$\mu_{\mu}^2$</td>
<td>1.000</td>
<td></td>
<td>0.663</td>
<td>0.757</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.390</td>
</tr>
<tr>
<td>$\mu_{\mu}^2$</td>
<td>0.325</td>
<td></td>
<td>0.624</td>
<td>0.568</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.396</td>
</tr>
<tr>
<td>$\mu_{\mu}^2$</td>
<td>0.325</td>
<td></td>
<td>0.386</td>
<td>0.373</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.106</td>
</tr>
<tr>
<td>$\mu_{\mu}^3$</td>
<td>1.414</td>
<td></td>
<td>-4.626</td>
<td>0.864</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>125.409</td>
</tr>
<tr>
<td>$\mu_{\mu}^3$</td>
<td>0.325</td>
<td></td>
<td>6.212</td>
<td>0.876</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>125.456</td>
</tr>
<tr>
<td>$\mu_{\mu}^4$</td>
<td>6.000</td>
<td></td>
<td>3.482</td>
<td>3.243</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.881</td>
</tr>
<tr>
<td>$\mu_{\mu}^4$</td>
<td>0.804</td>
<td></td>
<td>3.009</td>
<td>2.934</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.886</td>
</tr>
<tr>
<td>$J$ Rejection freq</td>
<td>0.068</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Convergence freq</td>
<td>0.621</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3: Simulation Results Based on a Smaller Sample ($n = 1,000$)

This Table reports parameter means and standard deviations (below each mean) across 1,000 simulation iterations. Column 2 corresponds to the regression estimator, columns 3-4 to the differential measurement error closed-form estimators, and columns 5-6 to the constrained GMM estimators. The last 2 rows report the Sargen-Hansen J test rejection frequency based on the Chi-squared critical value and the frequency with which the GMM numerical optimization converges.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>Regression</th>
<th>Closed form</th>
<th>GMM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\delta &gt; 0$</td>
<td>$\delta &lt; 0$</td>
</tr>
<tr>
<td>$\phi + \delta$</td>
<td>0.2</td>
<td>0.447</td>
<td>1.363</td>
<td>0.239</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.027</td>
<td>3.115</td>
<td>1.927</td>
</tr>
<tr>
<td>$\phi$</td>
<td>1.200</td>
<td>0.239</td>
<td>1.363</td>
<td>0.151</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.429</td>
<td>3.115</td>
<td>0.379</td>
</tr>
<tr>
<td>$\delta$</td>
<td>-1.000</td>
<td>1.124</td>
<td>-1.124</td>
<td>0.901</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.849</td>
<td>3.849</td>
<td>0.511</td>
</tr>
<tr>
<td>$\mu_{U}^2$</td>
<td>1.000</td>
<td>0.378</td>
<td>0.942</td>
<td>0.405</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.264</td>
<td>0.265</td>
<td>0.216</td>
</tr>
<tr>
<td>$\mu_{\epsilon}^2$</td>
<td>0.325</td>
<td>0.942</td>
<td>0.378</td>
<td>0.899</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.265</td>
<td>0.264</td>
<td>0.230</td>
</tr>
<tr>
<td>$\mu_{\eta}^2$</td>
<td>0.325</td>
<td>0.287</td>
<td>0.287</td>
<td>0.335</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.620</td>
<td>0.620</td>
<td>0.069</td>
</tr>
<tr>
<td>$\mu_{U}^3$</td>
<td>1.414</td>
<td>0.429</td>
<td>1.293</td>
<td>0.407</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.379</td>
<td>0.408</td>
<td>0.372</td>
</tr>
<tr>
<td>$\mu_{\epsilon}^3$</td>
<td>0.325</td>
<td>1.293</td>
<td>0.429</td>
<td>1.233</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.408</td>
<td>0.379</td>
<td>0.499</td>
</tr>
<tr>
<td>$\mu_{U}^4$</td>
<td>6.000</td>
<td>1.322</td>
<td>3.832</td>
<td>1.254</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.648</td>
<td>5.189</td>
<td>1.512</td>
</tr>
<tr>
<td>$\mu_{\epsilon}^4$</td>
<td>0.804</td>
<td>3.832</td>
<td>1.322</td>
<td>5.037</td>
</tr>
<tr>
<td></td>
<td></td>
<td>51.189</td>
<td>2.648</td>
<td>2.421</td>
</tr>
</tbody>
</table>

The last 3 columns in Table 2 report the results for 3 constrained GMM estimators. The third from last column reports the results for the classical measurement error GMM estimator which sets $\phi = 0$. The last two are the GMM estimators which restrict the parameter space such that $\delta$ is positive (resp. negative) in the second to last (resp. last) column. To help the numerical optimization solver converge, in each iteration, we set the initial point equal to the corresponding closed-form estimate. This improves the numerical performance of the GMM estimators, which now converge in every simulation iteration. As shown in Table 2, once properly guided
by the identification analysis in Theorem 3.2, the GMM estimators perform well. The Sargen-Hansen $J$ statistic exceeds the Chi-squared 5% critical value in around 6.8% of the (converging) simulation iterations for the unconstrained GMM estimator, in every iteration for the classical error GMM estimator, and in 6% of the iterations for the constrained GMM differential error estimators.

Table 3 examines the performance of the closed-form and constrained GMM estimators in a smaller sample, with $n = 1,000$. Here too, we use the closed-form estimates as initial points for the GMM numerical optimization. Naturally, both types of estimators now exhibit larger variation across iterations but the main patterns from Table 2 are generally preserved.

6 Estimating the Firm Investment Equation: Tobin’s q with Mismeasurement and Feedback

The q theory of investment (see, e.g., Abel and Eberly (1994)) postulates that the firm’s expected marginal return of capital $U$, “marginal q,” determines the firm’s investment $Y$. Tobin (1969) and Tobin and Brainard (1977) argue that the financial sector exerts an influence on the firm investment decision, such that investment is an increasing function of the ratio of the firm’s market value to its assets’ replacement value, “Tobin’s q.” While marginal q and Tobin’s average q generally differ, Hayashi (1982) gives conditions under which they coincide.6 As such, researchers often use the publicly observed Tobin’s q (measured by, e.g., the “market-to-book” ratio) as a proxy $W$ for the latent marginal q.

The literature imposes various assumptions on the measurement error in Tobin’s q and reports mixed findings (see, e.g., the discussion in Chalak and Kim, 2020). For example, Erickson and Whited (2000, 2012) use higher order moments to point identify the investment equation coefficients, and they report evidence in support of q theory. Almeida, Campello, and Galvao (2010) restrict the serial correlation of the error term in Tobin’s q to generate instrumental variables, and provide evidence that contradicts classical q theory (see also Fazzari, Hubbard, and Petersen, 1988; Gilchrist and Himmelberg 1995).

6This includes quadratic investment adjustment costs, constant return to scale, perfect competition, and an efficient financial market.
A common feature of these papers is that they assume non-differential measurement error. This implies that firm managers base their investment decisions on their expectations of the future economic conditions and the frictions, such as the irreversibility and cost, involved in adjusting the capital stock. However, Tobin and Brainard (1990) argue that “managers may make decisions they think the market will like and avoid those they fear the market will not like.” Indeed, there is ample theory regarding the role of the financial markets in producing and aggregating information through market prices via the trading process (e.g., Grossman and Stiglitz (1980) and Glosten and Milgrom (1985)). Moreover, several papers provide empirical evidence that managers extract information from stock prices and use it in forming their investment decisions (see, e.g., Chen, Goldstein, and Jiang (2007), Foucault and Fresard (2012), and Bond, Edmans, and Goldstein (2012)). Nevertheless, the literature on the “feedback” from the financial markets does not explicitly account for the consequences of measurement error in marginal q in its empirical analysis.

We extend the standard q theory investment equation to simultaneously account for the measurement error in marginal q and for the feedback from the financial markets. Specifically, we estimate the following equation for the firm’s investment Y:

\[ Y = X'\beta + W\phi + U\delta + \eta \quad \text{with} \quad W = U + \varepsilon, \]

where Tobin’s q, denoted by W, serves as an error-laden proxy for the firm’s marginal q, denoted by U, and can directly impact the firm’s investment decision. Following the literature, X includes additional controls, namely cash flow (CF) and firm size (Size), and we consider a specification without fixed effects and another with firm and year fixed effects. The standard q theory specification restricts \( \phi \) to be zero. In contrast, the feedback literature implies that \( \phi \) is positive but it generally ignores the measurement error in the proxies it uses for U. By letting Tobin’s q serve as an error-laden proxy for marginal q that can directly affect investment, our framework nests aspects of both literatures. In particular, we envision a setting in which managers act based on their private information encoded in U as well as on the information they extract from the financial market via W. This notion dates back at least to Keynes (1936, p. 135) who states that “the daily revaluations of the Stock Exchange, though they are primarily made to facilitate transfers of old investments between one
individual and another, inevitably exert a decisive influence on the rate of current
investment.”

