Measurement Error in Multiple Equations: Tobin’s q and Corporate Investment, Saving, and Debt

Karim Chalak†
University of Virginia

Daniel Kim
BI Norwegian Business School

July 25, 2019

Abstract

We characterize the sharp identification regions for the coefficients in a system of linear equations that share an explanatory variable measured with classical error. We demonstrate the identification gain from analyzing the equations jointly. We derive the sharp identification regions under any configuration of three auxiliary assumptions. These restrict the “noise-to-signal” ratio, the coefficients of determination, and the signs of the correlations among the cross-equation disturbances. For inference, we implement results on intersection bounds. The application studies the effects of cash flow on the investment, saving, and debt of firms when Tobin’s q serves as a proxy for marginal q.

Keywords: cash flow, measurement error, multiple equations, partial identification, sensitivity analysis, Tobin’s q.

JEL codes: C31, G30.

1 Introduction

Sometimes a mismeasured explanatory variable appears in multiple linear equations of interest which are nonetheless estimated separately. This paper studies the identification gain that results from analyzing the system’s equations jointly when this common explanatory

*Karim Chalak (corresponding author), Department of Economics, University of Virginia, chalak@virginia.edu. Daniel Kim, BI Norwegian Business School, daniel.s.kim@bi.no.

†Acknowledgments: Karim Chalak acknowledges the support from the Bankard Fund for Political Economy. We thank the participants in the 2018 North American Summer Meeting of the Econometric Society, 93rd Annual Conference of the Western Economic Association International, 88th Annual Meeting of the Southern Economic Association, 2018 Triangle Econometrics Conference, New York Camp Econometrics XIV, Warwick Econometrics Workshop, 2019 Africa Meeting of the Econometric Society, and the seminars at Boston College, University of Virginia, and the Wharton School of the University of Pennsylvania, as well as Brendan Kline and Luke Taylor for helpful feedback. Any errors are the responsibility of the authors.
variable suffers from classical measurement error. In addition, the paper puts forward useful auxiliary assumptions that can help identify the system’s coefficients and it studies how jointly sensitive are the system’s coefficients to deviations from these assumptions. To develop these results, we show how the identification region for the coefficients in each equation depends on the extent of the measurement error in the proxy for the common latent variable. When analyzing each equation separately, researchers might forgo information about the accuracy of the proxy that obtains when using the other equations. Further, they may reach incoherent conclusions that implicitly rest on different inference, derived from each equation separately, on the extent of the measurement error in the proxy. In contrast, we demonstrate how analyzing the system of equations jointly can yield tighter sharp identification regions for the system’s coefficients than the single equation analysis. Further, by analyzing the system of equations jointly, the paper’s framework guides the researcher toward employing useful but also compatible identifying assumptions and it facilitates testing hypotheses that involve coefficients from multiple equations.

Specifically, we consider a system of linear equations that share a mismeasured explanatory variable. Building on partial identification results in the presence of measurement error in e.g. Klepper and Leamer (1984), Leamer (1987), Bollinger (2003), and Chalak and Kim (2019), we characterize the sharp identification regions for the coefficients on the latent variable and the (correctly measured) covariates under the classical measurement error assumption and demonstrate the identification gain that results from analyzing the equations jointly as opposed to separately. Roughly speaking, this is akin to studying the efficiency gain that results from jointly estimating seemingly unrelated regressions (e.g. Zellner, 1962). Further, to tighten the bounds and conduct a sensitivity analysis, we derive the sharp identification regions under any configuration of the following three auxiliary assumptions. As we show, each of these assumptions weakens a stronger benchmark assumption that point identifies the system’s coefficients. The first auxiliary assumption weakens the assumption of “no measurement error” by imposing an upper bound on the “noise to signal” ratio (i.e. the ratio of the variance of the measurement error to that of the latent variable net-of-the covariates). The second controls the fit of the model by imposing upper bounds on the coefficients of determination that would obtain in each equation had there been no measurement error. The third weakens the assumption that the variance matrix of the equation disturbances is diagonal by specifying the signs of the correlations among the cross-equation disturbances, if
at all. We do not require a particular configuration of these auxiliary assumptions. Instead, we characterize the mapping from each configuration of the assumptions to the identification regions of the coefficients. We then conduct a sensitivity analysis that studies the consequences of deviating from the benchmark point-identifying assumptions. To facilitate inference, we express the identification regions for the coefficients in terms of intersection bounds. We then combine and implement results from Chernozhukov, Rigobon, and Stoker (2010) and Chernozhukov, Lee, and Rosen (2013). The resulting framework delivers a specification test for the imposed assumptions and enables inference under sequentially stronger identifying assumptions, whereby a researcher can gain confidence in results that hold true under weaker assumptions.

To illustrate our framework, we study the effects of a firm’s cash flow (internal funds) on its investment, saving, and debt. After accounting for the firm’s marginal $q$ (the firm’s expected marginal return of capital), various theories offer contradictory predictions about the sign of the effect of cash flow on each of these outcomes. Because researchers do not directly observe marginal $q$, it is common to use the observed Tobin’s $q$ (the ratio of the firm’s market value to its assets’ replacement value, measured by e.g. the “market-to-book” ratio) as an error-laden proxy for marginal $q$. To proceed, the literature employs various econometric methods that impose different assumptions on the measurement error in Tobin’s $q$. These methods yield mixed empirical conclusions, sometimes corroborating contradictory theoretical predictions, about the direction of the effects of cash flow on investment (e.g. Erickson and Whited (2000, 2012) and Almeida, Campello, and Galvao (2010)), saving (e.g. Almeida, Campello, and Weisbach (2004) and Riddick and Whited (2009)), and debt (e.g. Rajan and Zingales (1995) and Erickson, Jiang and Whited (2014)). Importantly, the literature estimates each of the investment, saving, and debt equations separately. Using data from Compustat, we apply our framework to study the joint effects of cash flow on the investment, saving, and debt of corporate firms in the US when Tobin’s $q$ serves as an error-laden proxy for a firm’s marginal $q$. Analyzing the equations jointly, as opposed to separately, tightens the identification regions considerably and sometimes determines the sign of the effects of cash flow without imposing stronger assumptions. In particular, the joint effects of cash flow on investment, saving, and debt can be zero if and only if Tobin’s $q$ is possibly a noisy proxy for marginal $q$, with a low reliability ratio. Otherwise, if Tobin’s $q$ is a moderately accurate proxy then, consistent with the signs of the linear regression
estimates, cash flow affects investment and saving positively and debt negatively.

More broadly, this paper’s framework can be useful in any context in which an error-laden proxy for a latent variable appears in multiple equations. For example, individual latent “ability” may affect multiple labor market outcomes, such as wage and hours worked, and is often proxied by a test score, such as IQ. Similarly, a medical test score may serve as a proxy for a latent health status that may affect multiple aspects of a patient’s behavior.

The paper is organized as follows. Section 2 introduces the data generating assumptions and notation. Section 3 derives the sharp identification regions under the classical measurement error assumption and any configuration of the auxiliary assumptions. Section 4 illustrates the identification results using a numerical example. Section 5 describes the estimation and inference procedure. Section 6 applies the paper’s framework to study the effects of cash flow on corporate behavior. Section 7 concludes. Supplementary material and mathematical proofs are gathered in the Online Appendix.

2 Data Generation and Assumptions

We assume that the data satisfies the following assumptions.

**Assumption A**

**A1 Linearity:** (i) Let \( (X', W, Y')' \) be a random vector with a finite variance. (ii) Let the random variables \( \eta, \varepsilon, U, X, W, \) and \( Y \) satisfy

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

(1)

with constant slope coefficients. The researcher observes realizations of \((X', W, Y')'\) but not \((U, \eta', \varepsilon)\).

**A1** specifies a linear equation for \( Y \) but does not restrict \( \eta \). Next, we maintain that the “disturbance” \( \eta \) is uncorrelated with \((X', U)\).

**Assumption A2 Uncorrelated Disturbance:** \( \text{Cov}[\eta, (X', U)'] = 0 \).

**A1** and **A2** may be viewed as data generating assumptions. Alternatively, the linear restriction on the \( Y \) equation in **A1** and the covariance restriction in **A2** hold by construction.

---

1The Online Appendix contains supplementary material on identification (Section A), inference (Section B), panel data (Section C), and the empirical application (Figure 4 and Table 7) as well as the mathematical proofs (Section D).
for the best linear prediction of \( Y \) given \((X, U)\), with \((\delta', \beta')'\) coinciding with this linear regression’s estimand. \( A_1 \) denotes the latent variable by \( U \) and the observed covariates that drive \( Y \) by \( X \). Further, \( A_1 \) decomposes the proxy \( W \) into the “signal” component\(^3\) \( U \) and the “noise” or measurement error \( \varepsilon \). \( A_3 \) assumes that \( \varepsilon \) is uncorrelated with \((X', U, \eta)\). Note that \( Cov(\varepsilon, U) = 0 \) holds by construction when \( U \) is the best linear prediction of \( W \).

**Assumption A\(_3\)** Uncorrelated measurement error: \( Cov[\varepsilon, (X', U, \eta')'] = 0 \).

We are interested in identifying the effects \( \delta_j \) and \( \beta_j \) of \( U \) and \( X \) on \( Y_j \) for \( j = 1, ..., p \) as encoded in the \( j \)th outcome equation,

\[
Y_j = X'\beta_j + U\delta_j + \eta_j. \tag{2}
\]

Or one may be interested in the best linear prediction of \( Y \) given \((X, U)\) when \( W \) measures \( U \) with classical error. The challenge in identifying \((\delta, \beta)\) is due to \( U \) being unobserved and possibly correlated with \( X \). Our framework does not require the presence of covariates, so \( X \) may be empty. When present, we allow \( X \) to enter all the \( Y_j \) equations, as can often occur in systems where multiple outcomes are determined jointly. Given \( Cov[(\eta', \varepsilon)', X] = 0 \), excluding an element \( X_l \) from a \( Y_j \) equation by setting \( \beta_{jl} = 0 \) can point identify the system coefficients since the excluded variable can serve as an instrumental variable (see the discussion following Theorem 3.1). We do not require such exclusion restrictions.

Assumptions \( A_1-A_3 \) are the classical error-in-variables assumptions (see e.g. Wooldridge, 2002, p. 80). We briefly comment on certain related papers that either weaken or strengthen \( A_1-A_3 \). In the case of a single equation with \( p = 1 \), Lewbel (1997) and Erickson and Whited (2002) strengthen \( A_1-A_3 \) by imposing additional restrictions on the higher order moments of \((\eta, \varepsilon, U, X')\) that may point identify\(^4\) \((\beta, \delta)\). We do not require these stronger assumptions\(^4\).

Instead, we impose the uncorrelation assumptions \( A_2-A_3 \) and study partially identifying \( \delta \) and \( \beta \). DiTraglia and Garcia-Jimeno (2017) relax \( A_2 \) to allow \( X \) (or its instrument) to be endogenous and, similarly to this paper’s joint equation analysis, they advocate analyzing

\(^2\)The structure \( Y' = X'\beta + V\gamma + \eta' \) and \( W = V\psi + \varepsilon \), with \( V \) unobserved, is observationally equivalent to \( A_1 \). Provided the scale \( \psi \neq 0 \), only the ratio \( \delta = \frac{\gamma}{\psi} \) of the coefficients on \( V \) may be (partially) identified. To ease the notation, we use the simpler representation in which \( U \equiv V\psi \).

\(^3\)Note that if \( X = (X_1', X_2')' \) and one further requires \( E[(\eta', \varepsilon')' | X_1] = E[(\eta', \varepsilon')'] \) then it may be possible to point identify \((\beta', \delta')' \) in \( Y' = X'\beta + W\delta + \eta' - \varepsilon\delta \) by generating an instrument for \( W \) as a function of \( X_1 \) that is excluded from \( X_2 \).

\(^4\)For instance, unlike in Erickson and Whited (2002), \( A_1-A_3 \) allow the system variables to be jointly normally distributed.
jointly the assumptions imposed on instrument exogeneity and measurement error. Krasker and Pratt (1986) and Erickson and Whited (2005) relax $A_3$ and study how highly correlated should $W$ and $U$ be in order to identify the sign of $\delta$ or of a component of $\beta$. Klepper and Leamer (1984) and Bollinger (2003) characterize the sharp identification regions for $\delta$ and $\beta$ under $A_1$-$A_3$. Chalak and Kim (2019) extend these results when $U$ is a scalar to relax the proxy exclusion restriction in $A_1$ by allowing $W$ to affect $Y$ directly. Whereas the papers discussed above consider a scalar outcome with $p = 1$, Leamer (1987) studies the identification of the coefficients under $A_1$-$A_3$ when $X$ is empty and $Y$ and $U$ are vectors of arbitrary dimensions. We build on these papers and study the identification gain that results from imposing the auxiliary assumptions $A_4$-$A_6$ discussed below. For concreteness and to gain analytical tractability, we focus on the case where $U$ and $W$ are scalars and $Y$ is a $p \times 1$ vector, as we maintain in the empirical application when studying the firm investment, saving, and debt equations. This enables us to operate in a simpler context and to demonstrate how this type of sensitivity analysis can be usefully implemented in empirical work.

2.1 Notation

To shorten the notation, for generic random vectors $A$ and $B$, we write:

$$\sigma^2_A \equiv \text{Var}(A) \quad \text{and} \quad \sigma_{A,B} \equiv \text{Cov}(A, B).$$

When $A$ and $B$ are nondegenerate scalars, $r_{A,B} \equiv \frac{\sigma_{A,B}}{\sigma_A \sigma_B}$ denotes the correlation between $A$ and $B$. Further, when $\sigma_{C,B}$ is square and nonsingular, we use the following succinct notation for the linear instrumental variable (IV) regression estimand and residual

$$b_{A,B|C} \equiv \sigma^{-1}_{C,B} \sigma_{C,A} \quad \text{and} \quad \epsilon'_{A,B|C} \equiv [A - E(A)]' - [B - E(B)]' b_{A,B|C}$$

so that by construction $E(\epsilon_{A,B|C}) = 0$ and $\text{Cov}(C, \epsilon_{A,B|C}) = 0$. In particular, $b_{A,B|C}$ is the vector of slope coefficients associated with $B$ in a linear IV regression of $A$ on $(1, B')'$ using instruments $(1, C')'$. If $B = C$, we obtain the linear regression estimand and residual $b_{A,B} \equiv b_{A,B|B}$ and $\epsilon_{A,B} \equiv \epsilon_{A,B|B}$. Last, for a scalar $A$, we denote by

$$R^2_{A,B} \equiv \sigma^2_A (\sigma_{A,B} \sigma_B^{-2} \sigma_{B,A}) \equiv b_{B,A} b_{A,B}$$

the population coefficient of determination (R-squared) from a regression of $A$ on $B$.

\footnote{If $\sigma^2_B$ is singular, we set $R^2_{A,B} = R^2_{A,B_{o}}$ where $B_{o}$ is a maximal linearly independent subset of $B$. Further, if either $\sigma^2_A = 0$ or $\sigma^2_B = 0$ then we set $r_{A,B} = 0$ and $R^2_{A,B} = 0$.}
2.2 Linear Projection

Using \( \tilde{A} \equiv \epsilon_{A,X} \) as a shorthand notation for the residual from the regression of a vector \( A \)
on \( X \) (with \( \sigma_X^2 \) nonsingular), we employ the following system of projected linear equations

\[
\tilde{Y}' = \tilde{U}\delta + \tilde{\eta}' \quad \text{and} \quad \tilde{W} = \tilde{U} + \tilde{\varepsilon}
\] (3)

to study identifying \( \delta \). The identification region for \( \beta \) then obtains using equation (4) since, by \( A_2-A_3 \), \( \text{Cov}[(\eta, \varepsilon)', X] = 0 \) and projecting \( W \) and \( Y \) onto \( X \) gives \( b_{W,X} = b_{U,X} \) and

\[
b_{Y,X} = \beta + b_{W,X}\delta.
\] (4)

2.3 Auxiliary Assumptions

To tighten the identification regions obtained under \( A_1-A_3 \) and conduct a sensitivity analysis, we consider the auxiliary assumptions \( A_4-A_6 \) that weaken three benchmark assumptions. We emphasize that we do not require \( A_4-A_6 \). Instead, we characterize the identification gain that results from imposing any configuration of these auxiliary assumptions.

The first auxiliary assumption weakens the “no measurement error” assumption \( \sigma_\varepsilon^2 = 0 \) by imposing an upper bound \( \kappa \) on the net-of-X “noise to signal ratio.”

**Assumption A_4**  *Bounded Net-of-X Noise to Signal Ratio:* \( \sigma_\varepsilon^2 \leq \kappa \sigma_U^2 \) where \( 0 \leq \kappa \).

