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

Karim Chalak*[†]
University of Virginia

Daniel Kim
University of Pennsylvania

January 3, 2019

Abstract

This paper studies identifying the coefficients in a system of linear equations that share a mismeasured explanatory variable. We characterize the sharp identification regions for the coefficients under the classical measurement error assumption and demonstrate the identification gain that results from analyzing the equations jointly as opposed to separately. Further, to conduct a sensitivity analysis, we derive 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 "noise to signal" ratio. 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 disturbances is diagonal by specifying the signs of the correlations among the cross-equation disturbances, if at all. For inference, the paper implements results on intersection bounds. Using data from COMPUSTAT, the paper applies its framework to study the effects of cash flow on the investment, saving, and debt of corporate firms in the US when Tobin's q is used as an error-laden proxy for a firm's marginal q.

Keywords: *debt, investment, measurement error, multiple equations, partial identification, saving, sensitivity analysis, Tobin's q.*

JEL codes: C31, G30.

---

*Karim Chalak (corresponding author), Department of Economics, University of Virginia, chalak@virginia.edu. Daniel Kim, The Wharton School, University of Pennsylvania, kimdanie@wharton.upenn.edu.

†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, 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.

1 Introduction

Sometimes a mismeasured explanatory variable appears in multiple linear equations of interest which are nonetheless estimated separately. What, if any, is the identification gain that results from analyzing the system's equations jointly, as opposed to separately, when a common explanatory variable suffers from classical measurement error? What are useful auxiliary assumptions that can help identify the system's coefficients? How are the system's coefficients jointly sensitive to deviations from these auxiliary assumptions? To address these questions, we show how the identification region for each equation's coefficients 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.

Specifically, we study identifying the coefficients in 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 (2018), 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 (net-of-the covariates) "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 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 estimating 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 Tobin's q (the ratio of the firm's market value to its assets' replacement value, 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 a noisy proxy for marginal q, with a low reliability ratio. Otherwise, if Tobin's q is a moderately accurate proxy then cash flow affects investment and saving positively and debt negatively.

More broadly, this paper's econometrics 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 is generated as follows.

Assumption $\mathbf{A_1}$ *Data Generation: (i) Let* $(\underset{k\times 1}{X}', \underset{1\times 1}{W}, \underset{p\times 1}{Y}')'$ *be a random vector with a finite variance. (ii) Let a structural system generate the random variables* $\underset{p\times 1}{\eta}, \underset{1\times 1}{\varepsilon}, \underset{1\times 1}{U}, X, W,$ *and* Y *such that*

$$Y' = X'\beta + U\delta + \eta' \quad \text{and} \quad W = U + \varepsilon \tag{1}$$

with constant slope coefficients. The researcher observes realizations of $(X', W, Y')'$ *but not* *of* (U, η', ε).

A_1 decomposes the proxy W into the "signal" component[1] U and the "noise" or error ε. We are interested in identifying the effects δ_j and β_j of U and X on Y_j for $j = 1, ..., p$ as encoded in the j^{th} outcome equation,

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

X denotes the observed determinants that drive Y. Our framework does not require the presence of these 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. When $Cov[(\eta', \varepsilon)', X] = 0$, as we will assume shortly, excluding a component of X from a Y_j equation 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. Here, the challenge in identifying (δ, β) is due to U being unobserved and possibly correlated with X. In particular, we maintain two standard assumptions about the other unobservables η and ε. First, the "disturbance" η is uncorrelated with $(X', U)'$.

Assumption $\mathbf{A_2}$ *Uncorrelated Disturbance:* $Cov[\eta, (X', U)'] = 0$.

Second, the measurement error ε is uncorrelated with (X', U, η).

Assumption $\mathbf{A_3}$ *Uncorrelated measurement error:* $Cov[\varepsilon, (X', U, \eta)'] = 0$.

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[2] (β, δ). We do not require these stronger assumptions[3].

[1]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 $\delta \equiv \frac{\gamma}{\psi}$ of the coefficients on V may be (partially) identified. To ease the notation, we use the simpler representation in which $U \equiv V\psi$.

[2]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.

[3]For instance, unlike in Erickson and Whited (2002), A_1-A_3 allow the system variables to be jointly normally distributed.

Instead, we impose the uncorrelation assumptions A_2-A_3 and study partially identifying δ and β. 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 jointly the assumptions imposed on instrument exogeneity and measurement error. Krasker and Pratt (1986) and Erickson and Whited (2005) relax A_3 and characterize how highly correlated should W and U be in order to identify the sign of δ or of a component of β. Klepper and Leamer (1984) and Bollinger (2003) characterize the sharp identification regions for δ and β under A_1-A_3. Chalak and Kim (2018) 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. Here, 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:

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

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

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

so that by construction $E(\epsilon_{A.B|C}) = 0$ and $Cov(C, \epsilon_{A.B|C}) = 0$. 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

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

the population coefficient of determination (R-squared) from a regression[4] of A on B.

2.2 Linear Projection

Recall that under A_2-A_3, $Cov[(\eta, \varepsilon)', X] = 0$. Thus, provided σ^2_X is nonsingular, projecting W and Y onto X gives $b_{W.X} = b_{U.X}$ and

$$b_{Y.X} = \beta + b_{W.X}\delta. \tag{3}$$

Further, using $\tilde{A} \equiv \epsilon_{A.X}$ as a shorthand notation for the residual from the regression of a vector A on X, we employ the following convenient system of projected linear equations:

$$\tilde{Y}' = \tilde{U}\delta + \tilde{\eta}' \quad \text{and} \quad \tilde{W} = \tilde{U} + \tilde{\varepsilon} \tag{4}$$

to study identifying δ. The identification region for β then obtains using equation (3).

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 do not require A_4-A_6. Instead, we characterize the identification gain that results from imposing any configuration of these auxiliary assumptions.

Klepper and Leamer (1984), Klepper (1988), and Chalak and Kim (2018) employ assumptions similar to A_4 and A_5 when $p = 1$. Since we consider multiple equations, $p \geq 1$, we also study assumption A_6, introduced below. Specifically, the first auxiliary assumption weakens the "no measurement error" assumption $\sigma^2_\varepsilon = 0$ by imposing an upper bound κ on the net-of-X "noise to signal ratio."

Assumption A_4 *Bounded Net-of-X Noise to Signal Ratio:* $\sigma^2_\varepsilon \leq \kappa\sigma^2_{\tilde{U}}$ *where* $0 \leq \kappa$.

[4] If σ^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 $R^2_{A.B} = 0$.

For example, A_4 reduces to the "no measurement error" assumption $\sigma_\varepsilon^2 = 0$ when $\kappa = 0$ whereas setting $\kappa = 1$ assumes that, after projecting on X, the variance of the measurement error is at most as large as the variance of U, $\sigma_\varepsilon^2 \leq \sigma_{\tilde{U}}^2$. Given A_1-A_3, A_4 equivalently imposes a lower bound $\frac{1}{1+\kappa}$ on ρ, the net-of-X "signal to total variance ratio":

$$\frac{1}{1+\kappa} \leq \rho \equiv \frac{\sigma_{\tilde{U}}^2}{\sigma_{\tilde{W}}^2} = \frac{\sigma_{\tilde{U}}^2}{\sigma_{\tilde{U}}^2 + \sigma_\varepsilon^2}.$$

Further, since $\rho \equiv \frac{\sigma_{\tilde{U}}^2}{\sigma_{\tilde{W}}^2} = \frac{R_{W.U}^2 - R_{W.X}^2}{1 - R_{W.X}^2}$ (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 $R_{\tilde{Y}_j.\tilde{U}}^2 \equiv 1 - \frac{\sigma_{\eta_j}^2}{\sigma_{\tilde{Y}_j}^2}$ in the $\tilde{Y}_j$ equation from display (4). By A_1-A_3 and since W measures U with error, Lemma D.1 in the Online Appendix gives that $R_{\tilde{Y}_j.\tilde{W}}^2 \leq R_{\tilde{Y}_j.\tilde{U}}^2$. The second auxiliary assumption controls the fit of the model by imposing a bound τ_j on how large can $R_{\tilde{Y}_j.\tilde{U}}^2$ be.

Assumption A_5 *Bounded Net-of-X Coefficient of Determination:* $R_{\tilde{Y}_j.\tilde{U}}^2 \leq \tau_j$ *where* $0 < \tau_j$ *and* $R_{\tilde{Y}_j.\tilde{W}}^2 \leq \tau_j \leq 1$ *for* $j = 1, .., p$.

Since $R_{A.(X',B)'}^2 = \frac{\sigma_{\tilde{A}}^2}{\sigma_A^2}(R_{\tilde{A}.\tilde{B}}^2 - 1) + 1$, A_5 equivalently imposes an upper bound $\tau_j^* \equiv \frac{\sigma_{\tilde{Y}_j}^2}{\sigma_{Y_j}^2}(\tau_j - 1) + 1$ on the coefficient of determination $R_{Y_j.(X',U)'}^2 \equiv 1 - \frac{\sigma_{\eta_j}^2}{\sigma_{Y_j}^2}$ in the Y_j equation from display (2). We let $\tau \equiv (\tau_1, ..., \tau_p)'$ and $\tau^* \equiv (\tau_1^*, ..., \tau_p^*)'$.

The third auxiliary assumption weakens the assumption that σ_η^2 is diagonal by specifying the sign of the correlation r_{η_j,η_h} among the cross-equation disturbances, if at all.

Assumption A_6 *Disturbance Correlation Sign Restriction:* $\underline{c}_{jh} \leq r_{\eta_j,\eta_h} \leq \overline{c}_{jh}$ *where* $(\underline{c}_{jh}, \overline{c}_{jh}) \in \{(-1,0), (0,1), (0,0), (-1,1)\}$.

A_6 encodes the sign restrictions (if any) imposed in A_6 on the $\frac{1}{2}p(p-1)$ off-diagonal elements of σ_η^2. For example, $(\underline{c}_{jh}, \overline{c}_{jh}) = (-1,0)$ encodes that $r_{\eta_j,\eta_h} \leq 0$ whereas if A_6 does not restrict the sign of r_{η_j,η_h} then we set $(\underline{c}_{jh}, \overline{c}_{jh}) = (-1,1)$. We collect these restrictions in the matrix $\underset{\frac{1}{2}p(p-1)\times 2}{\mathbf{c}} = (\underline{c}, \overline{c})$ where $\underline{c} = (\underline{c}_{12}, ..., \underline{c}_{(p-1)p})'$ and $\overline{c} = (\overline{c}_{12}, ..., \overline{c}_{(p-1)p})'$. For example, when σ_η^2 is assumed to be diagonal, we set $\mathbf{c} = 0$.

Online Appendix A extends A_6 to A'_6 which sets $\underline{c}_{jh} \leq r_{\eta_j,\eta_h} \leq \overline{c}_{jh}$ with $-1 \leq \underline{c}_{jh} \leq \overline{c}_{jh} \leq 1$. In particular, A'_6 may restrict the sign and/or magnitude of the correlation r_{η_j,η_h}. While A'_6 is conceptually similar to A_6, the expression for the identification region under A_1-A'_6 is more complex. To ease the exposition, we report these results in the Online Appendix. Here and in the empirical analysis in Section 6, we focus on specifying the sign of r_{η_j,η_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 A_4 directly restricts the net-of-X signal to total variance ratio ρ (i.e. the extent of the measurement error), A_5 and A_6 indirectly restrict ρ. We vary κ, τ, and $\mathbf{c}$ in A_4-A_6 to conduct a sensitivity analysis that weakens the no measurement error assumption $\kappa = 0$ (or $R^2_{\tilde{Y}_j.\tilde{W}} = \tau_j$ in A_5), controls the fit of the model ($R^2_{\tilde{Y}_j.\tilde{W}} \leq \tau_j$), and weakens the assumption that σ^2_η is diagonal ($\mathbf{c} = 0$). Conversely, we study for what configuration of $(\kappa, \tau, \mathbf{c})$ does the identification region admit a plausible value or range e.g. for a component of δ or β. 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 $\kappa \to +\infty$, $\tau = (1, ..., 1)'$, or $\mathbf{c}$ is such that $(\underline{c}_{jh}, \overline{c}_{jh}) = (-1, 1)$ for all $j < h$.

3 Identification

We study identifying δ, and consequently $\beta = b_{Y.X} - b_{W.X}\delta$, 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 $Var[(\tilde{Y}', \tilde{W})']$ can be expressed as

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

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

$$b_{\tilde{Y}.\tilde{W}} = \rho\delta \qquad \text{where } \rho \equiv \frac{\sigma^2_{\tilde{U}}}{\sigma^2_{\tilde{W}}} = \frac{\sigma^2_{\tilde{U}}}{\sigma^2_{\tilde{U}} + \sigma^2_{\varepsilon}}. \tag{5}$$

Since the (net-of-X) "noise to signal ratio" ρ satisfies $0 \leq \rho \leq 1$, we obtain the classic "attenuation bias" whereby $b_{\tilde{Y}_j.\tilde{W}}$ understates the magnitude of δ_j and has its sign. If there is no measurement error ($\sigma^2_\varepsilon = 0$) then $\rho = 1$ and $b_{\tilde{Y}.\tilde{W}} = \delta$. If U and X are perfectly collinear ($\sigma^2_{\tilde{U}} = 0$) then $\rho = 0$ and $b_{\tilde{Y}.\tilde{W}}$ does not identify δ. Similarly, normalizing $\sigma^2_{\tilde{Y}}$ by $\sigma^2_{\tilde{W}}$ gives that

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

where we have that

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

For example, the normalized covariance of the cross-equation disturbances is given by

$$\Gamma_{jh} \equiv \sigma^{-2}_{\tilde{W}}\sigma_{\eta_j,\eta_h} = \sigma^{-2}_{\tilde{W}}\sigma_{\tilde{Y}_j,\tilde{Y}_h} - \delta_j\rho\delta_h. \tag{8}$$

As we show in Corollary 3.2, the system of (in)equalities (3,5,6,7) exhausts the information on $(\rho, \delta, \beta, \Gamma)$ 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 $\frac{\sigma^2_{\tilde{Y}_j}}{\sigma^2_{\tilde{W}_j}}(1-\tau_j) \leq \Gamma_{jj}$, and A_6 may specify the (weak) sign of Γ_{jh}.

When U and X are not perfectly collinear, i.e. $\rho \neq 0$, Theorem 3.1 uses equations (3,5,6) to express δ, β, and Γ as functions D, B, and G of ρ. This mapping enables characterizing the identification region for $(\rho, \delta, \beta, \Gamma)$ in terms of restrictions on ρ 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 $Var[(X', U)']$ be nonsingular so that $0 < \rho$. Then*

$$\delta = D(\rho) \equiv \frac{1}{\rho} b_{\tilde{Y}.\tilde{W}},$$

$$\beta = B(\rho) \equiv b_{Y.X} - b_{W.X}\frac{1}{\rho} b_{\tilde{Y}.\tilde{W}}, \text{ and}$$

$$\Gamma = G(\rho) \equiv \sigma^{-2}_{\tilde{W}}\sigma^2_{\tilde{Y}} - b'_{\tilde{Y}.\tilde{W}}\frac{1}{\rho} b_{\tilde{Y}.\tilde{W}}.$$

Theorem 3.1 reveals how if there is no measurement error ($\rho = 1$) then (δ, β, Γ) is point identified. Further, even when $\rho < 1$, $b_{\tilde{Y}_j.\tilde{W}} = 0$ if and only if $(\delta_j, \beta_j, \Gamma_{jh}) = (0, b_{Y_j.X}, \sigma^{-2}_{\tilde{W}}\sigma_{\tilde{Y}_j,\tilde{Y}_h})$. Similarly, if the l^{th} element $b_{W.X,l}$ of $b_{W.X}$ is 0 then $\beta_l = b_{Y.X,l}$.

Last, as discussed in Section 2, if X_l is excluded from the Y_j equation so that $\beta_{jl} = b_{Y_j.X,l} - b_{W.X,l}\frac{1}{\rho}b_{\tilde{Y}_j.\tilde{W}} = 0$ then, provided $b_{Y_j.X,l} \neq 0$, ρ is point identified and it follows that (δ, β, Γ) is also point identified.

3.2 Identification Regions

We characterize the sharp identification regions for $(\rho, \delta, \beta, \Gamma)$ under A_1-A_3 and any configuration of the auxiliary assumptions A_4-A_6 (i.e. any $(k, \tau, \mathbf{c})$ value). Corollary 3.2 states the general result. We then discuss several special cases.

Corollary 3.2 *Under the conditions of Theorem 3.1 and A_4-A_6 for $j, h = 1, ..., p$ with $j < h$, $(\rho, \delta, \beta, \Gamma)$ is partially identified in the sharp set*[5]

$$\mathcal{J}^{k,\tau,\mathbf{c}} \equiv \{(r, D(r), B(r), G(r)) : 0 \preceq G(r),\ \ \frac{1}{1+\kappa} \leq r \leq 1,\ \ \frac{\sigma^2_{\tilde{Y}_j}}{\sigma^2_{\tilde{W}_j}}(1-\tau_j) \leq G_{jj}(r),\ and$$
$$\underline{c}_{jh} \leq sgn(G_{jh}(r)) \leq \bar{c}_{jh}\ for\ j, h = 1, ..., p\ and\ j < h\}.$$

Further, ρ, δ, β, and Γ are partially identified in the sharp sets

$$\mathcal{R}^{k,\tau,\mathbf{c}} = [R^2_{\tilde{W}.\tilde{Y}}, 1] \cap [\frac{1}{1+\kappa}, 1] \cap_{j=1}^{p} [\frac{1}{\tau_j} R^2_{\tilde{W}.\tilde{Y}_j}, 1] \bigcap_{\substack{j,h=1 \\ j<h}}^{p} \mathcal{R}^{\mathbf{c}}_{jh},$$

where

$$\mathcal{R}^{\mathbf{c}}_{jh} = \begin{cases} \frac{b_{\tilde{Y}_j.\tilde{W}} b_{\tilde{Y}_h.\tilde{W}}}{\sigma^{-2}_{\tilde{W}}\sigma_{\tilde{Y}_j,\tilde{Y}_h}} & if\ (\underline{c}_{jh}, \bar{c}_{jh}) = (0,0)\ and\ \sigma^{-2}_{\tilde{W}}\sigma_{\tilde{Y}_j,\tilde{Y}_h} \neq 0 \\ (-\infty, \frac{b_{\tilde{Y}_j.\tilde{W}} b_{\tilde{Y}_h.\tilde{W}}}{\sigma^{-2}_{\tilde{W}}\sigma_{\tilde{Y}_j,\tilde{Y}_h}}] & if\ (\underline{c}_{jh}, \bar{c}_{jh}) \in \{(-1,0),(0,1)\}\ and\ sgn(\sigma^{-2}_{\tilde{W}}\sigma_{\tilde{Y}_j,\tilde{Y}_h}) \notin [\underline{c}_{jh}, \bar{c}_{jh}] \\ [\frac{b_{\tilde{Y}_j.\tilde{W}} b_{\tilde{Y}_h.\tilde{W}}}{\sigma^{-2}_{\tilde{W}}\sigma_{\tilde{Y}_j,\tilde{Y}_h}}, \infty) & if\ (\underline{c}_{jh}, \bar{c}_{jh}) \in \{(-1,0),(0,1)\}\ and\ sgn(\sigma^{-2}_{\tilde{W}}\sigma_{\tilde{Y}_j,\tilde{Y}_h}) \in [\underline{c}_{jh}, \bar{c}_{jh}]\backslash\{0\} \\ \emptyset & if\ (\underline{c}_{jh}, \bar{c}_{jh}) \neq (-1,1),\ -sgn(b_{\tilde{Y}_j.\tilde{W}} b_{\tilde{Y}_h.\tilde{W}}) \notin [\underline{c}_{jh}, \bar{c}_{jh}],\ and\ \sigma^{-2}_{\tilde{W}}\sigma_{\tilde{Y}_j,\tilde{Y}_h} = 0 \\ (-\infty, \infty) & otherwise \end{cases}$$

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

Using the system of (in)equalities (3,5,6,7) and the mappings in Theorem 3.1, Corollary 3.2 characterizes the identification region $\mathcal{J}^{k,\tau,\mathbf{c}}$ for $(\rho, \delta, \beta, \Gamma)$. As shown in the proof of

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

Corollary 3.2, $\mathcal{J}^{k,\tau,\mathbf{c}}$ is sharp since for every $(r,d,b,g) \in \mathcal{J}^{k,\tau,\mathbf{c}}$ there exists $(U^*, \eta^*, \varepsilon^*)$, with $\frac{\sigma^2_{\tilde{U}^*}}{\sigma^2_{\tilde{W}}} = r$ and $G(r) = \sigma^{-2}_{\tilde{W}} \sigma^2_{\tilde{\eta}^*}$, that satisfy A_2-A_6 and that could have generated Y and W according to A_1. Further, Corollary 3.2 derives the projections $\mathcal{R}^{k,\tau,\mathbf{c}}$, $\mathcal{D}^{k,\tau,\mathbf{c}}$, $\mathcal{B}^{k,\tau,\mathbf{c}}$, and $\mathcal{G}^{k,\tau,\mathbf{c}}$ of the joint region $\mathcal{J}^{k,\tau,\mathbf{c}}$ onto the support of the components ρ, δ, β, and Γ. Each of these projected regions is sharp - for example, for every $d \in \mathcal{D}^{k,\tau,\mathbf{c}}$ there exists $(r,d,b,g) \in \mathcal{J}^{k,\tau,\mathbf{c}}$. Thus, Corollary 3.2 exhausts the information on $(\rho, \delta, \beta, \Gamma)$ in A_1-A_6.

It is useful to examine the projected identification regions in Corollary 3.2 under sequentially stronger configurations of $(k, \tau, \mathbf{c})$. First, suppose that $(\underline{c}_{jh}, \overline{c}_{jh}) = (-1, 1)$ for all $j, h = 1, ..., p$ with $j < h$, so that A_6 is not binding. Then $\mathcal{R}^{\mathbf{c}}_{jh} = (-\infty, \infty)$. In this case, we sometimes drop the superfluous superscript $\mathbf{c}$ and obtain $\mathcal{R}^{k,\tau,\mathbf{c}} = \mathcal{R}^{k,\tau} \equiv [R^2_{\tilde{W}.\tilde{Y}}, 1] \cap [\frac{1}{1+\kappa}, 1] \cap_{j=1}^{p} [\frac{1}{\tau_j} R^2_{\tilde{W}.\tilde{Y}_j}, 1]$. If $\kappa \to \infty$ and $\tau = (1, ..., 1)'$ then A_4 and A_5 are also not binding and we sometimes drop the κ and τ superscripts. Provided $R^2_{\tilde{W}.\tilde{Y}} \neq 0$, we obtain $\mathcal{R}^{k,\tau,\mathbf{c}} = \mathcal{R} \equiv [R^2_{\tilde{W}.\tilde{Y}}, 1]$ since $\max\{R^2_{\tilde{W}.\tilde{Y}_1}, ..., R^2_{\tilde{W}.\tilde{Y}_p}\} \leq R^2_{\tilde{W}.\tilde{Y}}$. In this case, Corollary 3.2 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 3.2 gives the single-equation bounds for ρ, δ, and β under classical measurement error (see e.g. Chalak and Kim, 2018, corollary 3.5). As the dimension of Y increases, $R^2_{\tilde{W}.\tilde{Y}}$ may increase and the joint-equations bounds $\mathcal{R}$ for ρ 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. In this case, if $R^2_{\tilde{W}.\tilde{Y}} \leq \max\{\frac{1}{1+\kappa}, \frac{1}{\tau_1} R^2_{\tilde{W}.\tilde{Y}_1}, ..., \frac{1}{\tau_p} R^2_{\tilde{W}.\tilde{Y}_p}\}$ then imposing A_4 and A_5 increases the lower bound on ρ. In turn, this leads to tighter bounds on (δ, β, Γ) via the mappings in Theorem 3.1. In the limit, setting $\kappa = 0$ or $\tau_j = R^2_{\tilde{W}.\tilde{Y}_j}$ yields $\rho = 1$ and therefore point identifies (δ, β, Γ).

Next, consider imposing A_6. To illustrate how restricting the sign of the off-diagonal elements in σ^2_η can help identify δ and β, consider the Y_j and Y_h equations and substitute for $U = W - \varepsilon$ in the Y_j equation:

$$Y_j = X'\beta_j + W\delta_j - \varepsilon\delta_j + \eta_j,$$

$$Y_h = X'\beta_h + U\delta_h + \eta_h.$$

Under A$_1$-A$_3$, if $\sigma_{\eta_j,\eta_h} = 0$ then $Cov[(\varepsilon, \eta_j)', Y_h] = 0$. In this case, analyzing the Y_j and Y_h equations jointly reveals how Y_h may serve as an instrument for W to point identify $(\delta_j, \beta_j')' = b_{Y_j.(W,X')'|(Y_h,X')'}$. Indeed, Corollary 3.2 shows that, even when A$_4$-A$_5$ are not binding, if $\sigma_{\eta_j,\eta_h} = 0$ (i.e. $(\underline{c}_{jh}, \overline{c}_{jh}) = (0,0)$) then, provided $\sigma_{\tilde{W}}^{-2}\sigma_{\tilde{Y}_j,\tilde{Y}_h} \neq 0$, $\rho = \frac{b_{\tilde{Y}_j.\tilde{W}} b_{\tilde{Y}_h.\tilde{W}}}{\sigma_{\tilde{W}}^{-2}\sigma_{\tilde{Y}_j,\tilde{Y}_h}}$ is point identified. When $b_{\tilde{Y}_j.\tilde{W}|\tilde{Y}_h}$ exists and is nonzero, we can express $\rho = \frac{b_{\tilde{Y}_j.\tilde{W}}}{b_{\tilde{Y}_j.\tilde{W}|\tilde{Y}_h}}$ as the ratio of the regression and IV regression estimands. It follows from the mappings in Theorem 3.1 that the full vector of system coefficients $(\rho, \delta, \beta, \Gamma)$ is point identified, with $\delta_j = b_{\tilde{Y}_j.\tilde{W}|\tilde{Y}_h}$ and $\beta_j = b_{Y_j.X} - b_{W.X} b_{\tilde{Y}_j.\tilde{W}|\tilde{Y}_h}$ as obtains via the IV regression $b_{Y_j.(W,X')'|(Y_h,X')'}$.

What if $\sigma_{\eta_j,\eta_h} = 0$ fails? Corollary 3.2 answers this question by deriving the identification regions for ρ, δ, and β under weaker restriction in A$_6$ on the sign of $\Gamma_{jh} \equiv \sigma_W^{-2}\sigma_{\eta_j,\eta_h}$. First, if the identification region $\mathcal{G}_{jh}^{k,\tau}$ identifies the sign of Γ_{jh} when A$_6$ is not binding (i.e. when $(\underline{c}_{jh}, \overline{c}_{jh}) = (-1,1)$ for all $j < h$) then imposing the (correct) sign restriction on Γ_{jh} in A.6 is uninformative about ρ, δ, and β. Otherwise, restricting the sign of Γ_{jh} in A$_6$ can rule out a region of $\mathcal{R}^{k,\tau}$. Specifically, recall that $\mathcal{G}_{jh}^{k,\tau}$ is given by

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

Thus, provided $\sigma_{\tilde{W}}^{-2}\sigma_{\tilde{Y}_j,\tilde{Y}_h}$ is nonzero[6], $0 \in int(\mathcal{G}_{jh}^{k,\tau})$ if and only if

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

Corollary 3.2 demonstrates how restricting the sign of Γ_{jh} can rule out elements of $\mathcal{R}^{k,\tau}$ that are either smaller or larger than $\frac{b_{\tilde{Y}_j.\tilde{W}} b_{\tilde{Y}_h.\tilde{W}}}{\sigma_{\tilde{W}}^{-2}\sigma_{\tilde{Y}_j.\tilde{Y}_h}}$, as encoded in $\mathcal{R}_{jh}^{\mathbf{c}}$. In turn, this can tighten the identification regions for δ and β. Last, if Corollary 3.2 yields $\mathcal{R}^{k,\tau,\mathbf{c}} = \emptyset$ then the model is misspecified and we reject the assumptions imposed in A$_1$-A$_6$.