As summarized in Crotty (1990, p. 523, emphasis in original), “Keynes [1936, Chapter 12] analyzes two potentially conflicting indicators of the expected rate of profit on capital. The first is the direct estimate made by the managers of the enterprise - the traditional MEC schedule. The second is the index of expected profitability implicit in the financial market value of the firm. The higher the value of the firm’s stock, ceteris paribus, the higher the ‘market’s’ implicit estimate of enterprise profitability - a thesis later formalized in Tobin’s Theory Q.” Substituting for $W$ in the $Y$ equation, we have that

$$Y = X' \beta + (\phi + \delta)U + \phi \varepsilon + \eta.$$ 

Here, $\phi + \delta$ captures the total sensitivity of investment to marginal $q$, $U$. Moreover, $\phi$ captures the sensitivity of investment to the additional market information $\varepsilon$ encoded in Tobin’s $q$. We apply our econometric framework to identify these and other key parameters.

### 6.1 Data

We closely follow Erickson and Whited (2012), Erickson, Jiang, and Whited (2014), and Chalak and Kim (2021) in selecting the sample and constructing the variables. We use data from Compustat on industrial firms\(^7\) between 1970 to 2022.\(^8\) We remove financial firms (Standard Industrial Classification (SIC) code 6000 to 6999) and regulated firms (SIC code 4900 to 4999). We also exclude small firms\(^9\) and we remove firms that appear in only one year. We deflate all the Compustat items that enter into the construction of the variables by the Federal Reserve Economic Data’s (yearly average) Producer Price Index, with 1982 as a base year. We define investment as capital expenditure (CAPX) normalized by total assets $AT$ at the beginning of the period. We measure (lagged) Tobin’s $q$ at the beginning of the period by

\(^7\)We apply 4 firm filters: INDFMT=INDL (industrial), DATAFMT=STD (standardized data reporting), POPSRC=D (domestic (North American)), and CONSOL=C (consolidated).

\(^8\)The raw data is from Compustat and accessed via Wharton Research Data Services.

\(^9\)We delete observations in which a firm has at most $2$ million in real total assets (Compustat item: AT) or $5$ million in real capital (Compustat item: PPEGT) at either the end or the beginning of a time period.
\[(PRCC_F \times CSHO) + AT \cdot CEQ - TXDB \]

where PRCC$_F$ is stock price, CSHO is number of common shares outstanding, CEQ is common equity, and TXDB is deferred taxes. We define cash flow as the sum of income before extraordinary items (IB) and depreciation and amortization (DP) normalized by AT at the beginning of the period. Further, we define firm size as the natural logarithm of real net sales (SALE). We delete firm-year observations with missing data on one of these variables. Last, following the literature, we winsorize the smallest and largest percentile of the variables in the panel. The final sample is an unbalanced panel of 176,376 firm-year observations, with 14,708 firms per year on average. Table 4 reports the summary statistics for the panel variables.

Table 4: Summary Statistics: This table reports summary statistics based on 176,376 firm-year observations in an unbalanced panel from year 1970 to 2022, with an average of 14,708 firms per year. In each year, investment and cash flow are normalized by the firm’s total assets, Tobin’s q is measured by the market-to-book ratio, and firm size is the log of the firm’s sales.

<table>
<thead>
<tr>
<th></th>
<th>Investment</th>
<th>Tobin’s q</th>
<th>Cash Flow</th>
<th>Firm Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>0.081</td>
<td>1.675</td>
<td>0.069</td>
<td>5.398</td>
</tr>
<tr>
<td>std dev</td>
<td>0.095</td>
<td>1.221</td>
<td>0.139</td>
<td>1.982</td>
</tr>
<tr>
<td>min</td>
<td>0.002</td>
<td>0.525</td>
<td>-0.552</td>
<td>0.297</td>
</tr>
<tr>
<td>max</td>
<td>0.580</td>
<td>7.891</td>
<td>0.393</td>
<td>10.322</td>
</tr>
</tbody>
</table>

6.2 Empirical Results

Tables 5 and 6 report the point estimates and bootstrapped 95% confidence intervals. We pay particular attention to the total sensitivity \((\phi + \delta)\) of investment to marginal q as well as to the sensitivity \(\phi\) of investment to the additional information \(\varepsilon\) encoded in Tobin’s q. As in Erickson and Whited (2012), we report the results for the pooled panel in Table 5 and after accounting for firm and year fixed effects in Table 6. We focus the discussion on the results with fixed effects. However, the results from the pooled sample are, for the most part, qualitatively similar.

The first column of Table 6 reports the regression coefficients. Consistent with the literature, the coefficients on Tobins’ q and cash flow are positive and statistically significant. The second column reports the closed-form estimates when we restrict \(\phi = 0\). The estimate for \(\delta\) is positive and statistically significant whereas the coefficient on cash flow is no longer statistically significant. We obtain qualitatively similar results

\[^{10}\text{For the bootstrap, we draw (time series of) firms, with replacement, from the realized panel.}\]
when using the constrained GMM estimator in column 5. These results are consistent with those reported in Erickson and Whited (2000, 2012).

Columns 3-4 and 6-7 in Table 6 report the closed-form and constrained GMM estimators\textsuperscript{11} when we relax the restriction $\phi = 0$. Column 3 reports the closed-form estimates when $\delta$ is positive. The estimate for $\phi$ is modest, positive, and statistically significant. Here, investment is more sensitive to marginal q ($U$) than to the additional information ($\varepsilon$) encoded in Tobin’s q. Concomitantly, $\mu^2_U$ is estimated to be significantly smaller than $\mu^2_\varepsilon$ suggesting that Tobin’s q is a relatively poor proxy for marginal q. Specifically, the net-of-X reliability ratio (after projecting on X and the fixed effects) is 0.101. That is, marginal q accounts for only 10.1% of the residual variation in Tobin’s q. Last, the coefficient on cash flow is modest, yet significantly negative. We obtain similar estimates using the GMM estimator in column 6.

Column 4 reports the closed-form estimates when $\delta$ is negative. Here, $\phi$ is sizeable, positive, and statistically significant. In particular, investment is less sensitive to marginal q ($U$) than to the additional information ($\varepsilon$) in the market proxy. Tobin’s q is now estimated to be a relatively accurate proxy for marginal q. Specifically, the (net-of-X) reliability ratio is 0.899. Here, the coefficient on cash flow is positive and significant. We obtain similar estimates using the GMM estimator in column 7.

The results in columns 3-4 (resp. 6-7) fit the data equally well under the imposed assumptions; these models are observationally equivalent. In column 3, Tobin’s q is a noisy proxy for firms’ investment opportunities. In this case, the managers attach more weight to their private information ($U$) rather than to the additional information ($\varepsilon$) in the market proxy when forming their investment decision. In column 4, Tobin’s q proxies marginal q relatively accurately. Here, the managers attach less weight in their investment decision to their private information ($U$) than to the additional information ($\varepsilon$) in the market price. In both cases, we estimate $\phi$ to be positive and statistically significant, although it is modest in the first case and substantial in the second. Our estimates thus show that the more precise the signal from the financial market is, the more sensitive the management investment decision to it will be.

\textsuperscript{11}For the numerical optimization of each GMM estimator, we set the initial point to the corresponding closed-form estimate and we set the range of each parameter such that its magnitude does not exceed 10 times that of the closed form estimate (and such that the even moments parameters are nonnegative). The GMM numerical optimization solver converges across all bootstrap iterations in Tables 5 and 6, except for 0.6% of the iterations for the classical GMM estimator in Table 6.
Table 5: Investment equation estimates using the pooled sample: Tobin’s q as an error-laden non-excluded proxy

Column 1 reports the regression results. Columns 2 and 3-4 report the closed-form estimates for classical and differential measurement error respectively. Columns 5-7 report the analogous GMM estimates. \( RR (RR_{net}) \) refers to the reliability ratio of Tobin’s q (net of the covariates). 95% bootstrapped confidence intervals appear below the point estimates. The last row reports the Sargen-Hansen J test bootstrapped p-value.