By varying \( \kappa \), \( A_4 \) enables studying how the extent of the measurement error affects the results. For example, \( A_4 \) reduces to the “no measurement error” assumption \( \sigma_\varepsilon^2 = 0 \) when \( \kappa = 0 \) whereas setting \( \kappa = 1 \) assumes that the variance of the measurement error is at most as large as the variance of \( \tilde{U} \), \( \sigma_\varepsilon^2 \leq \sigma_U^2 \). Given \( A_1-A_3 \), \( A_4 \) equivalently imposes a lower bound \( \frac{1}{1+\kappa} \) on \( \rho \), the net-of-X “signal to total variance ratio”:

\[
\frac{1}{1+\kappa} \leq \rho \equiv \frac{\sigma_U^2}{\sigma_W^2} = \frac{\sigma_U^2}{\sigma_U^2 + \sigma_\varepsilon^2}.
\]

Further, since \( \rho \equiv \frac{\sigma_U^2}{\sigma_W^2} = \frac{R_{W,U}^2}{1-R_{W,X}^2} \) (e.g. DiTraglia and Garcia-Jimeno, 2017, eq. (20)), \( A_4 \) equivalently sets a lower bound \( \kappa^* \equiv \frac{1+\kappa R_{W,X}^2}{1+\kappa} \) on the “reliability ratio” \( R_{W,U}^2 \), so that \( R_{W,X}^2 \leq \kappa^* \leq R_{W,U}^2 \). One may resort to any of these equivalent interpretations of \( A_4 \).

Consider the coefficient of determination \( R_{Y_U}^2 \equiv 1 - \frac{\sigma_\varepsilon^2}{\sigma_{Y_U}^2} \) in the \( \tilde{Y}_j \) equation in (3). By \( A_1-A_3 \), Lemma \( D.1 \) in the Online Appendix gives that \( R_{Y_{j,W}}^2 \leq R_{Y_{j,U}}^2 \). The second auxiliary assumption controls the fit of the model by imposing a bound \( \tau_j \) on how large \( R_{Y_{j,U}}^2 \) can be.
Assumption A₅ Bounded Net-of-X Coefficient of Determination: \( R_{Y_j,\tilde{U}}^2 \leq \tau_j \) where \( 0 < \tau_j \) and \( R_{Y_j,\tilde{W}}^2 \leq \tau_j \leq 1 \) for \( j = 1, \ldots, p \).

Since \( R_{A,(X',B)'}^2 = \frac{\sigma_{\eta}^2}{\sigma_{\eta}^2}(R_{A,B}^2 - 1) + 1 \), A₅ equivalently imposes an upper bound \( \tau_j^* \equiv \frac{\sigma_{Y_j}^2}{\sigma_{Y_j}^2}(\tau_j - 1) + 1 \) on the coefficient of determination \( R_{Y_j,(X',U)'}^2 \equiv 1 - \frac{\sigma_{\eta}^2}{\sigma_{\eta}^2} \) in the \( Y_j \) equation (2). We let \( \tau \equiv (\tau_1, \ldots, \tau_p)' \) and \( \tau^* \equiv (\tau_1^*, \ldots, \tau_p^*)' \).

Klepper and Leamer (1984), Klepper (1988), and Chalak and Kim (2019) employ assumptions similar to A₄ and A₅ when \( p = 1 \). Since we consider multiple equations, \( p \geq 1 \), we introduce a third auxiliary assumption A₆ that weakens the assumption that \( \sigma_{\eta}^2 \) is diagonal by specifying the sign of the correlation \( r_{\eta_j,\eta_h} \) among the cross-equation disturbances, if at all.

Assumption A₆ Disturbance Correlation Sign Restriction: \( \zeta_{jh} \leq r_{\eta_j,\eta_h} \leq \bar{\zeta}_{jh} \) where \( (\zeta_{jh}, \bar{\zeta}_{jh}) \in \{(-1,0), (0,1), (0,0), (-1,1)\} \) for \( j, h = 1, \ldots, p \) and \( j < h \).

A₆ encodes the sign restrictions (if any) imposed on the \( \frac{1}{2}p(p-1) \) off-diagonal elements of \( \sigma_{\eta}^2 \). For example, \( (\zeta_{jh}, \bar{\zeta}_{jh}) = (-1,0) \) encodes that \( r_{\eta_j,\eta_h} \leq 0 \) whereas if A₆ does not restrict the sign of \( r_{\eta_j,\eta_h} \) then we set \( (\zeta_{jh}, \bar{\zeta}_{jh}) = (-1,1) \). We collect these restrictions in the matrix \( c = (\zeta, \bar{\zeta}) \) where \( \zeta = (\zeta_{12}, \ldots, \zeta_{(p-1)p})' \) and \( \bar{\zeta} = (\bar{\zeta}_{12}, \ldots, \bar{\zeta}_{(p-1)p})' \). For example, we set \( c = 0 \) when \( \sigma_{\eta}^2 \) is assumed to be diagonal.

Online Appendix A extends A₆ to \( A_6' \) which sets \( \zeta_{jh} \leq r_{\eta_j,\eta_h} \leq \bar{\zeta}_{jh} \) with \( -1 \leq \zeta_{jh} \leq \bar{\zeta}_{jh} \leq 1 \). In particular, \( A_6' \) may restrict the sign and/or magnitude of the correlation \( r_{\eta_j,\eta_h} \). While \( A_6' \) is conceptually similar to A₆, the expression for the identification region under A₁-A₆ is more complex. To ease the exposition, we report these results in Section A.1 of the Online Appendix. Here and in the empirical analysis in Section 6, we focus on specifying the sign of \( r_{\eta_j,\eta_h} \), if at all, which can be more salient in empirical work and is sometimes more easily inferred from economic theory.

As we show in Section 3, whereas A₄ directly restricts the net-of-X signal to total variance ratio \( \rho \), A₅ and A₆ indirectly restrict \( \rho \). We vary \( \kappa, \tau, \) and \( c \) in A₄-A₆ to conduct a sensitivity analysis that weakens the no measurement error assumption \( \kappa = 0 \) (or \( R_{Y_j,\tilde{W}}^2 = \tau_j \) in A₅), controls the fit of the model \( (R_{Y_j,\tilde{W}}^2 \leq \tau_j) \), and weakens the assumption that \( \sigma_{\eta}^2 \) is diagonal \( (c = 0) \). Conversely, we study for what configuration of \( (\kappa, \tau, c) \) does the identification region admit a plausible value or range for a component of \( \delta \) or \( \beta \) for example. To keep
the exposition concise, we impose A_4-A_6 throughout and obtain the results when A_4, A_5, or A_6 is not binding as a special case in which κ → +∞, τ = (1, ..., 1)', or c is such that (c_jh, c_jh) = (−1, 1) for all j < h.

3 Identification

We study identifying δ, and consequently β = b_{Y,X} − b_{W,X}δ, under A_1-A_3 and demonstrate how considering the Y equations jointly can improve on the bounds that obtain when analyzing each Y_j equation separately. Moreover, we study the consequences of imposing any configuration of A_4-A_6 on the identification regions for δ and β.

3.1 Characterization Theorem

From Theorem [3.1] under A_1-A_3, the moments in Var[(Ŷ', W')] can be expressed as

\[ \sigma^2_{W} = \sigma^2_{U} + \sigma^2_{\varepsilon}, \quad \sigma_{W,Y} = \sigma_{W,U} \delta = \sigma^2_{U} \delta, \quad \text{and} \quad \sigma^2_{Y} = \delta' \sigma^2_{U} \delta + \sigma^2_{\eta}. \]

Dividing \( \sigma_{W,Y} \) by \( \sigma^2_{W} \), we obtain that

\[ b_{Y,W} = \rho \delta \quad \text{where} \quad \rho \equiv \frac{\sigma^2_{U}}{\sigma^2_{W}} = \frac{\sigma^2_{U}}{\sigma^2_{U} + \sigma^2_{\varepsilon}}. \tag{5} \]

Since the net-of-X “signal to total variance ratio” \( \rho \) satisfies \( 0 \leq \rho \leq 1 \), we obtain the classic “attenuation bias” whereby \( b_{Y,W} \) underestimates the magnitude of \( \delta_j \) and has its sign. If there is no measurement error (\( \sigma^2_{\varepsilon} = 0 \)) then \( \rho = 1 \) and \( b_{Y,W} = \delta \). If U and X are perfectly collinear (\( \sigma^2_{U} = 0 \)) then \( \rho = 0 \) and \( b_{Y,W} \) does not identify \( \delta \). Similarly, normalizing \( \sigma^2_{Y} \) by \( \sigma^2_{W} \) gives

\[ \sigma^{-2}_{Y} \sigma^2_{\eta} = \delta' \rho \delta + \sigma^{-2}_{W} \sigma^2_{\eta}, \tag{6} \]

where we have that

\[ \Gamma \equiv \sigma^{-2}_{W} \sigma^2_{\eta} \text{ is positive semi-definite (denoted by } 0 \preceq \Gamma). \tag{7} \]

As we show in Corollary [3.2] the system of (in)equalities (4-7) exhausts the information on (\( \rho, \delta, \beta, \Gamma \)) that is implied by A_1-A_3. The auxiliary assumptions A_4-A_6 impose additional restrictions on the parameters. A_4 requires that \( \frac{1}{1+\kappa} \leq \rho \), A_5 imposes the lower bound \( \frac{\sigma^2_{j}}{\sigma^2_{W_j}} (1 - \tau_j) \leq \Gamma_{jj} \), and A_6 may specify the (weak) sign of \( \Gamma_{jh} \).
When $U$ and $X$ are not perfectly collinear, i.e. $\rho \neq 0$, Theorem 3.1 uses equations (4-6) to express $\delta$, $\beta$, and $\Gamma$ as functions of $D$, $B$, and $G$ of $\rho$. This mapping enables characterizing the identification region for $(\rho, \delta, \beta, \Gamma)$ in terms of restrictions on $\rho$ only and facilitates a sensitivity analysis that studies the consequences of deviating from the “no measurement error” assumption $\rho = 1$.

**Theorem 3.1** Assume $A_1$-$A_3$ and let $\text{Var}[(X', U)']$ be nonsingular so that $0 < \rho$. Then

$$
\delta = D(\rho) \equiv \frac{1}{\rho} b_{Y, \tilde{W}}, \\
\beta = B(\rho) \equiv b_{Y,X} - b_{W,X} \frac{1}{\rho} b_{Y, \tilde{W}}, \text{ and}\\n\Gamma = G(\rho) \equiv \sigma_{\tilde{W}}^2 - b'_{\tilde{Y}, \tilde{W}} \frac{1}{\rho} b_{\tilde{Y}, \tilde{W}}.
$$

Theorem 3.1 reveals how if there is no measurement error ($\rho = 1$) then $(\delta, \beta, \Gamma)$ is point identified. Further, even when $\rho < 1$, $b_{\tilde{Y}, \tilde{W}} = 0$ if and only if $(\delta_j, \beta_j, \Gamma_{jh}) = (0, b_{Y, X, j}, \sigma_{\tilde{W}}^2 \sigma_{\tilde{Y}, \tilde{Y}}^2)$. Similarly, if the $i^{th}$ element $b_{W,X, i}$ of $b_{W,X}$ is 0 then $\beta_i = b_{Y,X, i}$. Last, as discussed in Section 2, if $X_i$ is excluded from the $Y_j$ equation so that $\beta_{jl} = b_{Y_j, X, i} - b_{W,X} \frac{1}{\rho} b_{\tilde{Y}, \tilde{W}} = 0$ then, provided $b_{Y_j, X, i} \neq 0$, $\rho$ is point identified and so is $(\delta, \beta, \Gamma)$.

### 3.2 Identification Regions

Corollary 3.2 characterizes the sharp joint identification regions $J^{k,\tau,c}$ for $(\rho, \delta, \beta, \Gamma)$ under $A_1$-$A_3$ and any configuration of the auxiliary assumptions $A_4$-$A_6$ (i.e. any $(k, \tau, c)$ value).

**Corollary 3.2** Under the conditions of Theorem 3.1 and $A_4$-$A_6$ for $j, h = 1, \ldots, p$ with $j < h$, $(\rho, \delta, \beta, \Gamma)$ is partially identified in the sharp set

$$
J^{k,\tau,c} \equiv \{(r, D(r), B(r), G(r)) : 0 \leq G(r), \frac{1}{1 + \kappa} \leq r \leq 1, \frac{\sigma_{\tilde{Y}}^2}{\sigma_{\tilde{W}}^2} (1 - \tau_j) \leq G_{jj}(r), \text{ and} \}
$$

$$
\epsilon_{jh} \leq \text{sgn}(G_{jh}(r)) \leq \bar{\epsilon}_{jh} \text{ for } j, h = 1, \ldots, p \text{ and } j < h \}.\n$$

In characterizing $J^{k,\tau,c}$, Corollary 3.2 uses the mappings in Theorem 3.1 to encode restrictions (7), $A_4$ (with $\rho \leq 1$), $A_5$, and $A_6$ respectively as the constraints $0 \leq G(r), \frac{1}{1 + \kappa} \leq r \leq 1, \frac{\sigma_{\tilde{Y}}^2}{\sigma_{\tilde{W}}^2} (1 - \tau_j) \leq G_{jj}(r), \text{ and} \epsilon_{jh} \leq \text{sgn}(G_{jh}(r)) \leq \bar{\epsilon}_{jh}$ that involve only one unknown parameter, $\rho$. Further, the proof of Corollary 3.2 shows that $J^{k,\tau,c}$ is sharp since for every

---

6For $a \in \mathbb{R}$, define the sign function: $\text{sgn}(a) = -1$ if $a < 0$, $\text{sgn}(a) = 0$ if $a = 0$, and $\text{sgn}(a) = 1$ if $a > 0$.  

---
(r, d, b, g) ∈ \mathcal{J}^{k,r,e} there exists (U^*, \eta^*, \varepsilon^*) with \frac{\sigma^2_{\omega}}{\sigma^2_{\eta}} = r and G(r) = \sigma^2_W \sigma^2_{\eta}, that satisfy A_2-A_6 and that could have generated Y and W according to A_1. Thus, Corollary 3.2 exhausts the information on (\rho, \delta, \beta, \Gamma) in A_1-A_6.

Next, Corollary 3.3 characterizes analytically the projections \mathcal{R}^{k,r,e}, \mathcal{D}^{k,r,e}, \mathcal{B}^{k,r,e}, and \mathcal{G}^{k,r,e} of the joint identification region \mathcal{J}^{k,r,e} onto the support of the components \rho, \delta, \beta, and \Gamma. To illustrate these identification regions in a simple context, suppose that X is empty and consider identifying \delta_j in the Y_j equation where we substitute for U = W - \varepsilon:

\[ Y_j = U \delta_j + \eta_j = W \delta_j - \varepsilon \delta_j + \eta_j \quad \text{for} \quad j = 1, ..., p. \tag{8} \]

Recall that, when analyzing the Y_j equation separately, \delta_j is bounded under A_1-A_3 using the forward and reverse regressions (see e.g. Klepper and Leamer, 1984; Bollinger, 2003):

\[ \delta_j \in \{ \rho b_{Y_j,W} + (1 - \rho) \frac{1}{b_{W,Y_j}} \lambda : 0 \leq \lambda \leq 1 \}. \]

The bound \( b_{Y_j,W} \) or \( \frac{1}{b_{W,Y_j}} \) is attained if there is no measurement error (\sigma^2 = 0) or if the constraint \( 0 \leq \sigma^2_W \sigma^2_{\eta} \) binds respectively. From Theorem 3.1’s mapping \( \delta_j = D_j(\rho) = \frac{1}{\rho} b_{Y_j,W} \), the bounds \( \frac{1}{b_{W,Y_j}} \) and \( b_{Y_j,W} \) on \( \delta_j \) correspond respectively to the lower and upper bounds \( R^2_{W,Y_j} \) and 1 on \( \rho \). Corollary 3.3 demonstrates how, by analyzing the Y equations jointly, the constraint \( 0 \leq G(r) = \sigma^2_W \sigma^2_{\eta} \) yields a tighter lower bound on \( \rho \) which can be conveniently expressed by \( R^2_{W,Y_j} \leq R^2_{W,Y} \leq \rho \leq 1 \). In turn, this can lead to tighter bounds on \( \delta_j \). A_4 and A_5 yield the additional constraints \( \frac{1}{1+\varepsilon} \leq \rho \) and \( \frac{1}{\tau_j} R^2_{W,Y_j} \leq \rho \) for \( j = 1, ..., p \) which can tighten the bounds on \( \rho \), and thus \( \delta_j \), further.