To conclude this section, we point out that imposing restrictions on the signs and/or magnitudes of some of the coefficients δ_j or β_{jl} may tighten the identification region of ρ, and therefore of δ, β, and Γ 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 δ or β.

[6] If $\sigma_{\tilde{W}}^{-2}\sigma_{\tilde{Y}_j,\tilde{Y}_h} = 0$ then restricting the sign of Γ_{jh} is either contradictory or uninformative about ρ, depending on the sign of $b_{\tilde{Y}_j.\tilde{W}} b_{\tilde{Y}_h.\tilde{W}}$, as encoded in $\mathcal{R}_{jh}^{\mathbf{c}}$.

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 $\underset{2\times 1}{\eta_X} \equiv (\eta_{X_1}, \eta_{X_2})'$, $\underset{2\times 1}{X} \equiv (X_1, X_2)'$, $\underset{3\times 1}{\eta} \equiv (\eta_1, \eta_2, \eta_3)'$, and $\underset{3\times 1}{Y} \equiv (Y_1, Y_2, Y_3)'$. We let η_X, U, ε, and η be jointly independent and normally distributed with mean 0 so that A_2 and A_3 hold. We allow the components of η_X (respectively η) to be correlated. It follows that (X', W, Y') is normally distributed and we can analytically express the identification regions for ρ, δ, and β in terms of the elements of $Var[(\eta'_X, U, \varepsilon, \eta')']$. In this example, we set the equation coefficients to

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

and the variances of η_X, U, ε, and η to

$$\sigma^2_\epsilon = 3, \quad \sigma^2_U = 5, \quad \sigma_{\eta X} = \begin{bmatrix} 1 & 0.14 \\ 0.14 & 1 \end{bmatrix}, \text{ and } \sigma^2_\eta = \begin{bmatrix} 1.1 & -0.31 & 0.63 \\ -0.31 & 1.99 & -0.59 \\ 0.63 & -0.59 & 2.25 \end{bmatrix}.$$

We obtain that $\rho = 0.53$ and thus any restriction $0.89 = \frac{\sigma^2_\epsilon}{\sigma^2_{\tilde{U}}} \leq \kappa$ in A_4 is valid. Further, we obtain that $R^2_{\tilde{W}.\tilde{Y}_1} = 0.31$, $R^2_{\tilde{W}.\tilde{Y}_2} = 0.34$, $R^2_{\tilde{W}.\tilde{Y}_3} = 0.27$, and $R^2_{\tilde{W}.\tilde{Y}} = 0.44$.

Using a grid search, we approximate 4 types of identification regions, illustrated in Figure 1. The first is the single-equation identification regions $\mathcal{S}_j$ that consider each Y_j equation separately. The second is the joint-equations region $\mathcal{J}$ that considers the Y equations jointly. $\mathcal{S}_j$ and $\mathcal{J}$ obtain under A_1-A_3 only (i.e. when $\kappa = \infty$, $\tau = (1, 1, 1)'$, and $(\underline{c}_{jh}, \overline{c}_{jh}) = (-1, 1)$ for all $j < h$). The third identification region is the joint-equations bounds $\mathcal{J}^{\kappa,\tau}$ that obtains under A_1-A_5, with $\kappa = 1$ and $\tau = (0.7, 0.7, 0.7)'$. The fourth region $\mathcal{J}^{\kappa,\tau,\mathbf{c}}$ obtains under A_1-A_6, with κ and τ as in $\mathcal{J}^{\kappa,\tau}$, where $\mathbf{c}$ 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 (ρ, δ_j), (ρ, β_{j1}), and (ρ, β_{j2}) spaces for $j = 1, 2, 3$. The plus sign denotes the true parameter values. Further, the asterisk corresponds to the regression estimand $b_{Y.(W,X')'}$ and the cross sign corresponds to the identification region $\mathcal{J}^{\kappa,\tau,\mathbf{c}^*}$ (the IV regression estimand) where $\mathbf{c}^*$ incorrectly sets $\sigma_{\eta_1,\eta_2} = 0$ and leaves σ_{η_1,η_3} and σ_{η_2,η_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 the single-equation and joint-equations identification regions $\mathcal{S}_j$ and $\mathcal{J}$. The second darkest region corresponds to the joint-equations region $\mathcal{J}^{\kappa,\tau}$ and yields the lower bound $\frac{1}{1+\kappa} = 0.5$ in $\mathcal{R}^{\kappa,\tau}$. Last, the darkest region corresponds to the joint-equations region $\mathcal{J}^{\kappa,\tau,\mathbf{c}}$.

Table 1 uses the analytical expressions in Section 3 to report several bounds, including those that correspond to the projections in Figure 1. The first and second columns report the sharp projections of the single-equation and joint-equations identification regions $\mathcal{S}_j^{\kappa,\tau}$ for $j = 1, 2, 3$ and $\mathcal{J}^{\kappa,\tau}$ respectively under A_1-A_5. Note that projecting $\mathcal{S}_j^{\kappa,\tau}$ yields different bounds for ρ, depending on j. The third column reports the joint-equations bounds $\mathcal{J}^{\kappa,\tau,\mathbf{c}}$ under A_1-A_6 with 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 fourth column reports the (IV regression) point estimand $\mathcal{J}^{\kappa,\tau,\mathbf{c}^*}$ where $\mathbf{c}^*$ incorrectly assumes that $\sigma_{\eta_1,\eta_2} = 0$, with σ_{η_1,η_3} and σ_{η_2,η_3} unrestricted. The last column reports the regression estimand $b_{Y.(W,X')'}$ which would point identify δ and β if there is no measurement error in W. 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 true parameter values are elements of the nested sets $\mathcal{J}^{\kappa,\tau,\mathbf{c}} \subseteq \mathcal{S}_1^{\kappa,\tau} \times ... \times \mathcal{S}_p^{\kappa,\tau}$ which become tighter as stricter valid restrictions on κ, τ, and/or $\mathbf{c}$ 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 ρ, δ_j, β_{jl}, and Γ_{jh} for $j, h = 1, ..., p$ and $l = 1, ..., k$. The procedure consists of three steps. First, we express each of the bounds in Corollary 3.2 as a function[7] of the vector of estimands[8]

$$\pi \equiv (vec(b_{Y.(W,X')'})', b'_{W.(Y,X')'}, b'_{W.(Y_1,X')'}, ..., b'_{W.(Y_p,X')'}, vec(b_{Y.X})', b'_{W.X}, \sigma^{-2}_{\tilde{W}} vec(\sigma^2_{\tilde{Y}})')',$$

where[9] $vec(\sigma^2_{\tilde{Y}})$ collects the $\frac{1}{2}p(p+1)$ variance and covariance elements of $\sigma^2_{\tilde{Y}}$. Further, we construct an estimator $\hat{\pi}$ for π 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 ρ that is partially identified in $\mathcal{R}^{\kappa,\tau,\mathbf{c}}$ for any $(\kappa, \tau, \mathbf{c})$ configuration. The last step uses the mappings, given in Theorem 3.1, that express δ_j, β_{jl}, and Γ_{jh} as functions of (π, ρ) to construct $1-\alpha$ confidence regions for the partially identified parameters δ_j, β_{jl}, and Γ_{jh}.

5.1 Estimation of π

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

$$\hat{\pi} \equiv (vec(\hat{b}_{Y.(W,X')'})', \hat{b}'_{W.(Y,X')'}, \hat{b}'_{W.(Y_1,X')'}, ..., \hat{b}'_{W.(Y_p,X')'}, vec(\hat{b}_{Y.X})', \hat{b}'_{W.X}, \hat{\sigma}^{-2}_{\tilde{W}} vec(\hat{\sigma}^2_{\tilde{Y}})'),$$

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 $\hat{\sigma}^2_A = \hat{\sigma}_{A,A}$) and the linear regression estimator and sample residuals by:

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

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 π is $\sqrt{n}$ consistent and asymptotically normally distributed. For this, let $\mu^2_A = E(AA')$ and define the square block-diagonal matrix Q:

$$Q \equiv diag\{\underset{p\times p}{I} \otimes \mu^2_{(1,W,X')'}, \mu^2_{(1,Y,X')'}, \mu^2_{(1,Y_1,X')'} ..., \mu^2_{(1,Y_p,X')'}, \underset{p\times p}{I} \otimes \mu^2_{(1,X')'}, \mu^2_{(1,X')'}, \underset{\frac{1}{2}p(p+1)\times\frac{1}{2}p(p+1)}{I} \otimes \sigma^2_{\tilde{W}}\},$$

[7] An alternative would express the bounds in Corollary 3.2 as a function of $Var[(1, Y', W, X')']$ and constructs an estimator for these moments.

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

[9] Let $\underset{m\times q}{A} = [\underset{m\times 1}{A_1}, ..., \underset{m\times 1}{A_q}]$. Then $\underset{mq\times 1}{vec(A)} \equiv (A'_1, ..., A'_q)'$. Further, if $q = m$ and A is symmetric then we let $\underset{\frac{1}{2}m(m+1)\times 1}{vec(A)} \equiv [A_{11}, ..., A_{mm}, A_{12}, ..., A_{1m}, ..., A_{(m-1)1}, ..., A_{(m-1)m}]'$ collect the diagonal and upper-diagonal elements of A.

where the moments in the diagonal blocks correspond to the estimands in π.

Theorem 5.1 *Assume $A_1(i)$ and that Q is nonsingular. Suppose further that:*
(i) $\frac{1}{n}\sum_{i=1}^{n}(1,Y_i',W_i,X_i')'(1,Y_i',W_i,X_i')\xrightarrow{p}\mu^2_{(1,Y',W,X')'}$ *and*

$$(ii)\quad n^{-1/2}\sum_{i=1}^{n}\begin{bmatrix} vec[(1,W_i,X_i')'\epsilon_{Y.(W,X')',i}] \\ (1,Y_i,X_i')'\epsilon_{W.(Y,X')',i} \\ (1,Y_{1i},X_i')'\epsilon_{W.(Y_1,X')',i} \\ \vdots \\ (1,Y_{pi},X_i')'\epsilon_{W.(Y_p,X')',i} \\ vec[(1,X_i')'\epsilon_{Y.X,i}] \\ (1,X_i')'\epsilon_{W.X,i} \\ vec(\epsilon_{Y.X,i}\epsilon_{Y.X,i}'-\sigma^2_{\tilde{Y}}) \end{bmatrix} \xrightarrow{d} N(0,\Xi) \quad \text{where } \Xi \equiv Var\begin{bmatrix} vec[(1,W,X')'\epsilon_{Y.(W,X')'}] \\ (1,Y,X')'\epsilon_{W.(Y,X')'} \\ (1,Y_1,X')'\epsilon_{W.(Y_1,X')'} \\ \vdots \\ (1,Y_p,X')'\epsilon_{W.(Y_p,X')'} \\ vec[(1,X')'\epsilon_{Y.X}] \\ (1,X')'\epsilon_{W.X} \\ vec(\epsilon_{Y.X}\epsilon_{Y.X}') \end{bmatrix}.$$

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

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

5.2 Inference on ρ

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

$$\mathcal{R}^{\kappa,\tau,\mathbf{c}}(\lambda)\equiv[\rho_o^l(\lambda),\rho_o^u(\lambda)]\equiv\cap_{v=1}^{M}[\rho_v^l(\lambda),\rho_v^u(\lambda)]\equiv\cap_{v=1}^{M}\mathcal{R}_v(\lambda),$$

which may depend on a vector of nuisance parameters $\underset{2T\times1}{\lambda}$ ($T\equiv\frac{1}{2}p(p-1)$), a function of π:

$$\underset{2T\times1}{\lambda}=(\sigma^{-2}_{\tilde{W}}\sigma_{\tilde{Y}_1,\tilde{Y}_2},\sigma^{-2}_{\tilde{W}}\sigma_{\tilde{Y}_1,\tilde{Y}_3},...,\sigma^{-2}_{\tilde{W}}\sigma_{\tilde{Y}_{p-1},\tilde{Y}_p},b_{\tilde{Y}_1.\tilde{W}}b_{\tilde{Y}_2.\tilde{W}},b_{\tilde{Y}_1.\tilde{W}}b_{\tilde{Y}_3.\tilde{W}},...,b_{\tilde{Y}_{p-1}.\tilde{W}}b_{\tilde{Y}_p.\tilde{W}}).$$

Further, for a given λ, each of the bounds $\rho_v^l(\lambda)$ and $\rho_v^u(\lambda)$ can be expressed as a function of π. For example, in the numerical example in Section 4, the identification region $\mathcal{R}^{\kappa,\tau,\mathbf{c}}$

under A_1-A_6 (with $\Gamma_{12} \leq 0$, $\Gamma_{13} \geq 0$, and $\Gamma_{23} \leq 0$) for ρ is

$$\begin{aligned}\mathcal{R}^{\kappa,\tau,\mathbf{c}}(\lambda) &= \cap_{v=1}^{8}[\rho_v^l(\lambda), \rho_v^u(\lambda)] \\ &= [R^2_{\tilde{W}.\tilde{Y}}, 1] \cap [\frac{1}{1+\kappa}, 1] \cap [\frac{1}{\tau_1} R^2_{\tilde{W}.\tilde{Y}_1}, 1] \cap [\frac{1}{\tau_2} R^2_{\tilde{W}.\tilde{Y}_2}, 1] \cap [\frac{1}{\tau_3} R^2_{\tilde{W}.\tilde{Y}_3}, 1] \\ &\cap (-\infty, \frac{b_{\tilde{Y}_1.\tilde{W}} b_{\tilde{Y}_2.\tilde{W}}}{\sigma^{-2}_{\tilde{W}} \sigma_{\tilde{Y}_1,\tilde{Y}_2}}] \cap [\frac{b_{\tilde{Y}_1.\tilde{W}} b_{\tilde{Y}_3.\tilde{W}}}{\sigma^{-2}_{\tilde{W}} \sigma_{\tilde{Y}_1,\tilde{Y}_3}}, \infty) \cap (-\infty, \frac{b_{\tilde{Y}_2.\tilde{W}} b_{\tilde{Y}_3.\tilde{W}}}{\sigma^{-2}_{\tilde{W}} \sigma_{\tilde{Y}_2,\tilde{Y}_3}}]\end{aligned}$$

where the last three intersected regions $\mathcal{R}^{\mathbf{c}}_{12}(\lambda)$, $\mathcal{R}^{\mathbf{c}}_{13}(\lambda)$, and $\mathcal{R}^{\mathbf{c}}_{23}(\lambda)$ in $\mathcal{R}^{\kappa,\tau,\mathbf{c}}(\lambda)$ obtain from Corollary 3.2 based on the signs of the nuisance parameters (here $T = 3$)

$$\underset{2T\times 1}{\lambda} = (\sigma^{-2}_{\tilde{W}} \sigma_{\tilde{Y}_1,\tilde{Y}_2}, \sigma^{-2}_{\tilde{W}} \sigma_{\tilde{Y}_1,\tilde{Y}_3}, \sigma^{-2}_{\tilde{W}} \sigma_{\tilde{Y}_2,\tilde{Y}_3}, b_{\tilde{Y}_1.\tilde{W}} b_{\tilde{Y}_2.\tilde{W}}, b_{\tilde{Y}_1.\tilde{W}} b_{\tilde{Y}_3.\tilde{W}}, b_{\tilde{Y}_2.\tilde{W}} b_{\tilde{Y}_3.\tilde{W}}).$$

Thus, λ determines whether each $\mathcal{R}^{\mathbf{c}}_{jh}(\lambda)$ is $\emptyset$, $(-\infty, \infty)$, $(-\infty, \frac{b_{\tilde{Y}_j.\tilde{W}} b_{\tilde{Y}_h.\tilde{W}}}{\sigma^{-2}_{\tilde{W}} \sigma_{\tilde{Y}_j,\tilde{Y}_h}})$, or $[\frac{b_{\tilde{Y}_j.\tilde{W}} b_{\tilde{Y}_h.\tilde{W}}}{\sigma^{-2}_{\tilde{W}} \sigma_{\tilde{Y}_j,\tilde{Y}_h}}, \infty)$.

5.2.1 Known Nuisance Parameters

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

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

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 $c^l_{1-\alpha_{21}}(\lambda)$ ($c^u_{1-\alpha_{21}}(\lambda)$) is a suitably selected critical value, discussed below, such that

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

In particular, setting $\alpha_{21} = \frac{1}{2}$ yields half-median-unbiased estimators $\hat{\rho}_o^l(\lambda; \frac{1}{2})$ and $\hat{\rho}_o^u(\lambda; \frac{1}{2})$. Using Bonferroni's inequality yields the confidence region $CI^{\mathcal{R}}_{1-\alpha_{21}}(\lambda)$ for the set $\mathcal{R}^{\kappa,\tau,\mathbf{c}}(\lambda)$:

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

$CI^{\mathcal{R}}_{1-\alpha_{21}}(\lambda)$ is a valid, but conservative, confidence region for $\rho \in \mathcal{R}^{\kappa,\tau,\mathbf{c}}(\lambda)$. To conduct inference on ρ 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 $CI^{\rho}_{1-\alpha_{21}}(\lambda)$ for the parameter ρ that is partially identified in $\mathcal{R}^{\kappa,\tau,\mathbf{c}}(\lambda)$:

$$\liminf_{n\to\infty} \Pr[\rho \in CI^{\rho}_{1-\alpha_{21}}(\lambda)] \geq 1-\alpha_{21}.$$

In particular, we apply the results in[10] 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 $c^{l}_{1-\alpha_{21}}(\lambda)$ and $c^{u}_{1-\alpha_{21}}(\lambda)$ and construct the bias-adjusted estimates $\hat{\rho}^{l}_{o}(\lambda; 1-\alpha_{21})$ and $\hat{\rho}^{u}_{o}(\lambda; 1-\alpha_{21})$ and the confidence region $CI^{\rho}_{1-\alpha_{21}}(\lambda)$, 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

In practice, λ must be estimated and the confidence regions must be adjusted to account for this estimation. Since λ is a function of π, we have that $\sqrt{n}(\hat{\lambda}-\lambda)\xrightarrow{d} N(0,\Sigma_\lambda)$, where Σ_λ 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 $\underset{2T\times 1}{\lambda}$ by inverting the Wald statistic which has an asymptotic χ^2_{2T} distribution:

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

where $c^{\lambda}_{1-\alpha_{22}}$ is the $1-\alpha_{22}$ quantile of χ^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

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

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

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

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

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

[10] See also Chernozhukov, Kim, Lee, and Rosen (2015).

Note that if $CR^{\rho}_{1-\alpha_2} = \emptyset$ then we reject, at the $1-\alpha_2$ significance level, the assumptions imposed in A_1-A_6. For example, if $CR^{\rho}_{0.95} = \emptyset$ when $\mathbf{c} = \mathbf{0}$ then one rejects (under A_1-A_5) that σ^2_{η} 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 A_4-A_6 as well as on the precision of the estimates, including the nuisance parameters λ. For example, if the sign of $\sigma^{-2}_{\tilde{W}}\sigma_{\tilde{Y}_j,\tilde{Y}_h}$ is imprecisely estimated then forming the union over $\Lambda_{1-\alpha_{22}}$ may effectively mute the impact of the σ_{η_j,η_h} restriction in A_6 on $CR^{\rho}_{1-\alpha_2}$.

In the empirical application, we report the confidence regions $CR^{\rho}_{0.5}$, which conveys similar information to the half-median-unbiased bound estimates, as well as $CR^{\rho}_{0.95}$. For this, we set $\alpha_{22} = 0.02$ and let $\alpha_{21} = 0.48$ or $\alpha_{21} = 0.03$ respectively.

5.3 Inference on δ_j, β_{jl}, and Γ_{jh}

Each identification region for δ_j, β_{jl}, and Γ_{jh} for $j, h = 1, ..., p$, $j < h$, and $l = 1, ..., k$ in Corollary 3.2 is of the form

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

where $\mathcal{R}^{\kappa,\tau,\mathbf{c}}$ is the identification region for ρ under any given $(\kappa, \tau, \mathbf{c})$ configuration, and $H(\cdot; r)$ is a function of π, given in Theorem 3.1. For example,

$$\mathcal{D}^{\kappa,\tau,\mathbf{c}}_j = \{\frac{1}{r} b_{\tilde{Y}_j.\tilde{W}} : r \in \mathcal{R}^{\kappa,\tau,\mathbf{c}}\}.$$

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:

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

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

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

In the empirical application, we report the confidence regions $CR^{\theta}_{0.5}$ and $CR^{\theta}_{0.95}$ for δ_j and β_{jl} (or the vector $(\beta_{1l}, ..., \beta_{pl})'$). For this, we set $\alpha_{21} = \alpha_{22} = 0.02$ and let $\alpha_1 = 0.46$ or $\alpha_1 = 0.01$ respectively.

6 Tobin's q in Corporate Investment, Saving, and Debt Equations

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[11], 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 either positively or negatively. For instance, Almeida, Campello, and Weisbach (2004) study a model, in which cash flow is not related 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 Tobin's q as a proxy for it. The literature imposes different assumptions on the measurement error in Tobin's q and reports contradictory findings. For example, Erickson and Whited (2000, 2012) apply the econometric method in Erickson and Whited (2002), which uses higher order moments[12]

[11] This result assumes quadratic investment adjustment costs, constant return to scale, perfect competition, and an efficient financial market (see Hayashi, 1982).

[12] Erickson and Whited (2002) strengthen A_2-A_3 to require ε, η, and $(X', U)'$ to be jointly independent.

to point identify the equation coefficients, and cannot reject that the effect of cash flow on investment may be 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 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 negatively affects debt (see also Gomes and Schmid (2010)) and 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 conduct a sensitivity analysis that studies the robustness of the empirical estimates to deviations from the no measurement error assumption, restrictions on the fit of the model, or deviations from the assumption that the variance matrix of the disturbances is diagonal.

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[13] 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

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

the beginning of a time period. Further, 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[14] them by the firm's total assets[15]. 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 the period by $\frac{(PRCC\_F \times CSHO)+AT-CEQ-TXDB}{AT}$ where PRCC_F is stock price, CSHO is number of common shares outstanding, CEQ is common equity, and TXDB is deferred taxes. We define[16] cash flow as the sum of income before extraordinary items (IB) and depreciation and amortization (DP) normalized by the beginning-of-the-period AT. Last, we define firm size as the natural logarithm of real net sales (SALE). In what follows, Y_1, Y_2, and Y_3 denote investment, saving, and debt respectively, U denotes the unobserved[17] marginal q, W denotes Tobin's q and serves as a proxy for U, X_1 denotes cash flow, and X_2 denotes firm size[18]. Section 6.6 considers including asset tangibility X_3 in X, 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 $161,960$ firm-year observations, with $3,375$ firms per year on average. Table 2 reports the summary statistics

[14]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.

[15]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)).

[16]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.

[17]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), 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.

[18]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)).

for the panel variables.

6.2 Bounds under Sequentially Stronger Assumptions

We begin our analysis by applying our framework to each cross section in our sample. 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 6.2 and 6.3 focus on the middle year in our sample, 1993. Sections 6.4 and 6.5 report the results for all the cross sections and for the full panel respectively. Table 3 reports the 50% and 95% confidence regions in 1993 for ρ, δ_j, and β_{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 $\mathcal{S}_j^{\kappa,\tau}$. This yields different identification regions for ρ across the investment, saving, and debt equations and wide bounds on the cash flow coefficients in each of these equations. Column 2 reports the results for the joint-equations bounds $\mathcal{J}^{\kappa,\tau}$. Considering 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 $[-\infty, 0.181]$ and $(-\infty, 0.223)$ respectively whereas the corresponding joint-equation confidence regions are $[-0.115, 0.181]$ and $(-0.270, 0.223)$. Nevertheless, in year 1993, the 95% confidence region for each of the effects of cash flow on investment, saving and debt in $\mathcal{J}^{\kappa,\tau}$ contains 0. For example, the 95% confidence region for the effect β_{11} of a \$1 increase in cash flow on investment is $(-0.397, 0.278)$. Column 3 reports the results for $\mathcal{J}^{\kappa,\tau,\mathbf{c}}$ when A_6 sets $\mathbf{c}$ 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 yields comparable confidence regions to $\mathcal{J}^{\kappa,\tau}$. In this case, the identification gain from imposing A_6 (with the $\mathbf{c}$ configuration above) is offset by the decrease in the precision of the estimates. Column 4 reports the (IV-type) results under $\mathcal{J}^{\kappa,\tau,\mathbf{c}^*}$ when A_6 sets $\mathbf{c}^* = 0$ so that the variance matrix of the disturbances is diagonal. For instance, $\mathbf{c}^* = 0$ rules out that the disturbances contain a common component (a fixed effect) that simultaneously influences the firm's investment, saving, and debt. 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 ρ (note that the 50% confidence region for ρ 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 δ_1, δ_2, and δ_3 are possibly attenuated relative to the bounds in $\mathcal{J}^{\kappa,\tau}$, $\mathcal{J}^{\kappa,\tau,\mathbf{c}}$, and $\mathcal{J}^{\kappa,\tau,\mathbf{c}^*}$. 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 $\kappa = 1.166$ and $\tau = (0.886, 0.896, 0.898)'$, so that the estimated κ^* and τ^* are 0.5 and $(0.9, 0.9, 0.9)'$. Setting $\kappa^* = 0.5$ assumes that at least half of the variance of Tobin's q is 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 $\tau_j^* = 0.9$ assumes that, in each equation, the coefficient of determination[19] would not exceed 0.9 had there been no measurement error. Under these settings, the A_4 restriction that the reliability ratio $R^2_{W.U}$ is at least as large as 50% forces the identification regions $\mathcal{S}_j^{\kappa,\tau}$, $\mathcal{J}^{\kappa,\tau}$, and $\mathcal{J}^{\kappa,\tau,\mathbf{c}}$ (with $\mathbf{c}$ encoding the same sign restrictions as in Table 3) to coincide. The bounds imply that cash flow affects investment and saving positively and debt negatively. Last, the 95% confidence region for $\mathcal{J}^{\kappa,\tau,\mathbf{c}^*}$, when A_6 assumes that σ^2_η is diagonal, is empty and the data rejects this specification at the 5% level. Thus, in year 1993, under A_1-A_3 and A_5, imposing a moderate lower bound $\kappa^* = 0.5$ on the reliability ratio of Tobin's q is incompatible with the assumption that the cross-equation disturbances are uncorrelated.