<table>
<thead>
<tr>
<th></th>
<th>Reg</th>
<th>Closed form</th>
<th>GMM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \phi + \delta )</td>
<td>( \phi = 0 )</td>
<td>( \delta &gt; 0 )</td>
</tr>
<tr>
<td>( \phi )</td>
<td>0.007</td>
<td>0.132</td>
<td>0.167</td>
</tr>
<tr>
<td></td>
<td>[0.006, 0.008]</td>
<td>[0.123, 0.141]</td>
<td>[0.154, 0.185]</td>
</tr>
<tr>
<td>( \delta )</td>
<td>0.132</td>
<td>0.165</td>
<td>-0.165</td>
</tr>
<tr>
<td></td>
<td>[0.002, 0.003]</td>
<td>[0.154, 0.185]</td>
<td>[-0.183, -0.152]</td>
</tr>
<tr>
<td>( \mu_{\eta} )</td>
<td>0.250</td>
<td>0.175</td>
<td>1.651</td>
</tr>
<tr>
<td>( \mu_{\xi} )</td>
<td>0.226, 0.273</td>
<td>0.152, 0.197</td>
<td>1.587, 1.714</td>
</tr>
<tr>
<td>( \mu_{\lambda} )</td>
<td>1.576</td>
<td>1.651</td>
<td>0.175</td>
</tr>
<tr>
<td></td>
<td>[1.514, 1.639]</td>
<td>[1.587, 1.714]</td>
<td>[0.152, 0.197]</td>
</tr>
<tr>
<td>( \beta_{CF} )</td>
<td>0.194</td>
<td>0.133</td>
<td>0.116</td>
</tr>
<tr>
<td></td>
<td>[0.185, 0.202]</td>
<td>[0.198, 0.307]</td>
<td>[5.056, 5.561]</td>
</tr>
<tr>
<td>( \beta_{Size} )</td>
<td>-0.009</td>
<td>-0.005</td>
<td>-0.004</td>
</tr>
<tr>
<td></td>
<td>[-0.010, -0.009]</td>
<td>[-0.007, -0.004]</td>
<td>[-0.006, -0.003]</td>
</tr>
<tr>
<td>( RR_{net} )</td>
<td>0.137</td>
<td>0.096</td>
<td>0.904</td>
</tr>
<tr>
<td></td>
<td>[0.123, 0.150]</td>
<td>[0.083, 0.108]</td>
<td>[0.892, 0.917]</td>
</tr>
<tr>
<td>( RR )</td>
<td>0.140</td>
<td>0.100</td>
<td>0.904</td>
</tr>
<tr>
<td></td>
<td>[0.127, 0.154]</td>
<td>[0.087, 0.112]</td>
<td>[0.893, 0.917]</td>
</tr>
<tr>
<td>J-test p</td>
<td>5.21</td>
<td>5.10</td>
<td>5.10</td>
</tr>
</tbody>
</table>
Table 6: Investment equation estimates with year and firm fixed effects: Tobin’s q as an error-laden non-excluded proxy

Column 1 reports the regression results. Columns 2 and 3-4 report the closed-form estimates for classical and differential measurement error respectively. Columns 5-7 report the analogous GMM estimates. \( RR_{\text{within}} \) (\( RR_{\text{net}} \)) refers to the reliability ratio of Tobin’s q based on the within transformed data (after further projecting on the covariates). 95% bootstrapped confidence intervals appear below the point estimates. The last row reports the Sargen-Hansen J test bootstrapped p-value.

|                | Reg | Closed form | GMM               |               |               |               |               |
|----------------|-----|-------------|-------------------|---------------|---------------|---------------|
| \( \phi + \delta \) |    | \( \phi = 0 \) \( \delta > 0 \) \( \delta < 0 \) | \( \phi = 0 \) \( \delta > 0 \) \( \delta < 0 \) |               |               |               |
| \( \phi \)     | 0.017 | 0.085 | 0.121 | 0.005 | 0.040 | 0.149 | 0.007 |
| \( \phi \)     | [0.016, 0.018] | [0.080, 0.091] | [0.112, 0.132] | [0.004, 0.006] | [0.020, 0.214] | [0.123, 0.182] | [0.006, 0.008] |
| \( \delta \)   | 0.085 | 0.116 | -0.116 | 0.040 | 0.142 | -0.142 |               |
| \( \delta \)   | [0.080, 0.091] | [0.107, 0.127] | [-0.127, -0.107] | [0.020, 0.214] | [0.116, 0.175] | [-0.175, -0.116] |               |
| \( \mu_{\epsilon}^{2} \) | 0.117 | 0.060 | 0.535 | 0.201 | 0.039 | 0.547 |               |
| \( \mu_{\epsilon}^{2} \) | [0.108, 0.127] | [0.051, 0.069] | [0.515, 0.557] | [0.038, 0.388] | [0.030, 0.050] | [0.526, 0.568] |               |
| \( \mu_{U}^{2} \)   | 0.478 | 0.535 | 0.060 | 0.378 | 0.547 | 0.039 |               |
| \( \mu_{U}^{2} \)   | [0.459, 0.498] | [0.515, 0.557] | [0.051, 0.069] | [0.187, 0.527] | [0.526, 0.568] | [0.030, 0.050] |               |
| \( \mu_{U}^{3} \) | 0.142 | 0.069 | 0.742 | 0.188 | 0.036 | 0.725 |               |
| \( \beta_{CF} \) | 0.125 | -0.004 | -0.073 | 0.147 | 0.081 | -0.126 | 0.144 |
| \( \beta_{CF} \) | [0.118, 0.132] | [-0.019, 0.010] | [-0.097, -0.051] | [0.140, 0.155] | [-0.251, 0.119] | [-0.189, -0.074] | [0.136, 0.152] |
| \( \beta_{Size} \) | -0.005 | -0.001 | 0.001 | -0.006 | -0.003 | 0.003 | -0.005 |
| \( \beta_{Size} \) | [-0.006, -0.004] | [-0.003, 0.001] | [-0.001, 0.004] | [-0.007, -0.004] | [-0.005, 0.008] | [0.000, 0.006] | [-0.007, -0.004] |
| \( RR_{\text{net}} \) | 0.197 | 0.101 | 0.899 | 0.347 | 0.066 | 0.934 |               |
| \( RR_{\text{within}} \) | [0.182, 0.213] | [0.085, 0.115] | [0.885, 0.915] | [0.068, 0.674] | [0.051, 0.086] | [0.914, 0.949] |               |
| \( J\text{-test p} \) | 0.278 | 0.191 | 0.909 | 0.413 | 0.160 | 0.940 |               |
| \( J\text{-test p} \) | [0.265, 0.291] | [0.176, 0.206] | [0.896, 0.923] | [0.162, 0.706] | [0.145, 0.178] | [0.923, 0.954] |               |
| \( J\text{-test p} \) | 0.458 | 0.497 | 0.497 | 0.458 | 0.497 | 0.497 |               |
7 Conclusion

The econometrics literature on measurement error focuses primarily on the classical assumptions under which agents act based on latent variables that the econometrician proxies using an error-laden redundant (excluded) proxy. The literature also sometimes considers the Berkson (1950) setting in which agents act based on their publicly observed optimal predictions of the latent variables. This paper describes several economic contexts in which the outcome of interest depends on both the latent variables, privately observed by the agents, and on publicly observed proxies for these variables. General econometrics results accommodating this more general setting have been lacking so far. We fill in this gap by studying a leading setting for differential measurement error in which a latent variable and its error-laden proxy can jointly affect the outcome of interest. Here, the “error” is construed broadly and may very well convey valuable economic information. We demonstrate that restricting the dependence among the higher order moments of the unobservables partially identifies the model’s parameters. In the leading case of a scalar latent variable, the identification set consists of two points. The lack of point identification arises because switching the labels of the latent variables and the proxy errors can yield observationally equivalent models. However, simple restrictions, either signing the effects of the latent variables or distinguishing between the moments of the latent variables and the moments of the errors, can secure point identification. We put forward a closed-form solution estimator and a generalized method of moment estimator, and we illustrate their performance using simulations. We apply our framework to estimate the firm investment equation using Tobin’s q as a proxy for marginal q. We extend the standard q theory specification to allow Tobin’s q to directly affect investment, reflecting the influence that the financial market exerts on the firms’ management decisions. We identify the equation parameters in a set of two points. We estimate that the influence of the additional information extracted from the financial market proxy on firm investment is greater when Tobin’s q proxies marginal q accurately, and is modest yet significant otherwise.

An interesting topic for future research would examine the asymptotic behavior of estimators based on higher order moments as the number of moments increases. Last, less is known about the consequences of differential measurement error in nonlinear
and nonparametric systems. The results in this paper about linear systems constitute a step toward understanding such more general cases.

References


Levinsohn, J. and A. Petrin (2003), “Estimating Production Functions Using In-


A Mathematical Proofs

Proof of Theorem 3.1: We use the system of equations (3) to express all the unknowns as a function of $\phi$. First, using the moments of order 3, we have that

$$\mu_{W,Y}^{2,1} - \phi \mu_{\varepsilon}^{3} = (\phi + \delta) \mu_{U}^{3} \quad \text{and} \quad \mu_{W,Y}^{1,2} - \phi^{2} \mu_{\varepsilon}^{3} = (\phi + \delta)^{2} \mu_{U}^{3}.$$  

Given that $(\phi + \delta)$ and $\mu_{U}^{3}$ are nonzero, we obtain

$$(\phi + \delta) = \frac{\mu_{W,Y}^{1,2} - \phi^{2} \mu_{\varepsilon}^{3}}{\mu_{W,Y}^{2,1} - \phi \mu_{\varepsilon}^{3}}, \quad \mu_{U}^{3} = \frac{\mu_{W,Y}^{2,1} - \phi \mu_{\varepsilon}^{3}}{(\phi + \delta)}, \quad \text{and} \quad \mu_{\varepsilon}^{3} = \mu_{W}^{3} - \frac{\mu_{W,Y}^{2,1} - \phi \mu_{\varepsilon}^{3}}{(\phi + \delta)}.$$  

Combining the equations for $(\phi + \delta)$ and $\mu_{U}^{3}$ then gives

$$\mu_{W}^{3} - \mu_{\varepsilon}^{3} = (\mu_{W,Y}^{2,1} - \phi \mu_{\varepsilon}^{3}) \frac{\mu_{W,Y}^{2,1} - \phi \mu_{\varepsilon}^{3}}{\mu_{W,Y}^{2,1} - \phi^{2} \mu_{\varepsilon}^{3}}.$$  