Moreover, analyzing the Y_j and Y_h equations jointly reveals how restricting the sign of the cross-equation disturbance \( \sigma_{\eta_j,\eta_h} \) can tighten the bounds on \( \delta_j \). In particular, under A_1-A_3, if \( \sigma_{\eta_j,\eta_h} = 0 \) then \( Cov[(\varepsilon, \eta_j)', Y_h] = 0 \) and Y_h may serve as an instrument for W in the Y_j equation \( (8) \) to point identify \( \delta_j = b_{Y_j,W|Y_h} \). Since \( \delta_j = \frac{1}{\rho} b_{Y_j,W} \), provided \( \sigma^2_W \sigma_{Y_j,Y_h} \neq 0 \), we obtain \( \rho = \frac{b_{Y_j,W} b_{Y_h,W}}{\sigma^2_W \sigma_{Y_j,Y_h}} \). More generally, using Theorem 3.1’s mapping

\[ \Gamma_{jh} \equiv \sigma^{-2}_W \sigma_{\eta_j,\eta_h} = G_{jh}(\rho) = \sigma^{-2}_W \sigma_{Y_j,Y_h} - b_{Y_j,W} \frac{1}{\rho} b_{Y_h,W}, \]

we characterize the identification regions for \( \rho \), and therefore \( \delta_j \), under weaker restriction in A_6 on the sign of \( \sigma_{\eta_j,\eta_h} \). Depending on \( \sigma^{-2}_W \sigma_{Y_j,Y_h} \) and \( b_{Y_j,W} b_{Y_h,W} \), restricting the sign of \( \sigma_{\eta_j,\eta_h} \), as encoded by \( (\varepsilon_{jh}, \tau_{jh}) \) in A_6, may be either redundant or contradictory given the other restrictions on \( \rho \). Otherwise, restricting the sign of \( \sigma_{\eta_j,\eta_h} \) can determine whether \( \rho \) is smaller or larger than \( \frac{b_{Y_j,W} b_{Y_h,W}}{\sigma^2_W \sigma_{Y_j,Y_h}} \) which may in turn tighten the bounds on \( \delta_j \).
Corollary 3.3 formalizes this discussion and extends it to accommodate the covariates X.

**Corollary 3.3** Under the conditions of Theorem 3.1 and $A_4$-$A_6$ for $j, h = 1, ..., p$ with $j < h$, $\rho, \delta, \beta$, and $\Gamma$ are partially identified in the sharp sets

$$\mathcal{R}^{k, \tau, c} = [R_W^2, Y, 1] \cap \left[ \frac{1}{1 + \kappa}, 1 \right] \cap \prod_{j=1}^{p} \mathcal{R}_{jh},$$

where

$$\mathcal{R}_{jh}^c = \begin{cases} \frac{b_j, \tilde{w}, b_h, \tilde{w}}{\sigma_W^2 \sigma_Y^2 \tilde{y}_j, \tilde{y}_h} & \text{if } (\zeta_{j h}, \tilde{c}_{j h}) = (0, 0) \text{ and } \sigma_W^2 \sigma_Y^2 \tilde{y}_j, \tilde{y}_h \neq 0 \\ (-\infty, \frac{b_j, \tilde{w}, b_h, \tilde{w}}{\sigma_W^2 \sigma_Y^2 \tilde{y}_j, \tilde{y}_h}) & \text{if } (\zeta_{j h}, \tilde{c}_{j h}) \in \{(-1, 0), (0, 1)\} \text{ and } \text{sgn}(\sigma_W^2 \sigma_Y^2 \tilde{y}_j, \tilde{y}_h) \notin [\zeta_{j h}, \tilde{c}_{j h}] \\ \left[ \frac{b_j, \tilde{w}, b_h, \tilde{w}}{\sigma_W^2 \sigma_Y^2 \tilde{y}_j, \tilde{y}_h}, \infty \right) & \text{if } (\zeta_{j h}, \tilde{c}_{j h}) \in \{(-1, 0), (0, 1)\} \text{ and } \text{sgn}(\sigma_W^2 \sigma_Y^2 \tilde{y}_j, \tilde{y}_h) \in [\zeta_{j h}, \tilde{c}_{j h}] \setminus \{0\} \\ \emptyset & \text{if } (\zeta_{j h}, \tilde{c}_{j h}) \neq (-1, 1), \text{sgn}(-b_j, \tilde{w}, b_h, \tilde{w}) \notin [\zeta_{j h}, \tilde{c}_{j h}], \text{ and } \sigma_W^2 \sigma_Y^2 \tilde{y}_j, \tilde{y}_h = 0 \\ (-\infty, \infty) & \text{otherwise} \end{cases},$$

$$\mathcal{D}^{k, \tau, c} = \{D(r) : r \in \mathcal{R}^{k, \tau, c}\}, \mathcal{B}^{k, \tau, c} = \{B(r) : r \in \mathcal{R}^{k, \tau, c}\}, \text{ and } \mathcal{G}^{k, \tau, c} = \{G(r) : r \in \mathcal{R}^{k, \tau, c}\}.$$

Corollary 3.3 characterizes the identification regions $\mathcal{R}^{k, \tau, c}$, $\mathcal{D}^{k, \tau, c}$, $\mathcal{B}^{k, \tau, c}$, and $\mathcal{G}^{k, \tau, c}$ for $\rho, \delta, \beta$, and $\Gamma$ under $A_4$-$A_6$. Further, it shows that each of these projected regions is sharp - for example, for every $d \in \mathcal{D}^{k, \tau, c}$ there exists $(r, d, b, g) \in \mathcal{J}^{k, \tau, c}$.

It is instructive to examine the identification regions in Corollary 3.3 under sequentially stronger configurations of $(k, \tau, c)$. First, suppose that $(\zeta_{j h}, \tilde{c}_{j h}) = (0, 0)$ for all $j, h = 1, ..., p$ with $j < h$, so that $A_6$ is not binding. Then $\mathcal{R}_{j h}^c = (-\infty, \infty)$. In this case, we sometimes drop the superfluous superscript $c$ and obtain $\mathcal{R}^{k, \tau, c} = \mathcal{R}^{k, \tau} \equiv [R_W^2, Y, 1] \cap \left[ \frac{1}{1 + \kappa}, 1 \right] \cap \prod_{j=1}^{p} \frac{1}{\tau_j} \left[ R_W^2, Y, 1 \right]$.

If $\kappa \to \infty$ and $\tau = (1, ..., 1)'$ then $A_4$ and $A_5$ are also not binding and we sometimes drop the $\kappa$ and $\tau$ superscripts. Provided $R_W^2 \neq 0$, we obtain $\mathcal{R}^{k, \tau, c} = \mathcal{R} \equiv [R_W^2, Y, 1]$ since $\max\{R_W^2, Y_1, ..., R_W^2, Y_p\} \leq R_W^2, Y$. In this case, Corollary 3.3 reduces to the bounds in Leamer (1987), specialized to a scalar mismeasured $U$, after projecting on the covariates $X$. As discussed in Leamer (1987), the joint-equations bounds improve on the single-equation bounds that obtain using each $Y_j$ equation separately. In particular, if the dimension of $Y$ is $p = 1$ then Corollary 3.3 gives the single-equation bounds for $\rho, \delta$, and $\beta$ (see e.g. Chalak and Kim, 2019, corollary 3.5). As the dimension of $Y$ increases, $R_W^2, Y$ may increase and the joint-equations bounds $\mathcal{R}$ for $\rho$ may become tighter. Instead, if $\kappa < \infty$ or $\tau_j < 1$ for some $j$ (or both) then $A_4$ or $A_5$ (or both) is in effect. If $R_W^2 < \max\{\frac{1}{1 + \kappa}, \frac{1}{\tau_j} R_W^2, Y_1, ..., \frac{1}{\tau_p} R_W^2, Y_p\}$ then imposing $A_4$ and $A_5$ increases the lower bound on $\rho$. In turn, this leads to tighter bounds.
on $\delta$, $\beta$, and $\Gamma$ via the mappings in Theorem 3.1. In the limit, setting $\kappa = 0$ or $\tau_j = R^2_{\tilde{W}, \tilde{Y}_h}$ yields $\rho = 1$ and therefore point identifies $(\delta, \beta, \Gamma)$.

Next, consider imposing $A_6$. Corollary 3.3 shows that, even when $A_4$-$A_5$ are not binding, if $\sigma_{\eta_j, \gamma_h} = 0$ (i.e. $(c_{jh}, \tau_{jh}) = (0, 0)$) and $\sigma^{-2}_{\tilde{W}} \sigma_{\tilde{Y}_h, \tilde{Y}_h} \neq 0$ then $\rho = \frac{b_{\tilde{Y}_j, \tilde{W}} b_{\gamma_h, \tilde{W}}}{\sigma^{-2}_{\tilde{W}} \sigma_{\tilde{Y}_j, \tilde{Y}_h}}$ is point identified. When $b_{\tilde{Y}_j, \tilde{W}} \gamma_h$ exists and is nonzero, we can express $\rho = \frac{b_{\tilde{Y}_j, \tilde{W}}}{b_{\tilde{Y}_j, \tilde{W}} \gamma_h}$ as the ratio of the regression and IV regression estimands. It follows from the mappings in Theorem 3.1 that $(\rho, \delta, \beta, \Gamma)$ is point identified, with $\delta_j = b_{\tilde{Y}_j, \tilde{W}} \gamma_h$ and $\beta_j = b_{\gamma_j, \tilde{X}} - b_{\tilde{W}, \tilde{X}} b_{\tilde{Y}_j, \tilde{W}} \gamma_h$ as obtains via the IV regression that uses $Y_h$ as an instrument for $W$, $(\delta_j, \beta'_j) = b_{\gamma_j, \tilde{X}}(\gamma_h, \tilde{X}) \gamma$. More generally, Corollary 3.3 derives the identification regions for $\delta, \beta,$ and $\Gamma$ under weaker restriction in $A_6$ on the sign of $\Gamma_{jh} \equiv \sigma^{-2}_{W} \gamma_{\eta_j, \gamma_h}$. First, if the identification region $G_{jh}^{k, \tau}$ identifies the sign of $\Gamma_{jh}$ when $A_6$ is not binding (i.e. when $(c_{jh}, \tau_{jh}) = (-1, 1)$ for all $j < h$) then imposing the (correct) sign restriction on $\Gamma_{jh}$ in $A_6$ is redundant. Otherwise, restricting the sign of $\Gamma_{jh}$ in $A_6$ can rule out a region of $R^{k, \tau}$. Specifically, recall that $G_{jh}^{k, \tau}$ is given by

$$G_{jh}^{k, \tau} = \{ \sigma^{-2}_{\tilde{W}} \sigma_{\tilde{Y}_j, \tilde{Y}_h} - b_{\tilde{Y}_j, \tilde{W}} \frac{1}{\rho} b_{\gamma_h, \tilde{W}} : r \in R^{k, \tau} \}.$$ 

Thus, provided $\sigma^{-2}_{\tilde{W}} \sigma_{\tilde{Y}_j, \tilde{Y}_h}$ is nonzero\footnote{If $\sigma^{-2}_{\tilde{W}} \sigma_{\tilde{Y}_j, \tilde{Y}_h} = 0$ then restricting the sign of $\Gamma_{jh}$ is either contradictory or uninformative about $\rho$, depending on the sign of $b_{\tilde{Y}_j, \tilde{W}} b_{\gamma_h, \tilde{W}}$, as encoded in $R^{c}_{jh}$.} $0 \in int(G_{jh}^{k, \tau})$ if and only if

$$\frac{b_{\tilde{Y}_j, \tilde{W}} b_{\gamma_h, \tilde{W}}}{\sigma^{-2}_{\tilde{W}} \sigma_{\tilde{Y}_j, \tilde{Y}_h}} \in int(R^{k, \tau}).$$

Corollary 3.3 demonstrates how restricting the sign of $\Gamma_{jh}$ can rule out elements of $R^{k, \tau}$ that are either smaller or larger than $\frac{b_{\tilde{Y}_j, \tilde{W}} b_{\gamma_h, \tilde{W}}}{\sigma^{-2}_{\tilde{W}} \sigma_{\tilde{Y}_j, \tilde{Y}_h}}$, as encoded in $R^{c}_{jh}$. In turn, this can tighten the identification regions for $\delta, \beta$, and $\Gamma$. Last, if Corollary 3.2 yields $R^{k, \tau, c} = \emptyset$ then the model is misspecified and we reject the assumptions imposed in $A_1$-$A_6$.

We note that imposing restrictions on the signs and/or magnitudes of some of the coefficients $\delta_j$ or $\beta_j$ may tighten the bounds on $\rho$, and therefore on $\delta$, $\beta$, and $\Gamma$ using Theorem 3.1’s mappings. We do not pursue this here; instead, we focus on the auxiliary assumptions $A_4$-$A_6$ which do not directly restrict $\delta$ or $\beta$. Finally, we comment on the consequences of relaxing the classical measurement error assumption $A_3$ on the paper’s bounds. Section A.2 of the Online Appendix dispenses with $X$ in $A_4$ for simplicity and discusses how the restrictions in $A_2$ alone do not identify $\delta_j$. This holds even if $A_3$ is imposed partially, with $\sigma_{\eta, \epsilon} = 0$ and $\sigma_{U, \epsilon}$ unrestricted. To keep the scope of the paper manageable, we leave a detailed study of
the sharp identification regions for the coefficients without $A_3$, under restrictions analogous to $A_4$-A$_6$, to other work.

4 Numerical Example

To illustrate the shape of the identification regions in Section 3, we consider the following numerical example. We generate $X, W,$ and $Y$ according to $A_1$, as follows:

$$X = U\varphi + \eta_X, \quad W = U + \varepsilon, \quad \text{and} \quad Y_j = X_1\beta_{j1} + X_2\beta_{j2} + U\delta_j + \eta_j \quad \text{for} \ j = 1, 2, 3,$$

where $\eta_X \equiv (\eta_{X_1}, \eta_{X_2})'$, $X \equiv (X_1, X_2)'$, $\eta \equiv (\eta_1, \eta_2, \eta_3)'$, and $Y \equiv (Y_1, Y_2, Y_3)'$. We let $\eta_X, U, \varepsilon,$ and $\eta$ be jointly independent and normally distributed with mean 0 so that $A_2$ and $A_3$ hold. We allow the components of $\eta_X$ (respectively $\eta$) to be correlated. It follows that $(X', W, Y')$ is normally distributed and we can analytically express the identification regions for $\rho$, $\delta$, and $\beta$ in terms of the elements of $\text{Var}[(\eta'_X, U, \varepsilon, \eta')']$. In this example, we set the equation coefficients to

$$\beta = \begin{bmatrix} 1 & 0.7 \\ 0.85 & 0.95 \\ 1.1 & 1.2 \end{bmatrix}, \quad \delta = \begin{bmatrix} 0.7 \\ 1.05 \\ 0.84 \end{bmatrix}, \quad \text{and} \quad \varphi = \begin{bmatrix} 0.3 \\ 0.14 \end{bmatrix},$$

and the variances of $\eta_X, U, \varepsilon,$ and $\eta$ to

$$\sigma_u^2 = 3, \quad \sigma_v^2 = 5, \quad \sigma_{\eta_X} = \begin{bmatrix} 1 & 0.14 \\ 0.14 & 1 \end{bmatrix}, \quad \text{and} \quad \sigma_{\eta}^2 = \begin{bmatrix} 1.1 & -0.31 & 0.63 \\ -0.31 & 1.99 & -0.59 \\ 0.63 & -0.59 & 2.25 \end{bmatrix}.$$

We obtain that $\rho = 0.53$ and, therefore, any restriction $0.89 = \frac{1 - \rho}{\rho} = \frac{\sigma_u^2}{\sigma_v^2} \leq \kappa$ in $A_4$ is valid. Further, we obtain that $R_{W.Y_1}^2 = 0.31$, $R_{W.Y_2}^2 = 0.34$, $R_{W.Y_3}^2 = 0.27$, and $R_{W.Y}^2 = 0.44$.