6.3 Sensitivity Analysis

If $\kappa = 0$ then there is no measurement error and (δ, β, Γ) is point identified. Next, we study the sensitivity of the identification regions for the cash flow coefficients β_{j1} for $j = 1, 2, 3$ to deviations from $\kappa = 0$. For this, we set $\tau = (1, 1, 1)'$ in A_5 and directly control the extent of the measurement error by varying κ in A_4. Using the sample from the middle year 1993, Figure 2 plots the 50% and 95% confidence regions for the partially identified β_{11}, β_{21} and β_{31} as κ ranges from 0 to ∞ (or equivalently as κ^* ranges from 1 to $R^2_{W.X}$). It plots the

[19] $R^2_{\tilde{Y}_1.\tilde{W}}$, $R^2_{\tilde{Y}_2.\tilde{W}}$, and $R^2_{\tilde{Y}_3.\tilde{W}}$ are estimated to be 3.66%, 0.9%, and 3.56% respectively.

regions under A_1-A_4 when each equation is analyzed separately ($\mathcal{S}_j^\kappa$), the three equations are analyzed jointly ($\mathcal{J}^\kappa$), and the three equations are analyzed jointly and A_6 is imposed ($\mathcal{J}^{\kappa,\mathbf{c}}$ with $\mathbf{c}$ set as in Tables 3 and 4). The 95% confidence region for the effect β_{11}, β_{21}, or β_{31} of cash flow contains 0 only when the reliability ratio $R^2_{W.U}$ is at least as small as 16.6% ($\kappa \geq 6.02$), 17.3% ($\kappa \geq 5.66$), and 24.1% ($\kappa \geq 3.55$) respectively. Otherwise, β_{11} and β_{21} are each estimated to be significantly positive and β_{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 β_{11}, β_{21}, and β_{31} that rest on different implicit 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 construct a 95% confidence region for $(\beta_{11}, \beta_{21}, \beta_{31})$ under A_1-A_4 and report the smallest κ, and the corresponding largest κ^*, 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 $\overline{\kappa}^*$, the reported threshold value of κ^*, then the null hypothesis is rejected. Table 5 considers the 8 possible null hypotheses corresponding to the possible signs of the elements of $(\beta_{11}, \beta_{21}, \beta_{31})$. In one extreme, for all values of κ^*, the hypothesis that cash flow affects investment and saving positively and debt negatively is not rejected. In another extreme, if the reliability ratio of Tobin's q is larger than 15.7% then any hypothesis on $(\beta_{11}, \beta_{21}, \beta_{31})$ in which the effect of cash flow on investment and saving is zero (or nonpositive) is rejected. Further, under the maintained assumptions, the joint effects of cash flow on investment, saving, and debt can be zero if and only if Tobin's q is a noisy proxy for marginal q, with a reliability ratio less than 17.7%. Otherwise, if Tobin's q is a moderately accurate proxy for marginal q, with $R^2_{W.U} \geq 25.7\%$, then any joint theory of investment, saving, and debt that does not predict $(0 < \beta_{11},\ 0 < \beta_{21},\ \beta_{31} < 0)$ is rejected under the maintained assumptions.

6.4 Results for all the Cross Sections

Whereas Sections 6.2 and 6.3 focus on the middle year 1993, Figure 3 plots the 50% and 95% confidence regions for β_{j1} (the coefficients on cash flow) that is partially identified in $\mathcal{B}_{j1}$, $\mathcal{B}_{j1}^{\mathbf{c}^*}$, or $\mathcal{B}_{j1}^{\kappa,\tau,\mathbf{c}}$, $j = 1, 2, 3$, for each cross section in our sample (years 1970 to 2017). The first column reports the joint equations bounds for $\beta_{j1} \in \mathcal{B}_{j1}$ 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. Otherwise, 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 $\beta_{j1} \in \mathcal{B}_{j1}^{\mathbf{c}^*}$ under A_1-A_3 and the A_6 diagonal restriction $\mathbf{c}^* = 0$. 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 $\beta_{j1} \in \mathcal{B}_{j1}^{\mathbf{c}^*}$ yield mixed results across different years. The third column reports the bounds for $\beta_{j1} \in \mathcal{B}_{j1}^{\kappa,\tau,\mathbf{c}}$ under A_1-A_6 where, for each year, we set κ and τ such that the estimated κ^* and τ^* are 0.5 and $(0.9, 0.9, 0.9)$. Setting $\kappa^* = 0.5$ assumes that Tobin's q is a moderately accurate proxy for marginal q. Further, as before, $\mathbf{c}$ sets $r_{\eta_1,\eta_2} \leq 0$, $r_{\eta_1,\eta_3} \geq 0$, $r_{\eta_2,\eta_3} \leq 0$ in A_6. We reject this specification at the 96% level in 5 years (1973, 1974, 1978, 1984, and 2008). For most of the remaining years, under this specification, 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 last column of 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.

6.5 Results for 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 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 (see e.g. Almeida, Campello, and Galvao (2010) who employ similar panel data methods to estimate the coefficient on cash flow in the investment equation). To keep the scope of the paper focused and manageable, we leave a detailed study of using panel data to estimate a system of equations with mismeasured variables to other work. Instead, our goal here is to 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. In particular, the robust standard errors for π are clustered at the firm level. We consider the panel 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[20]. We note that in this case the auxiliary assumptions A_4-A_6 should be interpreted relative to the within-transformed variables[21]. 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 $\beta_{j1} \in \mathcal{B}_{j1}$ falls in the positive range for investment and in the negative range for debt. Imposing the A_6 sign restrictions encoded in $\mathbf{c}$ tightens the bounds further ($\beta_{j1} \in \mathcal{B}^{\mathbf{c}}_{j1}$) and the effect of cash flow on saving is now estimated to be positive at the 95% level. On the other hand, we reject at the 96% level the specification that imposes the diagonal restriction in A_6 ($\beta_{j1} \in \mathcal{B}^{\mathbf{c}^*}_{j1}$). We also report the bounds when A_4-A_5 set κ and τ such that κ^* and τ^* 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, similar results obtain when including year and firm fixed effects in the equations, except that the sign of the effect of cash flow on investment is no longer recovered under only A_1-A_3 but remains significantly positive under the sign restrictions in A_6.

[20] Alternatively, one can consider (first-)differencing the data.

[21] 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 $(\kappa, \tau, \mathbf{c})$ to equivalent (or sufficient) restrictions on the level variables.

6.6 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 some specifications for 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. Here too, we do not require that X_3 is excluded from the investment and saving equations - instead, we allow X_3 to freely 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. Here too, the bounds for $\beta_{j1} \in \mathcal{B}_{j1}$ sometimes 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_6 ($\beta_{j1} \in \mathcal{B}_{j1}^{\mathbf{c}^*}$), reported in the second column, in 5 (as opposed to 19) of the 48 years. Further, the bounds under A_1-A_6 ($\beta_{j1} \in \mathcal{B}_{j1}^{\kappa,\tau,\mathbf{c}}$) 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 under A_1-A_3 now contains zero but remains in the positive range when the disturbance sign restrictions are imposed in A_6. Also, for the specification with year and firm fixed effects, the estimates for the effect of debt remain negative whereas the sign of the effect of cash flow on investment and saving is no longer recovered under A_1-A_3 nor after imposing the sign restrictions in A_6. The bounds for $\beta_{j1} \in \mathcal{B}_{j1}^{\kappa,\tau,\mathbf{c}}$ when A_1-A_6 assume that Tobin's q is a moderately accurate proxy of marginal q continue to be close to the regression estimates, with cash flow estimated to affect investment and saving positively and debt negatively.

7 Conclusion

This paper studies the identification of the coefficients in a system of linear equations that share a mismeasured explanatory variable. We characterize the sharp identification regions for the coefficients under the classical measurement error assumption 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 variance matrix of the disturbances is diagonal by specifying the signs of the covariances of the cross-equation disturbances, if at all. 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 Tobin's q is used 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 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 further investigate the estimated decrease over time in the magnitude of the effect of cash flow on debt.

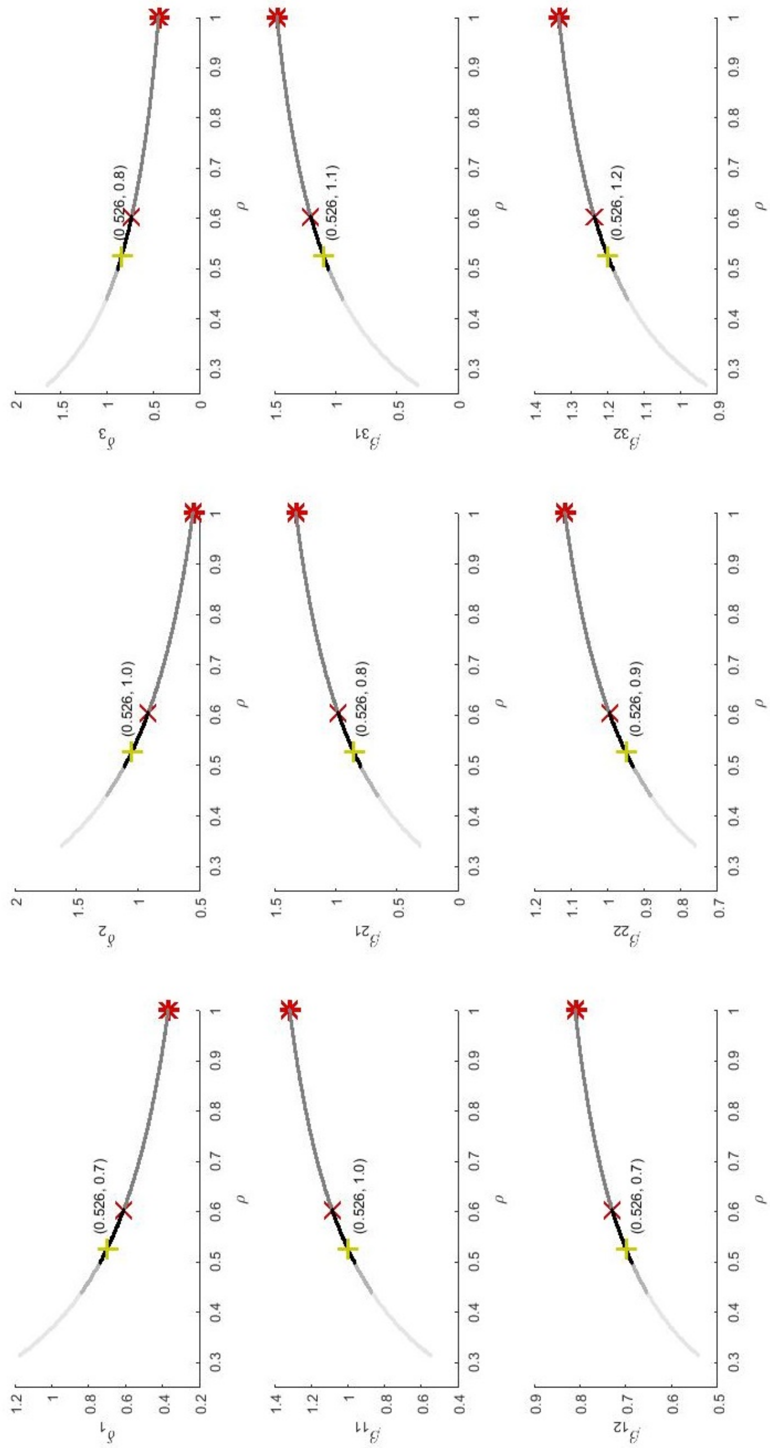

Figure 1: Identification regions $\mathcal{S}_j$ (light) for $j = 1, 2, 3$, $\mathcal{J}$, $\mathcal{J}^{\kappa,\tau}$, and $\mathcal{J}^{\kappa,\tau,\mathbf{c}}$ (dark) for $\kappa = 1$, $\tau = (0.7, 0.7, 0.7)'$ and $\mathbf{c}$ set to the (correct) sign restrictions $r_{\eta_1,\eta_2} \leq 0$, $r_{\eta_1,\eta_3} \geq 0$, $r_{\eta_2,\eta_3} \leq 0$. The plus, asterisk, and cross signs correspond to the true parameter values, $b_{Y.(W,X')'}$, and $\mathcal{J}^{\kappa,\tau,\mathbf{c}^*}$ respectively, where $\mathbf{c}^*$ incorrectly sets $\sigma_{\eta_1,\eta_2} = 0$ (with σ_{η_1,η_3} and σ_{η_2,η_3} unrestricted).

Table 1: Numerical Example

	DGP	$\mathcal{S}_j^{\kappa,\tau}$	$\mathcal{J}^{\kappa,\tau}$	$\mathcal{J}^{\kappa,\tau,\mathbf{c}}$	$\mathcal{J}^{\kappa,\tau,\mathbf{c}^*}$	$b_{Y.(W,X')'}$
	$\kappa \to \infty$ and $\tau = (1,1,1)'$					
ρ	0.527	[0.315 , 1]	[0.441 , 1]	[0.441 , 0.604]	0.604	
δ_1	0.700	[0.369 , 1.171]	[0.369 , 0.835]	[0.610 , 0.835]	0.610	0.369
β_{11}	1	[0.551 , 1.316]	[0.871 , 1.316]	[0.871 , 1.086]	1.086	1.316
β_{12}	0.700	[0.543 , 0.811]	[0.655 , 0.811]	[0.655 , 0.730]	0.730	0.811
ρ	0.527	[0.342 , 1]	[0.441 , 1]	[0.441 , 0.604]	0.604	
δ_2	1.050	[0.553 , 1.618]	[0.553 , 1.253]	[0.915 , 1.253]	0.915	0.553
β_{21}	0.850	[0.308 , 1.324]	[0.656 , 1.324]	[0.656 , 0.979]	0.979	1.324
β_{22}	0.950	[0.761 , 1.116]	[0.882 , 1.116]	[0.882 , 0.995]	0.995	1.116
ρ	0.527	[0.269 , 1]	[0.441 , 1]	[0.441 , 0.604]	0.604	
δ_3	0.840	[0.442 , 1.643]	[0.442 , 1.003]	[0.732 , 1.003]	0.732	0.442
β_{31}	1.100	[0.334 , 1.479]	[0.945 , 1.479]	[0.945 , 1.203]	1.203	1.479
β_{32}	1.200	[0.932 , 1.333]	[1.146 , 1.333]	[1.146 , 1.236]	1.236	1.333
	$\kappa = 1$ and $\tau = (0.7, 0.7, 0.7)'$					
ρ	0.527	[0.500 , 1]	[0.500 , 1]	[0.500 , 0.604]	0.604	
δ_1	0.700	[0.369 , 0.737]	[0.369 , 0.737]	[0.610 , 0.737]	0.610	0.369
β_{11}	1	[0.965 , 1.316]	[0.965 , 1.316]	[0.965 , 1.086]	1.086	1.316
β_{12}	0.700	[0.688 , 0.811]	[0.688 , 0.811]	[0.688 , 0.730]	0.730	0.811
ρ	0.527	[0.500 , 1]	[0.500 , 1]	[0.500 , 0.604]	0.604	
δ_2	1.050	[0.553 , 1.106]	[0.553 , 1.106]	[0.915 , 1.106]	0.915	0.553
β_{21}	0.850	[0.797 , 1.324]	[0.797 , 1.324]	[0.797 , 0.979]	0.979	1.324
β_{22}	0.950	[0.931 , 1.116]	[0.931 , 1.116]	[0.931 , 0.995]	0.995	1.116
ρ	0.527	[0.500 , 1]	[0.500 , 1]	[0.500 , 0.604]	0.604	
δ_3	0.840	[0.442 , 0.884]	[0.442 , 0.884]	[0.732 , 0.884]	0.732	0.442
β_{31}	1.100	[1.058 , 1.479]	[1.058 , 1.479]	[1.058 , 1.203]	1.203	1.479
β_{32}	1.200	[1.185 , 1.333]	[1.185 , 1.333]	[1.185 , 1.236]	1.236	1.333

This table reports population identification regions and point estimands. $\frac{\sigma_\varepsilon^2}{\sigma_{\tilde{U}}^2} = 0.89$ and $R^2_{\tilde{W}.\tilde{Y}} = 0.44$. $\mathbf{c}$ correctly sets $(\underline{c}_{12}, \overline{c}_{12}) = (\underline{c}_{23}, \overline{c}_{23}) = (-1, 0)$ and $(\underline{c}_{13}, \overline{c}_{13}) = (0, 1)$ whereas $\mathbf{c}^*$ incorrectly sets $(\underline{c}^*_{12}, \overline{c}^*_{12}) = (0, 0)$ and $(\underline{c}^*_{13}, \overline{c}^*_{13}) = (\underline{c}^*_{23}, \overline{c}^*_{23}) = (-1, 1)$.

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 the market-to-book ratio, and firm size is the log of the firm's sales.

	Investment	Saving	Debt	Tobin's Q	Cash Flow	Firm Size	Tangibility
mean	0.084	0.009	0.261	1.633	0.071	5.327	0.354
std dev	0.098	0.106	0.214	1.147	0.137	1.949	0.236
min	0.002	-0.298	0.000	0.526	-0.536	0.391	0.022
max	0.593	0.541	1.016	7.364	0.390	10.216	0.925

Table 3: Bounds for the Investment, Saving, and Debt Equations for Year 1993. $\kappa = \infty$ ($\kappa^* = R^2_{W.X} = 2.71\%$) and $\tau = (1,1,1)'$.

	$\mathcal{S}_j^{\kappa,\tau}$	$\mathcal{J}^{\kappa,\tau}$	$\mathcal{J}^{\kappa,\tau,\mathbf{c}}$	$\mathcal{J}^{\kappa,\tau,\mathbf{c}^*}$	$b_{Y.(W,X')'}$
ρ	[0.037 , 1]	[0.076 , 1]	[0.069 , 1]	-	-
	(0.019 , 1)	(0.055 , 1)	(0.051 , 1)	(0.050 , 0.244)	-
δ_1	[0.016 , 1.136]	[0.016 , 0.360]	[0.016 , 0.409]	[0.055 , 0.400]	0.018
	(0.012 , 1.382)	(0.012 , 0.438)	(0.012 , 0.497)	(0.041 , 0.486)	(0.013 , 0.022)
β_{11}	[-1.308 , 0.242]	[-0.243 , 0.242]	[-0.309 , 0.242]	[-0.297 , 0.187]	0.228
	(-1.800 , 0.278)	(-0.397 , 0.278)	(-0.484 , 0.278)	(-0.467 , 0.229)	(0.190 , 0.266)
β_{12}	[-0.010 , 0.052]	[-0.010 , 0.009]	[-0.010 , 0.012]	[-0.008 , 0.012]	-0.009
	(-0.012 , 0.073)	(-0.012 , 0.016)	(-0.012 , 0.020)	(-0.010 , 0.019)	(-0.011 , -0.007)
ρ	[0.009 , 1]	[0.076 , 1]	[0.069 , 1]	-	-
	(0.000 , 1)	(0.055 , 1)	(0.051 , 1)	(0.050 , 0.244)	-
δ_2	[0.007 , ∞]	[0.007 , 0.213]	[0.007 , 0.242]	[0.025 , 0.237]	0.009
	(0.002 , ∞)	(0.002 , 0.309)	(0.002 , 0.351)	(0.008 , 0.343)	(0.004 , 0.015)
β_{21}	[-∞ , 0.181]	[-0.115 , 0.181]	[-0.154 , 0.181]	[-0.146 , 0.154]	0.165
	(-∞ , 0.223)	(-0.270 , 0.223)	(-0.328 , 0.223)	(-0.317 , 0.203)	(0.120 , 0.209)
β_{22}	[-0.003 , ∞]	[-0.003 , 0.009]	[-0.003 , 0.010]	[-0.002 , 0.010]	-0.002
	(-0.004 , ∞)	(-0.004 , 0.014)	(-0.004 , 0.016)	(-0.003 , 0.016)	(-0.004 , -0.000)
ρ	[0.035 , 1]	[0.076 , 1]	[0.069 , 1]	-	-
	(0.024 , 1)	(0.055 , 1)	(0.051 , 1)	(0.050 , 0.244)	-
δ_3	[-1.641 , -0.033]	[-0.697 , -0.033]	[-0.791 , -0.033]	[-0.773 , -0.115]	-0.035
	(-1.873 , -0.028)	(-0.796 , -0.028)	(-0.903 , -0.028)	(-0.883 , -0.097)	(-0.041 , -0.030)
β_{31}	[-0.311 , 1.994]	[-0.311 , 0.654]	[-0.311 , 0.787]	[-0.202 , 0.762]	-0.287
	(-0.372 , 2.728)	(-0.372 , 0.961)	(-0.372 , 1.136)	(-0.276 , 1.103)	(-0.352 , -0.221)
β_{32}	[-0.078 , 0.012]	[-0.026 , 0.012]	[-0.031 , 0.012]	[-0.030 , 0.008]	0.011
	(-0.107 , 0.016)	(-0.038 , 0.016)	(-0.045 , 0.016)	(-0.044 , 0.012)	(0.008 , 0.015)

The sample size is 3,454 observations. Y_1, Y_2, and Y_3 denote Investment, Saving, and Debt respectively and $X = [\text{Cash Flow}, \text{Firm Size}]$. $\mathbf{c}$ sets $(\underline{c}_{12}, \bar{c}_{12}) = (\underline{c}_{23}, \bar{c}_{23}) = (-1, 0)$ and $(\underline{c}_{13}, \bar{c}_{13}) = (0, 1)$ whereas $\mathbf{c}^* = 0$. 50% and 95% confidence regions are in brackets and parentheses respectively.

Table 4: Bounds for the Investment, Saving, and Debt Equations for Year 1993. $\kappa = 1.1658$ ($\hat{\kappa}^* = 50\%$) and $\tau = (0.8857, 0.8955, 0.8978)'$ ($\hat{\tau}^* = (0.9, 0.9, 0.9)'$).

	$\mathcal{S}_j^{\kappa,\tau}$	$\mathcal{J}^{\kappa,\tau}$	$\mathcal{J}^{\kappa,\tau,\mathbf{c}}$	$\mathcal{J}^{\kappa,\tau,\mathbf{c}^*}$	$b_{Y.(W,X')'}$
ρ	[0.486 , 1]	[0.486 , 1]	[0.486 , 1]	-	-
	(0.486 , 1)	(0.486 , 1)	(0.486 , 1)	-	-
δ_1	[0.016 , 0.040]	[0.016 , 0.040]	[0.016 , 0.040]	-	0.018
	(0.012 , 0.049)	(0.012 , 0.049)	(0.012 , 0.049)	-	(0.013 , 0.022)
β_{11}	[0.188 , 0.242]	[0.188 , 0.242]	[0.188 , 0.242]	-	0.228
	(0.151 , 0.278)	(0.151 , 0.278)	(0.151 , 0.278)	-	(0.190 , 0.266)
β_{12}	[-0.010 , -0.008]	[-0.010 , -0.008]	[-0.010 , -0.008]	-	-0.009
	(-0.012 , -0.006)	(-0.012 , -0.006)	(-0.012 , -0.006)	-	(-0.011 , -0.007)
ρ	[0.486 , 1]	[0.486 , 1]	[0.486 , 1]	-	-
	(0.486 , 1)	(0.486 , 1)	(0.486 , 1)	-	-
δ_2	[0.007 , 0.024]	[0.007 , 0.024]	[0.007 , 0.024]	-	0.009
	(0.002 , 0.034)	(0.002 , 0.034)	(0.002 , 0.034)	-	(0.004 , 0.015)
β_{21}	[0.134 , 0.181]	[0.134 , 0.181]	[0.134 , 0.181]	-	0.165
	(0.090 , 0.223)	(0.090 , 0.223)	(0.090 , 0.223)	-	(0.120 , 0.209)
β_{22}	[-0.003 , -0.001]	[-0.003 , -0.001]	[-0.003 , -0.001]	-	-0.002
	(-0.004 , 0.001)	(-0.004 , 0.001)	(-0.004 , 0.001)	-	(-0.004 , -0.000)
ρ	[0.486 , 1]	[0.486 , 1]	[0.486 , 1]	-	-
	(0.486 , 1)	(0.486 , 1)	(0.486 , 1)	-	-
δ_3	[-0.077 , -0.033]	[-0.077 , -0.033]	[-0.077 , -0.033]	-	-0.035
	(-0.088 , -0.028)	(-0.088 , -0.028)	(-0.088 , -0.028)	-	(-0.041 , -0.030)
β_{31}	[-0.311 , -0.212]	[-0.311 , -0.212]	[-0.311 , -0.212]	-	-0.287
	(-0.372 , -0.148)	(-0.372 , -0.148)	(-0.372 , -0.148)	-	(-0.352 , -0.221)
β_{32}	[0.008 , 0.012]	[0.008 , 0.012]	[0.008 , 0.012]	-	0.011
	(0.005 , 0.016)	(0.005 , 0.016)	(0.005 , 0.016)	-	(0.008 , 0.015)

The sample size is 3,454 observations. Y_1, Y_2, and Y_3 denote Investment, Saving, and Debt respectively and $X = [\text{Cash Flow}, \text{Firm Size}]$. $\mathbf{c}$ sets $(\underline{c}_{12}, \overline{c}_{12}) = (\underline{c}_{23}, \overline{c}_{23}) = (-1, 0)$ and $(\underline{c}_{13}, \overline{c}_{13}) = (0, 1)$ whereas $\mathbf{c}^* = 0$. 50% and 95% confidence regions are in brackets and parentheses respectively.

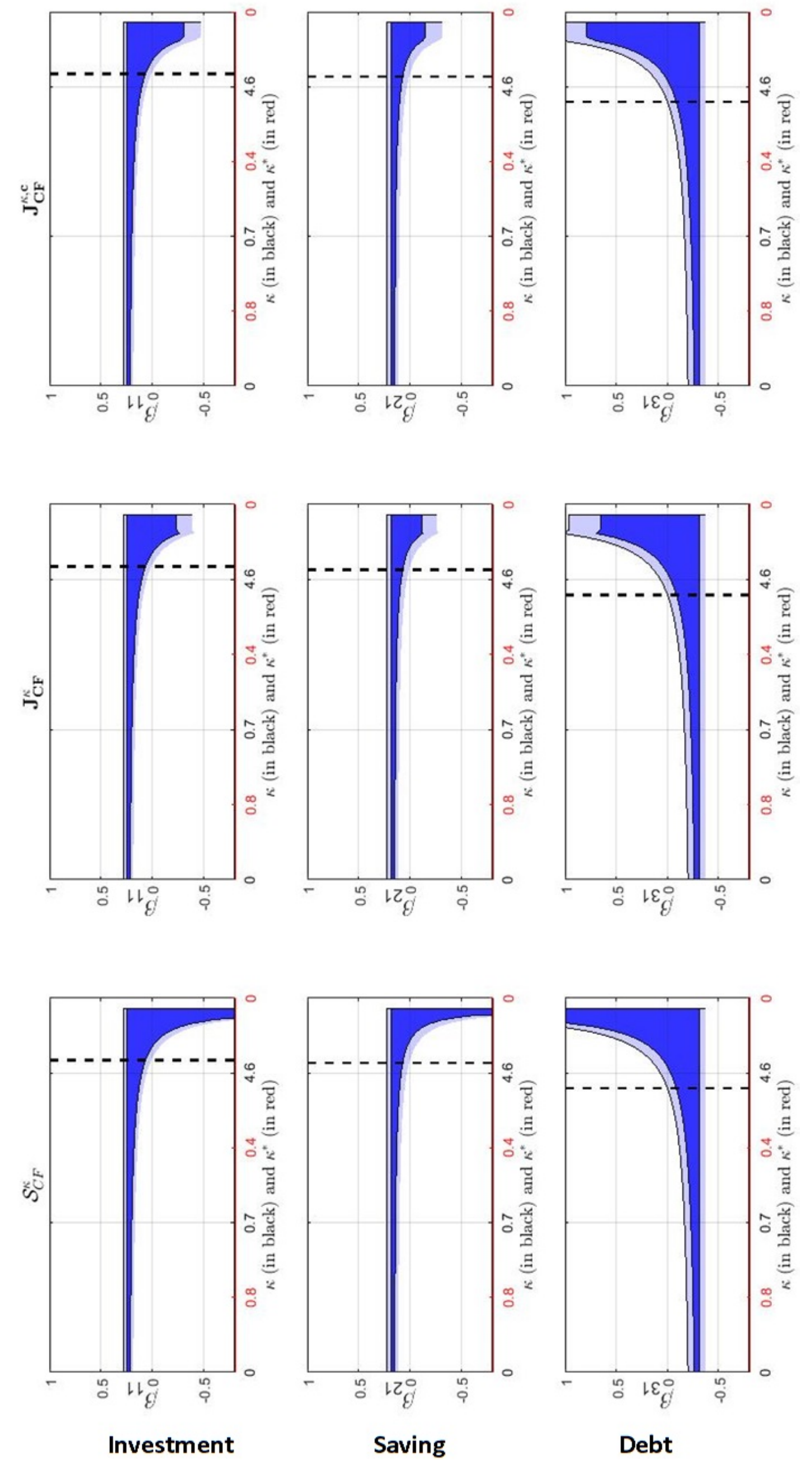

Figure 2: 50% (dark) and 95% (light) confidence regions for the partially identified coefficients β_{j1} on cash flow for $j = 1, 2, 3$ (investment, saving, and debt) for year 1993. We set $\tau = (1, 1, 1)'$ and consider the regions $\mathcal{S}_j^\kappa$, $\mathcal{J}^\kappa$, and $\mathcal{J}^{\kappa,\mathbf{c}}$ when $\kappa \in [0, \infty)$ and $\mathbf{c}$ sets $(\underline{c}_{12}, \overline{c}_{12}) = (\underline{c}_{23}, \overline{c}_{23}) = (-1, 0)$ and $(\underline{c}_{13}, \overline{c}_{13}) = (0, 1)$. The vertical dashed line indicates the smallest κ (largest κ^*) value such that the 95% confidence region contains zero. This corresponds to 6.018 (0.1658), 5.662 (0.1732), and 3.552 (0.2409) for $j = 1, 2, 3$ respectively.