Rearranging this expression gives

$$\left(\mu_{W}^{3} - \mu_{\varepsilon}^{3}\right) \left(\mu_{W,Y}^{2,1} - \phi \mu_{\varepsilon}^{3}\right) - \left(\mu_{W,Y}^{2,1} - \phi \mu_{\varepsilon}^{3}\right)^{2}$$

$$= \mu_{W}^{3} \mu_{W,Y}^{2,1} - \phi^{2} \mu_{W}^{3} \mu_{\varepsilon}^{3} - \mu_{\varepsilon}^{3} \mu_{W,Y}^{2,1} + \phi^{2} (\mu_{\varepsilon}^{3})^{2} - (\mu_{W,Y}^{2,1})^{2} + 2 \phi \mu_{\varepsilon}^{3} \mu_{W,Y}^{2,1} - \phi^{2} (\mu_{\varepsilon}^{3})^{2}$$

$$= \mu_{W}^{3} \mu_{W,Y}^{2,1} - (\mu_{W,Y}^{2,1})^{2} - \mu_{\varepsilon}^{3} (2 \phi \mu_{W}^{3} + \mu_{W,Y}^{2,1} - 2 \phi \mu_{W,Y}^{2,1})$$

$$= \mu_{W}^{3} \mu_{W,Y}^{2,1} - (\mu_{W,Y}^{2,1})^{2} - \mu_{\varepsilon}^{3} [\phi (\mu_{W}^{3} - \mu_{W,Y}^{2,1}) - (\mu_{W,Y}^{2,1} - \mu_{W,Y}^{1,2})] = 0.$$  

From system (3), since $\delta \neq 0$ and $\mu_{U}^{3} \neq 0$, we have

$$\phi (\phi \mu_{W}^{3} - \mu_{W,Y}^{2,1}) - (\phi \mu_{W,Y}^{2,1} - \mu_{W,Y}^{1,2})$$

$$= \phi (\mu_{U}^{3} + \phi \mu_{\varepsilon}^{3} - (\phi + \delta) \mu_{U}^{3} - \phi \mu_{\varepsilon}^{3}) - \phi (\mu_{W,Y}^{2,1} - \phi \mu_{\varepsilon}^{3}) + (\mu_{W,Y}^{2,1} - \phi \mu_{\varepsilon}^{3})$$

$$= \phi \mu_{U}^{3} [\phi - (\phi + \delta)] - \phi (\phi + \delta) \mu_{U}^{3} + (\phi + \delta)^{2} \mu_{U}^{3}$$

$$= [\phi - (\phi + \delta)]^{2} \mu_{U}^{3} = 0.$$  

We can then express $\mu_{\varepsilon}^{3}$, $(\phi + \delta)$, and $\mu_{U}^{3}$ as functions of $\phi$:

$$\mu_{\varepsilon}^{3} = S_{\varepsilon}(\phi) \equiv \frac{\mu_{W}^{3} \mu_{W,Y}^{2,1} - (\mu_{W,Y}^{2,1})^{2}}{\phi (\phi \mu_{W}^{3} - \mu_{W,Y}^{2,1}) - (\phi \mu_{W,Y}^{2,1} - \mu_{W,Y}^{1,2})},$$

$$(\phi + \delta) = T(\phi) \equiv \frac{\mu_{W,Y}^{2,1} - \phi^{2} S_{\varepsilon}(\phi)}{\mu_{W,Y}^{2,1} - \phi S_{\varepsilon}(\phi)} , \quad \text{and}$$

$$\mu_{U}^{3} = S_{U}(\phi) \equiv \frac{\mu_{W}^{2,1} - \phi S_{\varepsilon}(\phi)}{T(\phi)}.$$
Next, using the moments of order 2 in system (3), we have

\[ \mu_{W,Y}^{1,1} = (\phi + \delta)(\mu_{W}^2 - \mu_{\varepsilon}^2) + \phi \mu_{\varepsilon}^2 = T(\phi)\mu_{W}^2 + (\phi - T(\phi))\mu_{\varepsilon}^2 \]

and we can express \( \mu_{\varepsilon}^2, \mu_{\eta}^2 \), and \( \mu_{\eta}^2 \) as a function of \( \phi \). Given \( \delta \neq 0 \), we have

\[ \mu_{\varepsilon}^2 = V_\varepsilon(\phi) \equiv \frac{\mu_{W,Y}^{1,1} - T(\phi)\mu_{W}^2}{\phi - T(\phi)} \quad \text{and} \quad \mu_{U}^2 = V_U(\phi) \equiv \mu_{W}^2 - V_\varepsilon(\phi), \]

\[ \mu_{\eta}^2 = V_\eta(\phi) \equiv \mu_{Y}^2 - T(\phi)^2 V_U(\phi) - \phi^2 V_\varepsilon(\phi). \]

Last, the moments \( \mu_{W,Y}^{3,1} \) and \( \mu_{W,Y}^{2,2} \) of order 4 in system (3) can be rewritten as

\[ \mu_{W,Y}^{3,1} = (\phi + \delta)\mu_{U}^4 + \phi \mu_{\varepsilon}^4 + 3[(\phi + \delta) + \phi]\mu_{U}^2\mu_{\varepsilon}^2 \quad \text{and} \]

\[ \mu_{W,Y}^{2,2} = (\phi + \delta)^2\mu_{U}^4 + \phi^2 \mu_{\varepsilon}^4 + [\phi^2 + 4(\phi + \delta)\phi + (\phi + \delta)^2]\mu_{U}^2\mu_{\varepsilon}^2 + \mu_{U}^2\mu_{\eta}^2 + \mu_{\varepsilon}^2\mu_{\eta}^2. \]

Since \( \phi + \delta \neq 0 \), we have

\[ \mu_{U}^4 = \frac{\mu_{W,Y}^{3,1} - \phi \mu_{\varepsilon}^4 - 3[(\phi + \delta) + \phi]\mu_{U}^2\mu_{\varepsilon}^2}{(\phi + \delta)} \quad \text{and} \]

\[ \mu_{W,Y}^{2,2} = (\phi + \delta)\mu_{W,Y}^{3,1} - \phi(\phi + \delta)\mu_{\varepsilon}^4 - 3(\phi + \delta)[(\phi + \delta) + \phi]\mu_{U}^2\mu_{\varepsilon}^2 + \phi^2\mu_{\eta}^4 \]

\[ + [\phi^2 + 4(\phi + \delta)\phi + (\phi + \delta)^2]\mu_{U}^2\mu_{\varepsilon}^2 + \mu_{U}^2\mu_{\eta}^2 + \mu_{\varepsilon}^2\mu_{\eta}^2, \]

so that

\[ \mu_{W,Y}^{2,2} = (\phi + \delta)\mu_{W,Y}^{3,1} + \phi[\phi - (\phi + \delta)]\mu_{\varepsilon}^4 + [\phi - (\phi + \delta)][\phi + 2(\phi + \delta)]\mu_{U}^2\mu_{\varepsilon}^2 + \mu_{U}^2\mu_{\eta}^2 + \mu_{\varepsilon}^2\mu_{\eta}^2, \]

where we make use of

\[ \phi^2 + 4(\phi + \delta)\phi + (\phi + \delta)^2 - 3(\phi + \delta)[(\phi + \delta) + \phi] = [\phi - (\phi + \delta)][\phi + 2(\phi + \delta)]. \]

Given that \( \phi \neq 0 \) and \( \delta \neq 0 \), it follows that

\[ \mu_{\varepsilon}^4 = K_\varepsilon(\phi) \equiv \frac{1}{\phi[\phi - T(\phi)]} \times \{\mu_{W,Y}^{2,2} - T(\phi)\mu_{W,Y}^{3,1} \]

\[ - [\phi - T(\phi)][\phi + 2T(\phi)]V_U(\phi)V_\varepsilon(\phi) - V_U(\phi)V_\eta(\phi) - V_\varepsilon(\phi)V_\eta(\phi)\}, \]

and

\[ \mu_{U}^4 = K_U(\phi) \equiv \frac{\mu_{W,Y}^{3,1} - \phi K_\varepsilon(\phi) - 3[T(\phi) + \phi]V_U(\phi)V_\varepsilon(\phi)}{T(\phi)}. \]