Using a grid search, we approximate 4 types of identification regions, illustrated in Figure 1. The first is the single-equation identification regions $S_j$ that consider each $Y_j$ equation separately. The second is the joint-equations region $J$ that considers the $Y$ equations jointly. $S_j$ and $J$ obtain under $A_1$-$A_3$ only (i.e. when $\kappa = \infty, \tau = (1, 1, 1)'$, and $(\xi_{jh}, \bar{\xi}_{jh}) = (-1, 1)$ for all $j < h$). The third identification region is the joint-equations bounds $J^{\kappa, \tau}$ that obtains under $A_1$-$A_5$, with $\kappa = 1$ and $\tau = (0.7, 0.7, 0.7)'$. The fourth region $J^{\kappa, \tau, c}$ obtains under $A_1$-$A_6$ where $\kappa$ and $\tau$ are as in $J^{\kappa, \tau}$ and $c$ imposes the (correct) sign restrictions $r_{\eta_1, \eta_2} \leq 0$, $r_{\eta_1, \eta_3} \geq 0$, $r_{\eta_2, \eta_3} \leq 0$. Figure 1 illustrates these regions by plotting their two dimensional projections onto the $(\rho, \delta_j)$, $(\rho, \beta_{j1})$, and $(\rho, \beta_{j2})$ spaces for $j = 1, 2, 3$. The plus sign denotes
the population parameters. Further, the asterisk corresponds to the regression estimand $b_{Y,(W,X)'}$ and the cross sign corresponds to the identification region $J^{c^*}$ (the IV regression estimand) where $c^*$ incorrectly sets $\sigma_{n_1,n_2} = 0$ and leaves $\sigma_{n_1,n_3}$ and $\sigma_{n_2,n_3}$ unrestricted. Each graph in Figure 1 superimposes 4 identification regions represented in different shades. The darker regions are nested within the lighter regions. The lightest and second lightest shades correspond respectively to $S_j$ and $J$. The second darkest region corresponds to $J^{\kappa,\tau}$ (here $R^{\kappa,\tau} = \frac{1}{1+\kappa} = 0.5, 1$). Last, the darkest region corresponds to $J^{\kappa,\tau,c}$.

Table 1 uses the analytical expressions in Section 3 to report several bounds, including the one-dimensional projections of the regions in Figure 1. The first and second columns report the sharp projections of the single-equation and joint-equations identification regions $S_j^{\kappa,\tau}$ for $j = 1, 2, 3$ and $J^{\kappa,\tau}$ respectively under $A_1$-$A_5$. Note that $S_j^{\kappa,\tau}$ yields bounds for $\rho$ that vary with $j$. The third column reports the joint-equations bounds $J^{\kappa,\tau,c}$ under $A_1$-$A_6$ with the (correct) sign restrictions in $c$. The fourth column reports the (IV regression) region $J^{\kappa,\tau,c^*}$ where $c^*$ incorrectly sets $\sigma_{n_1,n_2} = 0$. The last column reports the regression estimand $b_{Y,(W,X)'}$ which would point identify $(\delta', \beta')'$ if $W$ is correctly measured. Table 1 reports the bounds when $\kappa = \infty$ and $\tau = (1, 1, 1)'$ (i.e. when $A_4$-$A_5$ are not binding) in the upper panel as well as when $\kappa = 1$ and $\tau = (0.7, 0.7, 0.7)'$ in the lower panel. Figure 1 and Table 1 illustrate how the population parameters are elements of the nested sets $J^{\kappa,\tau,c} \subseteq S_1^{\kappa,\tau} \times \ldots \times S_p^{\kappa,\tau}$ which become tighter as stricter valid restrictions on $\kappa$, $\tau$, and/or $c$ are imposed.

5 Estimation and Inference

For inference, we implement a procedure that delivers $1 - \alpha$ (e.g. 50% or 95%) confidence regions for each of the partially identified parameters $\rho$, $\delta_j$, $\beta_j l$, and $\Gamma_{jh}$ for $j, h = 1, \ldots, p$ and $l = 1, \ldots, k$. The procedure consists of three steps. First, we express each of the bounds in Corollary 3.3 as a function of the vector of estimands $^8$

$$\pi \equiv (vec(b_{Y,(W,X)'}), b_{W,(Y'X')}', b_{W,(Y_1,X')}', \ldots, b_{W,(Y_p,X')}', vec(b_{Y,X}'), b_{W,X}', \sigma_{W}^{-2}vec(\sigma_{Y}^2))'.$$

An alternative would express the bounds in Corollary 3.3 as a function of $\text{Var}[(1,Y',W,X')']$ and constructs an estimator for these moments.

Throughout this discussion, we assume that $\sigma_{Y}^2$ is nonsingular. Otherwise, we drop the redundant $Y$ elements from $(Y,X')'$ in $b'_{W,(Y,X)'}$ and $R_{W,Y}^2$.

---

8An alternative would express the bounds in Corollary 3.3 as a function of $\text{Var}[(1,Y',W,X')']$ and constructs an estimator for these moments.

9Throughout this discussion, we assume that $\sigma_{Y}^2$ is nonsingular. Otherwise, we drop the redundant $Y$ elements from $(Y,X')'$ in $b'_{W,(Y,X)'}$ and $R_{W,Y}^2$. 

---

15
where $\text{vec}(\sigma^2_\rho)$ collects the $\frac{1}{2}p(p+1)$ variance and covariance elements of $\sigma^2_\rho$. Further, we construct an estimator $\hat{\pi}$ for $\pi$ and give conditions under which $\hat{\pi}$ is $\sqrt{n}$ consistent and asymptotically normally distributed. Second, we employ results on intersection bounds to construct a $1 - \alpha$ confidence region $CR^{\rho}_{1 - \alpha}$ for the parameter $\rho$ that is partially identified in $R^{\kappa,\tau,c}$ for any $(\kappa, \tau, c)$ configuration. The last step uses the mappings, given in Theorem 3.1, that express $\delta_j, \beta_{jl}$, and $\Gamma_{jh}$ as functions of $(\pi, \rho)$ to construct $1 - \alpha$ confidence regions for the partially identified parameters $\delta_j, \beta_{jl}$, and $\Gamma_{jh}$.

5.1 Estimation of $\pi$

We estimate $\pi$ using the plug-in estimator $\hat{\pi}$:

$$\hat{\pi} \equiv (\text{vec}(\hat{b}_{Y,(W,X)})', \hat{b}_{W,(Y',X)',} \hat{b}_{W,(Y,X)'}', \text{vec}(\hat{b}_{Y,X}'), \hat{b}_{W,X}', \hat{\sigma}_{W}^{-2}\text{vec}(\hat{\sigma}^2_Y)'),$$

Specifically, given observations $\{A_i, B_i\}_{i=1}^n$ corresponding to random column vectors $A$ and $B$, let $\bar{A} \equiv \frac{1}{n} \sum_{i=1}^n A_i$ and denote the sample covariance (with $\hat{\sigma}_A^2 = \hat{\sigma}_{A,A}$) and the linear regression estimator and sample residuals by:

$$\hat{\sigma}_{A,B} \equiv \frac{1}{n} \sum_{i=1}^n (B_i - \bar{B})(A_i - \bar{A})', \quad \hat{b}_{A,B} \equiv \hat{\sigma}_{B}^{-2}\hat{\sigma}_{A,B}, \quad \text{and } \hat{\varepsilon}^2_{A,B} \equiv (A_i - \bar{A})' - (B_i - \bar{B})' \hat{b}_{A,B}.$$

Under conditions sufficient for the law of large numbers and central limit theorem (see e.g. White (2001) for primitive conditions), the estimator $\hat{\pi}$ for $\pi$ is $\sqrt{n}$ consistent and asymptotically normally distributed. For this, let $\mu_A^2 = E(AA')$ and define the square block-diagonal matrix $Q$:

$$Q \equiv \text{diag}\{ I \otimes \mu^2(1,W,X)' , \mu^2(1,Y',X) , \mu^2(1,Y,X) , \mu^2(1,Y,W) , I \otimes \mu^2(1,Y')', \mu^2(1,X)' , I \otimes \mu^2(1,X)', \mu^2(1,X)' , \frac{1}{4}p(p+1) \times \frac{1}{4}p(p+1) \otimes \sigma^2_W \} ,$$

where the moments in the diagonal blocks correspond to the estimands in $\pi$.

---

10Let $A = [A_1, ..., A_q] \in \mathbb{R}^{m \times q}$. Then $\text{vec}(A) \equiv (A_1', ..., A_q')'$. Further, if $q = m$ and $A$ is symmetric then we let $\text{vec}(A) \equiv [A_{11}, ..., A_{mm}, A_{12}, ..., A_{1m}, ..., A_{m-1,m}, ..., A_{m-1,m}']$ collect the diagonal and upper-diagonal elements of $A$.

11The framework in Kline and Tamer (2016) provides a useful alternative for inference which makes direct use of the mappings from the estimand $\pi$ to the identification regions. Following Kline and Tamer (2016, example 1), the intersection bounds can be expressed by $R^{k,\tau,c} = \{ \rho : C(\rho, \pi) = 0 \}$ where $C(\rho, \pi)$ is a criterion function (one may consider incorporating the nuisance parameters, discussed below, in this representation). Analogous expressions for $J^{k,\tau,c}, D^{k,\tau,c}, B^{k,\tau,c}$, and $G^{k,\tau,c}$ obtain using Theorem 3.1's mappings. One can then draw from the posterior for $\pi$ and apply the procedures in Kline and Tamer (2016) to generate posterior probability statements concerning the partially identified parameters $\rho, \delta, \beta,$ and $\Gamma$. 

---
Theorem 5.1 Assume $A_1(i)$ and that $Q$ is nonsingular. Suppose further that:

(i) $\frac{1}{n} \sum_{i=1}^{n} (1, Y'_i, W_i, X'_i)'(1, Y'_i, W_i, X'_i) \rightarrow_{P} \mu_1^{2}(Y,Y',W,X)'$ and

(ii) $n^{-1/2} \sum_{i=1}^{n} \begin{bmatrix}
 vec[(1, W_i, X'_i)'(Y'_i)'(X'_i)']
 (1, Y'_i, X'_i)'(Y'_i)'(X'_i)'
 (1, Y'_i, X'_i)'(Y'_i)'(W_i, X_i)'
 vec(\epsilon_i Y_i, \epsilon_i X_i, \epsilon_i Y_i, \epsilon_i X_i)
 vec(\epsilon_i Y_i, \epsilon_i X_i)
 \end{bmatrix} \rightarrow_{D} N(0, \Xi)$ where $\Xi \equiv \text{Var}(\epsilon_i Y_i, \epsilon_i X_i, \epsilon_i Y_i, \epsilon_i X_i)$.

Then $\sqrt{n}(\hat{\pi} - \pi) \rightarrow_{D} N(0, \Sigma)$ where $\Sigma$ obtains by removing from $\Sigma^* \equiv Q^{-1}\Xi Q'^{-1}$ the rows and columns corresponding to the regression intercepts.

We estimate $\Sigma$ using the relevant submatrix of the heteroskedasticity-robust estimator $\hat{\Sigma}^* \equiv \hat{Q}^{-1}\hat{\Xi} \hat{Q}'^{-1}$ for $\Sigma^*$ (see e.g. White, 1980). For example, we estimate the component $\text{Cov}(X\epsilon_{Yj,X}, X\epsilon_{Yh,X})$ of $\Xi$ using its counterpart $\frac{1}{n} \sum_{i=1}^{n} X_i \epsilon_{Yj,X} \epsilon_{Yh,X} X'_i$ in $\hat{\Xi}$.

5.2 Inference on $\rho$

To form a $1 - \alpha$ confidence region for the parameter $\rho$ that is partially identified in $\mathcal{R}^{k,\tau,c}$, we express the identification region for $\rho$ as a finite number of intersection bounds

$$\mathcal{R}^{k,\tau,c}(\lambda) \equiv [\rho^l(\lambda), \rho^u(\lambda)] \equiv \bigcap_{v=1}^{M} [\rho_v^l(\lambda), \rho_v^u(\lambda)] \equiv \bigcap_{v=1}^{M} \mathcal{R}_v(\lambda),$$

which may depend on a vector of nuisance parameters $\lambda \in \mathbb{R}^{2T}$, a function of $\pi$:

$$\lambda \equiv (\sigma^2_W, \sigma_{\tilde{Y}1, \tilde{Y}2}, \sigma^2_W, \sigma_{\tilde{Y}1, \tilde{Y}3}, ..., \sigma^2_W, \sigma_{\tilde{Y}_{p-1}, \tilde{Y}p}, b_{\tilde{Y}1, \tilde{W}}, b_{\tilde{Y}2, \tilde{W}}, b_{\tilde{Y}1, \tilde{W}}, b_{\tilde{Y}3, \tilde{W}}, ..., b_{\tilde{Y}_{p-1}, \tilde{W}}, b_{\tilde{Y}_{p}, \tilde{W}}).$$

Further, for a given $\lambda$, each of the bounds $\rho^l(\lambda)$ and $\rho^u(\lambda)$ can be expressed as a function of $\tau$. For example, in the numerical example in Section 4, the identification region $\mathcal{R}^{k,\tau,c}$ under $A_1$-$A_6$ (with $\Gamma_{12} \leq 0$, $\Gamma_{13} \geq 0$, and $\Gamma_{23} \leq 0$) for $\rho$ is

$$\mathcal{R}^{k,\tau,c}(\lambda) = \bigcap_{v=1}^{8} [\rho^l_v(\lambda), \rho^u_v(\lambda)]$$

$$= \left[ \frac{1}{\tau_1} R^2_{W,\tilde{Y}1}, 1 \right] \cap \left[ \frac{1}{1 + \kappa}, 1 \right] \cap \left[ \frac{1}{\tau_2} R^2_{W,\tilde{Y}2}, 1 \right] \cap \left[ \frac{1}{\tau_2} R^2_{W,\tilde{Y}3}, 1 \right] \cap \left[ \frac{1}{\tau_3} R^2_{W,\tilde{Y}3}, 1 \right]$$

$$\cap \left( -\frac{b_{\tilde{Y}_1, \tilde{W}} b_{\tilde{Y}_2, \tilde{W}}}{\sigma^2_W \sigma_{\tilde{Y}_1, \tilde{Y}_2}}, \infty \right) \cap \left( -\frac{b_{\tilde{Y}_1, \tilde{W}} b_{\tilde{Y}_2, \tilde{W}}}{\sigma^2_W \sigma_{\tilde{Y}_1, \tilde{Y}_2}}, \infty \right) \cap \left( -\frac{b_{\tilde{Y}_1, \tilde{W}} b_{\tilde{Y}_2, \tilde{W}}}{\sigma^2_W \sigma_{\tilde{Y}_1, \tilde{Y}_2}}, \infty \right).$$
where the last three intersected regions $\mathcal{R}_{r_2}^c(\lambda)$, $\mathcal{R}_{r_3}^c(\lambda)$, and $\mathcal{R}_{r_4}^c(\lambda)$ in $\mathcal{R}^{\kappa,\tau,c}(\lambda)$ obtain from Corollary 3.3 based on the signs of the nuisance parameters (here $T = 3$)

$$
\lambda_{2T \times 1} = \left(\sigma_{\tilde{Y}_1,\tilde{Y}_2,\tilde{Y}_3}, \sigma_{\tilde{Y}_1,\tilde{Y}_2,\tilde{Y}_3}, \sigma_{\tilde{Y}_1,\tilde{Y}_2,\tilde{Y}_3}, b_{\tilde{Y}_1,\tilde{Y}_2,\tilde{Y}_3}, b_{\tilde{Y}_1,\tilde{Y}_2,\tilde{Y}_3}, b_{\tilde{Y}_1,\tilde{Y}_2,\tilde{Y}_3}, b_{\tilde{Y}_1,\tilde{Y}_2,\tilde{Y}_3}.\right)
$$

Thus, $\lambda$ determines whether each $\mathcal{R}_{r_{2j}}^c(\lambda)$ is $\emptyset$, $(-\infty, \infty)$, $(-\infty, \frac{b_{\tilde{Y}_j,\tilde{Y}_h,\tilde{Y}_k}}{\sigma_{\tilde{Y}_j,\tilde{Y}_h,\tilde{Y}_k}})$, or $[\frac{b_{\tilde{Y}_j,\tilde{Y}_h,\tilde{Y}_k}}{\sigma_{\tilde{Y}_j,\tilde{Y}_h,\tilde{Y}_k}}, \infty]$.