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. $\overline{\kappa}^*$ is the largest κ^* such that H_0 is not rejected. $\underline{\kappa}$ is the smallest κ such that H_0 is not rejected.

	β_{11}	$-$	$-$	$-$	$-$	$+$	$+$	$+$	$+$
H_0	β_{21}	$-$	$-$	$+$	$+$	$-$	$-$	$+$	$+$
	β_{31}	$+$	$-$	$+$	$-$	$+$	$-$	$+$	$-$
$\overline{\kappa}^*$		0.157	0.157	0.177	0.177	0.197	0.197	0.257	1
$\underline{\kappa}$		6.483	6.483	5.486	5.486	4.723	4.723	3.230	0

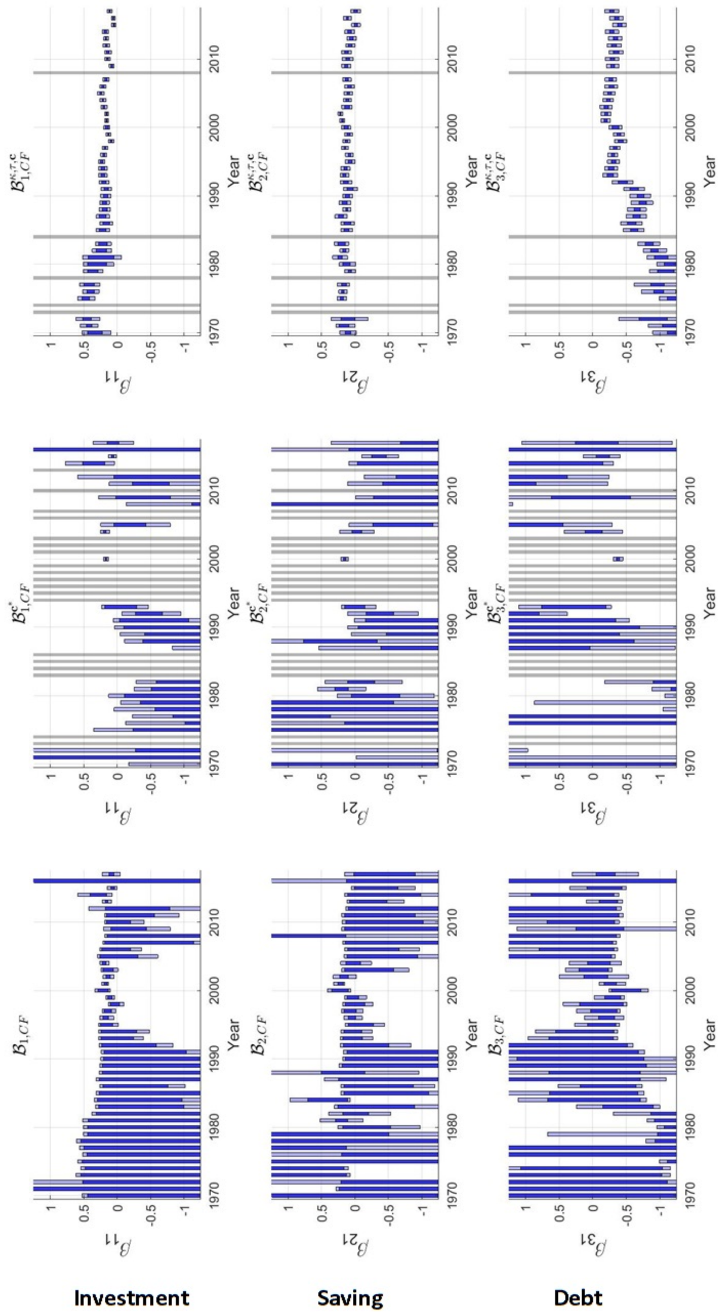

Figure 3: 50% (dark) and 95% (light) confidence regions for β_{j1} (cash flow) for $j = 1, 2, 3$ (investment, saving, and debt) from year 1970 to 2017. We consider the regions $\mathcal{B}_{j1}$, $\mathcal{B}_{j1}^{\mathbf{c}^*}$, and $\mathcal{B}_{j1}^{\kappa,\tau,\mathbf{c}}$ where $\mathbf{c}^* = 0$, κ and τ are such that $\hat{\kappa}^* = 0.5$ and $\hat{\tau}^* = (0.9, 0.9, 0.9)'$, and $\mathbf{c}$ is such that $(\underline{c}_{12}, \overline{c}_{12}) = (\underline{c}_{23}, \overline{c}_{23}) = (-1, 0)$ and $(\underline{c}_{13}, \overline{c}_{13}) = (0, 1)$. 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

	$\mathcal{S}_j^{\kappa,\tau}$	$\mathcal{J}^{\kappa,\tau}$	$\mathcal{J}^{\kappa,\tau,\mathbf{c}}$	$\mathcal{J}^{\kappa,\tau,\mathbf{c}^*}$	$b_{Y.(W,X')'}$
	Results without fixed effects for $\kappa = \infty$ and $\tau = (1,1,1)'$				
β_{11}	[-0.242 , 0.197]	[0.096 , 0.197]	[0.193 , 0.197]	-	0.196
	(-0.258 , 0.198)	(0.092 , 0.198)	(0.192 , 0.198)	-	(0.187 , 0.205)
β_{21}	[-0.279 , 0.124]	[-0.010 , 0.124]	[0.120 , 0.124]	-	0.124
	(-0.296 , 0.126)	(-0.016 , 0.126)	(0.119 , 0.126)	-	(0.116 , 0.131)
β_{31}	[-0.352 , 0.453]	[-0.352 , -0.087]	[-0.352 , -0.345]	-	-0.351
	(-0.355 , 0.485)	(-0.355 , -0.076)	(-0.355 , -0.341)	-	(-0.369 , -0.332)
	Results without fixed effects for $\kappa^* = 0.5$ and $\tau^* = (0.9, 0.9, 0.9)'$				
β_{11}	[0.189 , 0.197]	[0.189 , 0.197]	[0.193 , 0.197]	-	0.196
	(0.187 , 0.198)	(0.187 , 0.198)	(0.192 , 0.198)	-	(0.187 , 0.205)
β_{21}	[0.115 , 0.124]	[0.115 , 0.124]	[0.120 , 0.124]	-	0.124
	(0.113 , 0.126)	(0.113 , 0.126)	(0.119 , 0.126)	-	(0.116 , 0.131)
β_{31}	[-0.352 , -0.333]	[-0.352 , -0.333]	[-0.352 , -0.345]	-	-0.351
	(-0.355 , -0.330)	(-0.355 , -0.330)	(-0.355 , -0.341)	-	(-0.369 , -0.332)
	Results with year and firm fixed effects for $\kappa = \infty$ and $\tau = (1,1,1)'$				
β_{11}	[-0.627 , 0.130]	[-0.493 , 0.130]	[0.017 , 0.130]	-	0.129
	(-0.632 , 0.131)	(-0.497 , 0.131)	(0.016 , 0.131)	-	(0.122 , 0.137)
β_{21}	[-2.504 , 0.172]	[-0.248 , 0.172]	[0.096 , 0.172]	-	0.172
	(-2.542 , 0.173)	(-0.255 , 0.173)	(0.094 , 0.173)	-	(0.161 , 0.182)
β_{31}	[-0.367 , 17.982]	[-0.367 , -0.270]	[-0.367 , -0.349]	-	-0.366
	(-0.368 , 18.855)	(-0.368 , -0.265)	(-0.368 , -0.347)	-	(-0.381 , -0.351)
	Results with year and firm fixed effects for $\kappa^* = 0.5$ and $\tau^* = (0.9, 0.9, 0.9)'$				
β_{11}	[0.084 , 0.130]	[0.084 , 0.130]	[0.084 , 0.130]		0.129
	[0.083 , 0.131]	[0.083 , 0.131]	[0.083 , 0.131]	-	(0.122 , 0.137)
β_{21}	[0.141 , 0.172]	[0.141 , 0.172]	[0.141 , 0.172]	-	0.172
	(0.139 , 0.173)	(0.139 , 0.173)	(0.139 , 0.173)	-	(0.161 , 0.182)
β_{31}	[-0.367 , -0.359]	[-0.367 , -0.359]	[-0.367 , -0.359]	-	-0.366
	(-0.368 , -0.357)	(-0.368 , -0.357)	(-0.368 , -0.357)	-	(-0.381 , -0.351)

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. $\mathbf{c}$ sets $(\underline{c}_{12}, \overline{c}_{12}) = (\underline{c}_{23}, \overline{c}_{23}) = (-1, 0)$ and $(\underline{c}_{13}, \overline{c}_{13}) = (0, 1)$ whereas $\mathbf{c}^* = 0$. Robust standard errors for π are clustered by firm. 50% and 95% confidence regions are in brackets and parentheses respectively.

References

Almeida, H., M. Campello, and M. Weisbach (2004), "The Cash Flow Sensitivity of Cash," *Journal of Finance*, 59, 1777-1804.

Almeida, H. and M. Campello (2007), "Financial Constraints, Asset Tangibility, and Corporate Investment," *Review of Financial Studies*, 20, 1429-1460.

Almeida, H., M. Campello, and A. Galvao (2010), "Measurement Error in Investment Equations," *Review of Financial Studies*, 23, 3279- 3328.

Bollinger, C. (2003), "Measurement Error in Human Capital and the Black-White Wage Gap," *Review of Economics and Statistics*, 85, 578-585.

Chalak, K. and D. Kim (2018), "Measurement Error without Exclusion: the Returns to College Selectivity and Characteristics," working paper.

Chernozhukov, V., R. Rigobon, and T. Stoker (2010), "Set Identification and Sensitivity Analysis with Tobin Regressors," *Quantitative Economics*, 1, 255–277.

Chernozhukov, V., S. Lee, and A. Rosen (2013), "Intersection Bounds: Estimation and Inference," *Econometrica*, 81, 667–737.

Chernozhukov, V., W. Kim, S. Lee, and A. Rosen (2015) "Implementing Intersection Bounds in Stata," *The Stata Journal*, 15, 21-44.

DiTraglia, F. and C. Garcia-Jimeno (2017), "A Framework for Eliciting, Incorporating, and Disciplining Identification Beliefs in Linear Models," working paper.

Erickson, T. and T. Whited (2000), "Measurement error and the relationship between investment and q," *Journal of Political Economy*, 108, 1027-1057.

Erickson, T. and T. Whited (2002), "Two-Step GMM Estimation of the Errors-in-Variables Model Using High-Order Moments," *Econometric Theory*, 18, 776-799.

Erickson, T. and T. Whited (2005), "Proxy-Quality Thresholds: Theory and Applications," *Finance Research Letters*, 2, 131-151.

Erickson, T. and T. Whited (2008), "On the Accuracy of Different Measures of q," *Financial Management*, 35, 5-33.

Erickson T. and T. Whited (2012), "Treating Measurement Error in Tobin's q," *Review of Financial Studies*, 25, 1286-1329.

Erickson, T., C. Jiang, and T. Whited (2014), "Minimum Distance Estimation of the Errors-in-Variables Model Using Linear Cumulant Equations," *Journal of Econometrics*, 183, 211-221.

Fazzari, S., G. Hubbard, and B. Petersen (1988), "Financing Constraints and Corporate Investment," *Brookings Papers on Economic Activity*, 1, 141-195.

Gilchrist, S. and C. P. Himmelberg (1995), "Evidence on the Role of Cash Flow for Investment," *Journal of Monetary Economics*, 36, 541-572.

Gomes J. and L. Schmid (2010), "Levered Returns," *Journal of Finance,* 65, 467-494.

Hayashi, F. (1982), "Tobin's marginal q and average q: Neoclassical interpretation," *Econometrica*, 50, 213-224.

Hennessy, C. and T. Whited (2005), "Debt Dynamics," *Journal of Finance*, 60, 1129-1165.

Klepper, S. and E. Leamer (1984), "Consistent Sets of Estimates for Regressions with Errors in All Variables," *Econometrica*, 52, 163-184.

Klepper, S. (1988), "Regressor Diagnostics for the Classical Errors-in-Variables Model," *Journal of Econometrics*, 37, 225-250.

Krasker W. and J. Pratt (1986), "Bounding the Effects of Proxy Variables on Regression Coefficients," *Econometrica*, 54, 641-655.

Leamer, E. (1987), "Errors in Variables in Linear Systems," *Econometrica*, 55, 893-909.

Lewbel, A. (1997), "Constructing Instruments for Regressions with Measurement Error when No Additional Data are Available, with an Application to Patents and R&D," *Econometrica*, 65, 1201-1213.

Love, I. (2003), "Financial Development and Financing Constraints: International Evidence from the Structural Investment Model," *Review of Financial Studies,* 16, 765-791.

Manski, C. and J. Pepper (2009), "More on Monotone Instrumental Variables," *Econometrics Journal*, 12, S200–S216.

Miller, M. (1977), "Debt and Taxes," *Journal of Finance*, 32, 261-275.

Myers, S. and N. Majluf (1984), "Corporate Financing and Investment Decisions when Firms Have Information that Investors Do Not Have," *Journal of Financial Economics*, 13, 187-221.

Peters, R. and L. Taylor (2017), "Intangible Capital and the Investment-q Relation," *Journal of Financial Economics*, 123, 251–272.

Rajan, R. and L. Zingales (1995), "What Do We Know about Capital Structure? Some Evidence from International Data," *Journal of Finance*, 50, 1421-1460.

Riddick, L. and T. Whited (2009), "The Corporate Propensity to Save," *Journal of Finance*, 4, 1729-1766.

White, H. (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817-838.

White, H. (2001). *Asymptotic Theory for Econometricians.* New York: Academic Press.

Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA: MIT Press.

Zellner, A. (1962), "An Efficient Method of Estimating Seemingly Unrelated Regression Equations and Tests for Aggregation Bias," *Journal of the American Statistical Association*, 57, 348–368.

Supplementary Material for "Measurement Error in Multiple Equations: Tobin's q and Corporate Investment, Saving, and Debt"

Karim Chalak[1]
University of Virginia

Daniel Kim
University of Pennsylvania

A Restricting the Correlations among the Disturbances

We extend A_6 to A'_6 which restricts the sign and/or magnitude of the correlation r_{η_j,η_h} between η_j and η_h.

Assumption A'_6 *Disturbance Correlation Restriction:* $\underline{c}_{jh} \leq r_{\eta_j,\eta_h} \leq \bar{c}_{jh}$ *where* $-1 \leq \underline{c}_{jh} \leq \bar{c}_{jh} \leq 1$.

In particular, provided $\sigma^2_{\tilde{Y}_j}\sigma^2_{\tilde{Y}_h} \neq 0$, from the proof of Corollary A.1 we have that

$$r_{\eta_j,\eta_h} = \frac{\rho r_{\tilde{Y}_j,\tilde{Y}_h} - r_{\tilde{W},\tilde{Y}_j} r_{\tilde{W},\tilde{Y}_h}}{(\rho - R^2_{\tilde{W}.\tilde{Y}_j})^{\frac{1}{2}}(\rho - R^2_{\tilde{W}.\tilde{Y}_h})^{\frac{1}{2}}}.$$

A'_6 may restrict the sign of r_{η_j,η_h} as encoded by the sign of the function

$$S_{jh}(r) \equiv r \times r_{\tilde{Y}_j,\tilde{Y}_h} - r_{\tilde{W},\tilde{Y}_j} r_{\tilde{W},\tilde{Y}_h}.$$

Further, A'_6 may restrict the magnitude of r_{η_j,η_h} (either $r^2_{\eta_j,\eta_h} \leq c^2$ or $c^2 \leq r^2_{\eta_j,\eta_h}$) as encoded by the sign of the function

$$M_{jh}(r;c) \equiv (r \times r_{\tilde{Y}_j,\tilde{Y}_h} - r_{\tilde{W},\tilde{Y}_j} r_{\tilde{W},\tilde{Y}_h})^2 - c^2(r - R^2_{\tilde{W}.\tilde{Y}_j})(r - R^2_{\tilde{W}.\tilde{Y}_h}).$$

As shown in the proof of Corollary A.1, when $R^2_{\tilde{Y}_j.\tilde{Y}_h} \neq 1$, the discriminant of the quadratic function $M_{jh}(\cdot;c)$ is given by

$$\Delta_{jh}(c) \equiv c^2[R^4_{\tilde{W}.(\tilde{Y}_j.\tilde{Y}_h)'}(1 - R^2_{\tilde{Y}_j.\tilde{Y}_h})^2 - (1 - c^2)(R^2_{\tilde{W}.\tilde{Y}_j} - R^2_{\tilde{W}.\tilde{Y}_h})^2],$$

[1] Karim Chalak (corresponding author), Department of Economics, University of Virginia, chalak@virginia.edu. Daniel Kim, The Wharton School, University of Pennsylvania, kimdanie@wharton.upenn.edu.

and, when $R^2_{\tilde{Y}_j.\tilde{Y}_h} \neq c^2$, the roots of $M_{jh}(\cdot; c)$ are given by

$$\rho^-_{jh}(c) \equiv \frac{F_{jh}(c) - \Delta_{jh}(c)^{\frac{1}{2}}}{2(R^2_{\tilde{Y}_j.\tilde{Y}_h} - c^2)} \quad \text{and} \quad \rho^+_{jh}(c) \equiv \frac{F_{jh}(c) + \Delta_{jh}(c)^{\frac{1}{2}}}{2(R^2_{\tilde{Y}_j.\tilde{Y}_h} - c^2)}$$

where

$$F_{jh}(c) \equiv -R^2_{\tilde{W}.(\tilde{Y}_j,\tilde{Y}_h)'}(1 - R^2_{\tilde{Y}_j.\tilde{Y}_h}) + (1 - c^2)(R^2_{\tilde{W}.\tilde{Y}_j} + R^2_{\tilde{W}.\tilde{Y}_h}).$$

Corollary A.1 uses $S_{jh}(r)$ and $M_{jh}(r)$ to encode the sign and magnitudes restrictions in A$'_6$ and to express the identification region for $(\rho, \delta, \beta, \Gamma)$ under A$_1$-A$'_6$.

Corollary A.1 *Under the conditions of Theorem 3.1, A_4, A_5, and $A^{'}_6$ for $j, h = 1, ..., p$ with $j < h$, $(\rho, \delta, \beta, \Gamma)$ is partially identified in the sharp set*

$$\mathcal{J}^{k,\tau,\mathbf{c}} \equiv \{(r, D(r), B(r), G(r)) : 0 \preceq G(r), \ \frac{1}{1+\kappa} \leq r \leq 1, \ \frac{\sigma^2_{\tilde{Y}_j}}{\sigma^2_{\tilde{W}_j}}(1 - \tau_j) \leq G_{jj}(r),$$
$$\text{and } \underline{c}_{jh} \leq \frac{G_{jh}(r)}{[G_{jj}(r)G_{hh}(r)]^{\frac{1}{2}}} \leq \bar{c}_{jh} \text{ for } j, h = 1, ..., p \text{ and } j < h\}.$$

Further, ρ is partially identified in the sharp set

$$\mathcal{R}^{k,\tau,\mathbf{c}} = [R^2_{\tilde{W}.\tilde{Y}}, 1] \cap [\frac{1}{1+\kappa}, 1] \cap^p_{j=1} [\frac{1}{\tau_j} R^2_{\tilde{W}.\tilde{Y}_j}, 1] \bigcap_{\substack{j,h=1 \\ j<h}}^{p} \mathcal{R}^{\mathbf{c}}_{jh},$$

with

$$\mathcal{R}^{\mathbf{c}}_{jh} = \left\{ r : \begin{array}{cc} S_{jh}(r) \leq 0 \ \text{ and } M_{jh}(r; \underline{c}_{jh}) \leq 0 \leq M_{jh}(r; \bar{c}_{jh}) & \text{if } \underline{c}_{jh} \leq \bar{c}_{jh} \leq 0 \text{ and } \sigma^2_{\tilde{Y}_j}\sigma^2_{\tilde{Y}_h} \neq 0 \\ \{S_{jh}(r) \leq 0 \text{ and } M_{jh}(r; \underline{c}_{jh}) \leq 0\} \text{ or} & \\ \{0 \leq S_{jh}(r) \text{ and } M_{jh}(r; \bar{c}_{jh}) \leq 0\} & \text{if } \underline{c}_{jh} < 0 < \bar{c}_{jh} \text{ and } \sigma^2_{\tilde{Y}_j}\sigma^2_{\tilde{Y}_h} \neq 0 \\ 0 \leq S_{jh}(r) \ \text{ and } M_{jh}(r; \bar{c}_{jh}) \leq 0 \leq M_{jh}(r; \underline{c}_{jh}) & \text{if } 0 \leq \underline{c}_{jh} \leq \bar{c}_{jh} \text{ and } \sigma^2_{\tilde{Y}_j}\sigma^2_{\tilde{Y}_h} \neq 0 \\ r \in \emptyset & \text{if } 0 \notin [\underline{c}_{jh}, \bar{c}_{jh}] \text{ and } \sigma^2_{\tilde{Y}_j}\sigma^2_{\tilde{Y}_h} = 0. \\ -\infty < r < \infty & \text{if } 0 \in [\underline{c}_{jh}, \bar{c}_{jh}] \text{ and } \sigma^2_{\tilde{Y}_j}\sigma^2_{\tilde{Y}_h} = 0. \end{array} \right\},$$

where, provided $\sigma^2_{\tilde{Y}_j}\sigma^2_{\tilde{Y}_h} \neq 0$, we have

$$0 \leq S_{jh}(r) \Leftrightarrow \left\{ \begin{array}{ll} \frac{r_{\tilde{W},\tilde{Y}_j} r_{\tilde{W},\tilde{Y}_h}}{r_{\tilde{Y}_j,\tilde{Y}_h}} \leq r & \text{when } 0 < r_{\tilde{Y}_j,\tilde{Y}_h} \\ r_{\tilde{W},\tilde{Y}_j} r_{\tilde{W},\tilde{Y}_h} \leq 0 & \text{when } r_{\tilde{Y}_j,\tilde{Y}_h} = 0 \\ r \leq \frac{r_{\tilde{W},\tilde{Y}_j} r_{\tilde{W},\tilde{Y}_h}}{r_{\tilde{Y}_j,\tilde{Y}_h}} & \text{when } r_{\tilde{Y}_j,\tilde{Y}_h} < 0 \end{array} \right. ,$$

and if $R^2_{\tilde{Y}_j.\tilde{Y}_h} = 1$ *then* $0 \le M_{jh}(r;c) = (1-c^2)(r - R^2_{\tilde{W}.\tilde{Y}_j})^2$ *whereas if* $R^2_{\tilde{Y}_j.\tilde{Y}_h} \neq 1$ *then*

$$0 \le M_{jh}(r;c) \Leftrightarrow$$

$$\begin{cases} -\infty < r < \infty & \textit{when } 0 < c^2 < 1 - \frac{R^4_{\tilde{W}.(\tilde{Y}_j.\tilde{Y}_h)'}(1-R^2_{\tilde{Y}_j.\tilde{Y}_h})^2}{(R^2_{\tilde{W}.\tilde{Y}_j}-R^2_{\tilde{W}.\tilde{Y}_h})^2} \\ r \in (-\infty, \rho^-_{jh}(c)] \cup [\rho^+_{jh}(c), \infty) & \textit{when } c^2 = 0 < R^2_{\tilde{Y}_j.\tilde{Y}_h} \textit{ or } 1 - \frac{R^4_{\tilde{W}.(\tilde{Y}_j.\tilde{Y}_h)'}(1-R^2_{\tilde{Y}_j.\tilde{Y}_h})^2}{(R^2_{\tilde{W}.\tilde{Y}_j}-R^2_{\tilde{W}.\tilde{Y}_h})^2} \le c^2 < R^2_{\tilde{Y}_j.\tilde{Y}_h} \\ r \le \frac{-R^2_{\tilde{W}.\tilde{Y}_j} R^2_{\tilde{W}.\tilde{Y}_h}}{R^2_{\tilde{W}.(\tilde{Y}_j,\tilde{Y}_h)'} - (R^2_{\tilde{W}.\tilde{Y}_j} + R^2_{\tilde{W}.\tilde{Y}_h})} & \textit{when } c^2 = R^2_{\tilde{Y}_j.\tilde{Y}_h} \textit{ and } R^2_{\tilde{W}.(\tilde{Y}_j,\tilde{Y}_h)'} < R^2_{\tilde{W}.\tilde{Y}_j} + R^2_{\tilde{W}.\tilde{Y}_h} \\ 0 \le (1 - R^2_{\tilde{Y}_j.\tilde{Y}_h}) R^2_{\tilde{W}.\tilde{Y}_j} R^2_{\tilde{W}.\tilde{Y}_h} & \textit{when } c^2 = R^2_{\tilde{Y}_j.\tilde{Y}_h} \textit{ and } R^2_{\tilde{W}.(\tilde{Y}_j,\tilde{Y}_h)'} = R^2_{\tilde{W}.\tilde{Y}_j} + R^2_{\tilde{W}.\tilde{Y}_h} \\ \frac{-R^2_{\tilde{W}.\tilde{Y}_j} R^2_{\tilde{W}.\tilde{Y}_h}}{R^2_{\tilde{W}.(\tilde{Y}_j,\tilde{Y}_h)'} - (R^2_{\tilde{W}.\tilde{Y}_j} + R^2_{\tilde{W}.\tilde{Y}_h})} \le r & \textit{when } c^2 = R^2_{\tilde{Y}_j.\tilde{Y}_h} \textit{ and } R^2_{\tilde{W}.\tilde{Y}_j} + R^2_{\tilde{W}.\tilde{Y}_h} < R^2_{\tilde{W}.(\tilde{Y}_j,\tilde{Y}_h)'} \\ r \in [\rho^+_{jh}(c), \rho^-_{jh}(c)] & \textit{when } R^2_{\tilde{Y}_j.\tilde{Y}_h} < c^2 \end{cases}$$

Last, δ, β, *and* Γ *are partially identified in the sharp sets* $\mathcal{D}^{k,\tau,\mathbf{c}} = \{D(r) : r \in \mathcal{R}^{k,\tau,\mathbf{c}}\}$, $\mathcal{B}^{k,\tau,\mathbf{c}} = \{B(r) : r \in \mathcal{R}^{k,\tau,\mathbf{c}}\}$, *and* $\mathcal{G}^{k,\tau,\mathbf{c}} = \{G(r) : r \in \mathcal{R}^{k,\tau,\mathbf{c}}\}$.