We make use of lemmas A.1 and A.2 in the proof of Theorem 3.2.
Lemma A.1: Under the conditions of Theorem 3.1, we have:

\[
T(\phi) = \frac{\phi\mu_{W,Y}^{2,1} - \mu_{W,Y}^{1,2}}{\phi \mu_{W}^{2,1} - \mu_{W,Y}^{1,2}},
\]

\[
V_{\epsilon}(\phi) = \frac{\mu_{W,Y}^{1,1}(\phi \mu_{W}^{3} - \mu_{W,Y}^{2,1}) - (\phi \mu_{W,Y}^{2,1} - \mu_{W,Y}^{1,2})\mu_{W}^{2,1}}{\phi(\phi \mu_{W}^{3} - \mu_{W,Y}^{2,1}) - (\phi \mu_{W,Y}^{2,1} - \mu_{W,Y}^{1,2})},
\]

\[
V_{U}(\phi) = \frac{(\phi \mu_{W}^{3} - \mu_{W,Y}^{2,1})(\phi \mu_{W}^{2,1} - \mu_{W,Y}^{1,1})}{\phi(\phi \mu_{W}^{3} - \mu_{W,Y}^{2,1}) - (\phi \mu_{W,Y}^{2,1} - \mu_{W,Y}^{1,2})}, \text{ and}
\]

\[
V_{n}(\phi) = \mu_{W}^{2} + (\phi \mu_{W}^{2} - \mu_{W,Y}^{1,1})(\phi + \frac{\phi \mu_{W,Y}^{2,1} - \mu_{W,Y}^{1,2}}{\phi \mu_{W}^{3} - \mu_{W,Y}^{2,1}}) - \phi^{2}\mu_{W}^{2}.
\]

Proof of Lemma A.1: From the proof of Theorem 3.1, we have that

\[
\phi \mu_{W,Y}^{2,1} - \mu_{W,Y}^{1,2} = -\delta(\phi + \delta)\mu_{3}^{3} \neq 0 \text{ and } \phi \mu_{W}^{3} - \mu_{W,Y}^{2,1} = -\delta \mu_{U}^{3} \neq 0
\]

and

\[
\phi(\phi \mu_{W}^{3} - \mu_{W,Y}^{2,1}) - (\phi \mu_{W,Y}^{2,1} - \mu_{W,Y}^{1,2}) = \phi^{2}\mu_{W}^{3} + \mu_{W,Y}^{1,2} - 2\phi \mu_{W,Y}^{2,1} \neq 0.
\]

Thus, we write

\[
T(\phi) = \frac{1,2}{\phi \mu_{W,Y}^{2,1} - \mu_{W,Y}^{1,2}} = \frac{1,2}{\phi \mu_{W,Y}^{2,1} - \mu_{W,Y}^{1,2}} \cdot \frac{\mu_{W,Y}^{1,2} - \phi^{2}\mu_{W,Y}^{2,1} - (\mu_{W,Y}^{2,1})^{2}}{\phi \mu_{W,Y}^{3} + \mu_{W,Y}^{2,1} - \phi \mu_{W,Y}^{1,2} - (\mu_{W,Y}^{2,1})^{2}}
\]

\[
= \frac{1,2}{\mu_{W,Y}^{2,1} - \phi \mu_{W,Y}^{2,1} - (\mu_{W,Y}^{2,1})^{2}} \cdot \frac{\mu_{W,Y}^{1,2} - \phi \mu_{W,Y}^{2,1} - (\mu_{W,Y}^{2,1})^{2}}{\mu_{W,Y}^{2,1} - \phi \mu_{W,Y}^{2,1} - (\mu_{W,Y}^{2,1})^{2}}
\]

Further, using the expression for \(T(\phi)\) and given \(\phi \mu_{W}^{3} - \mu_{W,Y}^{2,1} \neq 0\), we have

\[
V_{\epsilon}(\phi) = \frac{1,1}{\phi - T(\phi)} = \frac{1,1}{\phi - T(\phi)} \cdot \frac{\mu_{W,Y}^{1,1} - (\phi \mu_{W}^{3} - \mu_{W,Y}^{2,1}) - (\phi \mu_{W,Y}^{2,1} - \mu_{W,Y}^{1,2})\mu_{W}^{2}}{\phi(\phi \mu_{W}^{3} - \mu_{W,Y}^{2,1}) - (\phi \mu_{W,Y}^{2,1} - \mu_{W,Y}^{1,2})}
\]

and

\[
V_{U}(\phi) = \mu_{W}^{2} - V_{\epsilon}(\phi) = \mu_{W}^{2} - \mu_{W}^{2} = \frac{1,1}{\phi(\phi \mu_{W}^{3} - \mu_{W,Y}^{2,1}) - (\phi \mu_{W,Y}^{2,1} - \mu_{W,Y}^{1,2})}
\]

\[
= \frac{(\phi \mu_{W}^{2} - \mu_{W,Y}^{1,1})(\phi \mu_{W}^{3} - \mu_{W,Y}^{1,1})}{\phi(\phi \mu_{W}^{3} - \mu_{W,Y}^{2,1}) - (\phi \mu_{W,Y}^{2,1} - \mu_{W,Y}^{1,2})}
\]

and

\[
V_{n}(\phi) = \mu_{W}^{2} + (\phi \mu_{W}^{2} - \mu_{W,Y}^{1,1})(\phi + \frac{\phi \mu_{W,Y}^{2,1} - \mu_{W,Y}^{1,2}}{\phi \mu_{W}^{3} - \mu_{W,Y}^{2,1}}) - \phi^{2}\mu_{W}^{2}.
\]
Last, we have that
\[ V_\eta(\phi) = \mu^2_Y - T(\phi)^2 V_U(\phi) - \phi^2 V_\varepsilon(\phi) \]
and since
\[ -T(\phi)^2 V_U(\phi) - \phi^2 V_\varepsilon(\phi) \]
\[ = [-(\phi + \delta)^3(\mu^4_U - 3\mu^2_U\mu^2_\varepsilon) - \phi^3(\phi^2\mu^2_U\mu^2_\varepsilon - 3\phi^2\mu^2_U\mu^2_\varepsilon) - 3(\phi + \delta)^2\phi^2\mu^2_U\mu^2_\varepsilon - 3\phi^2\mu^2_U\mu^2_\varepsilon]) \]
\[ = (\phi + \delta)^3(3\mu^2_U\mu^2_U - \mu^4_U) + \phi^3(3\mu^2_U\mu^2_U - \mu^4_U) \]
and
\[ \mu^3_{W,Y} - 3\mu^2_W\mu^1_{W,Y} \]
\[ = (\phi + \delta)\mu^4_U + \phi\mu^4_\varepsilon + 3(\phi + \delta)\mu^2_U\mu^2_\varepsilon + 3\mu^2_U\mu^2_\varepsilon - 3(\mu^2_U + \mu^2_\varepsilon)[\phi + \delta]\mu^2_U + \phi\mu^2_\varepsilon \]
\[ = (\phi + \delta)(\mu^4_U - 3\mu^2_U\mu^2_\varepsilon) + \phi(\mu^4_\varepsilon - 3\mu^2_\varepsilon\mu^2_\varepsilon). \]

**Lemma A.2**: Under the conditions of Theorem 3.2, we have:

\[ A = (\phi + \delta)^3[\mu^3_U(\mu^4_U - 3\mu^2_U\mu^2_\varepsilon) - \phi^3(\mu^4_\varepsilon - 3\mu^2_\varepsilon\mu^2_\varepsilon)], \]
\[ B = (\phi + \delta)^2[\phi\mu^4_U - 3\phi^2\mu^2_U\mu^2_\varepsilon], \] and
\[ C = (\phi + \delta)^2[\phi\mu^4_U - 3\phi^2\mu^2_U\mu^2_\varepsilon]. \]

**Proof of Lemma A.2**: We begin by examining the terms in the expression for \( B \).

We have
\[ \mu_{W,Y}^3 + 3\mu_{W,Y}^2 \]
\[ = -(\phi + \delta)^3(3\mu^2_U\mu^2_U - \mu^4_U) + \phi^3(3\mu^2_U\mu^2_U - \mu^4_U) \]
and
\[ \mu_{W,Y}^3 - 3\mu^2_W\mu_{W,Y} \]
\[ = (\phi + \delta)(\mu^4_U - 3\mu^2_U\mu^2_\varepsilon) + \phi(\mu^4_\varepsilon - 3\mu^2_\varepsilon\mu^2_\varepsilon). \]
It follows that

\[ B \equiv \mu_{W}^{3}(-\mu_{W,Y}^{1,3} + 3\mu_{W,Y}^{1,1}\mu_{Y}^{2}) + \mu_{W,Y}^{1,2}(\mu_{W,Y}^{3,1} - 3\mu_{W}^{2}\mu_{W,Y}^{1,1}) \]

\[ = (\mu_{U}^{3} + \mu_{\epsilon}^{3})(\phi + \delta)^{3}(3\mu_{U}^{2}\mu_{U}^{2} - \mu_{U}^{4}) + \phi^{3}(3\mu_{\epsilon}^{2}\mu_{\epsilon}^{2} - \mu_{\epsilon}^{4}) \]

\[ + [(\phi + \delta)^{2}\mu_{U}^{3} + \phi^{2}\mu_{\epsilon}^{3}][(\phi + \delta)(\mu_{U}^{4} - 3\mu_{U}^{2}\mu_{U}^{2}) + \phi(\mu_{\epsilon}^{4} - 3\mu_{\epsilon}^{2}\mu_{\epsilon}^{2})] \]

\[ = -(\phi + \delta)[(\phi + \delta)^{2} - \phi^{2}](\mu_{U}^{4} - 3\mu_{U}^{2}\mu_{U}^{2})\mu_{U}^{3} \]

\[ + \phi[(\phi + \delta)^{2} - \phi^{2}](\mu_{\epsilon}^{4} - 3\mu_{\epsilon}^{2}\mu_{\epsilon}^{2})\mu_{U}^{3} \]

\[ = [(\phi + \delta)^{2} - \phi^{2}][-(\phi + \delta)(\mu_{U}^{4} - 3\mu_{U}^{2}\mu_{U}^{2})\mu_{U}^{3} + \phi(\mu_{\epsilon}^{4} - 3\mu_{\epsilon}^{2}\mu_{\epsilon}^{2})\mu_{U}^{3}] . \]