### 5.2.1 Known Nuisance Parameters

First, suppose that the nuisance parameter $\lambda$ is known (or that $A_6$ is not imposed and $\lambda$ is irrelevant). As discussed in Manski and Pepper (2009) and Chernozhukov, Lee, and Rosen (2013), the sample analog estimator $\hat{\mathcal{R}}^{\kappa,\tau,c}(\lambda) \equiv \cap_{v=1}^{M} \left[\hat{\rho}_v^l(\lambda), \hat{\rho}_v^u(\lambda)\right]$ tends to be biased “inward” in finite samples, leading to estimates that are on average narrower than $\mathcal{R}^{\kappa,\tau,c}(\lambda)$. Further, the sampling error may vary with $v$, across the intersected regions $\mathcal{R}_v(\lambda)$, which complicates the inference on $\mathcal{R}^{\kappa,\tau,c}(\lambda)$. To overcome these difficulties, we follow Chernozhukov, Lee, and Rosen (2013) and use the “precision-corrected” estimators for $\rho_v^l(\lambda)$ and $\rho_v^u(\lambda)$, $v \in V \equiv \{1, \ldots, M\}$ in order to construct estimators for $\hat{\rho}_v^l(\lambda)$ and $\hat{\rho}_v^u(\lambda)$ as follows:

$$
\hat{\rho}_v^l(\lambda; 1-\alpha_{21}) \equiv \sup_{v \in V} [\hat{\rho}_v^l(\lambda) - c_{1-\alpha_{21}}(\lambda)se_v^l(\lambda)] \quad \text{and} \quad \hat{\rho}_v^u(\lambda; 1-\alpha_{21}) \equiv \inf_{v \in V} [\hat{\rho}_v^u(\lambda) + c_{1-\alpha_{21}}(\lambda)se_v^u(\lambda)]
$$

where $1-\alpha_{21}$ is a significance level with $\alpha_{21} \leq \frac{1}{2}$, $se_v^l(\lambda)$ ($se_v^u(\lambda)$) is the standard error for the plug-in estimators $\hat{\rho}_v^l(\lambda)$ ($\hat{\rho}_v^u(\lambda)$), and $c_{1-\alpha_{21}}(\lambda)$ ($c_{1-\alpha_{21}}(\lambda)$) is a suitably selected critical value, discussed below, such that

$$
\Pr[\hat{\rho}_v^l(\lambda; 1-\alpha_{21}) \leq \rho_v^l(\lambda)] \geq 1 - \alpha_{21} - o(1) \quad \text{and} \quad \Pr[\hat{\rho}_v^u(\lambda) \leq \rho_v^u(\lambda; 1-\alpha_{21})] \geq 1 - \alpha_{21} - o(1).
$$

In particular, setting $\alpha_{21} = \frac{1}{2}$ yields half-median-unbiased estimators $\hat{\rho}_v^l(\lambda; \frac{1}{2})$ and $\hat{\rho}_v^u(\lambda; \frac{1}{2})$.

Using Bonferroni’s inequality yields the confidence region $CI^{R}_{1-\alpha_{21}}(\lambda)$ for the set $\mathcal{R}^{\kappa,\tau,c}(\lambda)$:

$$
CI^{R}_{1-\alpha_{21}}(\lambda) \equiv \left[\hat{\rho}_v^l(\lambda; 1-\frac{\alpha_{21}}{2}), \hat{\rho}_v^u(\lambda; 1-\frac{\alpha_{21}}{2})\right]\text{ such that } \lim \inf_{n \to \infty} \Pr[\mathcal{R}^{\kappa,\tau,c}(\lambda) \subseteq CI^{R}_{1-\alpha_{21}}(\lambda)] \geq 1 - \alpha_{21}.
$$

$CI^{R}_{1-\alpha_{21}}(\lambda)$ is a valid, but conservative, confidence region for $\rho \in \mathcal{R}^{\kappa,\tau,c}(\lambda)$. To conduct inference on $\rho$ directly, we invert a test statistic that combines the lower and upper bounds. This yields an asymptotically valid $1-\alpha_{21}$ (e.g. 95%) confidence regions $CI_{1-\alpha_{21}}(\lambda)$ for the parameter $\rho$ that is partially identified in $\mathcal{R}^{\kappa,\tau,c}(\lambda)$:

$$
\lim \inf_{n \to \infty} \Pr[\rho \in CI_{1-\alpha_{21}}(\lambda)] \geq 1 - \alpha_{21}.
$$
In particular, we apply the results in Chernozhukov, Lee, and Rosen (2013, theorem 4 and example 1) for estimation and inference with parametrically estimated bounding functions in a “saturated” model with a finite number of intersections. To select \( c_{1-\alpha_{21}}(\lambda) \) and \( c_{1-\alpha_{21}}^u(\lambda) \) and construct the bias-adjusted estimates \( \hat{\rho}_o^l(\lambda; 1-\alpha_{21}) \) and \( \hat{\rho}_o^u(\lambda; 1-\alpha_{21}) \) and the confidence region \( CI_{1-\alpha_{21}}(\lambda) \), we implement their algorithm 1. For brevity, we describe the details of the algorithm in Online Appendix B.1.

5.2.2 Estimated Nuisance Parameters

When \( \Lambda_6 \) is imposed, \( \lambda \) must be estimated and the confidence regions must be adjusted to account for this estimation. Since \( \lambda \) is a function of \( \pi \), we have that \( \sqrt{n}(\hat{\lambda} - \lambda) \xrightarrow{d} N(0, \Sigma_\lambda) \), where \( \Sigma_\lambda \) obtains using the delta method, and the estimators \( \hat{\lambda} \) and \( \hat{\Sigma}_\lambda \) are the plug-in estimators that use \( \hat{\pi} \). We then construct a \( 1 - \alpha_{22} \) confidence region \( \Lambda_{1-\alpha_{22}} \) for \( \lambda \in \mathbb{R}^{2T \times 1} \) by inverting the Wald statistic which has an asymptotic \( \chi^2_{2T} \) distribution:

\[
\Lambda_{1-\alpha_{22}} = \{ \ell : \sqrt{n}(\hat{\lambda} - \ell)'\hat{\Sigma}_\lambda^{-1}\sqrt{n}(\hat{\lambda} - \ell) \leq c_{1-\alpha_{22}}^\lambda \}
\]

where \( c_{1-\alpha_{22}}^\lambda \) is the \( 1 - \alpha_{22} \) quantile of \( \chi^2_{2T} \). By Proposition 3 of Chernozhukov, Rigobon, and Stoker (2010), we form the union over \( \ell \in \Lambda_{1-\alpha_{22}} \) to obtain the bias-corrected estimators

\[
\hat{\rho}_o^l(1 - \alpha_2) = \min_{\ell \in \Lambda_{1-\alpha_{22}}} \hat{\rho}_o^l(\ell; 1 - \alpha_{21}) \quad \text{and} \quad \hat{\rho}_o^u(1 - \alpha_2) = \max_{\ell \in \Lambda_{1-\alpha_{22}}} \hat{\rho}_o^u(\ell; 1 - \alpha_{21})
\]

where \( \alpha_2 = \alpha_{21} + \alpha_{22} \), such that:

\[
\Pr[\hat{\rho}_o^l(1 - \alpha_2) \leq \rho_o^l] \geq 1 - \alpha_2 - o(1) \quad \text{and} \quad \Pr[\rho_o^u \leq \hat{\rho}_o^u(1 - \alpha_2)] \geq 1 - \alpha_2 - o(1),
\]

as well as the \( 1 - \alpha_2 \) (e.g. 95\%) confidence regions \( CI_{1-\alpha_2}^\rho \) for \( \rho \in \mathbb{R}^{\kappa, \tau, \mathbf{c}} \):

\[
CR_{1-\alpha_2}^\rho = \bigcup_{\ell \in \Lambda_{1-\alpha_{22}}} CI_{1-\alpha_{21}}^\rho(\ell) \text{ such that } \liminf_{n \to \infty} \Pr[\rho \in CR_{1-\alpha_2}^\rho] \geq 1 - \alpha_2
\]

Note that if \( CR_{1-\alpha_2}^\rho = \emptyset \) then we reject, at the \( 1 - \alpha_2 \) significance level, the assumptions imposed in \( \Lambda_1-\Lambda_6 \). For example, if \( CR_{0.95}^\rho = \emptyset \) when \( \mathbf{c} = 0 \) then one rejects (under \( \Lambda_1-\Lambda_5 \)) that \( \sigma_q^2 \) is diagonal. Otherwise, imposing tighter restrictions on \( (\kappa, \tau, \mathbf{c}) \) can yield a tighter confidence region. This depends on the extent of the identification gain from imposing \( \Lambda_4-\Lambda_6 \) as well as on the precision of the estimates, including the nuisance parameters \( \lambda \). For

\[\text{See also Chernozhukov, Kim, Lee, and Rosen (2015).}\]
example, if the sign of $\sigma^{-2}_W \sigma_{Y_j \hat{S}_h}$ is imprecisely estimated then forming the union over $\Lambda_{1-\alpha_{22}}$ may effectively mute the impact of the $\sigma_{\eta_j \eta_h}$ restriction in $A_6$ on $CR_{1-\alpha_2}^\rho$.

In the empirical application, we report the confidence regions $CR_{0.5}^\rho$, which conveys information similar to the half-median-unbiased bound estimates, as well as $CR_{0.95}^\rho$. For this, we set $\alpha_{22} = 0.02$ when $\lambda$ must be estimated (under $A_6$) and let $\alpha_{21} = 0.48$ or $\alpha_{21} = 0.03$ respectively. Otherwise, we set $\alpha_{22} = 0$ when $\lambda$ is irrelevant and let $\alpha_{21} = 0.5$ or $\alpha_{21} = 0.05$ respectively.

5.3 Inference on $\delta_j$, $\beta_{jl}$, and $\Gamma_{jh}$

Each identification region for $\delta_j$, $\beta_{jl}$, and $\Gamma_{jh}$ for $j, h = 1, ..., p$, $j < h$, and $l = 1, ..., k$ in Corollary 3.3 is of the form

$$\theta \in H^{k, r, c} = \{H(\pi; r) : r \in R^{k, r, c}\},$$

where $R^{k, r, c}$ is the identification region for $\rho$ under a $(k, r, c)$ configuration, and $H(\cdot; r)$ is a function of $\pi$, given in Theorem 3.1. For example,

$$D_j^{k, r, c} = \left\{\frac{1}{r} b_{\hat{Y}_j, \hat{W}} : r \in R^{k, r, c}\right\}.$$

Using the delta method, we have that for each $r \in (0, 1]$, the estimator $H(\hat{\pi}; r)$ for $H(\pi; r)$ is consistent and asymptotically normally distributed:

$$\sqrt{n}(H(\hat{\pi}; r) - H(\pi; r)) \xrightarrow{d} N(0, \nabla_\pi H(\pi; r) \Sigma \nabla_\pi H(\pi; r)^\prime).$$

For brevity, Online Appendix B.2 gives the expressions for $H(\pi; r)$ and $\nabla_\pi H(\pi; r)$ for each of the parameters $\delta_j$, $\beta_{jl}$, and $\Gamma_{jh}$. If $R^{k, r, c}$ is known then, by proposition 2 of Chernozhukov, Rigobon, and Stoker (2010), one can construct a confidence region for $\theta$ by forming the union of $CR_{1-\alpha_1}^\rho(r)$ over $r \in R^{k, r, c}$. When $R^{k, r, c}$ is estimated, the confidence region must be adjusted accordingly. Using the $1 - \alpha_2$ confidence region $CR_{1-\alpha_2}^\rho$ for $\rho \in R^{k, r, c}$, we construct an asymptotically valid $1 - \alpha_1 - \alpha_2$ confidence region $CR_{1-\alpha_1-\alpha_2}^\rho$ for $\theta \in H^{k, r, c}$ by applying Proposition 3 of Chernozhukov, Rigobon, and Stoker (2010) to form the union:

$$CR_{1-\alpha_1-\alpha_2}^\rho = \bigcup_{r \in CR_{1-\alpha_2}^\rho} CR_{1-\alpha_1}^\rho(r).$$

In the empirical application, we report the confidence regions $CR_{0.5}^\rho$ and $CR_{0.95}^\rho$ for $\delta_j$ and $\beta_{jl}$ (or the vector $(\beta_{jl}, ..., \beta_{pl})^\prime$). For this, we set $\alpha_{21} = \alpha_{22} = 0.02$ when $\lambda$ must be estimated and let $\alpha_1 = 0.46$ or $\alpha_1 = 0.01$ respectively. Otherwise, we set $(\alpha_{21}, \alpha_{22}) = (0.02, 0)$ when $\lambda$ is irrelevant and let $\alpha_1 = 0.48$ or $\alpha_1 = 0.03$ respectively.
6 Tobin’s q and Corporate Investment, Saving, and Debt

How does a firm’s cash flow affect its investment, saving, and debt? After accounting for a firm’s marginal q, q theory predicts that cash flow does not affect a firm’s investment (under the classical assumptions\textsuperscript{13} Tobin’s q is a sufficient statistic for the optimal investment policy). Further, given marginal q, various theoretical models predict that cash flow may affect a firm’s saving and debt. For instance, Almeida, Campello, and Weisbach (2004) study a model, in which cash flow is unrelated to productivity shocks and physical capital depreciates completely in a single period, that predicts that cash flow affects a firm’s saving positively. On the other hand, under the assumptions that cash flow may be related to productivity and that physical capital may depreciate partially in a single period, the model in Riddick and Whited (2009) predicts that the effect of cash flow on a firm’s saving is negative. Similarly, tradeoff theory (see e.g. Miller, 1977) predicts that a firm with a high cash flow faces a lower expected bankruptcy cost and borrows more whereas pecking order theory (see e.g. Myers and Majluf, 1984) postulates that a firm with a high cash flow borrows less because external financing is costly relative to internal funds.

Because marginal q is unobserved, researchers often employ a measure of Tobin’s q as a proxy for it. The literature imposes various assumptions on the measurement error in Tobin’s q and reports findings that are sometimes contradictory. For example, Erickson and Whited (2000, 2012) apply the econometric method in Erickson and Whited (2002), which uses higher order moments\textsuperscript{14} to point identify the equation coefficients, and cannot reject that the effect of cash flow on investment is zero, thereby corroborating the prediction of q theory. Almeida, Campello, and Galvao (2010) use lagged variables in a panel structure as instrumental variables to address the measurement error in Tobin’s q and find that cash flow affects investment positively, contradicting the theoretical prediction in the absence of financing frictions (see also Fazzari, Hubbard, and Petersen, 1988; Gilchrist and Himmelberg 1995; Love, 2003). Similarly, using regression analysis, Almeida, Campello, and Weisbach (2004) find that a firm’s cash flow affects its saving positively whereas Riddick and Whited (2009) use higher order moments to account for measurement error in Tobin’s q and find

\textsuperscript{13}This result assumes quadratic investment adjustment costs, constant return to scale, perfect competition, and an efficient financial market (see Hayashi, 1982).

\textsuperscript{14}Erickson and Whited (2002) strengthen $A_2$-$A_3$ to require $\varepsilon$, $\eta$, and $(X', U)'$ to be jointly independent.
that cash flow affects saving negatively. Last, Rajan and Zingales (1995) and Hennessy and Whited (2005) study firm profitability and cash flow respectively and find that either variable affects debt negatively (see also Gomes and Schmid (2010)). Erickson, Jiang and Whited (2014) corroborate this finding for profitability when using higher order moments to account for measurement error.

We build on this literature and apply this paper’s framework to examine the identification gain that results from considering the investment, saving, and debt equations jointly under the classical measurement error assumption. Further, we analyze the sensitivity of the empirical estimates to restrictions on the extent of the measurement error, the fit of the model, or the signs of the correlations among the cross-equation disturbances.

6.1 Data

We follow the literature closely in selecting the sample and constructing the variables (see e.g. Almeida and Campello, 2007; Erickson and Whited, 2012; Erickson, Jiang, and Whited, 2014). Specifically, we use data from Compustat on industrial firms between 1970 to 2017. We remove financial firms (Standard Industrial Classification (SIC) code 6000 to 6999) and regulated firms (SIC code 4900 to 4999). To exclude small firms, 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. 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. For each cross section, we construct the variables as follows and normalize them by the firm’s total assets. We define investment as capital expenditure (CAPX) normalized by the beginning-of-the-period total assets AT. Saving is defined as a one-year change in cash and short-term investments (CHE) normalized by the beginning-of-the-period AT. We use gross debt to define leverage as short and long-term debt (DLTT+DLC) divided by the current AT. We measure (lagged) Tobin’s Q at the beginning of

\[^{15}\text{Specifically, we apply 4 firm filters: INDFMT=INDL (industrial), DATAFMT=STD (standardized data reporting), POPSRC=D (domestic (North American)), and CONSOL=C (consolidated).}\]

\[^{16}\text{We deflate flow variables by the firm's beginning-of-the-period (i.e. lagged) total assets and stock variables by the current period's total assets.}\]

\[^{17}\text{The investment literature deflates the variables by either the firm's capital or its total assets (see e.g. Erickson and Whited, 2012). Since we also consider the saving and debt equations, we construct Tobin's q as the “market-to-book ratio” and deflate all the variables by total assets, as is common in these literatures (e.g. Riddick and Whited (2009) and Erickson, Jiang, and Whited, (2014)).}\]
the period by \( \frac{(PRCC_F \times \text{CSHO}) + AT - \text{CEQ} - TXDB}{AT} \) 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 the beginning-of-the-period AT. Further, we define firm size as the natural logarithm of real net sales (SALE). In Section 6.7, we also account for asset tangibility, defined by the total net property, plant and equipment (PPENT) divided by the current AT. We delete firm-year observations with missing data on one of these variables. Last, we winsorize the smallest and largest percentile of the variables in the panel in order to limit the impact of outliers. The final sample is an unbalanced panel of 161,959 firm-year observations, with 3,375 firms per year on average. Table 2 reports the summary statistics for the panel variables.