The bounds in Corollary A.1 yield those from Corollary 3.2 when $(\underline{c}_{jh}, \bar{c}_{jh})$ is set to $(-1,0)$, $(0,1)$, $(0,0)$, or $(-1,1)$. In particular, when $R^2_{\tilde{Y}_j.\tilde{Y}_h} \neq 0$, the proof of Corollary A.1 gives

$$\rho^-_{jh}(0) = \rho^+_{jh}(0) = \frac{r_{\tilde{W},\tilde{Y}_j} r_{\tilde{W},\tilde{Y}_h}}{r_{\tilde{Y}_j,\tilde{Y}_h}} = \frac{\sigma_{\tilde{W},\tilde{Y}_j}\sigma_{\tilde{W},\tilde{Y}_h}}{\sigma^2_{\tilde{W}}\sigma_{\tilde{Y}_j,\tilde{Y}_h}},$$

so that $0 \le M_{jh}(\rho;0) \Leftrightarrow \rho \in (-\infty,\infty)$ and $M_{jh}(\rho;0) \le 0 \Leftrightarrow \rho = \frac{\sigma_{\tilde{W},\tilde{Y}_j}\sigma_{\tilde{W},\tilde{Y}_h}}{\sigma^2_{\tilde{W}}\sigma_{\tilde{Y}_j,\tilde{Y}_h}}$. Also,

$$\rho^-_{jh}(-1) = \rho^-_{jh}(1) = R^2_{\tilde{W}.(\tilde{Y}_j,\tilde{Y}_h)'} \quad \text{and} \quad \rho^+_{jh}(-1) = \rho^+_{jh}(1) = 0.$$

Thus, when $R^2_{\tilde{Y}_j.\tilde{Y}_h} < 1$, $M_{jh}(\rho;1) = M_{jh}(\rho;-1) \le 0 \Leftrightarrow \rho \in (-\infty, 0] \cup [R^2_{\tilde{W}.(\tilde{Y}_j,\tilde{Y}_h)'}, \infty)$. Since $R^2_{\tilde{W}.(\tilde{Y}_j,\tilde{Y}_h)'} \le R^2_{\tilde{W}.\tilde{Y}}$, this magnitude restriction is not binding in $\mathcal{R}^{k,\tau,\mathbf{c}}$. It follows that Corollary A.1 yields the same bound $\mathcal{R}^{k,\tau,\mathbf{c}}$ from Corollary 3.2, with $\mathcal{R}^{\mathbf{c}}_{jh}$ determined by the magnitude restriction encoded in $M_{jh}(\rho;0) \le 0$ when $\underline{c}_{jh} = \bar{c}_{jh} = 0$ and by the sign restrictions, if any, encoded in $S_{jh}(r)$ otherwise.

B Supplementary Material on Inference

B.1 Algorithm for Inference on ρ

In order to apply only one algorithm that delivers $\hat{\rho}^l_o(\lambda; 1-\alpha_{21})$, $\hat{\rho}^u_o(\lambda; 1-\alpha_{21})$, and $CI^{\rho}_{1-\alpha_{21}}(\lambda)$, it is useful to adopt the following notation. For $r \in [0,1]$, we let

$$g^l(\pi; r, \lambda) = (g^l_1(\pi; r, \lambda), ..., g^l_M(\pi; r, \lambda)) \text{ where } g^l_v(\pi; r, \lambda) \equiv r - \rho^l_v(\lambda) \text{ for } v = 1, ..., M, \text{ and}$$
$$g^u(\pi; r, \lambda) = (g^u_1(\pi; r, \lambda), ..., g^u_M(\pi; r, \lambda)) \text{ where } g^u_v(\pi; r, \lambda) \equiv \rho^u_v(\lambda) - r \text{ for } v = 1, ..., M.$$

Thus, $\rho^l_v(\lambda) = -g^l_v(\pi; 0, \lambda)$ and[2] $\rho^u_v(\lambda) = g^u(\pi; 0, \lambda)$. Further, we collect all the lower and upper bounds, denoted by $g^c_v(\pi; r, \lambda)$ for $v = 1, ..., 2M$, into

$$g^c(\pi; r, \lambda) = (g^l(\pi; r, \lambda)', g^u(\pi; r, \lambda)')'.$$

We estimate $g^c(\pi; r, \lambda)$ using the consistent plug-in estimator $g^c(\hat{\pi}; r, \lambda)$. Using the delta method, the linearly independent subset $g^c_*(\hat{\pi}; r, \lambda)$ of $g^c(\hat{\pi}; r, \lambda)$ (recall that some of bounds in $g^c(\pi; r, \lambda)$ are constant or linearly dependent, e.g. in the single equation case or under the diagonal restrictions in A_6) is asymptotically normally distributed:

$$\sqrt{n}(g^c_*(\hat{\pi}; r, \lambda) - g^c_*(\pi; r, \lambda)) \xrightarrow{d} N(0, \nabla_\pi g^c_*(\pi; r, \lambda) \Sigma \nabla_\pi g^c_*(\pi; r, \lambda)').$$

Note that $\nabla_\pi g^c(\pi; r, \lambda)$ does not depend on r. Section B.2 collects the expressions for $g^c(\pi; r, \lambda)$, and $\nabla_\pi g^c(\pi; r, \lambda)$.

Next, for each $\ell \in \Lambda_{1-\alpha_{22}}$, we implement algorithm 1 in Chernozhukov, Lee, and Rosen (2013). To compute, $CI^{\rho}_{1-\alpha_{21}}(\ell)$, we invert a test statistic and perform a grid search over $(0, 1]$. For a thorough discussion of the algorithm[3], we refer the reader to Chernozhukov, Lee, and Rosen (2013) and Chernozhukov, Kim, Lee, and Rosen (2015).

1. Let $\alpha \leq \frac{1}{2}$ and $\mathcal{V}^c \equiv \mathcal{V}^l \cup \mathcal{V}^u \equiv \{1, ..., M\} \cup \{M+1, ..., 2M\}$.

   If the target output is:

[2] We employ $g^l(\pi; 0, \lambda)$ to transform the lower bounds for ρ into upper bounds for $-\rho$. We then use a single algorithm (for an upper bound) when estimating the lower and upper bounds for ρ.

[3] We adjust the algorithm in Chernozhukov, Lee, and Rosen (2013) slightly since some of our bounds are deterministic (e.g $\rho \leq 1$). Specifically, we use the estimated bounds to calculate the critical value. Then we report the smallest upper bound among the precision-corrected estimators and the deterministic bounds.

(a) $\hat{\rho}_o^l(\ell; 1-\alpha)$ or $\hat{\rho}_o^u(\ell; 1-\alpha)$ then set $m = l$ or u and $r = 0$.

(b) $CI_{1-\alpha}^{\rho}(\ell)$ then set $m = c$ and $r \in (0, 1]$.

2. Set $\tilde{\gamma} = 1 - \frac{0.1}{\log n}$. Simulate S draws $Z_1, .., Z_S$ from $N(0, I_{2M})$.

3. For each $v \in \mathcal{V}^c$, compute[4] $\hat{h}(v; \ell) = [\mathbf{1}(v = 1), ..., \mathbf{1}(v = 2M)][\nabla_\pi g^c(\hat{\pi}; r, \ell)\hat{\Sigma}\nabla_\pi g^c(\hat{\pi}; r, \ell)']^{\frac{1}{2}}$ and set $se(v; \ell) = \frac{1}{\sqrt{n}} \left\| \hat{h}(v; \ell) \right\|$.

4. Define $\mathcal{V}_+^m = \{v \in \mathcal{V}^m : se(v; \ell) \neq 0\}$. Compute

$$c_{\mathcal{V}^m}(\tilde{\gamma}; \ell) = \tilde{\gamma}\text{-quantile of } \{ \sup_{v \in \mathcal{V}_+^m} \frac{\hat{h}(v; \ell) Z_s}{\left\| \hat{h}(v; \ell) \right\|}, s = 1, ..., S\}$$

and

$$\hat{\mathcal{V}}^m = \{v \in \mathcal{V}_+^m : g_v^m(\hat{\pi}; 0, \ell) \leq \min_{v \in \mathcal{V}_+^m} [g_v^m(\hat{\pi}; 0, \ell) + c_{\mathcal{V}^m}(\tilde{\gamma}; \ell) se(v; \ell)] + 2c_{\mathcal{V}^m}(\tilde{\gamma}; \ell) se(v; \ell)\}.$$

5. Compute

$$c_{\hat{\mathcal{V}}^m}(\ell) = (1-\alpha)\text{-quantile of } \{ \sup_{v \in \hat{\mathcal{V}}^m} \frac{\hat{h}(v; \ell) Z_s}{\left\| \hat{h}(v; \ell) \right\|}, s = 1, ..., S\}.$$

6. Compute

$$g_o^m(\hat{\pi}; r, \ell) = \inf_{v \in \mathcal{V}^m} [g_v^m(\hat{\pi}; r, \ell)) + c_{\hat{\mathcal{V}}^m}(\ell) se(v; \ell)]$$

If $m = l$ or u then report

$$\hat{\rho}_o^l(\ell; 1-\alpha) = -g_o^l(\hat{\pi}; 0, \ell) \qquad \text{or} \qquad \hat{\rho}_o^u(\ell; 1-\alpha) = g_o^u(\hat{\pi}; 0, \ell)$$

Otherwise, if $m = c$ then report

$$CI_{1-\alpha}^{\rho}(\ell) = \{r \in (0, 1] : g_o^c(\hat{\pi}; r, \ell) \geq 0\}.$$

In the single equation bounds or when A_6 does not bind, the value ℓ of the nuisance parameters does not affect the bounds. Otherwise, let $t = 1, ..., T$ enumerate the $T \equiv \frac{1}{2}p(p-1)$ (j_t, h_t) pairs, $j_t, h_t = 1, ..., p$ with $j_t < h_t$, that correspond to the first T components

[4] $\nabla_\pi g^c(\hat{\pi}; r, \ell)\hat{\Sigma}\nabla_\pi g^c(\hat{\pi}; r, \ell)'$ may be positive semi-definite and its matrix square root is computed using a singular value decomposition.

of λ. From Corollary 3.2, we have that if ℓ is such that $(\underline{c}_{j_t h_t}, \overline{c}_{j_t h_t}) \neq (-1, 1)$, $-sgn(\ell_{T+t}) \notin [\underline{c}_{j_t h_t}, \overline{c}_{j_t h_t}]$, *and* $\ell_t = 0$ then $\mathcal{R}^{\mathbf{c}}_{j_t h_t}(\ell) = \emptyset$. As such, we drop from $\Lambda_{1-\alpha_{22}}$ the elements that satisfy these conditions since these have no effect on $CR^{\rho}_{1-\alpha_2} = \bigcup_{\ell \in \Lambda_{1-\alpha_{22}}} CI^{\rho}_{1-\alpha_{21}}(\ell)$. For the remaining components ℓ of $\Lambda_{1-\alpha_{22}}$, $CI^{\rho}_{1-\alpha_{21}}(\ell)$ depends only on the signs (negative, zero, or positive) of the elements of the first T components of $\underset{2T\times 1}{\ell}$. To speed up the computation, we remove from $\Lambda_{1-\alpha_{22}}$ the components that are redundant, so that each admissible sign configuration of the first T components of ℓ is represented only once in $\bigcup_{\ell \in \Lambda_{1-\alpha_{22}}} CI^{\rho}_{1-\alpha_{21}}(\ell)$.

B.2 Delta Method

Recall that the nuisance parameters $\lambda \equiv g^{\lambda}(\pi)$, the vector of lower and upper bounds $g^c_v(\pi; r, \lambda)$ in the intersection bounds algorithm for inference on ρ, and the parameters δ_j, β_{jl}, and Γ_{jh}, $j, h = 1, ..., p$ and $l = 1, ..., k$, (written in the form $\theta \equiv H(\pi; \rho)$) can all be expressed as functions of the estimands

$$\begin{aligned}\underset{1\times B}{\pi'} &\equiv (\underset{1\times p(k+1)}{\pi_1'}, \underset{1\times(p+k)}{\pi_2'}, \underset{1\times p(1+k)}{\pi_3'}, \underset{1\times pk}{\pi_4'}, \underset{1\times k}{\pi_5'}, \underset{1\times p}{\pi_6'}, \underset{1\times\frac{1}{2}p(p-1)}{\pi_7'}) \\ &= [vec(b_{Y.(W,X')'})', b'_{W.(Y,X')'}, (b'_{W.(Y_1,X')'}, ..., b'_{W.(Y_p,X')'}), vec(b_{Y.X})', \\ b'_{W.X}&, \sigma^{-2}_{\tilde{W}}(\sigma^2_{\tilde{Y}_1}, ..., \sigma^2_{\tilde{Y}_p}), \sigma^{-2}_{\tilde{W}}(\sigma_{\tilde{Y}_1,\tilde{Y}_2}, ..., \sigma^2_{\tilde{Y}_{p-1},\tilde{Y}_p})].\end{aligned}$$

Since the plug-in estimator $\hat{\pi}$ satisfies $\sqrt{n}(\hat{\pi} - \pi) \xrightarrow{d} N(0, \Sigma)$, the delta method gives

$$\begin{aligned}&\sqrt{n}(\hat{\lambda} - \lambda) \xrightarrow{d} N(0, \nabla_\pi g^{\lambda}(\pi)\Sigma\nabla_\pi g^{\lambda}(\pi)'), \\ &\sqrt{n}(g^c_*(\hat{\pi}; r, \lambda) - g^c_*(\pi; r, \lambda)) \xrightarrow{d} N(0, \nabla_\pi g^c_*(\pi; r, \lambda)\Sigma\nabla_\pi g^c_*(\pi; r, \lambda)'), \text{ and} \\ &\sqrt{n}(H(\hat{\pi}; r) - H(\pi; r)) \xrightarrow{d} N(0, \nabla_\pi H(\pi; r)\Sigma\nabla_\pi H(\pi; r)'),\end{aligned}$$

for any $r \in (0, 1]$. In what follows, we provide specific expressions for g^{λ}, $\nabla_\pi g^{\lambda}(\pi)$, $g^c(\pi; r, \lambda)$, $\nabla_\pi g^c(\pi; r, \lambda)$, $H(\pi; r)$ and $\nabla_\pi H(\pi; r)$.

B.2.1 Nuisance Parameters

The $2T = p(p-1)$ nuisance parameters are collected in

$$\lambda = (\lambda_1, \cdots, \lambda_{2T})' = g^{\lambda}(\pi) \equiv (\sigma^{-2}_{\tilde{W}}\sigma_{\tilde{Y}_1,\tilde{Y}_2}, ..., \sigma^{-2}_{\tilde{W}}\sigma^2_{\tilde{Y}_{p-1},\tilde{Y}_p}, b_{\tilde{Y}_1.\tilde{W}}b_{\tilde{Y}_2.\tilde{W}}, ..., b_{\tilde{Y}_{p-1}.\tilde{W}}b_{\tilde{Y}_p.\tilde{W}})'.$$

It follows that, for $t = 1, ..., T$, the components of $\underset{p(p-1)\times B}{\nabla_\pi g^\lambda(\pi)}$ are given by

$$\underset{1\times B}{\nabla_\pi g_t^\lambda(\pi)} = \left[\begin{array}{cc} \underset{1\times[p(k+1)+(p+k)+p(1+k)+pk+k+p]}{0} & \underset{1\times\frac{1}{2}p(p-1)}{\imath_t'} \end{array}\right]$$

where $\underset{\frac{1}{2}p(p-1)\times 1}{\imath_t}$ is the unit vector with 1 in the t^{th} position and 0 elsewhere, and for $t = T+1, ..., 2T$

$$\underset{1\times B}{\nabla_\pi g_t^\lambda(\pi)} = \left[\begin{array}{ccc} \underset{1\times p}{\imath_{jt}'} \otimes \left[\begin{array}{cc} b_{\tilde{Y}_{h_t}.\tilde{W}} & \underset{1\times k}{\mathbf{0}} \end{array}\right] + \underset{1\times p}{\imath_{h_t}'} \otimes \left[\begin{array}{cc} b_{\tilde{Y}_{j_t}.\tilde{W}} & \underset{1\times k}{\mathbf{0}} \end{array}\right] & \underset{1\times[(p+k)+p(1+k)+pk+k+p+\frac{1}{2}p(p-1)]}{0} \end{array}\right].$$

B.2.2 Lower and Upper Intersection bounds

Consider the joint equation bounds with $\lambda = \ell^*$ with $(\underline{c}_{j_t h_t}, \overline{c}_{j_t h_t}) \in \{(-1,0),(0,1)\}$ and $sgn(\ell_t^*) \in [\underline{c}_{j_t h_t}, \overline{c}_{j_t h_t}]\backslash\{0\}$ for $t = 1, ..., T$. In this case, we have $g^c(\pi; r, \lambda) = (g^l(\pi; r, \lambda)', g^u(\pi; r, \lambda)')'$ with

$$\underset{M\times 1}{g^l(\pi; r, \ell^*)} = \left[\begin{array}{c} r - R^2_{\tilde{W}.\tilde{Y}} \\ r - \frac{1}{1+\kappa} \\ r - \frac{1}{\tau_1} R^2_{\tilde{W}.\tilde{Y}_1} \\ \vdots \\ r - \frac{1}{\tau_p} R^2_{\tilde{W}.\tilde{Y}_p} \\ r - \frac{b_{\tilde{Y}_1.\tilde{W}} b_{\tilde{Y}_2.\tilde{W}}}{\sigma^{-2}_{\tilde{W}} \sigma_{\tilde{Y}_1,\tilde{Y}_2}} \\ \vdots \\ r - \frac{b_{\tilde{Y}_{p-1}.\tilde{W}} b_{\tilde{Y}_p.\tilde{W}}}{\sigma^{-2}_{\tilde{W}} \sigma_{\tilde{Y}_{p-1},\tilde{Y}_p}} \end{array}\right] \quad \text{and} \quad \underset{M\times 1}{g^u(\pi; r, \ell^*)} = \left[\begin{array}{c} 1-r \\ 1-r \\ 1-r \\ \vdots \\ 1-r \\ \infty \\ \vdots \\ \infty \end{array}\right]$$

where $M \equiv 2 + p + T$ (recall $T \equiv \frac{1}{2}p(p-1)$) and

$$R^2_{\tilde{W}.\tilde{Y}} = b_{\tilde{Y}.\tilde{W}} b_{\tilde{W}.\tilde{Y}} = \sum_{h=1}^p b_{Y_h.(W,X')',1} b_{W.(Y',X')',h} \quad \text{and} \quad R^2_{\tilde{W}.\tilde{Y}_j} = b_{\tilde{Y}_j.\tilde{W}} b_{\tilde{W}.\tilde{Y}_j} = b_{Y_j.(W,X')',1} b_{W.(Y_j,X')',1}.$$

In this case, the components of $\underset{2M\times B}{\nabla_\pi g^c(\pi; r, \ell^*)}$ are given by

1. For $v = 1$

$$\underset{1\times B}{\nabla_\pi g_1^c(\pi; r, \ell^*)} = \left[\begin{array}{ccc} \sum_{h=1}^p \underset{1\times p}{\imath_h'} \otimes \left[\begin{array}{cc} -b_{W.(Y',X')',h} & \underset{1\times k}{0} \end{array}\right] & \left[\begin{array}{cc} \underset{1\times p}{-b_{\tilde{Y}.\tilde{W}}} & \underset{1\times k}{0} \end{array}\right] & 0 \end{array}\right]$$

2. for $v = 2$,

$$\underset{1\times B}{\nabla_\pi g_2^c(\pi; r, \ell^*)} = 0,$$

3. for $v = 2 + j$ and $j = 1, ..., p$

$$\underset{1\times B}{\nabla_\pi g_v^c}(\pi; r, \ell^*) = \left[\begin{array}{cccc} \underset{1\times p}{\imath_j'} \otimes \left[\begin{array}{cc} -\frac{1}{\tau_j} b_{\tilde{W}.\tilde{Y}_j} & \underset{1\times k}{0} \end{array} \right] & \underset{1\times(p+k)}{0} & \underset{1\times p}{\imath_j'} \otimes \left[\begin{array}{cc} -\frac{1}{\tau_j} b_{\tilde{Y}_j.\tilde{W}} & \underset{1\times k}{0} \end{array} \right] & 0 \end{array} \right]$$

4. for $v = 2 + p + t$ and $t = 1, ..., T$,

$$\underset{1\times B}{\nabla_\pi g_v^c}(\pi; r, \ell^*) = \left[\begin{array}{ccc} \underset{1\times p(k+1)}{\nabla_{\pi_1} g_v^l(\pi; r, \ell^*)} & 0 & \underset{1\times \frac{1}{2}p(p-1)}{\nabla_{\pi_7} g_v^l(\pi; r, \ell^*)} \end{array} \right],$$

where

$$\begin{aligned} \underset{1\times p(k+1)}{\nabla_{\pi_1} g_v^l(\pi; r, \ell^*)} &= \nabla_{\pi_1}(r - \frac{b_{\tilde{Y}_{j_t}.\tilde{W}} b_{\tilde{Y}_{h_t}.\tilde{W}}}{\sigma_{\tilde{W}}^{-2} \sigma_{\tilde{Y}_{j_t}, \tilde{Y}_{h_t}}}) \\ &= (\sigma_{\tilde{W}}^{-2} \sigma_{\tilde{Y}_{j_t}, \tilde{Y}_{h_t}})^{-1} \{ \underset{1\times p}{\imath_{j_t}'} \otimes \left[\begin{array}{cc} -b_{\tilde{Y}_{h_t}.\tilde{W}} & \underset{1\times k}{\mathbf{0}} \end{array} \right] + \underset{1\times p}{\imath_{h_t}'} \otimes \left[\begin{array}{cc} -b_{\tilde{Y}_{j_t}.\tilde{W}} & \underset{1\times k}{\mathbf{0}} \end{array} \right] \} \end{aligned}$$

and

$$\underset{1\times \frac{1}{2}p(p-1)}{\nabla_{\pi_7} g_v^l(\pi; r, \ell^*)} = \nabla_{\pi_7}(r - \frac{b_{\tilde{Y}_{j_t}.\tilde{W}} b_{\tilde{Y}_{h_t}.\tilde{W}}}{\sigma_{\tilde{W}}^{-2} \sigma_{\tilde{Y}_{j_t}, \tilde{Y}_{h_t}}}) = \underset{1\times T}{\imath_t'} \otimes \frac{b_{\tilde{Y}_{j_t}.\tilde{W}} b_{\tilde{Y}_{h_t}.\tilde{W}}}{(\sigma_{\tilde{W}}^{-2} \sigma_{\tilde{Y}_{j_t}, \tilde{Y}_{h_t}})^2},$$

5. for $v = 2 + p + T + 1, ..., 2(2 + p + T)$

$$\underset{1\times B}{\nabla_\pi g_v^c}(\pi; r, \ell^*) = 0$$

Above, we set $\lambda = \ell^*$ where $(\underline{c}_{j_t h_t}, \overline{c}_{j_t h_t}) \in \{(-1, 0), (0, 1)\}$ and $sgn(\ell_t^*) \in [\underline{c}_{j_t h_t}, \overline{c}_{j_t h_t}] \backslash \{0\}$ for $t = 1, ..., T \equiv \frac{1}{2}p(p-1)$. More generally, we consider any arbitrary $\ell \in \Lambda_{1-\alpha_{22}}$ and define the matrix $\underset{2M\times 2M}{P}$ $(M \equiv 2 + p + T)$ to operationalize how the nuisance parameters λ determines whether $\mathcal{R}_{jh}^{\mathbf{c}}$ contains an upper or lower bound, if at all, according to Corollary 3.2. In particular, for $\lambda = \ell$, we let

$$\underset{2M\times 1}{g^c(\pi; r, \ell)} = P \underset{2M\times 1}{g^c(\pi; r, \ell^*)} \qquad \text{and} \qquad \underset{2M\times B}{\nabla_\pi g^c(\pi; r, \ell)} = P \underset{2M\times B}{\nabla_\pi g^c(\pi; r, \ell^*)}$$

where we set the v^{th} row P_v of P as follows, for $t = 1, ..., \frac{1}{2}p(p-1)$:

1. Set $P = \underset{2M\times 2M}{I}$.

2. If $(\underline{c}_{j_t h_t}, \overline{c}_{j_t h_t}) = (0, 0)$ and $\ell_t \neq 0$ then change $P_{M+(2+p+t)}$ to $-\imath_{2+p+t}$.

3. If $(\underline{c}_{j_t h_t}, \overline{c}_{j_t h_t}) \in \{(-1, 0), (0, 1)\}$ and $sgn(\ell_t) \notin [\underline{c}_{j_t h_t}, \overline{c}_{j_t h_t}]$ then change (a) P_{2+p+t} to $\imath_{M+(2+p+t)}$ and (b) $P_{M+(2+p+t)}$ to $-\imath_{2+p+t}$.

4. If $(\underline{c}_{j_t h_t}, \overline{c}_{j_t h_t}) \in \{(-1,0),(0,1)\}$ and $sgn(\ell_t) \in [\underline{c}_{j_t h_t}, \overline{c}_{j_t h_t}]\backslash\{0\}$ then keep (a) P_{2+p+t} as $\imath_{2+p+t}$ and (b) $P_{M+(2+p+t)}$ as $\imath_{M+(2+p+t)}$.