Similarly, we examine the expression for \( A \). We have that

\[ 3\mu_{W,Y}^{1,1}\mu_{W}^{2} - \mu_{W,Y}^{3,1} \]

\[ = 3(\mu_{U}^{2} + \mu_{\epsilon}^{2})[(\phi + \delta)\mu_{U}^{2} + \phi\mu_{\epsilon}^{2}] - (\phi + \delta)\mu_{U}^{4} - \phi\mu_{\epsilon}^{4} - 3(\phi + \delta)\mu_{U}^{2}\mu_{U}^{2} - 3\phi\mu_{U}^{2}\mu_{U}^{2} \]

\[ = -(\phi + \delta)(\mu_{U}^{4} - 3\mu_{U}^{2}\mu_{U}^{2}) - \phi(\mu_{\epsilon}^{4} - 3\mu_{\epsilon}^{2}\mu_{\epsilon}^{2}) , \]

and

\[ \mu_{W,Y}^{2,2} - 2\mu_{W,Y}^{1,1}\mu_{W,Y}^{1,1} - \mu_{W}^{2}\mu_{Y}^{2} \]

\[ = (\phi + \delta)^{2}\mu_{U}^{4} + \phi^{2}\mu_{U}^{2}\mu_{U}^{2} + \mu_{U}^{2}\mu_{U}^{2} + 4(\phi + \delta)\phi\mu_{U}^{2}\mu_{U}^{2} + (\phi + \delta)^{2}\mu_{U}^{2}\mu_{U}^{2} + \phi^{2}\mu_{U}^{2}\mu_{U}^{2} \]

\[ - 2[(\phi + \delta)\mu_{U}^{2} + \phi\mu_{\epsilon}^{2}][(\phi + \delta)\mu_{U}^{2} + \phi\mu_{\epsilon}^{2}] + (\phi + \delta)^{2}\mu_{U}^{2}\mu_{U}^{2} + \phi^{2}\mu_{U}^{2}\mu_{U}^{2} \]

\[ = (\phi + \delta)^{2}(\mu_{U}^{4} - 3\mu_{U}^{2}\mu_{U}^{2}) + \phi^{2}(\mu_{\epsilon}^{4} - 3\mu_{\epsilon}^{2}\mu_{\epsilon}^{2}) , \]

and it follows that

\[ A \equiv \mu_{W,Y}^{2,1}(3\mu_{W,Y}^{1,1}\mu_{W}^{2} - \mu_{W,Y}^{3,1}) + \mu_{W}^{3}(\mu_{W,Y}^{2,2} - 2\mu_{W,Y}^{1,1}\mu_{W,Y}^{1,1} - \mu_{W}^{2}\mu_{Y}^{2}) \]

\[ = -[(\phi + \delta)\mu_{U}^{3} + \phi\mu_{\epsilon}^{3}][(\phi + \delta)(\mu_{U}^{4} - 3\mu_{U}^{2}\mu_{U}^{2}) + \phi(\mu_{\epsilon}^{4} - 3\mu_{\epsilon}^{2}\mu_{\epsilon}^{2})] \]

\[ + (\mu_{U}^{3} + \mu_{\epsilon}^{3})[(\phi + \delta)^{2}\mu_{U}^{2} - 3\mu_{U}^{2}\mu_{U}^{2}) + \phi^{2}(\mu_{\epsilon}^{4} - 3\mu_{\epsilon}^{2}\mu_{\epsilon}^{2})] \]

\[ = [(\phi + \delta) - \phi][(\phi + \delta)\mu_{U}^{3}(\mu_{U}^{4} - 3\mu_{U}^{2}\mu_{U}^{2}) - \phi\mu_{\epsilon}^{3}(\mu_{\epsilon}^{4} - 3\mu_{\epsilon}^{2}\mu_{\epsilon}^{2})] . \]

Last, for the expression for \( C \), we have that

\[ \mu_{W,Y}^{1,3} - 3\mu_{W,Y}^{1,1}\mu_{Y}^{2} \]

\[ = (\phi + \delta)^{3}\mu_{U}^{4} + 3(\phi + \delta)\mu_{U}^{2}\mu_{U}^{2} + 3(\phi + \delta)\phi\mu_{U}^{2}\mu_{U}^{2} + \phi^{3}\mu_{U}^{2}\mu_{U}^{2} + 3(\phi + \delta)^{2}\phi\mu_{U}^{2}\mu_{U}^{2} + 3\phi\mu_{U}^{2}\mu_{U}^{2} \]

\[ - 3[(\phi + \delta)\mu_{U}^{2} + \phi\mu_{\epsilon}^{2}][(\phi + \delta)^{2}\mu_{U}^{2} + \phi^{2}\mu_{U}^{2} + \mu_{U}^{2}] \]

\[ = (\phi + \delta)^{3}(\mu_{U}^{4} - 3\mu_{U}^{2}\mu_{U}^{2}) + \phi^{3}(\mu_{\epsilon}^{4} - 3\mu_{\epsilon}^{2}\mu_{\epsilon}^{2}) \]
and it follows that

\[
C \equiv \mu_{W,Y}^{2,1} - 3\mu_{W,Y}^{1,1}\mu_{Y}^{2} - \mu_{W,Y}^{1,2}(\mu_{W,Y}^{2,2} - 2\mu_{W,Y}^{1,1}\mu_{W,Y}^{1,1} - \mu_{W}^{2}\mu_{Y}^{2})
\]

\[
= [(\phi + \delta)^3\phi_{W,Y}^3 + \phi\phi_{W,Y}^3][[(\phi + \delta)^3(\mu_{W}^4 - 3\mu_{U}^2\mu_{U}^2) + \phi^3(\mu_{\epsilon}^4 - 3\mu_{\epsilon}^2\mu_{\epsilon}^2)]
\]

\[-[(\phi + \delta)^2\mu_{U}^2 + \phi\phi_{W,Y}^3][[(\phi + \delta)^2(\mu_{U}^4 - 3\mu_{U}^2\mu_{U}^2) + \phi^2(\mu_{\epsilon}^4 - 3\mu_{\epsilon}^2\mu_{\epsilon}^2)]
\]

\[
= [(\phi + \delta) - \phi][\phi(\phi + \delta)][(\phi + \delta)^3(\mu_{U}^4 - 3\mu_{U}^2\mu_{U}^2) - \phi\mu_{U}^2(\mu_{\epsilon}^4 - 3\mu_{\epsilon}^2\mu_{\epsilon}^2)]
\]

\[
\text{Proof of Theorem 3.2: (i) We use the mapping } \theta^* = M(\phi) \text{ in Theorem 3.1 and the expression for the moment } \mu_{W,Y}^{1,3} \text{ to generate a nonlinear equation } N(\phi) = 0 \text{ in } \phi.
\]

Given \(\phi + \delta \neq 0\), substituting

\[
\mu_{U}^4 = \frac{\mu_{W,Y}^{3,1} - \phi\mu_{U}^4 - 3[(\phi + \delta) + \phi]\mu_{U}^2\mu_{U}^2}{(\phi + \delta)}
\]

into the expression for \(\mu_{W,Y}^{1,3}\) in system (3) gives

\[
\mu_{W,Y}^{1,3} = (\phi + \delta)^2\mu_{W,Y}^{3,1} + \phi[(\phi - (\phi + \delta)][\phi - (\phi + \delta)]\mu_{\epsilon}^4
\]

\[+\phi^3\mu_{\epsilon}^4 + 3(\phi + \delta)^2\mu_{U}^2\mu_{U}^2 + 3\phi(\phi + \delta)[\phi + (\phi + \delta)]\mu_{U}^2\mu_{\epsilon}^2 + 3\phi\mu_{\epsilon}^2\mu_{\epsilon}^2.
\]

Since

\[
\phi^3 - \phi(\phi + \delta)^2 = \phi[\phi^2 - (\phi + \delta)^2] = \phi[\phi - (\phi + \delta)]\phi + (\phi + \delta)
\]

and

\[-3(\phi + \delta)^2[(\phi + \delta) + \phi] + 3\phi(\phi + \delta)[\phi + (\phi + \delta)] = 3(\phi + \delta)[(\phi + \delta) + \phi][\phi - (\phi + \delta)],
\]

we rewrite the expression for \(\mu_{W,Y}^{1,3}\) as

\[
\mu_{W,Y}^{1,3} = (\phi + \delta)^2\mu_{W,Y}^{3,1} + \phi[\phi - (\phi + \delta)]\mu_{\epsilon}^4
\]

\[+3(\phi + \delta)[\phi + (\phi + \delta)][\phi - (\phi + \delta)]\mu_{U}^2\mu_{\epsilon}^2 + 3(\phi + \delta)\mu_{U}^2\mu_{\epsilon}^2 + 3\phi\mu_{\epsilon}^2\mu_{\epsilon}^2.
\]