### 6.2 Specification and Sequentially Stronger Assumptions

We impose the standard assumptions A₁-A₃ where \( Y₁, Y₂, \) and \( Y₃ \) denote investment, saving, and debt respectively, the measured Tobin's q \( W \) serves as a proxy for the unobserved marginal q \( U \), and \( X₁ \) and \( X₂ \) denotes cash flow and firm size respectively. Section 6.7 augments \( X \) to include asset tangibility \( X₃ \).

We begin by applying our framework to each cross section in our sample separately. This allows the equation coefficients to vary across years. For example, Erickson, Jiang and Whited (2014) provide evidence suggesting that the assumption that the slope coefficients are constant over time may not hold. To illustrate our results, Sections 6.3 and 6.4 focus on the middle year in our sample, 1993. We report the results for all the cross sections in Section 6.5 and when using the full panel in Section 6.6.

---

18 Alternatively, the literature sometimes examine the effect of profitability (defined by operating income before depreciation (OIBDP) normalized by the beginning-of-the-period AT) on e.g. debt. Cash flow and profitability are highly correlated in our sample.

19 Applying these standard sample selection and data cleaning rules facilitates comparing the paper's results, which analyze the investment, saving, and debt equations jointly, to the literature. Nevertheless, we note that trimming and winsorizing the data in the presence of measurement error may introduce other types of biases (see e.g. Bollinger and Chandra, 2005). It would be of interest to study the consequences of these common practices on the estimates in these three equations. To keep the scope of the paper focused and manageable, we leave this study to other work.

20 We follow the saving and debt literatures and condition on firm size (see e.g. Almedia, Campello, and Weisbach (2004), Riddick and Whited (2009), and Erickson, Jiang and Whited (2014)).

21 We treat \( Y \) and \( X \) as perfectly measured whereas we let \( W \) measure \( U \) with error. It is of interest to extend the analysis to allow several or all variables to be measured with error (see e.g. Erickson, Jiang, and Whited, 2014; Abel, 2018), e.g. due to different accounting practices. Nevertheless, we note that, unlike \( Y \) and \( X \), the expected marginal return on capital \( U \) (marginal q) is intrinsically unobserved.
We report the bounds $S_j^{\kappa,\tau}$, $j = 1, 2, 3$, that analyze each equation separately, and the bounds $J^{\kappa,\tau,c}$, that analyze the equations jointly, under sequentially stronger configurations of the auxiliary assumptions. We consider two configuration for $\kappa$ and $\tau$ in $A_4$ and $A_5$ (Section 6.4 conducts a sensitivity analysis that examines how the identification regions vary as $\kappa$ ranges from 0 to $\infty$). The first sets $\kappa = \infty$ and $\tau = (1, 1, 1)'$ so that $A_4$ and $A_5$ are not binding. For example, this yields the bounds $S_j$ and $J$ under $A_1$-$A_3$ only. The second sets $\kappa$ and $\tau$ such that the estimated $\kappa^*$ and $\tau^*$ are $\hat{\kappa}^* = 0.5$ and $\hat{\tau}^* = (0.9, 0.9, 0.9)'$. Imposing the lower bound $\kappa^* = 0.5 \leq R^2_{W,U}$ on the reliability ratio assumes that the measured Tobin’s $q$ is a moderately accurate proxy for marginal $q$, with at least half of its variance due to marginal $q$. For instance, this coincides with the largest reliability ratio estimate (0.473 with standard error 0.064) for the market-to-book ratio obtained using the fifth order cumulant estimator in Erickson, Jiang, and Whited (2014, table 5). Setting $\tau^*_j = 0.9$ is arguably conservative and assumes that, in each equation, the coefficient of determination would not exceed 0.9 had there been no measurement error. For each $(\kappa, \tau)$ configuration, we consider three configurations in $A_6$. The first leaves the covariances among the disturbances unrestricted so that $A_6$ does not bind. The second sets $c$ such that the investment and saving disturbances are negatively correlated, the investment and debt disturbances are positively correlated, and the saving and debt disturbances are negatively correlated. This holds if, given marginal $q$, cash flow, and size, firms that invest more than the average tend to hold more than the average debt and if firms that have relatively large savings tend to invest and borrow less than the average. The third and strongest configuration for $A_6$ sets $c^* = 0$ so that the variance matrix of the disturbances is diagonal. For instance, this rules out that the disturbances contain a firm-specific component that simultaneously influences the firm’s investment, saving, and debt.

6.3 Results Using the Middle Cross Section

Table 3 reports the 50% and 95% confidence regions in 1993 for $\rho$, $\delta_j$, and $\beta_{jl}$ when $\kappa = \infty$ and $\tau = (1, 1, 1)'$, with $A_4$ and $A_5$ not binding. Column 1 reports the results corresponding to the single equation bounds $S_j$. This yields different identification regions for $\rho$ across the investment, saving, and debt equations and wide bounds on the cash flow coefficients $\beta_{jl}$ in each of these equations. Column 2 reports the results for the joint-equations bounds $J$. Con-
sidering the three equations jointly yields considerably tighter identification regions than the
single-equations bounds. For example, the single-equation 50% and 95% confidence regions
for the effect of cash flow on saving are \([-\infty, 0.180]\) and \((-\infty, 0.214)\) respectively whereas
the corresponding joint-equation confidence regions are \([-0.112, 0.180]\) and \((-0.235, 0.214)\).
Nevertheless, in year 1993, the 95% confidence region for each of the effects of cash flow on
investment, saving and debt in \(J\) contains 0. For example, the 95% confidence region for
the effect \(\beta_{11}\) of cash flow on investment is \((-0.363, 0.270)\). Column 3 reports the results
for \(J^c\) where \(c\) in \(A_6\) restricts the signs of the correlations among the disturbances. This
yields confidence regions comparable to \(J\). In this case, the identification gain from im-
posing the sign restrictions in \(A_6\) is offset by the decrease in the precision of the estimates.
Column 4 reports the (IV-type) results under \(J^c^*\) where \(c^* = 0\) in \(A_6\). We do not reject
this specification in year 1993 and obtain the 95% confidence region \((0.050, 0.244)\) for the
net-of-X signal to total variance ratio \(\rho\) (note that the 50% confidence region for \(\rho\) is empty).
Last, column 5 reports the results from the regression estimator which would be consistent
if there is no measurement error in Tobin’s \(q\). The regression estimates for \(\delta_j, j = 1, 2, 3,\) are
possibly attenuated relative to the bounds in \(S_j, J, J^c,\) and \(J^c^*\). Further, the regression
estimates that cash flow affects investment and saving positively \((\beta_{11} > 0\) and \(\beta_{21} > 0)\) and
debt negatively \((\beta_{31} < 0)\).

Table 4 illustrates the consequences of imposing \(A_4\) and \(A_5\) and reports the results for
year 1993 when \(\kappa = 1.057\) and \(\tau = (0.888, 0.895, 0.895)'\), so that the estimated \(\kappa^*\) and \(\tau^*\) are
0.5 and \((0.9, 0.9, 0.9)\)' in this case, \(A_4 (0.5 \leq R^2_W)\) forces the identification regions \(S^\kappa,\tau_j\),
\(J^\kappa,\tau,\) and \(J^\kappa,\tau,c^*\) to essentially coincide. The bounds imply that cash flow affects investment
and saving positively and debt negatively. Last, the 95% confidence region for \(J^\kappa,\tau,c^*\) is
empty and the data rejects this specification at the 5% level. Thus, in year 1993, under
A1-A3 and \(A_5\), imposing a moderate lower bound \(\kappa^* = 0.5\) on the reliability ratio of Tobin’s
\(q\) is incompatible with the assumption \(c^* = 0\) that the cross-equation disturbances are uncorrelated.

6.4 Sensitivity Analysis Using the Middle Cross Section

If \(\kappa = 0\) then there is no measurement error and \((\delta, \beta, \Gamma)\) is point identified. Next, we study
the sensitivity of the identification regions for the cash flow coefficients \(\beta_{j1}\) for \(j = 1, 2, 3\) to
deviations from \(\kappa = 0\). For this, we set \(\tau = (1, 1, 1)'\) in \(A_5\) and control directly the extent
of the measurement error by varying \( \kappa \) in \( A_4 \). Using the sample from the middle year 1993, Figure 2 plots the 50\% and 95\% confidence regions for the partially identified \( \beta_{11}, \beta_{21} \) and \( \beta_{31} \) as \( \kappa \) ranges from 0 to \( \infty \) (or equivalently as \( \kappa^* \) ranges from 1 to \( R^2_{W,X} \)). It plots the regions under \( A_1-A_4 \) when each equation is analyzed separately (\( S^\kappa_j \)), the three equations are analyzed jointly (\( J^\kappa \)), and the three equations are analyzed jointly under the sign restrictions in \( A_6 \) (\( J^{\kappa,c} \)). The 95\% confidence region for the effect \( \beta_{11}, \beta_{21}, \) or \( \beta_{31} \) of cash flow contains 0 only if the reliability ratio \( R^2_{W,U} \) is possibly low. For instance, this occurs for \( J^\kappa \) when \( \kappa^* \) is at most 15.8\% (\( \kappa \geq 6.46 \)), 15.9\% (\( \kappa \geq 6.40 \)), and 22.7\% (\( \kappa \geq 3.87 \)) respectively. Otherwise, \( \beta_{11} \) and \( \beta_{21} \) are each significantly estimated to be positive and \( \beta_{31} \) to be negative.

Table 5 studies the joint consequences of measurement error on the identification of the coefficients (\( \beta_{11}, \beta_{21}, \beta_{31} \)) on cash flow in the three equations. This safeguards against maintaining empirical conclusions about \( \beta_{11}, \beta_{21}, \) and \( \beta_{31} \) that rest implicitly on contradictory inference, derived from each equation separately, on the extent of the measurement error in Tobin’s q. Further, it enables testing theories that consider multiple outcomes simultaneously. For this, we compute a 95\% joint confidence region for (\( \beta_{11}, \beta_{21}, \beta_{31} \)) under \( A_1-A_4 \) and report the smallest (in a fine grid) \( \kappa \), and the corresponding largest \( \kappa^* \), such that a null hypotheses about (\( \beta_{11}, \beta_{21}, \beta_{31} \)) is not rejected. In particular, if the reliability ratio of the proxy exceeds \( \kappa^* \), the reported threshold value for \( \kappa^* \), then the null hypothesis is rejected. Table 5 considers the 8 possible null hypotheses for the signs of the elements of (\( \beta_{11}, \beta_{21}, \beta_{31} \)).

In the one extreme, if the lower bound on the reliability ratio of Tobin’s q does not exceed 14.7\% then none of these 8 hypotheses is rejected. In the other extreme, if the reliability ratio of Tobin’s q is at least 24.7\% then any joint theory of investment, saving, and debt that does not predict \( 0 < \beta_{11}, 0 < \beta_{21}, \) and \( \beta_{31} < 0 \) is rejected under the maintained assumptions. For instance, under \( A_1-A_4 \), the joint effects of cash flow on investment, saving, and debt can be zero only if the measured Tobin’s q is possibly a noisy proxy for marginal q, with a reliability ratio less than 16.7\%.

Taken together, Figure 2 and Table 5 suggest that either the observed Tobin’s q is a moderately accurate proxy for marginal q and cash flow affects investment and saving positively and debt negatively or the observed Tobin’s q is possibly a noisy proxy for marginal q and the effect of cash flow on investment, saving, and debt may be zero. Under the maintained assumptions, the results reject hypotheses that presume e.g. q theory (\( \beta_{11} = 0 \)) while maintaining that the measured Tobin’s q is an accurate proxy for marginal q.
6.5 Results Using each Cross Section

For each cross section in our sample (years 1970 to 2017), Figure 3 plots the 50% and 95% confidence regions for $\beta_{j1}$ (the coefficients on cash flow) when it is partially identified in $B_{j1}$, $B_{j1}^{c*}$, or $B_{j1}^{\kappa,\tau,c}$, $j = 1, 2, 3$. The first column reports the joint equations bounds for $\beta_{j1} \in B_{j1}$ under $A_{1}-A_{3}$. In some years, the 95% confidence region for the coefficient on cash flow in the investment, saving, or debt equations contain zero. Nevertheless, for the other years, the 95% confidence region for the effect of cash flow falls in the positive range for investment or saving and in the negative range for debt. The second column report the bounds for $\beta_{j1} \in B_{j1}^{c*}$ under $A_{1}-A_{3}$ and the diagonal restriction $c^* = 0$ in $A_{6}$. We reject this specification in 19 of the 48 years at the 96% level (i.e. $CR_{0.96}^\rho = \emptyset$ and hence $CR_{0.95}^\theta = \emptyset$). When nonempty, the confidence regions for $\beta_{j1} \in B_{j1}^{c*}$ yield mixed results across different years. The third column reports the bounds for $\beta_{j1} \in B_{j1}^{\kappa,\tau,c}$ under $A_{1}-A_{6}$ where, for each year, we set $\kappa$ and $\tau$ such that the estimated $\kappa^*$ and $\tau^*$ are 0.5 and $(0.9, 0.9, 0.9)$ and we impose the sign restrictions $c$ described in Section 6.2. Setting the estimated $\kappa^*$ to 0.5 assumes that the observed Tobin’s $q$ proxy is moderately accurate. We reject this specification at the 96% level in 5 years (1973, 1974, 1978, 1984, and 2008). For most of the remaining years, the 95% confidence region for the effect of cash flow falls in the positive range for investment or saving and in the negative range for debt. Interestingly, we note that, in the graphs corresponding to $B_{j1}^{\kappa,\tau,c}$ in Figure 3, the time trends in the effects of cash flow on investment and saving appear relatively flat. In contrast, the magnitude of the effect of cash flow on debt diminishes over time. We leave investigating this time-series trend to other work.

6.6 Results Using the Full Panel

Although the paper’s framework does not require panel data, we illustrate how it can be applied to the full panel. As in e.g. Almeida, Campello, and Galvao (2010) and Erickson, Jiang, and Whited (2014), we now assume that the slope coefficients are constant over time and we maintain that the data on firms are missing at random from certain years of the unbalanced panel. We note that imposing assumptions on the serial correlation of the measurement error may generate instruments that can point identify the system coefficients (for example, Almeida, Campello, and Galvao (2010) employ similar panel data methods to estimate the coefficient on cash flow in the investment equation). To keep the scope of the paper focused, we leave a detailed study of using panel data to estimate a system of
equations with mismeasured variables to other work. Here, we provide a basic extension of our framework to the panel case, as summarized in Online Appendix C.

We treat the number of time periods in the panel as fixed and the number of firms to be large. After stacking each firm’s observations, our analysis proceeds analogously to the cross section case, with the robust standard errors for $\pi$ clustered at the firm level. We consider the case without fixed effects as well as when the outcome equations include year and firm fixed effects. In the latter case, we include the year indicator variables in $X$ and we remove the firm fixed effects by applying a within transformation\(^{23}\). We note that in this case the auxiliary assumptions $A_4-A_6$ should be interpreted relative to the within-transformed variables\(^{24}\). See Online Appendix C for further details.

Table 6 replicates the analysis in Tables 3 and 4 using the full panel and reports the results for the cash flow coefficients. As in the cross section analysis, the joint equation bounds improve substantially over the single equation bounds. Specifically, for the specification under $A_1$-$A_3$ without fixed effects, the 95% confidence regions for the effect of cash flow $\beta_{j1} \in B_{j1}$ falls in the positive range for investment. Imposing the $A_6$ sign restrictions encoded in $c$ tightens the bounds further ($\beta_{j1} \in B_{j1}^c$) and the effect of cash flow on saving (debt) is now estimated to be positive (negative) at the 95% level. On the other hand, we reject at the 96% level the specification that imposes the diagonal restriction in $A_6$ ($\beta_{j1} \in B_{j1}^{c^*}$). We also report the bounds when $A_4$-$A_5$ set $\kappa$ and $\tau$ such that $\kappa^*$ and $\tau^*$ are estimated to be 0.5 and $(0.9, 0.9, 0.9)'$ respectively. This yields 95% confidence regions that are close to the regression estimates, whereby cash flow is estimated to affect investment and saving positively and debt negatively. Last, when including year and firm fixed effects in the equations, $B_{j1}$ and $B_{j1}^c$ become wider whereas imposing the strong diagonal restriction in $A_6$ yield results ($B_{j1}^{c^*}$) that correspond to a very low net-of-$X$ signal to total variance ratio (with the 95% confidence region $[0.048, 0.071]$). Here too, further imposing $A_4$ and $A_5$ yields the empty $B_{j1}^{\kappa,\tau,c^*}$ region and the bounds $B_{j1}^{\kappa,\tau}$ and $B_{j1}^{\kappa,\tau,c}$ that are close to the regression estimates.