5. Otherwise, change P_{2+p+t} to $\imath_{M+(2+p+t)}$.

Moreover, for the j^{th} single equation bounds, P mutes the irrelevant bounds as follows:

1. Change P_1 to $\imath_{M+(2+p+1)}$

2. For $h = 1, ..., p$, if $h \neq j$ then change P_{2+h} to $\imath_{M+(2+p+h)}$

3. For $t = 1, ..., \frac{1}{2}p(p-1)$, change P_{2+p+t} to $\imath_{M+(2+p+t)}$.

B.2.3 δ_j, β_{jl}, and Γ_{jh}

We have that δ_j is given by:

$$\delta_j = D_j(\pi; r) \equiv \frac{1}{r} b_{\tilde{Y}_j.\tilde{W}} \text{ for } j = 1, ..., p,$$

where we let π enter explicitly in D_j. It follows that

$$\underset{1\times B}{\nabla_\pi D_j(\pi; r)} = \left[\begin{array}{cc} \underset{1\times p}{\imath'_j} \otimes \left[\begin{array}{cc} \frac{1}{r} & \underset{1\times k}{0} \end{array} \right] & 0 \end{array} \right].$$

Similarly, β_{jl} is given by:

$$\beta_{jl} = B_{jl}(\pi; r) \equiv b_{Y_j.X,l} - b_{W.X,l} \frac{1}{r} b_{\tilde{Y}_j.\tilde{W}} \text{ for } j = 1, ..., p \text{ and } l = 1, ..., k$$

It follows that

$$\underset{1\times B}{\nabla_\pi B_{jl}(r)} = \left[\begin{array}{cccccc} \underset{1\times p}{\imath'_j} \otimes \left[\begin{array}{cc} -\frac{1}{r} b_{W.X,l} & \underset{1\times k}{0} \end{array} \right] & \underset{1\times(p+k)}{0} & \underset{1\times p(k+1)}{0} & \underset{1\times p}{\imath'_j} \otimes \underset{1\times k}{\imath'_l} - \underset{1\times k}{\imath'_l} \otimes \frac{1}{r} b_{\tilde{Y}_j.\tilde{W}} & \underset{1\times p}{0} & \underset{1\times \frac{1}{2}p(p-1)}{0} \end{array} \right].$$

Last, Γ_{jh} is given by:

$$\Gamma_{jh} = G_{jh}(\pi; r) \equiv \sigma^{-2}_{\tilde{W}} \sigma_{\tilde{Y}_j,\tilde{Y}_h} - b_{\tilde{Y}_j.\tilde{W}} \frac{1}{r} b_{\tilde{Y}_h.\tilde{W}} \text{ for } j \leq h \text{ and } j, h = 1, ..., p.$$

Letting $\underset{1\times \frac{1}{2}p(p+1)}{\imath'_{(j,h)}}$ take the value 1 at the entry (j,h) corresponding to $\sigma^{-2}_{\tilde{W}} \sigma_{\tilde{Y}_j,\tilde{Y}_h}$, we have

$$\underset{1\times B}{\nabla_\pi G_{jh}(\pi; r)} = \left[\begin{array}{ccc} \underset{1\times p}{\imath'_j} \otimes \left[\begin{array}{cc} -\frac{1}{r} b_{\tilde{Y}_h.\tilde{W}} & \underset{1\times k}{0} \end{array} \right] + \underset{1\times p}{\imath'_h} \otimes \left[\begin{array}{cc} -\frac{1}{r} b_{\tilde{Y}_j.\tilde{W}} & \underset{1\times k}{0} \end{array} \right] & 0 & \underset{1\times \frac{1}{2}p(p+1)}{\imath'_{(j,h)}} \end{array} \right].$$

Joint Confidence Regions We sometimes construct a confidence region for the vector of parameters $\beta_{\cdot l} = (\beta_{1l}, \beta_{2l}, \beta_{3l})'$ associated with the variable X_l in the system of Y equations. For a given $(\kappa, \tau, \mathbf{c})$ and α_1, α_{21} and α_{22}, we construct a $1-\alpha_2$ confidence region $CR^{\rho}_{1-\alpha_2}$ for ρ. For each $r \in CR^{\rho}_{1-\alpha_2}$, the delta method gives the asymptotic distribution of the plug-in estimator $\hat{B}_{\cdot 1}(\pi; r)$ for

$$B_{\cdot 1}(\pi; r) = b_{Y.X,l} - b_{W.X,l}\frac{1}{r}b_{\tilde{Y}.\tilde{W}}.$$

Specifically, we have that

$$\sqrt{n}(B_{\cdot 1}(\hat{\pi}; r) - B_{\cdot 1}(\pi; r)) \xrightarrow{d} N(0, \Sigma_{B_{\cdot 1}}(r)) \text{ where } \Sigma_{B_{\cdot 1}}(r) = \nabla_{\pi} B_{\cdot 1}(\pi; r)\Sigma\nabla_{\pi} B_{\cdot 1}(\pi; r)'$$

and $\underset{p\times B}{\nabla_{\pi} B_{\cdot 1}(\pi; r)}$ stacks the expressions $\underset{1\times B}{\nabla_{\pi} B_{jl}(\pi; r)}$ for $j = 1, ..., p$ derived above. We construct a $1-\alpha_1$ confidence region $CR^{B_{\cdot 1}}_{1-\alpha_1}(r)$ for $B_{\cdot 1}(\pi; r)$, by inverting the following Wald statistic which has an asymptotic χ^2_p distribution:

$$CR^{B_{\cdot 1}}_{1-\alpha_1}(r) = \{b_{\cdot 1} \in \mathcal{B}^*_{\cdot 1} : \sqrt{n}(B_{\cdot 1}(\hat{\pi}; r) - b_{\cdot 1})'\Sigma^{-1}_{B_{\cdot 1}}(r)\sqrt{n}(B_{\cdot 1}(\hat{\pi}; r) - b_{\cdot 1}) \le c_{1-\alpha_1}\}$$

where $c_{1-\alpha_1}$ denotes the $1-\alpha_1$ quantile of χ^2_p and where we search over an initial neighborhood $\mathcal{B}^*_{\cdot 1}$. For instance, we let $\mathcal{B}^*_{\cdot 1}$ be the cube that contains each of the p unidimensional 95% confidence regions:

$$\mathcal{B}^*_{\cdot 1} = \{(b_{11}, ..., b_{p1}) : B_{j1}(\hat{\pi}; r) - 3se(B_{j1}(\hat{\pi}; r)) \le b_{j1} \le B_{j1}(\hat{\pi}; r) + 3se(B_{j1}(\hat{\pi}; r)) \text{ for } j = 1, ..., p\}.$$

Last, we construct the confidence region $CR^{\beta_{\cdot 1}}_{1-\alpha_1-\alpha_2}$ for $\beta_{\cdot 1}$ by forming the union:

$$CR^{\beta_{\cdot 1}}_{1-\alpha_1-\alpha_2} = \bigcup_{r \in CR^{\rho}_{1-\alpha_2}} CR^{B_{\cdot 1}}_{1-\alpha_1}(r)$$

and use $CR^{\beta_{\cdot 1}}_{1-\alpha_1-\alpha_2}$ to form decisions regarding a null hypothesis for $(\beta_{11}, ..., \beta_{p1})$.

C Extension of the Framework to Panel Data

Consider the unbalanced panel equations with firm fixed effects γ_i:

$$\underset{1\times p}{Y_{it}}' = \underset{1\times p}{\gamma_i}' + \underset{1\times k}{X'_{it}}\underset{k\times p}{\beta} + \underset{1\times 1}{U_{it}}\underset{1\times p}{\delta} + \underset{1\times p}{\eta_{it}}' \quad \text{and} \quad \underset{1\times 1}{W_{it}} = \underset{1\times 1}{U_{it}} + \underset{1\times 1}{\varepsilon_{it}} \quad \text{for } i = 1, ..., n \text{ and } t \in S_i.$$

We assume that the data is missing at random from certain time periods. Specifically, we let T denote[5] the total number of time periods in the panel and S_i, for $i = 1, ..., n$, denote the subset of T in which the data on firm i are observed, with T_i denoting the cardinality of S_i. When time fixed effects are included, X_{it} contains $T_i - 1$ indicator variables corresponding to the years in S_i. We let $E(\eta_{it}) = \mu_\eta$ and $E(\varepsilon_{it}) = \mu_\varepsilon$ for $i = 1, .., n$ and $t \in S_i$ and we consider the case where n is large relative to $T_1, ..., T_n$.

Letting $\bar{A}_i \equiv \frac{1}{T_i} \sum_{t \in S_i} A_{it}$ and $\ddot{A}_{it} \equiv A_{it} - \bar{A}_i$, the fixed effect γ_i drops out from the $\ddot{Y}_{it}$ equation:

$$\underset{1\times p}{\ddot{Y}'_{it}} = \underset{1\times k}{\ddot{X}'_{it}} \underset{k\times p}{\beta} + \underset{1\times 1}{\ddot{U}_{it}} \underset{1\times p}{\delta} + \underset{1\times p}{\ddot{\eta}'_{it}} \quad \text{and} \quad \underset{1\times 1}{\ddot{W}_{it}} = \underset{1\times 1}{\ddot{U}_{it}} + \underset{1\times 1}{\ddot{\varepsilon}_{it}}.$$

Letting $\underset{T_i\times k}{\ddot{A}_i} \equiv [\ddot{A}'_{i1}, ..., \ddot{A}'_{iT_i}]'$, we obtain the panel analogue of assumption A_1:

$$\underset{T_i\times p}{\ddot{Y}_i} = \underset{T_i\times k}{\ddot{X}_i} \underset{k\times p}{\beta} + \underset{T_i\times 1}{\ddot{U}_i} \underset{1\times p}{\delta} + \underset{T_i\times p}{\ddot{\eta}_i} \quad \text{and} \quad \underset{T_i\times 1}{\ddot{W}_i} = \underset{T_i\times 1}{\ddot{U}_i} + \underset{T_i\times 1}{\ddot{\varepsilon}_i} \quad \text{for } i = 1, ..., n.$$

Suppose that A_2-A_3 hold for this equation. Specifically, let

$$Cov[\eta_{it}, (X_{is}, U_{is})] = 0 \quad \text{and} \quad Cov[\varepsilon_{it}, (X_{is}, U_{is}, \eta_{is})] = 0 \text{ for } i = 1, ..., n \text{ and } t, s \in S_i.$$

This imposes "strict exogeneity" across time periods, as is common when applying a within transformation. Given that $\ddot{A}_{it} \equiv A_{it} - \frac{1}{T_i} \sum_{t \in S_i} A_{it}$, we obtain

$$Cov[\ddot{\eta}_{it}, (\ddot{X}_{is}, \ddot{U}_{is})] = 0 \quad \text{and} \quad Cov[\ddot{\varepsilon}_{it}, (\ddot{X}_{is}, \ddot{U}_{is}, \ddot{\eta}_{is})] = 0 \text{ for } i = 1, ..., n \text{ and } t, s \in S_i.$$

Let the binary indicator I_{it}, $i = 1, ..., n$ and $t = 1, ..., T$ denote whether the observation (Y_{it}, X_{it}, W_{it}) is missing (at random). Let $\underset{T\times 1}{I_i}$ stack I_{it} for $t = 1, ..., T$. Let

$$\sigma_{\ddot{A}_i, \ddot{B}_i} \equiv E(\underset{a\times T_i}{\ddot{A}'_i} \underset{T_i\times b}{\ddot{B}_i}) = E(\sum_{t \in S_i} \underset{a\times 1}{\ddot{A}_{it}} \underset{1\times b}{\ddot{B}'_{it}}) = \sum_{t=1}^{T} E(I_{it} \underset{a\times 1}{\ddot{A}_{it}} \underset{1\times b}{\ddot{B}'_{it}}) = \sum_{t=1}^{T} E(I_{it}) E(\underset{a\times 1}{\ddot{A}_{it}} \underset{1\times b}{\ddot{B}'_{it}}).$$

In particular, we have $\sigma_{\ddot{X}_i, \ddot{\eta}_i} = 0$ and $\sigma_{\ddot{X}_i, \ddot{\varepsilon}_i} = 0$. Further, let

$$b_{\ddot{A}_i . \ddot{B}_i} \equiv \sigma^{-2}_{\ddot{B}_i} \sigma_{\ddot{B}_i, \ddot{A}_i} \quad \text{and} \quad \epsilon'_{\ddot{A}_{it} . \ddot{B}_{it}} \equiv \ddot{A}'_{it} - \ddot{B}'_{it} b_{\ddot{A}_i . \ddot{B}_i}.$$

Then, provided $\sigma^2_{\ddot{X}_i}$ is nonsingular,

$$\beta = b_{\ddot{Y}_i . \ddot{X}_i} - b_{\ddot{W}_i . \ddot{X}_i} \delta.$$

[5] The number of time periods T should not be confused with the dimension of the nuisance parameter $\underset{2T\times 1}{\lambda}$ in Online Appendix B.

Let $\tilde{A}_{it} \equiv \epsilon_{\ddot{A}_{it}.\ddot{X}_{it}}$ and $\tilde{A}_i = [\tilde{A}'_{i1}, ..., \tilde{A}'_{iT_i}]'$. By A$_1$-A$_3$, we obtain

$$\underset{T_i \times p}{\tilde{Y}_i} = \underset{T_i \times 1}{\tilde{U}_i} \underset{1 \times p}{\delta} + \underset{T_i \times p}{\tilde{\eta}_i} \quad \text{and} \quad \underset{T_i \times 1}{\tilde{W}_i} = \underset{T_i \times 1}{\tilde{U}_i} + \underset{T_i \times 1}{\tilde{\varepsilon}_i}.$$

In particular, we have

$$\sigma^2_{\tilde{W}_i} = \sigma^2_{\tilde{U}_i} + \sigma^2_{\tilde{\varepsilon}_i}, \qquad \sigma_{\tilde{W}_i,\tilde{Y}_i} = \sigma_{\tilde{W}_i,\tilde{U}_i}\delta = \sigma^2_{\tilde{U}_i}\delta, \quad \text{and } \sigma^2_{\tilde{Y}_i} = \delta'\sigma^2_{\tilde{U}_i}\delta + \sigma^2_{\tilde{\eta}_i}.$$

Provided $\sigma^2_{\tilde{W}_i}$ is nonsingular, we have

$$b_{\tilde{W}_i.\tilde{Y}_i} = \sigma^{-2}_{\tilde{W}_i}\sigma_{\tilde{W}_i,\tilde{Y}_i} = \rho\delta \quad \text{and} \quad \sigma^{-2}_{\tilde{W}_i}\sigma^2_{\tilde{Y}_i} = \delta'\rho\delta + \sigma^{-2}_{\tilde{W}_i}\sigma^2_{\tilde{\eta}_i},$$

where

$$\rho = \sigma^{-2}_{\tilde{W}_i}\sigma^2_{\tilde{U}_i} = E(\sum_{t \in S_i} \tilde{W}_{it}\tilde{W}'_{it})^{-1} E(\sum_{t \in S_i} \tilde{U}_{it}\tilde{U}'_{it}).$$

Given $\rho \neq 0$, we obtain the representation from Theorem 3.1 and we can apply the results of the paper to the transformed variables. For inference, we use the robust standard errors that are clustered at the firm level. For example, we estimate $b_{\ddot{A}_i.\ddot{B}_i}$ and $\underset{T_i \times a}{\epsilon_{\ddot{A}_i.\ddot{B}_i}} = (\epsilon'_{\ddot{A}_{i1}.\ddot{B}_{i1}}, ..., \epsilon'_{\ddot{A}_{iT_i}.\ddot{B}_{iT_i}})'$ using their plug in sample analogues

$$\hat{b}_{\ddot{A}_i.\ddot{B}_i} \equiv (\frac{1}{n}\sum_{i=1}^{n} \underset{b \times T_i}{\ddot{B}'_i} \underset{T_i \times b}{\ddot{B}_i})^{-1}(\frac{1}{n}\sum_{i=1}^{n} \underset{b \times T_i}{\ddot{B}'_i} \underset{T_i \times a}{\ddot{A}_i}) \text{ and } \underset{1 \times a}{\hat{\epsilon}'_{\ddot{A}_{it}.\ddot{B}_{it}}} = \ddot{A}'_{it} - \ddot{B}'_{it}\hat{b}_{\ddot{A}_i.\ddot{B}_i}$$

and estimate the asymptotic variance of $\sqrt{n}(\hat{b}_{\ddot{A}_i.\ddot{B}_i} - b_{\ddot{A}_i.\ddot{B}_i})$ by

$$(\frac{1}{n}\sum_{i=1}^{n} \underset{b \times T_i}{\ddot{B}'_i} \underset{T_i \times b}{\ddot{B}_i})^{-1}(\frac{1}{n}\sum_{i=1}^{n} \underset{k \times T_i}{\ddot{B}'_i} \underset{T_i \times p}{\hat{\epsilon}_{\ddot{A}_i.\ddot{B}_i}} \underset{p \times T_i}{\hat{\epsilon}'_{\ddot{A}_i.\ddot{B}_i}} \underset{T_i \times k}{\ddot{B}'_i})(\frac{1}{n}\sum_{i=1}^{n} \underset{b \times T_i}{\ddot{B}'_i} \underset{T_i \times b}{\ddot{B}_i})^{-1}$$

Note that the interpretation of A$_4$-A$_6$ applies to the stacked and within-transformed variables. In particular, A$_4$ assumes that

$$\sigma^2_{\tilde{\varepsilon}_i} = E(\sum_{t \in S_i} \tilde{\varepsilon}^2_{it}) = \sum_{t=1}^{T} E(I_{it})E(\tilde{\varepsilon}^2_{it}) \leq \kappa\sigma^2_{\tilde{U}_i} = \kappa E(\sum_{t \in S_i} \tilde{U}^2_{it}) = \kappa \sum_{t=1}^{T} E(I_{it})E(\tilde{U}^2_{it}).$$

For this to hold, it suffices that $E(\tilde{\varepsilon}^2_{it}) \leq \kappa E(\tilde{U}^2_{it})$ for $t = 1, ..., T$. A$_5$ assumes that

$$R^2_{\tilde{Y}_{ji}.\tilde{U}_i} = 1 - \frac{\sigma^2_{\tilde{\eta}_{ji}}}{\sigma^2_{\tilde{Y}_{ji}}} = 1 - \frac{E(\sum_{t \in S_i} \tilde{\eta}^2_{jit})}{E(\sum_{t \in S_i} \tilde{Y}^2_{jit})} = 1 - \frac{\sum_{t=1}^{T} E(I_{it})E(\tilde{\eta}^2_{jit})}{\sum_{t=1}^{T} E(I_{it})E(\tilde{Y}^2_{jit})} \leq \tau_j,$$

and it suffices for this that $R^2_{\tilde{Y}_{jit}.\tilde{U}_{it}} = 1 - \frac{\sigma^2_{\tilde{\eta}_{jit}}}{\sigma^2_{\tilde{Y}_{jit}}} \leq \tau_j$ for $t = 1, ..., T$. And A_6 assumes that

$$\underline{c}_{jh} \leq r_{\tilde{\eta}_{ji},\tilde{\eta}_{hi}} = \frac{E(\sum_{t\in S_i} \tilde{\eta}_{jit}\tilde{\eta}_{hit})}{E(\sum_{t\in S_i} \tilde{\eta}^2_{jit})^{\frac{1}{2}} E(\sum_{t\in S_i} \tilde{\eta}^2_{hit})^{\frac{1}{2}}} = \frac{\sum_{t=1}^{T} E(I_{it})E(\tilde{\eta}_{jit}\tilde{\eta}_{hit})}{[\sum_{t=1}^{T} E(I_{it})E(\tilde{\eta}^2_{jit})]^{\frac{1}{2}} [\sum_{t=1}^{T} E(I_{it})E(\tilde{\eta}^2_{hit})]^{\frac{1}{2}}} \leq \overline{c}_{jh},$$

which holds if one imposes the same sign restriction on $Cov(\tilde{\eta}_{jit}, \tilde{\eta}_{hit})$ for $t = 1, ..., T$.

The panel analysis without fixed effects proceeds similarly but omits the within transformation (i.e. it sets $\gamma_i = \gamma$ for $i = 1, ..., n$ and $\ddot{A}_{it} = A_{it} - \frac{1}{n}\frac{1}{T_i}\sum_{i=1}^{n}\sum_{t\in S_i} A_{it}$).

D Mathematical Proofs

Proof of Theorem 3.1: By A_2-A_3, $Cov[(\eta', \varepsilon)', X] = 0$. Since $Var(X)$ is nonsingular, A_1 gives

$$\beta = b_{Y.X} - b_{W.X}\delta.$$

A_2-A_3 also give $\sigma_{\tilde{U},\varepsilon} = 0$ and $\sigma_{\tilde{U},\eta} = \sigma_{\varepsilon,\eta} = 0$. Using $\tilde{\varepsilon} = \varepsilon - E(\varepsilon)$ and $\tilde{\eta} = \eta - E(\eta)$ together with $\tilde{Y}' = \tilde{U}\delta + \tilde{\eta}'$ and $\tilde{W} = \tilde{U} + \tilde{\varepsilon}$, we have

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

Since $Var[(X', U)']$ is nonsingular, $\sigma^2_{\tilde{U}} \neq 0$. Thus, $\sigma^2_{\tilde{W}} \neq 0$ and

$$b_{\tilde{Y}.\tilde{W}} \equiv \sigma^{-2}_{\tilde{W}}\sigma_{\tilde{W},\tilde{Y}} = \rho\delta \quad \text{and} \quad \sigma^{-2}_{\tilde{W}}\sigma^2_{\tilde{Y}} = \delta'\rho\delta + \Gamma.$$

Since $\rho \neq 0$, we obtain

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

Lemma D.1 *Under the conditions of Theorem 3.1,* $R^2_{\tilde{Y}_j.\tilde{W}} \leq R^2_{\tilde{Y}_j.\tilde{U}}$.

Proof of Lemma D.1: If $\sigma^2_{\tilde{Y}_j} = 0$, set $R^2_{\tilde{Y}_j.\tilde{W}} = R^2_{\tilde{Y}_j.\tilde{U}} = 0$. If $0 < \sigma^2_{\tilde{Y}_j}$, we have

$$R^2_{\tilde{Y}_j.\tilde{W}} = \frac{\sigma^2_{\tilde{W}}}{\sigma^2_{\tilde{Y}_j}} b^2_{\tilde{Y}_j.\tilde{W}} = \frac{\sigma^2_{\tilde{W}}}{\sigma^2_{\tilde{Y}_j}} (\delta_j \rho)^2 \text{ and}$$
$$R^2_{\tilde{Y}_j.\tilde{U}} = 1 - \frac{\sigma^2_{\eta_j}}{\sigma^2_{\tilde{Y}_j}} = \frac{1}{\sigma^2_{\tilde{Y}_j}}(\sigma^2_{\tilde{Y}_j} - \sigma^2_{\eta_j}) = \frac{1}{\sigma^2_{\tilde{Y}_j}} \delta^2_j \sigma^2_{\tilde{U}} = \frac{\sigma^2_{\tilde{W}}}{\sigma^2_{\tilde{Y}_j}} \delta^2_j \rho$$

It follows that

$$R^2_{\tilde{Y}_j.\tilde{U}} - R^2_{\tilde{Y}_j.\tilde{W}} = \frac{\sigma^2_{\tilde{W}}}{\sigma^2_{\tilde{Y}_j}} (\delta^2_j \rho - \delta^2_j \rho^2) = \frac{\sigma^2_{\tilde{W}}}{\sigma^2_{\tilde{Y}}} \rho(1-\rho)\delta^2_j \geq 0.$$

Proof of Corollary 3.2: The identification set $\mathcal{J}^{k,\tau,\mathbf{c}}$ obtains from A_1-A_6 and the ($Var[(\tilde{Y}', \tilde{W})']$) moments given by (in)equalities (3,5,6,7), using the expressions in Theorem 3.1. To show that $\mathcal{J}^{k,\tau,\mathbf{c}}$ is sharp, let $d = D(r)$, $b = B(r)$, and $g = G(r)$. We show that for each $(r, d, b, g) \in \mathcal{J}^{k,\tau,\mathbf{c}}$ there exist random variables $(U^*, \eta^*, \varepsilon^*)$ such that $Y' = X'b + U^*d + \eta^*$ and $W = U^* + \varepsilon^*$ that satisfy A_2-A_6. Specifically, $(X, U^*, \varepsilon^*, \eta^*)$ satisfy A_2-A_3, $Cov[\eta^*, (X', U^*)'] = 0$, $Cov[\varepsilon^*, (\eta^*, X', U^*)'] = 0$. Further, $\frac{\sigma^2_{\tilde{U}^*}}{\sigma^2_{\tilde{W}}} = r$ and thus A_4 holds, $\sigma^2_{\varepsilon^*} \leq \kappa \sigma^2_{\tilde{U}^*}$. Last, $G(r) = \sigma^{-2}_{\tilde{W}} \sigma^2_{\tilde{\eta}^*}$ and therefore A_5 holds since, when $\sigma^2_{\tilde{Y}_j} \neq 0$,

$$1 - \frac{\sigma^2_{\eta^*_j}}{\sigma^2_{\tilde{Y}_j}} = \frac{\sigma^2_{\tilde{W}}}{\sigma^2_{\tilde{Y}_j}} (\frac{\sigma^2_{\tilde{Y}_j}}{\sigma^2_{\tilde{W}}} - G_{jj}(r)) \leq \frac{\sigma^2_{\tilde{W}}}{\sigma^2_{\tilde{Y}_j}} [\frac{\sigma^2_{\tilde{Y}_j}}{\sigma^2_{\tilde{W}}} - \frac{\sigma^2_{\tilde{Y}_j}}{\sigma^2_{\tilde{W}}}(1 - \tau_j)] = \tau_j,$$

and A_6 holds since $\underline{c}_{jh} \leq sgn(G_{jh}(r)) \leq \overline{c}_{jh}$.

To construct these variables we proceed similarly to Chalak and Kim (2018, proof of corollary 3.2). In particular, we let V be any random variable such that $\tilde{V} \equiv \epsilon_{V.X}$ is nondegenerate and satisfies

$$\sigma_{\tilde{W}.\tilde{V}} = \sqrt{r} \sigma_{\tilde{V}} \sigma_{\tilde{W}} \quad \text{and} \quad \sigma_{\tilde{Y}.\tilde{V}} = \frac{1}{\sqrt{r}} \sigma_{\tilde{V}} \sigma_{\tilde{W}} \frac{\sigma_{\tilde{Y}.\tilde{W}}}{\sigma^2_{\tilde{W}}}.$$

Note that these covariance restrictions are coherent. Specifically,

$$Var(\tilde{V}, \tilde{W}, \tilde{Y}') = \begin{bmatrix} \sigma^2_{\tilde{V}} & \sqrt{r} \sigma_{\tilde{V}} \sigma_{\tilde{W}} & \frac{\sigma_{\tilde{V}} \sigma_{\tilde{W}}}{\sqrt{r}} \frac{\sigma_{\tilde{W}.\tilde{Y}}}{\sigma^2_{\tilde{W}}} \\ \sqrt{r} \sigma_{\tilde{V}} \sigma_{\tilde{W}} & \sigma^2_{\tilde{W}} & \sigma_{\tilde{W}.\tilde{Y}} \\ \frac{\sigma_{\tilde{V}} \sigma_{\tilde{W}}}{\sqrt{r}} \frac{\sigma_{\tilde{Y}.\tilde{W}}}{\sigma^2_{\tilde{W}}} & \sigma_{\tilde{Y},\tilde{W}} & \sigma^2_{\tilde{Y}} \end{bmatrix}$$

is positive semi-definite because $0 < \sigma^2_{\tilde{V}}$ and its Schur complement

$$0 \preceq \sigma^2_{(\tilde{W}, \tilde{Y}')'} - \sigma_{(\tilde{W}, \tilde{Y}')', \tilde{V}} \sigma^{-2}_{\tilde{V}} \sigma_{\tilde{V}, (\tilde{W}, \tilde{Y}')'} = \begin{bmatrix} (1-r)\sigma^2_{\tilde{W}} & 0 \\ 0 & \sigma^2_{\tilde{W}} G(r) \end{bmatrix}$$

is positive semi-definite since it is block diagonal with $0 \leq (1-r)\sigma^2_{\tilde{W}}$ and $0 \preceq G(r)$.

For instance, to construct V, set $\sigma_{\tilde{V}}$ to some value (e.g. $\sigma_{\tilde{V}} = 1$) and let ϑ be any random variable that is uncorrelated with $(X', W, Y)'$ (e.g. a residual from a regression on $(X', W, Y'))$. When $\sigma^2_{(\tilde{W},\tilde{Y}')'}$ is nonsingular, one can use the above restrictions on $\sigma_{\tilde{W}.\tilde{V}}$ and $\sigma_{\tilde{Y}.\tilde{V}}$ to construct $b_{\tilde{V}.(\tilde{W},\tilde{Y}')'}$ and the scalar

$$\varkappa = \{\frac{1}{\sigma^2_\vartheta}[\sigma^2_{\tilde{V}} - b'_{\tilde{V}.(\tilde{W},\tilde{Y}')'}\sigma^2_{(\tilde{W},\tilde{Y}')}b_{\tilde{V}.(\tilde{W},\tilde{Y}')'}]\}^{\frac{1}{2}}$$

($\varkappa$ is set such that the variance of the generated $\tilde{V}$ is $\sigma^2_{\tilde{V}}$) in order to generate

$$\tilde{V} = (\tilde{W}, \tilde{Y})b_{\tilde{V}.(\tilde{W},\tilde{Y}')'} + \varkappa\vartheta.$$

If $\sigma^2_{(\tilde{W},\tilde{Y})'}$ is singular, one can generate $\tilde{V}$ by omitting the redundant $\tilde{Y}$ components from the above regression construction. Last, $V = X'b_{V.X} + \tilde{V} + E[V - X'b_{V.X}]$ obtains by setting $b_{V.X}$ and $E(V)$ to some value (e.g. zero).