Given \(\phi \neq 0\) and \(\delta \neq 0\), substituting further for

\[
\mu_{\epsilon}^4 = \frac{1}{\phi[\phi - (\phi + \delta)]} \times \left\{\mu_{W,Y}^{2,2} - (\phi + \delta)\mu_{W,Y}^{3,1}
\right. \]

\[\left. - [\phi - (\phi + \delta)]\phi + 2(\phi + \delta)]\mu_{U}^2\mu_{\epsilon}^2 - \mu_{U}^2\mu_{\epsilon}^2 - \mu_{\epsilon}^2\mu_{\epsilon}^2 \right\}
\]
in the $\mu_{W,Y}^{1,3}$ equation gives

$$
\mu_{W,Y}^{1,3} = (\phi + \delta)^2 \mu_{W,Y}^{3,1} + [\phi + (\phi + \delta)]
\times \{ \mu_{W,Y}^{2,2} - (\phi + \delta)\mu_{W,Y}^{3,1} + [\phi - (\phi + \delta)]\mu_{W,Y}^{2,2} - \mu_{W,Y}^{3,2} \}
+ 3(\phi + \delta)[\phi + (\phi + \delta)]\mu_{W,Y}^{2,2} + 3(\phi + \delta)^2 \mu_{W,Y}^{2,2} + 3\phi \mu_{W,Y}^{2,2}
= -\phi(\phi + \delta)\mu_{W,Y}^{3,1} + [\phi + (\phi + \delta)]\mu_{W,Y}^{2,2} - [\phi + (\phi + \delta)]\mu_{W,Y}^{2,2} - [\phi + 2(\phi + \delta)]
+ [2(\phi + \delta) - \phi] \mu_{W,Y}^{3,2} + [2\phi - (\phi + \delta)] \mu_{W,Y}^{2,2},
$$

where we make use of

$$
+ 3(\phi + \delta)[\phi + (\phi + \delta)][\phi - (\phi + \delta)] - [\phi + (\phi + \delta)][\phi - (\phi + \delta)][\phi + 2(\phi + \delta)]
= -[\phi + (\phi + \delta)][\phi - (\phi + \delta)]^2.
$$

We arrive therefore at the following nonlinear equation in $\phi$:

$$
\mu_{W,Y}^{1,3} = -\phi T(\phi) \mu_{W,Y}^{3,1} + [\phi + T(\phi)]\mu_{W,Y}^{2,2} - [\phi + T(\phi)][\phi - T(\phi)] \mu_{W,Y}^{2,2} V_U(\phi) V_\varepsilon(\phi) \quad (9)
+ [2T(\phi) - \phi] V_U(\phi) V_\eta(\phi) + [2\phi - T(\phi)] V_\varepsilon(\phi) V_\eta(\phi).
$$

Using the reduced-form expressions for $T(\phi)$, $V_U(\phi)$, $V_\varepsilon(\phi)$ and $V_\eta(\phi)$ in Lemma A.1, we rewrite the terms in equation (9). Since $V_\varepsilon(\phi) = \mu_{W,Y}^{3,1} - V_U(\phi)$, we have

$$
[2T(\phi) - \phi] V_U(\phi) + [2\phi - T(\phi)] V_\varepsilon(\phi)
= 3[T(\phi) - \phi] V_U(\phi) + [2\phi - T(\phi)] \mu_{W,Y}^{2,2}
= -3 \frac{\phi (\phi W - \mu_{W,Y}^{2,1}) - (\phi W^{3,1} - \mu_{W,Y}^{1,2})}{\mu_{W,Y}^{3} - \mu_{W,Y}^{2,1}} \frac{\phi (\phi W^{3} - \mu_{W,Y}^{2,1}) (\phi W^{2} - \mu_{W,Y}^{1,1})}{\phi (\phi W^{3} - \mu_{W,Y}^{2,1}) (\phi W^{2} - \mu_{W,Y}^{1,1})}
+ (2\phi - \frac{\phi W^{2,1} - \mu_{W,Y}^{1,2}}{\phi W^{3} - \mu_{W,Y}^{2,1}}) \mu_{W,Y}^{2,2}
= -3(\phi W^{2,1} - \mu_{W,Y}^{1,1}) + (2\phi - \frac{\phi W^{2,1} - \mu_{W,Y}^{1,2}}{\phi W^{3} - \mu_{W,Y}^{2,1}}) \mu_{W,Y}^{2,2}.
$$

Further, we have that

$$
[\phi - T(\phi)] V_U(\phi) = \frac{\phi (\phi W^{3} - \mu_{W,Y}^{2,1}) - (\phi W^{3,1} - \mu_{W,Y}^{1,2})}{\phi W^{3} - \mu_{W,Y}^{2,1}} \frac{\phi (\phi W^{3} - \mu_{W,Y}^{2,1}) (\phi W^{2} - \mu_{W,Y}^{1,1})}{\phi (\phi W^{3} - \mu_{W,Y}^{2,1}) (\phi W^{2} - \mu_{W,Y}^{1,1})}
= \phi W^{2} - \mu_{W,Y}^{1,1}.
$$
and since $V_\epsilon(\phi) = \mu^2_W - V_U(\phi)$, we obtain

$$[\phi - T(\phi)]V_\epsilon(\phi) = [\phi - T(\phi)]\mu^2_W - \phi\mu^2_W + \mu^1_{W,Y} = \mu^1_{W,Y} - \frac{\phi\mu^2_{W,Y} - \mu^1_{W,Y}}{\phi\mu^3_W - \mu^2_{W,Y}} \mu^2_W$$

so that

$$[\phi + T(\phi)][\phi - T(\phi)]^2 V_U(\phi)V_\epsilon(\phi) = (\phi + \frac{\phi\mu^2_{W,Y} - \mu^1_{W,Y}}{\phi\mu^3_W - \mu^2_{W,Y}})(\phi\mu^2_W - \mu^1_{W,Y})(\mu^1_{W,Y} - \frac{\phi\mu^2_{W,Y} - \mu^1_{W,Y}}{\phi\mu^3_W - \mu^2_{W,Y}} \mu^2_W).$$

Thus, we can express equation (9) as follows:

$$-\mu^1_{W,Y} - \phi T(\phi)\mu^3_{W,Y} + [\phi + T(\phi)]^2 \mu^2_{W,Y} - [\phi + T(\phi)][\phi - T(\phi)]^2 V_U(\phi)V_\epsilon(\phi) + [2T(\phi) - \phi]V_U(\phi)V_\eta(\phi) + [2\phi - T(\phi)]V_\epsilon(\phi)V_\eta(\phi)$$

$$= -\mu^1_{W,Y} - \phi \frac{\phi\mu^2_{W,Y} - \mu^1_{W,Y}}{\phi\mu^3_W - \mu^2_{W,Y}} \mu^3_{W,Y} + (\phi + \frac{\phi\mu^2_{W,Y} - \mu^1_{W,Y}}{\phi\mu^3_W - \mu^2_{W,Y}})\mu^2_{W,Y}$$

$$- (\phi + \frac{\phi\mu^2_{W,Y} - \mu^1_{W,Y}}{\phi\mu^3_W - \mu^2_{W,Y}})(\phi\mu^2_W - \mu^1_{W,Y})(\mu^1_{W,Y} - \frac{\phi\mu^2_{W,Y} - \mu^1_{W,Y}}{\phi\mu^3_W - \mu^2_{W,Y}} \mu^2_W)$$

$$+ [-3(\phi\mu^2_W - \mu^1_{W,Y}) + (2\phi - \frac{\phi\mu^2_{W,Y} - \mu^1_{W,Y}}{\phi\mu^3_W - \mu^2_{W,Y}})\mu^2_W]$$

$$\times [\mu^2_Y + (\phi\mu^2_W - \mu^1_{W,Y})(\phi + \frac{\phi\mu^2_{W,Y} - \mu^1_{W,Y}}{\phi\mu^3_W - \mu^2_{W,Y}}) - \phi^2 \mu^2_W]$$

$$= -\mu^1_{W,Y} - \phi \frac{\phi\mu^2_{W,Y} - \mu^1_{W,Y}}{\phi\mu^3_W - \mu^2_{W,Y}} \mu^3_{W,Y}$$

$$+ (\phi + \frac{\phi\mu^2_{W,Y} - \mu^1_{W,Y}}{\phi\mu^3_W - \mu^2_{W,Y}})\mu^2_{W,Y} + (\phi\mu^2_W - \mu^1_{W,Y})(-\phi^2 + 2\mu^1_{W,Y})]$$

$$+ [-3(\phi\mu^2_W - \mu^1_{W,Y}) + (2\phi - \frac{\phi\mu^2_{W,Y} - \mu^1_{W,Y}}{\phi\mu^3_W - \mu^2_{W,Y}})\mu^2_W](\mu^2_Y - \phi^2 \mu^2_W) = 0.$$