### 6.7 Accounting for Asset Tangibility

Similarly to e.g. Hennessy and Whited (2005), the specification for the debt equation above does not condition on the tangibility of the firm’s assets. Following certain specifications for

\(^{23}\)Alternatively, one can consider (first-)differencing the data.

\(^{24}\)One may consider imposing assumptions on the serial correlation of the variables to facilitate relating the (sensitivity analysis) restrictions imposed on the within-transformed variables via $(\kappa, \tau, c)$ to equivalent (or sufficient) restrictions on the level variables.
the debt equation (e.g. Rajan and Zingales (1995) and Erickson, Jiang and Whited (2014)), we replicate our analysis after augmenting $X$ to include $X_3$, the firm’s asset tangibility. We do not require that $X_3$ is excluded from the investment and saving equations - instead, we allow $X_3$ to affect all the system’s outcome variables. The analysis yields results that are qualitatively similar in certain respects to the results above. Specifically, Figure 4 in the Online Appendix replicates Figure 3, after augmenting $X$ with $X_3$, and yields results that share similar features. As in Figure 3, the bounds $B_{j1}$ are either inconclusive about the sign of $\beta_{j1}$ or they fall in the positive (negative) range for investment and saving (debt).

We note that, after accounting for asset tangibility, we reject the diagonal specification in $A_1$-$A_3$ and $A_6$ ($\beta_{j1} \in B_{j1}^c$), reported in the second column, in 5 (as opposed to 19) of the 48 years. Further, the bounds under $A_1$-$A_6$ ($\beta_{j1} \in B_{j1}^{c,\tau,c}$) in the last column remain close to the regression estimates and this specification is now rejected in 11 (as opposed to 5) years.

Last, Table 7 in the Online Appendix replicates the panel data analysis in Table 6 after augmenting $X$ with $X_3$ and, here too, the results share similar features to those in Table 6.

To summarize the differences, for the specification without fixed effects, the 95% confidence region for the effect of cash flow on investment (debt) in $B_{j1}$ now contains zero (falls in the negative range). Further, $B_{j1}^c$ and $B_{j1}^{c,\tau,c}$ are no longer empty and are comparable to $B_{j1}^c$ and $B_{j1}^{c,\tau,c}$ and to the regression estimates. The results for the specification with year and firm fixed effects are similar to those in Table 6.

7 Conclusion

This paper studies the identification of the coefficients in a system of linear equations that share an explanatory variable that is measured with classical error. We characterize the sharp identification regions for the coefficients and demonstrate the identification gain that results from analyzing the equations jointly as opposed to separately. To tighten these regions and conduct a sensitivity analysis, we characterize the sharp identification regions under any configuration of three auxiliary assumptions that weaken benchmark point-identifying assumptions. The first weakens the assumption of “no measurement error” by imposing an upper bound on the net-of-the covariates “noise to signal” ratio. The second controls the fit of the model by imposing an upper bound on the coefficients of determination that would obtain in each equation had there been no measurement error. The third weakens the assumption that the covariances among the cross-equation disturbances are zero by
specifying their signs, if at all. Further, we demonstrate how this joint-equation analysis facilitates testing hypotheses involving coefficients from multiple equations. For inference, we implement results on intersection bounds. Using data from Compustat, we apply our framework to study the effects of cash flow on the investment, saving, and debt of corporate firms in the US when the observed Tobin’s q serves as an error-laden proxy for marginal q. We find that analyzing the equations jointly, as opposed to separately, tightens the identification regions considerably and sometimes permits recovering the sign of the effects of cash flow without imposing stronger assumptions. Further, the effects of cash flow on investment, saving, and debt can be zero if and only if Tobin’s q is possibly a noisy proxy for marginal q, with a low reliability ratio. Otherwise, cash flow affects investment and saving positively and debt negatively.

Several extensions are of interest. It would be useful to extend this paper’s econometrics framework to accommodate multiple latent variables, a nonlinear specification, or weaker assumptions on the measurement error. Further, the paper’s empirical results call for the development of theoretical models that jointly determine the firm’s investment, saving, and debt. Also, the results stress the benefits of improved measures of Tobin’s q in identifying the investment, saving, and debt equation coefficients (see e.g. Erickson and Whited (2005, 2008) and Peters and Taylor (2017)). Another inquiry would investigate further the estimated decrease over time in the magnitude of the effect of cash flow on debt.
Figure 1: Identification regions $S_j$ (light) for $j = 1, 2, 3$, $J$, $J^{\kappa, \tau}$, and $J^{\kappa, \tau, c}$ (dark) for $\kappa = 1$, $\tau = (0.7, 0.7, 0.7)'$ and $c$ set to the (correct) sign restrictions $r_{\eta_1, \eta_2} \leq 0$, $r_{\eta_1, \eta_3} \geq 0$, $r_{\eta_2, \eta_3} \leq 0$. The plus, asterisk, and cross signs correspond to the population parameters, $b_{Y; (W, X)'}$, and $J^c$ respectively, where $c^*$ incorrectly sets $\sigma_{\eta_1, \eta_2} = 0$ (with $\sigma_{\eta_1, \eta_3}$ and $\sigma_{\eta_2, \eta_3}$ unrestricted).
Table 1: Numerical Example

<table>
<thead>
<tr>
<th>DGP</th>
<th>$S^s_{\tau}$</th>
<th>$J^{k,\tau}$</th>
<th>$J^{k,\tau,c}$</th>
<th>$J^{k,\tau,c}$</th>
<th>$b_{Y(W,X')}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa \to \infty$ and $\tau = (1,1)'$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.526</td>
<td>[0.315, 1.000]</td>
<td>[0.441, 1.000]</td>
<td>[0.441, 0.604]</td>
<td>0.604</td>
</tr>
<tr>
<td>$\delta_1$</td>
<td>0.700</td>
<td>[0.369, 1.171]</td>
<td>[0.369, 0.835]</td>
<td>[0.610, 0.835]</td>
<td>0.610</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>1.000</td>
<td>[0.551, 1.316]</td>
<td>[0.871, 1.316]</td>
<td>[0.871, 1.086]</td>
<td>1.086</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>0.700</td>
<td>[0.543, 0.811]</td>
<td>[0.655, 0.811]</td>
<td>[0.655, 0.730]</td>
<td>0.730</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.526</td>
<td>[0.342, 1.000]</td>
<td>[0.441, 1.000]</td>
<td>[0.441, 0.604]</td>
<td>0.604</td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>1.050</td>
<td>[0.553, 1.253]</td>
<td>[0.553, 1.253]</td>
<td>[0.915, 1.253]</td>
<td>0.915</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>0.850</td>
<td>[0.308, 1.324]</td>
<td>[0.656, 1.324]</td>
<td>[0.656, 0.979]</td>
<td>0.979</td>
</tr>
<tr>
<td>$\beta_{22}$</td>
<td>0.950</td>
<td>[0.761, 1.116]</td>
<td>[0.882, 1.116]</td>
<td>[0.882, 0.995]</td>
<td>0.995</td>
</tr>
<tr>
<td>$\kappa = 1$ and $\tau = (0.7,0.7,0.7)'$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.526</td>
<td>[0.500, 1.000]</td>
<td>[0.500, 1.000]</td>
<td>[0.500, 0.604]</td>
<td>0.604</td>
</tr>
<tr>
<td>$\delta_3$</td>
<td>0.840</td>
<td>[0.442, 1.643]</td>
<td>[0.442, 1.003]</td>
<td>[0.732, 1.003]</td>
<td>0.732</td>
</tr>
<tr>
<td>$\beta_{31}$</td>
<td>1.100</td>
<td>[0.334, 1.479]</td>
<td>[0.945, 1.479]</td>
<td>[0.945, 1.203]</td>
<td>1.203</td>
</tr>
<tr>
<td>$\beta_{32}$</td>
<td>1.200</td>
<td>[0.932, 1.333]</td>
<td>[1.146, 1.333]</td>
<td>[1.146, 1.236]</td>
<td>1.236</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.526</td>
<td>[0.500, 1.000]</td>
<td>[0.500, 1.000]</td>
<td>[0.500, 0.604]</td>
<td>0.604</td>
</tr>
<tr>
<td>$\delta_4$</td>
<td>1.050</td>
<td>[0.553, 1.256]</td>
<td>[0.553, 1.256]</td>
<td>[0.915, 1.253]</td>
<td>0.915</td>
</tr>
<tr>
<td>$\beta_{31}$</td>
<td>0.850</td>
<td>[0.797, 1.324]</td>
<td>[0.797, 1.324]</td>
<td>[0.797, 0.979]</td>
<td>0.979</td>
</tr>
<tr>
<td>$\beta_{32}$</td>
<td>0.950</td>
<td>[0.931, 1.116]</td>
<td>[0.931, 1.116]</td>
<td>[0.931, 0.995]</td>
<td>0.995</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.526</td>
<td>[0.500, 1.000]</td>
<td>[0.500, 1.000]</td>
<td>[0.500, 0.604]</td>
<td>0.604</td>
</tr>
<tr>
<td>$\delta_5$</td>
<td>0.840</td>
<td>[0.442, 0.884]</td>
<td>[0.442, 0.884]</td>
<td>[0.732, 0.884]</td>
<td>0.732</td>
</tr>
<tr>
<td>$\beta_{31}$</td>
<td>1.100</td>
<td>[1.058, 1.479]</td>
<td>[1.058, 1.479]</td>
<td>[1.058, 1.203]</td>
<td>1.203</td>
</tr>
<tr>
<td>$\beta_{32}$</td>
<td>1.200</td>
<td>[1.185, 1.333]</td>
<td>[1.185, 1.333]</td>
<td>[1.185, 1.236]</td>
<td>1.236</td>
</tr>
</tbody>
</table>

This table reports population identification regions and point estimates. $\frac{\beta^2}{\sigma^2} = 0.89$ and $R^2_{W,Y} = 0.44$. $c$ correctly sets $(c_{12}, c_{12}) = (c_{23}, c_{23}) = (-1,0)$ and $(c_{13}, c_{13}) = (0,1)$ whereas $c^*$ incorrectly sets $(c^*_{12}, c^*_{12}) = (0,0)$ and $(c^*_{13}, c^*_{13}) = (c^*_{23}, c^*_{23}) = (-1,1)$. 

32
Table 2: Summary statistics based on 161,959 firm-year observations in an unbalanced panel from year 1970 to 2017, with an average of 3,375 firms per year. In each year, investment, saving, debt, cash flow, and asset tangibility 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>Saving</th>
<th>Debt</th>
<th>Tobin’s Q</th>
<th>Cash Flow</th>
<th>Firm Size</th>
<th>Tangibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>0.084</td>
<td>0.009</td>
<td>0.261</td>
<td>1.633</td>
<td>0.071</td>
<td>5.327</td>
<td>0.354</td>
</tr>
<tr>
<td>std dev</td>
<td>0.098</td>
<td>0.106</td>
<td>0.214</td>
<td>1.147</td>
<td>0.137</td>
<td>1.949</td>
<td>0.236</td>
</tr>
<tr>
<td>min</td>
<td>0.002</td>
<td>-0.298</td>
<td>0.000</td>
<td>0.526</td>
<td>-0.536</td>
<td>0.391</td>
<td>0.022</td>
</tr>
<tr>
<td>max</td>
<td>0.593</td>
<td>0.541</td>
<td>1.016</td>
<td>7.364</td>
<td>0.390</td>
<td>10.216</td>
<td>0.925</td>
</tr>
</tbody>
</table>
Table 3: Bounds for the Investment, Saving, and Debt Equations for Year 1993. $\kappa = \infty$ and $\tau = (1, 1, 1)'$.

<table>
<thead>
<tr>
<th></th>
<th>$S_j$</th>
<th>$J$</th>
<th>$J^c$</th>
<th>$J_c^*$</th>
<th>$b_{Y,[W,X]'r}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho$</td>
<td>[0.037 , 1.000]</td>
<td>[0.076 , 1.000]</td>
<td>[0.069 , 1.000]</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.022 , 1.000)</td>
<td>(0.059 , 1.000)</td>
<td>(0.051 , 1.000)</td>
<td>(0.050 , 0.244)</td>
<td>-</td>
</tr>
<tr>
<td>$\delta_1$</td>
<td>[0.016 , 1.131]</td>
<td>[0.016 , 0.359]</td>
<td>[0.016 , 0.409]</td>
<td>[0.055 , 0.400]</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>(0.013 , 1.327)</td>
<td>(0.012 , 0.421)</td>
<td>(0.012 , 0.497)</td>
<td>(0.041 , 0.486)</td>
<td>(0.013 , 0.022)</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>[-1.299 , 0.242]</td>
<td>[-0.240 , 0.242]</td>
<td>[-0.309 , 0.242]</td>
<td>[-0.297 , 0.187]</td>
<td>0.228</td>
</tr>
<tr>
<td></td>
<td>(-1.692 , 0.270)</td>
<td>(-0.363 , 0.270)</td>
<td>(-0.484 , 0.278)</td>
<td>(-0.467 , 0.229)</td>
<td>(0.190 , 0.266)</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>[-0.010 , 0.051]</td>
<td>[-0.010 , 0.009]</td>
<td>[-0.010 , 0.012]</td>
<td>[-0.008 , 0.012]</td>
<td>-0.009</td>
</tr>
<tr>
<td></td>
<td>(-0.011 , 0.068)</td>
<td>(-0.011 , 0.015)</td>
<td>(-0.012 , 0.020)</td>
<td>(-0.010 , 0.019)</td>
<td>(-0.011 , -0.007)</td>
</tr>
<tr>
<td>$\rho$</td>
<td>[0.009 , 1.000]</td>
<td>[0.076 , 1.000]</td>
<td>[0.069 , 1.000]</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.001 , 1.000)</td>
<td>(0.059 , 1.000)</td>
<td>(0.051 , 1.000)</td>
<td>(0.050 , 0.244)</td>
<td>-</td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>[0.007 , $\infty$]</td>
<td>[0.007 , 0.212]</td>
<td>[0.007 , 0.242]</td>
<td>[0.025 , 0.237]</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>(0.003 , $\infty$)</td>
<td>(0.003 , 0.288)</td>
<td>(0.002 , 0.351)</td>
<td>(0.008 , 0.343)</td>
<td>(0.004 , 0.015)</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>[-$\infty$ , 0.180]</td>
<td>[-0.112 , 0.180]</td>
<td>[-0.154 , 0.181]</td>
<td>[-0.146 , 0.154]</td>
<td>0.165</td>
</tr>
<tr>
<td></td>
<td>(-$\infty$ , 0.214)</td>
<td>(-0.235 , 0.214)</td>
<td>(-0.328 , 0.223)</td>
<td>(-0.317 , 0.203)</td>
<td>(0.120 , 0.209)</td>
</tr>
<tr>
<td>$\beta_{22}$</td>
<td>[-0.003 , $\infty$]</td>
<td>[-0.003 , 0.008]</td>
<td>[-0.003 , 0.010]</td>
<td>[-0.002 , 0.010]</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>(-0.004 , $\infty$)</td>
<td>(-0.004 , 0.013)</td>
<td>(-0.004 , 0.016)</td>
<td>(-0.003 , 0.016)</td>
<td>(-0.004 , -0.000)</td>
</tr>
<tr>
<td>$\rho$</td>
<td>[0.036 , 1.000]</td>
<td>[0.076 , 1.000]</td>
<td>[0.069 , 1.000]</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.026 , 1.000)</td>
<td>(0.059 , 1.000)</td>
<td>(0.051 , 1.000)</td>
<td>(0.050 , 0.244)</td>
<td>-</td>
</tr>
<tr>
<td>$\delta_3$</td>
<td>[-1.637 , -0.033]</td>
<td>[-0.695 , -0.033]</td>
<td>[-0.791 , -0.033]</td>
<td>[-0.773 , -0.115]</td>
<td>-0.035</td>
</tr>
<tr>
<td></td>
<td>(-1.822 , -0.029)</td>
<td>(-0.774 , -0.029)</td>
<td>(-0.903 , -0.028)</td>
<td>(-0.883 , -0.097)</td>
<td>(-0.041 , -0.030)</td>
</tr>
<tr>
<td>$\beta_{31}$</td>
<td>[-0.310 , 1.981]</td>
<td>[-0.310 , 0.648]</td>
<td>[-0.311 , 0.787]</td>
<td>[-0.202 , 0.762]</td>
<td>-0.287</td>
</tr>
<tr>
<td></td>
<td>(-0.359 , 2.566)</td>
<td>(-0.359 , 0.893)</td>
<td>(-0.372 , 1.136)</td>
<td>(-0.276 , 1.103)</td>
<td>(-0.352 , -0.221)</td>
</tr>
<tr>
<td>$\beta_{32}$</td>
<td>[-0.077 , 0.012]</td>
<td>[-0.025 , 0.012]</td>
<td>[-0.031 , 0.012]</td>
<td>[-0.030 , 0.008]</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td>(-0.100 , 0.015)</td>
<td>(-0.035 , 0.015)</td>
<td>(-0.045 , 0.016)</td>
<td>(-0.044 , 0.012)</td>
<td>(0.008 , 0.015)</td>
</tr>
</tbody>
</table>

The sample size is 3,454 observations. $Y_1$, $Y_2$, and $Y_3$ denote Investment, Saving, and Debt respectively and $X = [\text{Cash Flow, Firm Size}]$. $c$ sets $(\zeta_{12}, \zeta_{13}) = (\zeta_{23}, \zeta_{23}) = (-1, 0)$ and $(\zeta_{13}, \zeta_{13}) = (0, 1)$ whereas $c^* = 0$. 50% and 95% confidence regions are in brackets and parentheses respectively.
Table 4: Bounds for the Investment, Saving, and Debt Equations for Year 1993. $\kappa = 1.0574$ ($\hat{\kappa}^* = 50\%$) and $\tau = (0.8880, 0.8954, 0.8949)'$ ($\hat{\tau}^* = (0.9, 0.9, 0.9)'$).