Then it suffices to construct U^*, ε^*, and η^* as follows

$$W \equiv (X', V)b_{W.(X',V)'} + \{\epsilon_{W.(X',V)'} + E[W - (X', V)b_{W.(X',V)'}\} \equiv U^* + \varepsilon^*,$$

and, if $r \neq 1$,

$$Y \equiv (X', V, \varepsilon^*)b_{Y.(X',V,\varepsilon^*)'} + \{\epsilon_{Y.(X',V,\varepsilon^*)'} + E[Y - (X', V, \varepsilon^*)b_{Y.(X',V,\varepsilon^*)'}]\} \equiv (X', V, \varepsilon^*)b_{Y.(X',V,\varepsilon^*)'} + \eta^*$$

whereas if $r = 1$ then $r_{\tilde{W},\tilde{V}} = 1$ and $\epsilon_{W.(X',V)'} = \epsilon_{\tilde{W}.\tilde{V}} = 0$ and

$$Y = (X', V)b_{Y.(X',V)'} + \{\epsilon_{Y.(X',V)'} + E[Y - (X', V)b_{Y.(X',V)'}]\} \equiv (X', V)b_{Y.(X',V)'} + \eta^*.$$

In particular, $(X, U^*, \varepsilon^*, \eta^*)$ satisfy A_2-A_3 since by construction $Cov[\eta^*, (X', U^*)'] = 0$ and $Cov[\varepsilon^*, (\eta^*, X', U^*)'] = 0$. To verify that A_1 holds, note that if $r \neq 1$,

$$\begin{aligned} Y &= Vb_{\tilde{Y}.\tilde{V}} + X'(b_{Y.X} - b_{V.X}b_{\tilde{Y}.\tilde{V}}) + \varepsilon^* b_{Y.\varepsilon^*} + \{\epsilon_{Y.(X',V,\varepsilon^*)'} + E[Y - (X', V, \varepsilon^*)b_{Y.(X',V,\varepsilon^*)'}]\} \\ &= Vb_{\tilde{W}.\tilde{V}}d + X'(b_{W.X} - b_{V.X}b_{\tilde{W}.\tilde{V}})d + X'(b_{Y.X} - b_{W.X}d) + \varepsilon^* b_{Y.\varepsilon^*} + \eta^* \\ &= (X', V)b_{W.(X',V)'}d + X'b + \eta^* \\ &\equiv U^*d + X'b + \eta^* \end{aligned}$$

where the first equality uses $Cov[\varepsilon^*, (X', V)'] = 0$ and partitioned regression, the second equality makes use of

$$b_{\tilde{Y}.\tilde{V}} = \sigma_{\tilde{W}}^{-2}\sigma_{\tilde{Y}.\tilde{V}} = \sigma_{\tilde{W}}^{-2}\frac{1}{\sqrt{r}}\sigma_{\tilde{W}}\sigma_{\tilde{V}}\frac{\sigma_{\tilde{Y}.\tilde{W}}}{\sigma_{\tilde{W}}^2} = \sigma_{\tilde{W}}^{-2}\frac{1}{r}\sigma_{\tilde{W}.\tilde{V}}b_{\tilde{Y}.\tilde{W}} = b_{\tilde{W}.\tilde{V}}d,$$

and the third equality uses partitioned regression, $b = b_{Y.X} - b_{W.X}d$, and

$$\begin{aligned} b_{Y.\varepsilon^*} = b_{\tilde{Y}.\epsilon_{W.(X',V)'}} &= \frac{\sigma_{\tilde{Y},\epsilon_{W.(X',V)'}}}{\sigma^2_{\epsilon_{W.(X',V)'}}} = \frac{1}{\sigma^2_{\epsilon_{\tilde{W}.\tilde{V}}}}Cov(\tilde{Y}, \tilde{W} - \tilde{V}b_{\tilde{W}.\tilde{V}}) \\ &= \frac{1}{(1-r)\sigma_{\tilde{W}}^2}[\sigma_{\tilde{Y}.\tilde{W}} - \frac{(\frac{1}{\sqrt{r}}\sigma_{\tilde{W}}\sigma_{\tilde{V}}\frac{\sigma_{\tilde{Y},\tilde{W}}}{\sigma_{\tilde{W}}^2})\sqrt{r}\sigma_{\tilde{V}}\sigma_{\tilde{W}}}{\sigma_{\tilde{V}}^2}] = 0. \end{aligned}$$

If $r = 1$, a similar calculation gives,

$$\begin{aligned} Y &= (X', V)b_{Y.(X',V)'} + \{\epsilon_{Y.(X',V)'} + E[Y - (X', V)b_{Y.(X',V)'}]\} \\ &= (X', V)b_{W.(X',V)'}d + X'b + \eta^* \equiv U^*d + X'b + \eta^*. \end{aligned}$$

Last, to verify that A_4-A_6 hold, it suffices to verify that

$$\begin{aligned} &\frac{\sigma_{\tilde{U}^*}^2}{\sigma_{\tilde{W}}^2} = \frac{Var(\tilde{V}b_{\tilde{W}.\tilde{V}})}{\sigma_{\tilde{W}}^2} = \frac{\sigma_{\tilde{W}.\tilde{V}}^2}{\sigma_{\tilde{V}}^2\sigma_{\tilde{W}}^2} = r, \text{ and} \\ &G(r) = \sigma_{\tilde{W}}^{-2}\sigma_{\tilde{Y}}^2 - b'_{\tilde{Y}.\tilde{W}}\frac{1}{r}b_{\tilde{Y}.\tilde{W}} = \sigma_{\tilde{W}}^{-2}(d'\sigma_{\tilde{U}^*}^2 d + \sigma_{\tilde{\eta}^*}^2) - b'_{\tilde{Y}.\tilde{W}}\frac{1}{r}b_{\tilde{Y}.\tilde{W}} = \sigma_{\tilde{W}}^{-2}\sigma_{\tilde{\eta}^*}^2. \end{aligned}$$

Next, we derive the identification region $\mathcal{R}^{k,\tau,\mathbf{c}}$ for ρ. First, we show that $R^2_{\tilde{W}.\tilde{Y}} \leq \rho \leq 1$. If $\sigma_{\tilde{Y},\tilde{W}} = 0$ then $R^2_{\tilde{W}.\tilde{Y}} = 0 \leq \rho \leq 1$. Suppose that $\sigma_{\tilde{Y},\tilde{W}} \neq 0$. Since $0 < \rho$ and $0 \preceq \Gamma$ then for any vector $\underset{p\times 1}{x}$, we have

$$0 \leq \rho x'\sigma_{\tilde{W}}^{-2}\sigma_{\tilde{Y}}^2 x - x'b'_{\tilde{Y}.\tilde{W}}b_{\tilde{Y}.\tilde{W}}x.$$

Suppose that $\sigma_{\tilde{Y}}^2$ is positive definite so that $0 < \sigma_{\tilde{W},\tilde{Y}}\sigma_{\tilde{Y}}^{-2}\sigma_{\tilde{Y},\tilde{W}}$ (this is without loss of generality since we can drop the redundant $\tilde{Y}$ components otherwise). In particular, for $x = \sigma_{\tilde{Y}}^{-2}\sigma_{\tilde{Y},\tilde{W}}$, we obtain

$$R^2_{\tilde{W}.\tilde{Y}} = \sigma_{\tilde{W}}^{-2}\sigma_{\tilde{W},\tilde{Y}}\sigma_{\tilde{Y}}^{-2}\sigma_{\tilde{Y},\tilde{W}} = \frac{(\sigma_{\tilde{W},\tilde{Y}}\sigma_{\tilde{Y}}^{-2})\sigma_{\tilde{Y},\tilde{W}}\sigma_{\tilde{W}}^{-2}\sigma_{\tilde{W}}^{-2}\sigma_{\tilde{W},\tilde{Y}}(\sigma_{\tilde{Y}}^{-2}\sigma_{\tilde{Y},\tilde{W}})}{(\sigma_{\tilde{W},\tilde{Y}}\sigma_{\tilde{Y}}^{-2})\sigma_{\tilde{W}}^{-2}\sigma_{\tilde{Y}}^2(\sigma_{\tilde{Y}}^{-2}\sigma_{\tilde{Y},\tilde{W}})} \leq \rho \leq 1.$$

Second, by A_4, we have $1 - \rho = \frac{\sigma_{\tilde{\varepsilon}}^2}{\sigma_{\tilde{W}}^2} \leq \kappa\frac{\sigma_{\tilde{U}}^2}{\sigma_{\tilde{W}}^2} = \kappa\rho$ and thus $\rho \in [\frac{1}{1+\kappa}, 1]$. Third, by A_5, we have that for $j = 1, ..., p$, $R^2_{\tilde{Y}_j.\tilde{U}} = (1 - \frac{\sigma_{\eta_j}^2}{\sigma_{\tilde{Y}_j}^2}) \leq \tau_j$ (recall that if $\sigma_{\tilde{Y}_j}^2 = 0$ then we set

$R^2_{\tilde{Y}_j.\tilde{U}} = R^2_{\tilde{W}.\tilde{Y}_j} = 0$). Multiplying by $\frac{\sigma^2_{\tilde{Y}_j}}{\sigma^2_{\tilde{W}}}$ and substituting for Γ_{jj} we obtain

$$b'_{\tilde{Y}_j.\tilde{W}}\frac{1}{\rho}b_{\tilde{Y}_j.\tilde{W}} = \frac{\sigma^2_{\tilde{Y}_j}}{\sigma^2_{\tilde{W}}} - (\sigma^{-2}_{\tilde{W}}\sigma^2_{\tilde{Y}_j} - b'_{\tilde{Y}_j.\tilde{W}}\frac{1}{\rho}b_{\tilde{Y}_j.\tilde{W}}) \leq \tau_j\frac{\sigma^2_{\tilde{Y}_j}}{\sigma^2_{\tilde{W}}}$$

and thus $\frac{1}{\tau_j}R^2_{\tilde{W}.\tilde{Y}_j} = \frac{1}{\tau_j}b^2_{\tilde{Y}_j.\tilde{W}}\frac{\sigma^2_{\tilde{W}}}{\sigma^2_{\tilde{Y}_j}} \leq \rho \leq 1$. Last, the set $\mathcal{R}^{\mathbf{c}}_{jh}$ obtains since $0 < \rho$ and $\Gamma_{jh} = G_{jh}(\rho) = \sigma^{-2}_{\tilde{W}}\sigma_{\tilde{Y}_j,\tilde{Y}_h} - b_{\tilde{Y}_j.\tilde{W}}\frac{1}{\rho}b_{\tilde{Y}_h.\tilde{W}}$ so that

$$G_{jh}(\rho) \leq 0 \text{ if and only if } \begin{cases} \frac{b_{\tilde{Y}_j.\tilde{W}}b_{\tilde{Y}_h.\tilde{W}}}{\sigma^{-2}_{\tilde{W}}\sigma_{\tilde{Y}_j.\tilde{Y}_h}} \leq \rho & \text{when } \sigma^{-2}_{\tilde{W}}\sigma_{\tilde{Y}_j.\tilde{Y}_h} < 0 \\ 0 \leq b_{\tilde{Y}_j.\tilde{W}}b_{\tilde{Y}_h.\tilde{W}} & \text{when } \sigma^{-2}_{\tilde{W}}\sigma_{\tilde{Y}_j.\tilde{Y}_h} = 0 \\ \rho \leq \frac{b_{\tilde{Y}_j.\tilde{W}}b_{\tilde{Y}_h.\tilde{W}}}{\sigma^{-2}_{\tilde{W}}\sigma_{\tilde{Y}_j.\tilde{Y}_h}} & \text{when } 0 < \sigma^{-2}_{\tilde{W}}\sigma_{\tilde{Y}_j.\tilde{Y}_h} \end{cases}.$$

Combining the results, we have $\rho \in \mathcal{R}^{k,\tau,\mathbf{c}} = [R^2_{\tilde{W}.\tilde{Y}}, 1] \cap [\frac{1}{1+\kappa}, 1] \cap_{j=1}^{p} [\frac{1}{\tau_j}R^2_{\tilde{W}.\tilde{Y}_j}, 1] \bigcap_{\substack{j,h=1 \\ j<h}}^{p} \mathcal{R}^{\mathbf{c}}_{jh}$.

To show that $\mathcal{R}^{k,\tau,\mathbf{c}}$ is sharp, it suffices to show that every $r \in \mathcal{R}^{k,\tau,\mathbf{c}}$ corresponds to a point $(r, d, b, g) \in \mathcal{J}^{k,\tau,\mathbf{c}}$. First, we show that $0 \preceq G(r)$. If $R^2_{\tilde{W}.\tilde{Y}} = 0$ then $G(r) = \sigma^{-2}_{\tilde{W}}\sigma^2_{\tilde{Y}} \succeq 0$. Otherwise, note that

$$G(1) = \sigma^{-2}_{\tilde{W}}\sigma^2_{\tilde{Y}} - b'_{\tilde{Y}.\tilde{W}}b_{\tilde{Y}.\tilde{W}} = \sigma^{-2}_{\tilde{W}}[\sigma^2_{\tilde{Y}} - \sigma_{\tilde{Y},\tilde{W}}\sigma^{-2}_{\tilde{W}}\sigma_{\tilde{W},\tilde{Y}}] = \sigma^{-2}_{\tilde{W}}E(\epsilon_{\tilde{Y}.\tilde{W}}\epsilon'_{\tilde{Y}.\tilde{W}}) \succeq 0.$$

Further, when $R^2_{\tilde{W}.\tilde{Y}} \neq 0$, $0 \preceq G(R^2_{\tilde{W}.\tilde{Y}})$. Specifically, $0 < \sigma^4_{\tilde{W}}R^2_{\tilde{W}.\tilde{Y}}$ and

$$\sigma^4_{\tilde{W}}R^2_{\tilde{W}.\tilde{Y}}G(R^2_{\tilde{W}.\tilde{Y}}) = (R^2_{\tilde{W}.\tilde{Y}}\sigma^2_{\tilde{W}})\sigma^2_{\tilde{Y}} - \sigma_{\tilde{Y},\tilde{W}}\sigma_{\tilde{W},\tilde{Y}} = Var(b'_{\tilde{W}.\tilde{Y}}\tilde{Y})\sigma^2_{\tilde{Y}} - \sigma_{\tilde{Y},\tilde{W}}\sigma_{\tilde{W},\tilde{Y}} \succeq 0$$

since and for any vector $\underset{p\times 1}{x}$, applying the Cauchy–Schwarz inequality gives

$$\begin{aligned} & x'Var(b'_{\tilde{W}.\tilde{Y}}\tilde{Y})\sigma^2_{\tilde{Y}}x - x'\sigma_{\tilde{Y},\tilde{W}}\sigma_{\tilde{W},\tilde{Y}}x \\ &= Var(b'_{\tilde{W}.\tilde{Y}}\tilde{Y})Var(x'\tilde{Y}) - [Cov(x'\tilde{Y}, \tilde{W})]^2 \\ &= Var(b'_{\tilde{W}.\tilde{Y}}\tilde{Y})Var(x'\tilde{Y}) - [Cov(x'\tilde{Y}, b'_{\tilde{W}.\tilde{Y}}\tilde{Y})]^2 \geq 0 \end{aligned}$$

where we make use of $\tilde{W}' = \tilde{Y}'b_{\tilde{W}.\tilde{Y}} + \epsilon'_{\tilde{W}.\tilde{Y}}$ and $Cov(\tilde{Y}, \epsilon_{\tilde{W}.\tilde{Y}}) = 0$ in the last equality. Then for any $r \in \mathcal{R}^{k,\tau,\mathbf{c}} \subseteq [R^2_{\tilde{W}.\tilde{Y}}, 1]$ there exists $0 \leq \lambda \leq 1$ such that $\frac{1}{r} = \lambda + (1-\lambda)\frac{1}{R^2_{\tilde{W}.\tilde{Y}}}$ and it follows that

$$0 \preceq G(r) = \lambda G(1) + (1-\lambda)G(R^2_{\tilde{W}.\tilde{Y}}).$$

Clearly, $\frac{1}{1+\kappa} \leq r \leq 1$. Further, for $j = 1, ..., p$, if $\sigma^2_{\tilde{Y}_j} = 0$ then we set $\frac{1}{\tau_j} R^2_{\tilde{W}.\tilde{Y}_j} = 0 \leq r$ whereas if $\sigma^2_{\tilde{Y}_j} \neq 0$ then $\frac{1}{\tau_j} b^2_{\tilde{Y}_j.\tilde{W}} \frac{\sigma^2_{\tilde{W}}}{\sigma^2_{\tilde{Y}_j}} = \frac{1}{\tau_j} R^2_{\tilde{W}.\tilde{Y}_j} \leq r$ implies that $\frac{\sigma^2_{\tilde{Y}_j}}{\sigma^2_{\tilde{W}}}(1-\tau_j) \leq \frac{\sigma^2_{\tilde{Y}_j}}{\sigma^2_{\tilde{W}}} - b^2_{\tilde{Y}_j.\tilde{W}} \frac{1}{r} = G_{jj}(r)$. Last, from the expression for $G_{jh}(r)$, we have that $\underline{c}_{jh} \leq sgn(G_{jh}(r)) \leq \bar{c}_{jh}$ for every $r \in \mathcal{R}^{k,\tau,\mathbf{c}}$ and $j, h = 1, ..., p$ with $j < h$.

The sharp bounds $\mathcal{D}^{k,\tau}$, $\mathcal{B}^{k,\tau}$, and $\mathcal{G}^{k,\tau}$ for δ, β, and Γ follow from the mappings $D(\cdot)$, $B(\cdot)$, and $G(\cdot)$ in Theorem 3.1.

Proof of Theorem 5.1: First, for random column vectors A and B, we collect the regression intercept and slope estimands as follows

$$A' = [E(A)' - E(B)'b_{A.B}] + B'b_{A.B} + \epsilon'_{A.B} \equiv (1, B')b^*_{A.B} + \epsilon'_{A.B}.$$

Given observations $\{A_i, B_i\}_{i=1}^n$, denote the linear regression intercept $(\hat{b}^0_{A.B})$ and slope $(\hat{b}_{A.B})$ estimators and the sample residual $(\hat{\epsilon}_{A.B,i})$ by:

$$\tilde{b}_{A.B} = (\hat{b}^0_{A.B}, \hat{b}'_{A.B})' \equiv (\frac{1}{n}\sum_{i=1}^n (1, B'_i)'(1, B'_i))^{-1}(\frac{1}{n}\sum_{i=1}^n (1, B'_i)'A'_i) \text{ and } \hat{\epsilon}'_{A.B,i} \equiv A'_i - (1, B'_i)\tilde{b}_{A.B}.$$

Further, we collect into π^* the following estimands

$$\pi^* \equiv [vec(b^*_{Y.(W,X')'})', b^{*\prime}_{W.(Y,X')'}, b^{*\prime}_{W.(Y_1,X')'}, ..., b^{*\prime}_{W.(Y_p,X')'}, vec(b^*_{Y.X})', b^{*\prime}_{W.X}, \sigma^{-2}_{\tilde{W}} vec(\sigma^2_{\tilde{Y}})'],$$

and into $\tilde{\pi}$ the corresponding estimators:

$$\tilde{\pi} \equiv [vec(\tilde{b}_{Y.(W,X')'})', \tilde{b}'_{W.(Y,X')'}, \tilde{b}'_{W.(Y_1,X')'}, ..., \tilde{b}'_{W.(Y_p,X')'}, vec(\tilde{b}_{Y.X})', \tilde{b}'_{W.X}, \hat{\sigma}^{-2}_{\tilde{W}} vec(\hat{\sigma}^2_{\tilde{Y}})'].$$

Last, let $\hat{\mu}^2_A = \frac{1}{n}\sum_{i=1}^n A_i A'_i$,

$$\hat{Q} \equiv diag\{\underset{p\times p}{I} \otimes \hat{\mu}^2_{(1,W,X')'}, \hat{\mu}^2_{(1,Y,X')'}, \hat{\mu}^2_{(1,Y_1,X')'}, ..., \hat{\mu}^2_{(1,Y_p,X')'}, \underset{p\times p}{I} \otimes \hat{\mu}^2_{(1,X')'}, \hat{\mu}^2_{(1,X')'}, \underset{\frac{1}{2}p(p+1)\times\frac{1}{2}p(p+1)}{I} \otimes \hat{\sigma}^2_{\tilde{W}}\}.$$

and

$$L \equiv \frac{1}{n}\sum_{i=1}^n [vec((1, W_i, X'_i)'\epsilon_{Y.(W,X')',i})', (1, Y_i, X'_i)\epsilon_{W.(Y,X')',i},$$
$$(1, Y_{1i}, X'_i)\epsilon_{W.(Y_1,X')',i}, ..., (1, Y_{pi}, X'_i)\epsilon_{W.(Y_p,X')',i}, vec((1, X'_i)'\epsilon_{Y.X,i})', (1, X'_i)'\epsilon_{W.X,i}, vec(\epsilon_{Y.X,i}\epsilon'_{Y.X,i} - \sigma^2_{\tilde{Y}})']'.$$

Recall that Q is finite (by $A_1(i)$) and nonsingular. For a symmetric matrix C and a vector D, let C_1 denote the submatrix that removes the last $\frac{1}{2}p(p+1)$ row and column of C and let D_1 be the subvector that removes the last $\frac{1}{2}p(p+1)$ row of D. Then

$$\sqrt{n}(\tilde{\pi}_1 - \pi^*_1) = \hat{Q}_1^{-1}\sqrt{n}L_1 = (\hat{Q}_1^{-1} - Q_1^{-1})\sqrt{n}L_1 + Q_1^{-1}\sqrt{n}L_1.$$

Since (i) gives $\hat{Q}_1^{-1} - Q_1^{-1} = o_p(1)$ and (ii) gives $\sqrt{n}L_1 \xrightarrow{d} N(0, \Xi_1)$, we obtain that $\sqrt{n}(\tilde{\pi}_1 - \pi_1^*) = Q_1^{-1}\sqrt{n}L_1 + o_p(1) \xrightarrow{d} N(0, \Sigma_1^*)$. Moreover, it follows from $\hat{\mu}^2_{(1,X')'} \xrightarrow{p} \mu^2_{(1,X')'}$, $\sqrt{n}(\tilde{b}_{Y_j.X} - b^*_{Y_j.X}) = O_p(1)$, and $\frac{1}{n}\sum_{i=1}^{n}\epsilon_{Y_j.X,i}(1, X_i')' = E[\epsilon_{Y_j.X}(1, X')'] + o_p(1) = o_p(1)$ for $j = 1, ..., p$ that for any $j, h = 1, ..., p$

$$
\begin{aligned}
& n^{-\frac{1}{2}}\sum_{i=1}^{n}\hat{\epsilon}_{Y_j.X,i}\hat{\epsilon}_{Y_h.X,i} \\
&= n^{-\frac{1}{2}}\sum_{i=1}^{n}(\epsilon_{Y_j.X,i} - (1, X_i')(\tilde{b}_{Y_j.X} - b^*_{Y_j.X}))(\epsilon_{Y_h.X,i} - (1, X_i')(\tilde{b}_{Y_h.X} - b^*_{Y_h.X})) \\
&= n^{-\frac{1}{2}}\sum_{i=1}^{n}\epsilon_{Y_j.X,i}\epsilon_{Y_h.X,i} - [\frac{1}{n}\sum_{i=1}^{n}\epsilon_{Y_j.X,i}(1, X_i')]\sqrt{n}(\tilde{b}_{Y_h.X} - b^*_{Y_h.X}) \\
&- [\frac{1}{n}\sum_{i=1}^{n}\epsilon_{Y_h.X,i}(1, X_i')]\sqrt{n}(\tilde{b}_{Y_j.X} - b^*_{Y_j.X}) + (\tilde{b}_{Y_h.X} - b^*_{Y_h.X})'\hat{\mu}^2_{(1,X')'}\sqrt{n}(\tilde{b}_{Y_j.X} - b^*_{Y_j.X}) \\
&= n^{-\frac{1}{2}}\sum_{i=1}^{n}\epsilon_{Y_j.X,i}\epsilon_{Y_h.X,i} + o_p(1).
\end{aligned}
$$

Similarly, by (i), we have that

$$
\frac{1}{n}\sum_{i=1}^{n}\hat{\epsilon}_{Y_j.X,i}\hat{\epsilon}_{Y_h.X,i} = E(\epsilon_{Y_j.X}\epsilon_{Y_h.X}) + o_p(1) = \sigma_{\tilde{Y}_j,\tilde{Y}_h} + o_p(1) \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n}\hat{\epsilon}^2_{W.X,i} = \sigma^2_{\tilde{W}} + o_p(1).
$$

Thus, since $n^{-1/2}\sum_{i=1}^{n}\epsilon_{Y_j.X,i}\epsilon_{Y_h.X,i}$ is $O_p(1)$ by (ii), we have that for $j, h = 1, ..., p$

$$
\sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^{n}\hat{\epsilon}_{Y_j.X,i}\hat{\epsilon}_{Y_h.X,i}}{\frac{1}{n}\sum_{i=1}^{n}\hat{\epsilon}^2_{W.X,i}} = (\sigma^2_{\tilde{W}})^{-1}n^{-\frac{1}{2}}\sum_{i=1}^{n}\epsilon_{Y_j.X,i}\epsilon_{Y_h.X,i} + o_p(1).
$$

Together with $\sqrt{n}(\tilde{\pi}_1 - \pi_1^*) = Q_1^{-1}\sqrt{n}L_1 + o_p(1)$, we obtain by (i) and (ii) that

$$
\sqrt{n}(\tilde{\pi} - \pi^*) = Q^{-1}\sqrt{n}L + o_p(1) \xrightarrow{d} N(0, \Sigma^*)
$$

and therefore that the subvector $\sqrt{n}(\hat{\pi} - \pi) \xrightarrow{d} N(0, \Sigma)$.

Proof of Corollary A.1: The identification set $\mathcal{J}^{k,\tau,\mathbf{c}}$ obtains from A_1-A_6' and the $(Var[(\tilde{Y}, \tilde{W})'])$ the moments given by (in)equalities (3,5,6,7), using the expressions in Theorem 3.1. The sharpness proof in Corollary 3.2 implies that $\mathcal{J}^{k,\tau,\mathbf{c}}$ is sharp. Specifically, since $G(r) = \sigma^{-2}_{\tilde{W}}\sigma^2_{\tilde{\eta}^*}$, we have that $\underline{c}_{jh} \leq r_{\tilde{\eta}_j^*,\tilde{\eta}_h^*} \leq \bar{c}_{jh}$.