Multiplying by $\phi\mu^3_W - \mu^2_{W,Y} = -\delta u^3 \neq 0$ gives

$$-\mu^1_{W,Y}(\phi\mu^3_W - \mu^2_{W,Y}) - \phi(\phi\mu^2_{W,Y} - \mu^1_{W,Y})\mu^3_{W,Y}$$

$$+ [\phi(\phi\mu^3_W - \mu^2_{W,Y}) + (\phi\mu^2_{W,Y} - \mu^1_{W,Y})]\mu^2_{W,Y} + (\phi\mu^2_W - \mu^1_{W,Y})(-\phi^2 + 2\mu^1_{W,Y})]$$

$$+ [-3(\phi\mu^2_W - \mu^1_{W,Y})(\phi\mu^3_W - \mu^2_{W,Y}) + [2\phi(\phi\mu^3_W - \mu^2_{W,Y}) - (\phi\mu^2_{W,Y} - \mu^1_{W,Y})]\mu^2_W](\mu^2_Y - \phi^2 \mu^2_W) = 0.$$
Note that
\[ \phi(\phi \mu_W^3 - \mu_{W,Y}^2) + (\phi \mu_{W,Y}^2 - \mu_{W,Y}^{1,2}) = \phi^2 \mu_W^3 - \mu_{W,Y}^{1,2} \]
and
\[
-3(\phi \mu_W^2 - \mu_{W,Y}^{1,1})(\phi \mu_W^3 - \mu_{W,Y}^2) + [2\phi(\phi \mu_W^3 - \mu_{W,Y}^2) - (\phi \mu_{W,Y}^2 - \mu_{W,Y}^{1,2})]\mu_W^2 \\
= \phi^2 \mu_W^3 + 3\phi^2 \mu_{W,Y}^3 + 3\mu_{W,Y}^{1,1}\phi \mu_W^3 - 3\mu_{W,Y}^{1,1}\mu_W^2 + (2\phi^2 \mu_W^3 - 3\phi \mu_{W,Y}^2 + \mu_{W,Y}^{1,2})\mu_W^2 \\
= -\phi^2 \mu_W^{2,3} + 3\phi \mu_{W,Y}^{1,1}\mu_W^3 - 3\mu_{W,Y}^{1,1}\mu_W^2 + \mu_{W,Y}^{1,2}\mu_W^2 \\
= -\mu_W(\phi^2 \mu_W^3 - \mu_{W,Y}^{1,2}) + 3\mu_{W,Y}^{1,1}(\phi \mu_W - \mu_{W,Y}^2)
\]
We can then rewrite equation (9) as
\[
-\mu_{W,Y}^{1,3}(\phi \mu_W^3 - \mu_{W,Y}^2) - \phi(\phi \mu_{W,Y}^2 - \mu_{W,Y}^{1,2})\mu_W^{3,1} \\
+(\phi^2 \mu_W^3 - \mu_{W,Y}^{1,2})[\mu_{W,Y}^{2,2} + (\phi \mu_{W,Y}^2 - \mu_{W,Y}^{1,1})(-\phi \mu_W^2 + 2\mu_{W,Y}^{1,1})] \\
+[-\mu_W(\phi^2 \mu_W^3 - \mu_{W,Y}^{1,2}) + 3\mu_{W,Y}^{1,1}(\phi \mu_W - \mu_{W,Y}^2)](\mu_W^2 - \phi^2 \mu_W^2) \\
= (\phi \mu_W^3 - \mu_{W,Y}^{2,1})[-\mu_{W,Y}^{1,3} + 3\mu_{W,Y}^{1,1}(\mu_W^2 - \phi^2 \mu_W^2)] - \phi(\phi \mu_{W,Y}^2 - \mu_{W,Y}^{1,2})\mu_W^{3,1} \\
+(\phi^2 \mu_W^3 - \mu_{W,Y}^{1,2})[\mu_{W,Y}^{2,2} + (\phi \mu_{W,Y}^2 - \mu_{W,Y}^{1,1})(-\phi \mu_W^2 + 2\mu_{W,Y}^{1,1}) - \mu_W^2(\mu_W^2 - \phi^2 \mu_W^2)] \\
= (\phi \mu_W^3 - \mu_{W,Y}^{2,1})[-\mu_{W,Y}^{1,3} + 3\mu_{W,Y}^{1,1}(\mu_W^2 - \phi^2 \mu_W^2)] - \phi(\phi \mu_{W,Y}^2 - \mu_{W,Y}^{1,2})\mu_W^{3,1} \\
+(\phi^2 \mu_W^3 - \mu_{W,Y}^{1,2})[\mu_{W,Y}^{2,2} + 3\phi \mu_{W,Y}^2 \mu_{W,Y}^{1,1} - 2\mu_{W,Y}^{1,1}\mu_W^2 - \mu_W^{2,2}] \\
= \phi^2 [3\mu_{W,Y}^{1,1}\mu_{W,Y}^2 - \mu_{W,Y}^{1,1}\mu_{W,Y}^{2,1} + \mu_{W,Y}^3(\mu_W^2 - 2\mu_{W,Y}^{1,1}\mu_W^2 - \mu_{W,Y}^{2,2})] \\
+\phi(-\mu_{W,Y}^{1,3} + 3\mu_{W,Y}^{1,1}\mu_W^2 + \mu_{W,Y}^{1,2}\mu_W^{3,1} - 3\mu_{W,Y}^{1,2}\mu_{W,Y}^{2,1}) \\
+\mu_{W,Y}^{1,3}\mu_W^2 - 3\mu_{W,Y}^{1,1}\mu_W^2 - \mu_{W,Y}^{1,1}\mu_W^{2,2} = 0.
\]
This is a quadratic equation of the form
\[
A\phi^2 + B\phi + C = \phi^2 \mu_W^3 - \mu_{W,Y}^{1,2} = 0.
\]
The discriminant of the equation is \( \Delta \equiv B^2 - 4AC \). Next, we show that \( \Delta \geq 0 \). From
that D unique root f variables for any given solution, there exits a second solution that interchanges the role of the system (3). Note that system (3) is symmetric in (V
\Delta
≡ the expressions for A, B, and C in Lemma A.2, we obtain

\[ \Delta \equiv B^2 - 4AC \]

\[ = [(\phi + \delta)^2 - \phi^2][-(\phi + \delta)(\mu_U^4 - 3\mu_U^2\mu_V^2)\mu_\varepsilon^3 + \phi(\mu_\varepsilon^4 - 3\mu_\varepsilon^2\mu_\varepsilon^2)\mu_U^3]^2 \]

\[ -4[(\phi + \delta) - \phi^2][\phi(\phi + \delta)][(\phi + \delta)\mu_\varepsilon^3(\mu_U^4 - 3\mu_U^2\mu_V^2) - \phi(\mu_\varepsilon^4 - 3\mu_\varepsilon^2\mu_\varepsilon^2)\mu_U^3]^2 \]

\[ = \{[(\phi + \delta)^2 - \phi^2]^2 - 4[(\phi + \delta) - \phi^2][\phi(\phi + \delta)][(\phi + \delta)(\mu_U^4 - 3\mu_U^2\mu_V^2)\mu_\varepsilon^3 - \phi(\mu_\varepsilon^4 - 3\mu_\varepsilon^2\mu_\varepsilon^2)\mu_U^3]^2 \]

\[ = [(\phi + \delta) - \phi]^4[(\phi + \delta)(\mu_U^4 - 3\mu_U^2\mu_V^2)\mu_\varepsilon^3 - \phi(\mu_\varepsilon^4 - 3\mu_\varepsilon^2\mu_\varepsilon^2)\mu_U^3]^2 \geq 0. \]

Given A \neq 0, the roots of the quadratic equation A\phi^2 + B\phi + C = 0 are given by

\[ f^- = \frac{-B - \sqrt{\Delta}}{2A} \quad \text{and} \quad f^+ = \frac{-B + \sqrt{\Delta}}{2A}. \]

It follows from Theorem 3.1 that \( \theta^* \) admits two roots given by \( M(f^-) \) and \( M(f^+) \).

(ii) Let \( \mu_U \) and \( \mu_\varepsilon \) collect respectively the moments for \( U \) and \( \varepsilon \) that appear in system (3). Note that system (3) is symmetric in (\( \phi, \mu_\varepsilon \)) and (\( \phi + \delta, \mu_U \)). In particular, for any given solution, there exits a second solution that interchanges the role of the variables \( \varepsilon \) and \( U \), as well as their associated coefficients \( \phi \) and \( \phi + \delta \). In particular, \( f^- - f^+ \) must correspond to either \( \phi - (\phi + \delta) = -\delta \) or \( (\phi + \delta) - \phi = \delta \). It follows that \( D(f^-) = -D(f^+) \) and, therefore, knowing the sign of \( \delta \) would identify the unique root \( f^* \in \{f^-, f^+\} \). Similarly, we have that \( V_U(f^-) = V_\varepsilon(f^+) \) and \( V_U(f^+) = V_\varepsilon(f^-) \). Therefore, knowing the sign of \( \mu_U^2 - \mu_\varepsilon^2 \) would identify the unique root \( f^* \).

An analogous comment applies to the signs of \( \mu_U^3 - \mu_\varepsilon^3 \) and \( \mu_U^4 - \mu_\varepsilon^4 \). \( \blacksquare \)