<table>
<thead>
<tr>
<th></th>
<th>$S^\kappa,\tau$</th>
<th>$J^\kappa,\tau$</th>
<th>$J^\kappa,\tau,c$</th>
<th>$J^\kappa,\tau,c'$</th>
<th>$b_{Y,(W,X)'}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho$</td>
<td>[0.486, 1.000]</td>
<td>[0.486, 1.000]</td>
<td>[0.486, 1.000]</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.486, 1.000)</td>
<td>(0.486, 1.000)</td>
<td>(0.486, 1.000)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\delta_1$</td>
<td>[0.016, 0.040]</td>
<td>[0.016, 0.040]</td>
<td>[0.016, 0.040]</td>
<td>-</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>(0.013, 0.047)</td>
<td>(0.013, 0.047)</td>
<td>(0.012, 0.049)</td>
<td>-</td>
<td>(0.013, 0.022)</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>[0.189, 0.242]</td>
<td>[0.189, 0.242]</td>
<td>[0.188, 0.242]</td>
<td>-</td>
<td>0.228</td>
</tr>
<tr>
<td></td>
<td>(0.160, 0.270)</td>
<td>(0.160, 0.270)</td>
<td>(0.151, 0.278)</td>
<td>-</td>
<td>(0.190, 0.266)</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>[-0.010, -0.008]</td>
<td>[-0.010, -0.008]</td>
<td>[-0.010, -0.008]</td>
<td>-</td>
<td>-0.009</td>
</tr>
<tr>
<td></td>
<td>(-0.011, -0.006)</td>
<td>(-0.011, -0.006)</td>
<td>(-0.012, -0.006)</td>
<td>-</td>
<td>(-0.011, -0.007)</td>
</tr>
<tr>
<td>$\rho$</td>
<td>[0.486, 1.000]</td>
<td>[0.486, 1.000]</td>
<td>[0.486, 1.000]</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.486, 1.000)</td>
<td>(0.486, 1.000)</td>
<td>(0.486, 1.000)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>[0.007, 0.023]</td>
<td>[0.007, 0.023]</td>
<td>[0.007, 0.024]</td>
<td>-</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>(0.003, 0.032)</td>
<td>(0.003, 0.032)</td>
<td>(0.002, 0.034)</td>
<td>-</td>
<td>(0.004, 0.015)</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>[0.135, 0.180]</td>
<td>[0.135, 0.180]</td>
<td>[0.134, 0.181]</td>
<td>-</td>
<td>0.165</td>
</tr>
<tr>
<td></td>
<td>(0.099, 0.214)</td>
<td>(0.099, 0.214)</td>
<td>(0.090, 0.223)</td>
<td>-</td>
<td>(0.120, 0.209)</td>
</tr>
<tr>
<td>$\beta_{22}$</td>
<td>[-0.003, -0.001]</td>
<td>[-0.003, -0.001]</td>
<td>[-0.003, -0.001]</td>
<td>-</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>(-0.004, 0.000)</td>
<td>(-0.004, 0.000)</td>
<td>(-0.004, 0.001)</td>
<td>-</td>
<td>(-0.004, -0.000)</td>
</tr>
<tr>
<td>$\rho$</td>
<td>[0.486, 1.000]</td>
<td>[0.486, 1.000]</td>
<td>[0.486, 1.000]</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.486, 1.000)</td>
<td>(0.486, 1.000)</td>
<td>(0.486, 1.000)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\delta_3$</td>
<td>[-0.077, -0.033]</td>
<td>[-0.077, -0.033]</td>
<td>[-0.077, -0.033]</td>
<td>-</td>
<td>-0.035</td>
</tr>
<tr>
<td></td>
<td>(-0.086, -0.029)</td>
<td>(-0.086, -0.029)</td>
<td>(-0.088, -0.028)</td>
<td>-</td>
<td>(-0.041, 0.030)</td>
</tr>
<tr>
<td>$\beta_{31}$</td>
<td>[-0.310, -0.213]</td>
<td>[-0.310, -0.213]</td>
<td>[-0.311, -0.212]</td>
<td>-</td>
<td>-0.287</td>
</tr>
<tr>
<td></td>
<td>(-0.359, -0.162)</td>
<td>(-0.359, -0.162)</td>
<td>(-0.372, -0.148)</td>
<td>-</td>
<td>(-0.352, -0.221)</td>
</tr>
<tr>
<td>$\beta_{32}$</td>
<td>[0.008, 0.012]</td>
<td>[0.008, 0.012]</td>
<td>[0.008, 0.012]</td>
<td>-</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td>(0.005, 0.015)</td>
<td>(0.005, 0.015)</td>
<td>(0.005, 0.016)</td>
<td>-</td>
<td>(0.008, 0.015)</td>
</tr>
</tbody>
</table>

The sample size is 3,454 observations. $Y_1$, $Y_2$, and $Y_3$ denote Investment, Saving, and Debt respectively and $X = [\text{Cash Flow}, \text{Firm Size}]$. $c$ sets $(\xi_{12}, \xi_{23}) = (0, 1)$ and $(\xi_{13}, \xi_{23}) = (0, 1)$ whereas $c^* = 0$. 50% and 95% confidence regions are in brackets and parentheses respectively.
Figure 2: 50% (dark) and 95% (light) confidence regions for $\beta_{j1}$ (cash flow coefficient) for $j = 1, 2, 3$ (investment, saving, and debt) for year 1993. We set $\tau = (1, 1, 1)'$ and consider the regions $S^\kappa$, $J^\kappa$, and $J^{\kappa,c}$ where $\kappa \in [0, \infty)$ and c sets $(\zeta_{12}, \bar{c}_{12}) = (\zeta_{23}, \bar{c}_{23}) = (-1, 0)$ and $(\zeta_{13}, \bar{c}_{13}) = (0, 1)$. The vertical dashed line marks the smallest $\kappa$ (largest $\kappa^*$) value such that the 95% confidence region contains zero. For $j = 1, 2, 3$, this corresponds respectively to 6.457 (0.158), 6.405 (0.159), and 3.866 (0.227) for $S^\kappa$ or $J^\kappa$ and to 6.018 (0.166), 5.662 (0.173), and 3.552 (0.241) for $J^{\kappa,c}$.
Table 5: Joint test for the possible signs of the components of $(\beta_{11}, \beta_{21}, \beta_{31})$ under $A_1$-$A_4$ at the 5% level for year 1993. $\pi^*$ is the largest $\kappa^*$ such that $H_0$ is not rejected. $\kappa$ is the smallest $\kappa$ such that $H_0$ is not rejected.

<table>
<thead>
<tr>
<th>$H_0$</th>
<th>$\beta_{11}$</th>
<th>$\beta_{21}$</th>
<th>$\beta_{31}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>$-$</td>
<td>$-$</td>
<td>$+$</td>
</tr>
<tr>
<td>$\beta_{31}$</td>
<td>$-$</td>
<td>$+$</td>
<td>$+$</td>
</tr>
<tr>
<td>$\pi^*$</td>
<td>0.147</td>
<td>0.157</td>
<td>0.167</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>7.107</td>
<td>6.483</td>
<td>5.949</td>
</tr>
</tbody>
</table>
Figure 3: 50% (dark) and 95% (light) confidence regions for $\beta_{j1}$ (cash flow) for $j = 1, 2, 3$ (investment, saving, and debt) from year 1970 to 2017. We consider the regions $B_{j1}$, $B_{j1}^c$, and $B_{j1}^{\kappa, \tau, c}$ where $c^* = 0$, $\kappa$ and $\tau$ are such that $\hat{\kappa}^* = 0.5$ and $\hat{\tau}^* = (0.9, 0.9, 0.9)'$, and $c$ is such that $(\zeta_{12}, \zeta_{12}) = (\zeta_{23}, \zeta_{23}) = (-1, 0)$ and $(\zeta_{13}, \zeta_{13}) = (0, 1)$. The shaded vertical bars indicate years in which the maintained assumptions are rejected at the 96% level.
Table 6: Bounds on the Cash Flow Coefficients in the Investment, Saving, and Debt Equations Using the Full Panel

<table>
<thead>
<tr>
<th></th>
<th>$S^\kappa,\tau$</th>
<th>$J^\kappa,\tau$</th>
<th>$J^\kappa,\tau, c^*$</th>
<th>$J^\kappa,\tau, c^*$</th>
<th>$b_{Y,(W,X')}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Results without fixed effects for $\kappa = \infty$ and $\tau = (1,1,1)'$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>[-0.366, 0.199]</td>
<td>[0.079, 0.199]</td>
<td>[0.189, 0.199]</td>
<td>-</td>
<td>0.196</td>
</tr>
<tr>
<td></td>
<td>(-0.464, 0.206)</td>
<td>(0.058, 0.206)</td>
<td>(0.181, 0.208)</td>
<td>-</td>
<td>(0.187, 0.205)</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>[-0.362, 0.126]</td>
<td>[-0.031, 0.126]</td>
<td>[0.116, 0.126]</td>
<td>-</td>
<td>0.124</td>
</tr>
<tr>
<td></td>
<td>(-0.443, 0.132)</td>
<td>(-0.056, 0.132)</td>
<td>(0.109, 0.133)</td>
<td>-</td>
<td>(0.116, 0.131)</td>
</tr>
<tr>
<td>$\beta_{31}$</td>
<td>[-0.358, 0.679]</td>
<td>[-0.358, -0.041]</td>
<td>[-0.358, -0.334]</td>
<td>-</td>
<td>-0.351</td>
</tr>
<tr>
<td></td>
<td>(-0.371, 0.875)</td>
<td>(-0.371, 0.020)</td>
<td>(-0.375, -0.316)</td>
<td>-</td>
<td>(-0.369, -0.332)</td>
</tr>
<tr>
<td>Results with year and firm fixed effects for $\kappa = \infty$ and $\tau = (1,1,1)'$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>[0.187, 0.199]</td>
<td>[0.187, 0.199]</td>
<td>[0.187, 0.199]</td>
<td>-</td>
<td>0.196</td>
</tr>
<tr>
<td></td>
<td>(0.180, 0.206)</td>
<td>(0.180, 0.206)</td>
<td>(0.181, 0.208)</td>
<td>-</td>
<td>(0.187, 0.205)</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>[0.113, 0.126]</td>
<td>[0.113, 0.126]</td>
<td>[0.116, 0.126]</td>
<td>-</td>
<td>0.124</td>
</tr>
<tr>
<td></td>
<td>(0.107, 0.132)</td>
<td>(0.107, 0.132)</td>
<td>(0.109, 0.133)</td>
<td>-</td>
<td>(0.116, 0.131)</td>
</tr>
<tr>
<td>$\beta_{31}$</td>
<td>[-0.358, -0.328]</td>
<td>[-0.358, -0.328]</td>
<td>[-0.358, -0.334]</td>
<td>-</td>
<td>-0.351</td>
</tr>
<tr>
<td></td>
<td>(-0.371, -0.313)</td>
<td>(-0.371, -0.313)</td>
<td>(-0.375, -0.316)</td>
<td>-</td>
<td>(-0.369, -0.332)</td>
</tr>
<tr>
<td>Results without fixed effects for $\hat{\kappa}^* = 0.5$ and $\hat{\tau}^* = (0.9,0.9,0.9)'$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>[-0.718, 0.132]</td>
<td>[-0.565, 0.132]</td>
<td>[-0.583, 0.132]</td>
<td>[-0.583, -0.252]</td>
<td>0.129</td>
</tr>
<tr>
<td></td>
<td>(-0.758, 0.138)</td>
<td>(-0.598, 0.138)</td>
<td>(-0.625, 0.139)</td>
<td>(-0.625, -0.228)</td>
<td>(0.122, 0.137)</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>[-3.262, 0.175]</td>
<td>[-0.304, 0.175]</td>
<td>[-0.317, 0.176]</td>
<td>[-0.317, -0.079]</td>
<td>0.172</td>
</tr>
<tr>
<td></td>
<td>(-3.533, 0.183)</td>
<td>(-0.344, 0.183)</td>
<td>(-0.368, 0.186)</td>
<td>(-0.368, -0.048)</td>
<td>(0.161, 0.182)</td>
</tr>
<tr>
<td>$\beta_{31}$</td>
<td>[-0.372, $\infty$]</td>
<td>[-0.372, -0.238]</td>
<td>[-0.372, -0.234]</td>
<td>[-0.323, -0.234]</td>
<td>-0.366</td>
</tr>
<tr>
<td></td>
<td>(-0.383, $\infty$)</td>
<td>(-0.383, -0.186)</td>
<td>(-0.388, -0.167)</td>
<td>(-0.386, -0.167)</td>
<td>(-0.381, -0.351)</td>
</tr>
<tr>
<td>Results with year and firm fixed effects for $\hat{\kappa}^* = 0.5$ and $\hat{\tau}^* = (0.9,0.9,0.9)'$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>[0.081, 0.132]</td>
<td>[0.081, 0.132]</td>
<td>[0.081, 0.132]</td>
<td>-</td>
<td>0.129</td>
</tr>
<tr>
<td></td>
<td>[0.076, 0.138]</td>
<td>[0.076, 0.138]</td>
<td>[0.074, 0.139]</td>
<td>-</td>
<td>(0.122, 0.137)</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>[0.137, 0.175]</td>
<td>[0.137, 0.175]</td>
<td>[0.137, 0.176]</td>
<td>-</td>
<td>0.172</td>
</tr>
<tr>
<td></td>
<td>(0.128, 0.183)</td>
<td>(0.128, 0.183)</td>
<td>(0.125, 0.186)</td>
<td>-</td>
<td>(0.161, 0.182)</td>
</tr>
<tr>
<td>$\beta_{31}$</td>
<td>[-0.372, -0.353]</td>
<td>[-0.372, -0.353]</td>
<td>[-0.372, -0.353]</td>
<td>-</td>
<td>-0.366</td>
</tr>
<tr>
<td></td>
<td>(-0.383, -0.341)</td>
<td>(-0.383, -0.341)</td>
<td>(-0.386, -0.337)</td>
<td>-</td>
<td>(-0.381, -0.351)</td>
</tr>
</tbody>
</table>

The sample is an unbalanced panel of 161,959 firm-year observations. $Y_1$, $Y_2$, and $Y_3$ denote Investment, Saving, and Debt respectively and $X = [\text{Cash Flow}, \text{Firm Size}]$. When year fixed effects are included, $X$ also includes year indicator variables. When firm fixed effects are included, the equations’ variables undergo a within transformation. $c$ sets $(c_{12}, c_{12}) = (c_{23}, c_{23}) = (-1, 0)$ and $(c_{13}, c_{13}) = (0, 1)$ whereas $c^* = 0$. Robust standard errors for $\pi$ are clustered by firm. 50% and 95% confidence regions are in brackets and parentheses respectively.
References