To derive $\mathcal{R}^{k,\tau,\mathbf{c}}$, for $j, h = 1, ..., p$ and $j < h$, consider the restriction

$$
\underline{c}_{jh} \leq \Gamma_{jh} = \frac{G_{jh}(\rho)}{[G_{jj}(\rho)G_{hh}(\rho)]^{\frac{1}{2}}} = \frac{\sigma^{-2}_{\tilde{W}}\sigma_{\tilde{Y}_j,\tilde{Y}_h} - b_{\tilde{Y}_j.\tilde{W}}\frac{1}{\rho}b_{\tilde{Y}_h.\tilde{W}}}{(\sigma^{-2}_{\tilde{W}}\sigma^2_{\tilde{Y}_j} - \frac{1}{\rho}b^2_{\tilde{Y}_j.\tilde{W}})^{\frac{1}{2}}(\sigma^{-2}_{\tilde{W}}\sigma^2_{\tilde{Y}_h} - \frac{1}{\rho}b^2_{\tilde{Y}_h.\tilde{W}})^{\frac{1}{2}}} \leq \bar{c}_{jh}.
$$

If $\sigma^2_{\tilde{Y}_j} = 0$ or $\sigma^2_{\tilde{Y}_h} = 0$ then $\sigma^2_{\eta_j} = 0$ or $\sigma^2_{\eta_h} = 0$ and $\underline{c}_{jh} \leq \sigma_{\eta_j,\eta_h} \leq \bar{c}_{jh}$ is either incorrect (if $0 \notin [\underline{c}_{jh}, \bar{c}_{jh}]$) or uninformative about ρ (if $0 \in [\underline{c}_{jh}, \bar{c}_{jh}]$). Suppose that $\sigma^2_{\tilde{Y}_j} \neq 0$ and $\sigma^2_{\tilde{Y}_h} \neq 0$. Multiplying the numerator and denominator by $0 < \rho\sigma^2_{\tilde{W}}\sigma^{-1}_{\tilde{Y}_j}\sigma^{-1}_{\tilde{Y}_h}$ gives

$$\underline{c}_{jh} \leq \frac{\rho r_{\tilde{Y}_j,\tilde{Y}_h} - r_{\tilde{W},\tilde{Y}_j} r_{\tilde{W},\tilde{Y}_h}}{(\rho - R^2_{\tilde{W}.\tilde{Y}_j})^{\frac{1}{2}}(\rho - R^2_{\tilde{W}.\tilde{Y}_h})^{\frac{1}{2}}} \leq \bar{c}_{jh}.$$

The expression for $\mathcal{R}^{\mathbf{c}}_{jh}$ then follows from encoding the sign of r_{η_j,η_h} via the function

$$S_{jh}(r) \equiv r \times r_{\tilde{Y}_j,\tilde{Y}_h} - r_{\tilde{W},\tilde{Y}_j} r_{\tilde{W},\tilde{Y}_h}$$

and the magnitude of r_{η_j,η_h} ($r^2_{\eta_j,\eta_h} \leq c^2$ or $c^2 \leq r^2_{\eta_j,\eta_h}$) via the quadratic function

$$M_{jh}(r;c) \equiv (r \times r_{\tilde{Y}_j,\tilde{Y}_h} - r_{\tilde{W},\tilde{Y}_j} r_{\tilde{W},\tilde{Y}_h})^2 - c^2(r - R^2_{\tilde{W}.\tilde{Y}_j})(r - R^2_{\tilde{W}.\tilde{Y}_h}).$$

By Corollary 3.2, we obtain that $\rho \in \mathcal{R}^{k,\tau,\mathbf{c}} = [R^2_{\tilde{W}.\tilde{Y}}, 1] \cap [\frac{1}{1+\kappa}, 1] \cap_{j=1}^{p} [\frac{1}{\tau_j} R^2_{\tilde{W}.\tilde{Y}_j}, 1] \bigcap_{\substack{j,h=1 \\ j<h}}^{p} \mathcal{R}^{\mathbf{c}}_{jh}$. In addition, $\mathcal{R}^{k,\tau,\mathbf{c}}$ is sharp since every $r \in \mathcal{R}^{k,\tau,\mathbf{c}}$ corresponds to a point $(r,d,b,g) \in \mathcal{J}^{k,\tau,\mathbf{c}}$. Specifically, if $r \in \mathcal{R}^{k,\tau,\mathbf{c}}$ then $\frac{1}{1+\kappa} \leq r \leq 1$, $0 \preceq G(r)$, and $R^2_{\tilde{Y}_j.\tilde{U}} \leq \tau_j$ for $j = 1, ..., p$ by Corollary 3.2. Further, from the sign and magnitude restrictions in $S_{jh}(r)$ and $M_{jh}(r;c)$, we have that $\underline{c}_{jh} \leq \frac{G_{jh}(r)}{[G_{jj}(r)G_{hh}(r)]^{\frac{1}{2}}} \leq \bar{c}_{jh}$ for every $r \in \mathcal{R}^{k,\tau,\mathbf{c}} \subseteq \mathcal{R}^{\mathbf{c}}_{jh}$ and $j,h = 1,...,p$ with $j < h$.

Next, we examine the behavior of $S_{jh}(r)$ and $M_{jh}(r;c)$ when $\sigma^2_{\tilde{Y}_j}\sigma^2_{\tilde{Y}_h} \neq 0$. First, we have that

$$0 \leq S_{jh}(r) \Leftrightarrow \begin{cases} \frac{r_{\tilde{W},\tilde{Y}_j} r_{\tilde{W},\tilde{Y}_h}}{r_{\tilde{Y}_j,\tilde{Y}_h}} \leq r & \text{when } 0 < r_{\tilde{Y}_j,\tilde{Y}_h} \\ r_{\tilde{W},\tilde{Y}_j} r_{\tilde{W},\tilde{Y}_h} \leq 0 & \text{when } r_{\tilde{Y}_j,\tilde{Y}_h} = 0 \\ r \leq \frac{r_{\tilde{W},\tilde{Y}_j} r_{\tilde{W},\tilde{Y}_h}}{r_{\tilde{Y}_j,\tilde{Y}_h}} & \text{when } r_{\tilde{Y}_j,\tilde{Y}_h} < 0 \end{cases}.$$

Further, if $R^2_{\tilde{Y}_j.\tilde{Y}_h} = 1$ then

$$M_{jh}(r;c) = (1-c^2)(r - R^2_{\tilde{W}.\tilde{Y}_j})^2 \geq 0.$$

Suppose instead that $R^2_{\tilde{Y}_j.\tilde{Y}_h} \neq 1$. We obtain

$$\begin{aligned} M_{jh}(r;c) &= r^2 R^2_{\tilde{Y}_j.\tilde{Y}_h} + R^2_{\tilde{W}.\tilde{Y}_j} R^2_{\tilde{W}.\tilde{Y}_h} - 2r \times r_{\tilde{Y}_j,\tilde{Y}_h} r_{\tilde{W},\tilde{Y}_j} r_{\tilde{W},\tilde{Y}_h} \\ &\quad - c^2 r^2 + c^2 r(R^2_{\tilde{W}.\tilde{Y}_j} + R^2_{\tilde{W}.\tilde{Y}_h}) - c^2 R^2_{\tilde{W}.\tilde{Y}_j} R^2_{\tilde{W}.\tilde{Y}_h} \\ &= r^2(R^2_{\tilde{Y}_j.\tilde{Y}_h} - c^2) + r[-2r_{\tilde{Y}_j,\tilde{Y}_h} r_{\tilde{W},\tilde{Y}_j} r_{\tilde{W},\tilde{Y}_h} + c^2(R^2_{\tilde{W}.\tilde{Y}_j} + R^2_{\tilde{W}.\tilde{Y}_h})] + (1-c^2) R^2_{\tilde{W}.\tilde{Y}_j} R^2_{\tilde{W}.\tilde{Y}_h} \\ &= r^2(R^2_{\tilde{Y}_j.\tilde{Y}_h} - c^2) + r[R^2_{\tilde{W}.(\tilde{Y}_j,\tilde{Y}_h)'}(1 - R^2_{\tilde{Y}_j.\tilde{Y}_h}) - (1-c^2)(R^2_{\tilde{W}.\tilde{Y}_j} + R^2_{\tilde{W}.\tilde{Y}_h})] \\ &\quad + (1-c^2) R^2_{\tilde{W}.\tilde{Y}_j} R^2_{\tilde{W}.\tilde{Y}_h}, \end{aligned}$$

where the last equality makes use of

$$R^2_{\tilde{W}.(\tilde{Y}_j,\tilde{Y}_h)'} = \begin{bmatrix} r_{\tilde{W},\tilde{Y}_j} & r_{\tilde{W},\tilde{Y}_h} \end{bmatrix} \begin{bmatrix} 1 & r_{\tilde{Y}_j,\tilde{Y}_h} \\ r_{\tilde{Y}_h,\tilde{Y}_j} & 1 \end{bmatrix}^{-1} \begin{bmatrix} r_{\tilde{W},\tilde{Y}_j} \\ r_{\tilde{W},\tilde{Y}_h} \end{bmatrix} = \frac{R^2_{\tilde{W},\tilde{Y}_j} + R^2_{\tilde{W},\tilde{Y}_h} - 2r_{\tilde{W},\tilde{Y}_j} r_{\tilde{Y}_j,\tilde{Y}_h} r_{\tilde{W},\tilde{Y}_h}}{1 - R^2_{\tilde{Y}_j.\tilde{Y}_h}}.$$

If $c^2 = R^2_{\tilde{Y}_j.\tilde{Y}_h}$ then $M_{jh}(\cdot;c)$ is a linear function

$$\begin{aligned} M_{jh}(r; r_{\tilde{Y}_j,\tilde{Y}_h}) &= r[R^2_{\tilde{W}.(\tilde{Y}_j,\tilde{Y}_h)'}(1 - R^2_{\tilde{Y}_j.\tilde{Y}_h}) - (1 - R^2_{\tilde{Y}_j.\tilde{Y}_h})(R^2_{\tilde{W}.\tilde{Y}_j} + R^2_{\tilde{W}.\tilde{Y}_h})] \\ &+ (1 - R^2_{\tilde{Y}_j.\tilde{Y}_h}) R^2_{\tilde{W}.\tilde{Y}_j} R^2_{\tilde{W}.\tilde{Y}_h} \\ &= (1 - R^2_{\tilde{Y}_j.\tilde{Y}_h})\{r[R^2_{\tilde{W}.(\tilde{Y}_j,\tilde{Y}_h)'} - (R^2_{\tilde{W}.\tilde{Y}_j} + R^2_{\tilde{W}.\tilde{Y}_h})] + R^2_{\tilde{W}.\tilde{Y}_j} R^2_{\tilde{W}.\tilde{Y}_h}\} \end{aligned}$$

and

$$0 \le M_{jh}(r;c) \Leftrightarrow$$
$$\begin{cases} r \le \frac{-R^2_{\tilde{W}.\tilde{Y}_j} R^2_{\tilde{W}.\tilde{Y}_h}}{R^2_{\tilde{W}.(\tilde{Y}_j,\tilde{Y}_h)'} - (R^2_{\tilde{W}.\tilde{Y}_j} + R^2_{\tilde{W}.\tilde{Y}_h})} & \text{when } c^2 = R^2_{\tilde{Y}_j.\tilde{Y}_h} \text{ and } R^2_{\tilde{W}.(\tilde{Y}_j,\tilde{Y}_h)'} < R^2_{\tilde{W}.\tilde{Y}_j} + R^2_{\tilde{W}.\tilde{Y}_h} \\ 0 \le (1 - R^2_{\tilde{Y}_j.\tilde{Y}_h}) R^2_{\tilde{W}.\tilde{Y}_j} R^2_{\tilde{W}.\tilde{Y}_h} & \text{when } c^2 = R^2_{\tilde{Y}_j.\tilde{Y}_h} \text{ and } R^2_{\tilde{W}.(\tilde{Y}_j,\tilde{Y}_h)'} = R^2_{\tilde{W}.\tilde{Y}_j} + R^2_{\tilde{W}.\tilde{Y}_h} \\ \frac{-R^2_{\tilde{W}.\tilde{Y}_j} R^2_{\tilde{W}.\tilde{Y}_h}}{R^2_{\tilde{W}.(\tilde{Y}_j,\tilde{Y}_h)'} - (R^2_{\tilde{W}.\tilde{Y}_j} + R^2_{\tilde{W}.\tilde{Y}_h})} \le r & \text{when } c^2 = R^2_{\tilde{Y}_j.\tilde{Y}_h} \text{ and } R^2_{\tilde{W}.\tilde{Y}_j} + R^2_{\tilde{W}.\tilde{Y}_h} < R^2_{\tilde{W}.(\tilde{Y}_j,\tilde{Y}_h)'} \end{cases}$$

Otherwise, if $c^2 \ne R^2_{\tilde{Y}_j.\tilde{Y}_h}$, the discriminant of $M_{jh}(\cdot;c)$ is

$$\begin{aligned} \Delta_{jh}(c) &= [R^2_{\tilde{W}.(\tilde{Y}_j,\tilde{Y}_h)'}(1 - R^2_{\tilde{Y}_j.\tilde{Y}_h}) - (1 - c^2)(R^2_{\tilde{W}.\tilde{Y}_j} + R^2_{\tilde{W}.\tilde{Y}_h})]^2 - 4(1 - c^2)(R^2_{\tilde{Y}_j.\tilde{Y}_h} - c^2) R^2_{\tilde{W}.\tilde{Y}_j} R^2_{\tilde{W}.\tilde{Y}_h} \\ &= [R^2_{\tilde{W}.(\tilde{Y}_j,\tilde{Y}_h)'}(1 - R^2_{\tilde{Y}_j.\tilde{Y}_h}) - (1 - c^2)(R^2_{\tilde{W}.\tilde{Y}_j} + R^2_{\tilde{W}.\tilde{Y}_h})]^2 \\ &- (1 - c^2) 4 R^2_{\tilde{Y}_j.\tilde{Y}_h} R^2_{\tilde{W}.\tilde{Y}_j} R^2_{\tilde{W}.\tilde{Y}_h} + 4c^2(1 - c^2) R^2_{\tilde{W}.\tilde{Y}_j} R^2_{\tilde{W}.\tilde{Y}_h} \\ &= [R^2_{\tilde{W}.(\tilde{Y}_j,\tilde{Y}_h)'}(1 - R^2_{\tilde{Y}_j.\tilde{Y}_h}) - (1 - c^2)(R^2_{\tilde{W}.\tilde{Y}_j} + R^2_{\tilde{W}.\tilde{Y}_h})]^2 \\ &- (1 - c^2)[R^2_{\tilde{W}.(\tilde{Y}_j,\tilde{Y}_h)'}(1 - R^2_{\tilde{Y}_j.\tilde{Y}_h}) - (R^2_{\tilde{W}.\tilde{Y}_h} + R^2_{\tilde{W}.\tilde{Y}_j})]^2 + 4c^2(1 - c^2) R^2_{\tilde{W}.\tilde{Y}_j} R^2_{\tilde{W}.\tilde{Y}_h} \\ &= c^2 R^4_{\tilde{W}.(\tilde{Y}_j.\tilde{Y}_h)'}(1 - R^2_{\tilde{Y}_j.\tilde{Y}_h})^2 - c^2(1 - c^2)(R^2_{\tilde{W}.\tilde{Y}_j} + R^2_{\tilde{W}.\tilde{Y}_h})^2 + 4c^2(1 - c^2) R^2_{\tilde{W}.\tilde{Y}_j} R^2_{\tilde{W}.\tilde{Y}_h} \\ &= c^2\{R^4_{\tilde{W}.(\tilde{Y}_j.\tilde{Y}_h)'}(1 - R^2_{\tilde{Y}_j.\tilde{Y}_h})^2 - (1 - c^2)[(R^2_{\tilde{W}.\tilde{Y}_j} + R^2_{\tilde{W}.\tilde{Y}_h})^2 - 4R^2_{\tilde{W}.\tilde{Y}_j} R^2_{\tilde{W}.\tilde{Y}_h}]\} \\ &= c^2[R^4_{\tilde{W}.(\tilde{Y}_j.\tilde{Y}_h)'}(1 - R^2_{\tilde{Y}_j.\tilde{Y}_h})^2 - (1 - c^2)(R^2_{\tilde{W}.\tilde{Y}_j} - R^2_{\tilde{W}.\tilde{Y}_h})^2]. \end{aligned}$$

In particular, $\Delta_{jh}(c) < 0$ if and only if

$$0 < c^2 < 1 - \frac{R^4_{\tilde{W}.(\tilde{Y}_j.\tilde{Y}_h)'}(1 - R^2_{\tilde{Y}_j.\tilde{Y}_h})^2}{(R^2_{\tilde{W}.\tilde{Y}_j} - R^2_{\tilde{W}.\tilde{Y}_h})^2}.$$

Further, we have that $1 - \frac{R^4_{\tilde{W}.(\tilde{Y}_j.\tilde{Y}_h)'}(1-R^2_{\tilde{Y}_j.\tilde{Y}_h})^2}{(R^2_{\tilde{W}.\tilde{Y}_j}-R^2_{\tilde{W}.\tilde{Y}_h})^2} \leq R^2_{\tilde{Y}_j.\tilde{Y}_h}$ since if $c^2 = R^2_{\tilde{Y}_j.\tilde{Y}_h}$ then

$$\begin{aligned}\Delta_{jh}(c) &= [R^2_{\tilde{W}.(\tilde{Y}_j,\tilde{Y}_h)'}(1-R^2_{\tilde{Y}_j.\tilde{Y}_h}) - (1-c^2)(R^2_{\tilde{W}.\tilde{Y}_j} + R^2_{\tilde{W}.\tilde{Y}_h})]^2 - 4(1-c^2)(R^2_{\tilde{Y}_j.\tilde{Y}_h} - c^2)R^2_{\tilde{W}.\tilde{Y}_j}R^2_{\tilde{W}.\tilde{Y}_h} \\ &= (1-R^2_{\tilde{Y}_j.\tilde{Y}_h})^2[R^2_{\tilde{W}.(\tilde{Y}_j,\tilde{Y}_h)'} - (R^2_{\tilde{W}.\tilde{Y}_j} + R^2_{\tilde{W}.\tilde{Y}_h})]^2 \geq 0\end{aligned}$$

and if $R^2_{\tilde{Y}_j.\tilde{Y}_h} = 0$ then

$$1 - \frac{R^4_{\tilde{W}.(\tilde{Y}_j.\tilde{Y}_h)'}(1-R^2_{\tilde{Y}_j.\tilde{Y}_h})^2}{(R^2_{\tilde{W}.\tilde{Y}_j} - R^2_{\tilde{W}.\tilde{Y}_h})^2} = 1 - \frac{(R^2_{\tilde{W},\tilde{Y}_j} + R^2_{\tilde{W},\tilde{Y}_h})^2}{(R^2_{\tilde{W}.\tilde{Y}_j} - R^2_{\tilde{W}.\tilde{Y}_h})^2} \leq 0 = R^2_{\tilde{Y}_j.\tilde{Y}_h}.$$

It follows that if $0 < c^2 < 1 - \frac{R^4_{\tilde{W}.(\tilde{Y}_j.\tilde{Y}_h)'}(1-R^2_{\tilde{Y}_j.\tilde{Y}_h})^2}{(R^2_{\tilde{W}.\tilde{Y}_j}-R^2_{\tilde{W}.\tilde{Y}_h})^2}$ then $c^2 < R^2_{\tilde{Y}_j.\tilde{Y}_h}$ and

$$0 \leq M_{jh}(r;c) \Leftrightarrow -\infty < r < \infty.$$

If $c^2 \neq R^2_{\tilde{Y}_j.\tilde{Y}_h}$ and $0 \leq \Delta_{jh}(c)$ then define

$$F_{jh}(c) \equiv -R^2_{\tilde{W}.(\tilde{Y}_j,\tilde{Y}_h)'}(1-R^2_{\tilde{Y}_j.\tilde{Y}_h}) + (1-c^2)(R^2_{\tilde{W}.\tilde{Y}_j} + R^2_{\tilde{W}.\tilde{Y}_h}),$$

so that $M_{jh}(\rho;c)$ has the two roots

$$\rho^-_{jh}(c) \equiv \frac{F_{jh}(c) - \Delta_{jh}(c)^{\frac{1}{2}}}{2(R^2_{\tilde{Y}_j.\tilde{Y}_h} - c^2)} \quad \text{and} \quad \rho^+_{jh}(c) \equiv \frac{F_{jh}(c) + \Delta_{jh}(c)^{\frac{1}{2}}}{2(R^2_{\tilde{Y}_j.\tilde{Y}_h} - c^2)}.$$

We then have that

$$0 \leq M_{jh}(r;c) \Leftrightarrow \begin{cases} r \in (-\infty, \rho^-_{jh}(c)] \cup [\rho^+_{jh}(c), \infty) & \text{when } c^2 < R^2_{\tilde{Y}_j.\tilde{Y}_h} \\ r \in [\rho^+_{jh}(c), \rho^-_{jh}(c)] & \text{when } R^2_{\tilde{Y}_j.\tilde{Y}_h} < c^2 \end{cases}.$$

Combining these results, yields the equivalence between $0 \leq M_{jh}(r;c)$ and the range of r.

The sharp bounds $\mathcal{D}^{k,\tau,\mathbf{c}}$, $\mathcal{B}^{k,\tau,\mathbf{c}}$, and $\mathcal{G}^{k,\tau,\mathbf{c}}$ follow from the mappings $D(\cdot)$, $B(\cdot)$, and $G(\cdot)$ in Theorem 3.1.

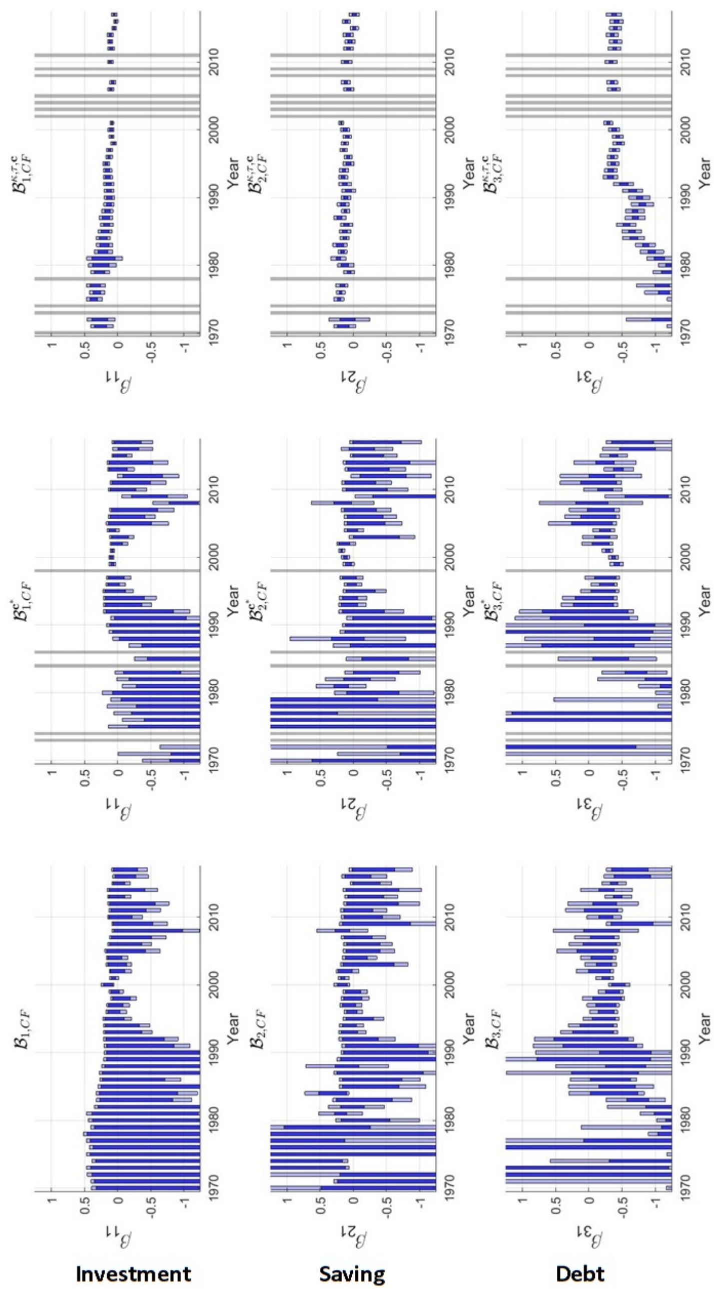

Figure 4: 50% (dark) and 95% (light) confidence regions for β_{j1} (cash flow) for $j = 1, 2, 3$ (investment, saving, and debt) from year 1970 to 2017, when X includes asset tangibility. We consider the regions $\mathcal{B}_{j1}$, $\mathcal{B}_{j1}^{\mathbf{c}^*}$, and $\mathcal{B}_{j1}^{\kappa,\tau,\mathbf{c}}$ where $\mathbf{c}^* = 0$, κ and τ are such that $\hat{\kappa}^* = 0.5$ and $\hat{\tau}^* = (0.9, 0.9, 0.9)'$, and $\mathbf{c}$ is such that $(\underline{c}_{12}, \overline{c}_{12}) = (\underline{c}_{23}, \overline{c}_{23}) = (-1, 0)$ and $(\underline{c}_{13}, \overline{c}_{13}) = (0, 1)$. The shaded vertical bars indicate years in which the maintained assumptions are rejected.

Table 7: Bounds on the Cash Flow Coefficients in the Investment, Saving, and Debt Equations Using the Full Panel and Accounting for Asset Tangibility

	$\mathcal{S}_j^{\kappa,\tau}$	$\mathcal{J}^{\kappa,\tau}$	$\mathcal{J}^{\kappa,\tau,\mathbf{c}}$	$\mathcal{J}^{\kappa,\tau,\mathbf{c}^*}$	$b_{Y.(W,X')'}$
	Results without fixed effects for $\kappa = \infty$ and $\tau = (1,1,1)'$				
β_{11}	[-0.134 , 0.144]	[-0.028 , 0.144]	[0.130 , 0.144]	-	0.143
	(-0.142 , 0.145)	(-0.033 , 0.145)	(0.129 , 0.145)	-	(0.136 , 0.150)
β_{21}	[-0.438 , 0.122]	[-0.014 , 0.122]	[0.111 , 0.122]	-	0.121
	(-0.456 , 0.124)	(-0.019 , 0.124)	(0.109 , 0.124)	-	(0.114 , 0.129)
β_{31}	[-0.419 , 1.242]	[-0.419 , -0.248]	[-0.419 , -0.405]	-	-0.418
	(-0.423 , 1.294)	(-0.423 , -0.242)	(-0.423 , -0.402)	-	(-0.436 , -0.400)
	Results without fixed effects for $\kappa^* = 0.5$ and $\tau^* = (0.9, 0.9, 0.9)'$				
β_{11}	[0.127 , 0.144]	[0.127 , 0.144]	[0.130 , 0.144]	-	0.143
	(0.125 , 0.145)	(0.125 , 0.145)	(0.129 , 0.145)	-	(0.136 , 0.150)
β_{21}	[0.108 , 0.122]	[0.108 , 0.122]	[0.111 , 0.122]	-	0.121
	(0.106 , 0.124)	(0.106 , 0.124)	(0.109 , 0.124)	-	(0.114 , 0.129)
β_{31}	[-0.419 , -0.401]	[-0.419 , -0.401]	[-0.419 , -0.405]	-	-0.418
	(-0.423 , -0.398)	(-0.423 , -0.398)	(-0.423 , -0.402)	-	(-0.436 , -0.400)
	Results with year and firm fixed effects for $\kappa = \infty$ and $\tau = (1,1,1)'$				
β_{11}	[-0.629 , 0.130]	[-0.480 , 0.130]	[-0.481 , 0.130]	-	0.129
	(-0.634 , 0.131)	(-0.484 , 0.131)	(-0.485 , 0.131)	-	(0.122 , 0.137)
β_{21}	[-2.267 , 0.170]	[-0.273 , 0.170]	[-0.274 , 0.170]	-	0.170
	(-2.300 , 0.172)	(-0.279 , 0.172)	(-0.280 , 0.172)	-	(0.159 , 0.181)
β_{31}	[-0.369 , 32.655]	[-0.369 , -0.315]	[-0.369 , -0.315]	-	-0.368
	(-0.370 , 35.443)	(-0.370 , -0.31)	(-0.370 , -0.309)	-	(-0.383 , -0.353)
	Results with year and firm fixed effects for $\kappa^* = 0.5$ and $\tau^* = (0.9, 0.9, 0.9)'$				
β_{11}	[0.083 , 0.130]	[0.083 , 0.130]	[0.083 , 0.130]	-	0.129
	(0.083 , 0.131)	(0.083 , 0.131)	(0.083 , 0.131)	-	(0.122 , 0.137)
β_{21}	[0.136 , 0.170]	[0.136 , 0.170]	[0.136 , 0.170]	-	0.170
	(0.135 , 0.172)	(0.135 , 0.172)	(0.135 , 0.172)	-	(0.159 , 0.181)
β_{31}	[-0.369 , -0.364]	[-0.369 , -0.364]	[-0.369 , -0.364]	-	-0.368
	(-0.370 , -0.362)	(-0.370 , -0.362)	(-0.370 , -0.362)	-	(-0.383 , -0.353)

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 $X = [\text{Cash Flow}, \text{Firm Size}, \text{Asset Tangibility}]$. 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. $\mathbf{c}$ sets $(\underline{c}_{12}, \bar{c}_{12}) = (\underline{c}_{23}, \bar{c}_{23}) = (-1, 0)$ and $(\underline{c}_{13}, \bar{c}_{13}) = (0, 1)$ whereas $\mathbf{c}^* = 0$. Robust standard errors for π are clustered by firm. 50% and 95% confidence regions are in brackets and parentheses respectively.