2019

Essays In Corporate Finance

Daniel Kim

University of Pennsylvania, danielkim116@gmail.com

Follow this and additional works at: https://repository.upenn.edu/edissertations

Part of the Finance and Financial Management Commons

Recommended Citation

https://repository.upenn.edu/edissertations/3281

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/edissertations/3281

For more information, please contact repository@pobox.upenn.edu.
Essays In Corporate Finance

Abstract
The first chapter "Shareholder Recovery and Leverage" estimates shareholder recovery rate and examines its implications on firm outcomes, including leverage. A positive recovery rate makes shareholders more willing to default, which increases borrowing costs. In response, firms lower leverage ex-ante. This channel helps to match distributions of both leverage and default probabilities. Structural estimation reveals a dramatic change over time in the U.S. bankruptcy system: shareholder recovery rate increased from roughly zero to 29% around the Bankruptcy Reform Act of 1978, and has gradually decreased back to zero. Finally, I show that a positive shareholder recovery rate has a quantitatively large effect on leverage, default probabilities, firm value, and government tax revenue.

In the second chapter "Measurement Error in Multiple Equations", co-authored with Karim Chalak, we econometrically characterize the identification regions for the coefficients in a system of linear equations under the classical measurement error assumptions. We demonstrate the identification gain that results from jointly considering the equations, as opposed to separately. We apply this framework to COMPUSTAT data and analyze the effects of cash flow on the investment, saving, and debt of firms when Tobin's q is used as an error-laden proxy for marginal q. Using our framework, we document a considerable identification gain that results from analyzing the investment, saving, and debt equations jointly. Further, the measurement error in Tobin's q can reconcile the discrepancy with the theories if, and only if, Tobin's q is a noisy proxy for marginal q. If a researcher maintains that Tobin's q is a moderately accurate proxy for marginal q, then we show that a more elaborate theory or specification must be considered.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Finance

First Advisor
Amir Yaron

Second Advisor
Lucian A. Taylor

Keywords
bankruptcy, creditor rights, leverage, measurement error, Tobin's q

Subject Categories
Finance and Financial Management

This dissertation is available at ScholarlyCommons: https://repository.upenn.edu/edissertations/3281
ESSAYS IN CORPORATE FINANCE

Daniel Kim

A DISSERTATION

in

Finance

For the Graduate Group in Managerial Science and Applied Economics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2019

Co-supervisor of Dissertation

Amir Yaron
Robert Morris Professor of Banking & Finance

Co-supervisor of Dissertation

Lucian A. Taylor
Associate Professor of Finance

Graduate Group Chairperson

Catherine Schrand
Celia Z. Moh Professor, Professor of Accounting

Dissertation Committee

Joao F. Gomes, Howard Butcher III Professor of Finance

Christian Opp, Assistant Professor of Finance

Nikolai Roussanov, Moise Y. Safra Associate Professor of Finance
ACKNOWLEDGEMENT

I am deeply indebted to my dissertation committee: Joao Gomes, Christian Opp, Nikolai Roussanov, Luke Taylor, and Amir Yaron for their insightful comments, guidance, and support. Moreover, I thank my co-author, Karim Chalak, for his comments, guidance, and support.

I am also grateful to my friends at Wharton for their help, support, and countless laughs that have made the years at Wharton much more enjoyable: Xingyi Chen, Anna Cororaton, Adrian Corum, Roberto Gomez Cram, Mehran Ebrahimian, Amora Elsaiy, Xiang Fang, Marco Grotteria, Pengfei Han, Nina Karnaukh, Sam Rosen, Lin Shen, Ruslan Sverchkov, Ram Yamarthy, and Junyuan Zou.

For always patiently and genuinely supporting me no matter what, I am deeply grateful to my wife, Seul-A, and my son, Jio. I will love you guys forever.

Finally, I am extremely lucky to my parents: Daeho Kim and Young Hee Kim. Thank you for your faith in me and unwavering support throughout my life.
ABSTRACT

ESSAYS IN CORPORATE FINANCE

Daniel Kim
Amir Yaron
Lucian A. Taylor

The first chapter “Shareholder Recovery and Leverage” estimates shareholder recovery rate and examines its implications on firm outcomes, including leverage. A positive recovery rate makes shareholders more willing to default, which increases borrowing costs. In response, firms lower leverage ex-ante. This channel helps to match distributions of both leverage and default probabilities. Structural estimation reveals a dramatic change over time in the U.S. bankruptcy system: shareholder recovery rate increased from roughly zero to 29% around the Bankruptcy Reform Act of 1978, and has gradually decreased back to zero. Finally, I show that a positive shareholder recovery rate has a quantitatively large effect on leverage, default probabilities, firm value, and government tax revenue.

In the second chapter “Measurement Error in Multiple Equations”, co-authored with Karim Chalak, we econometrically characterize the identification regions for the coefficients in a system of linear equations under the classical measurement error assumptions. We demonstrate the identification gain that results from jointly considering the equations, as opposed to separately. We apply this framework to COMPUSTAT data and analyze the effects of cash flow on the investment, saving, and debt of firms when Tobin’s q is used as an error-laden proxy for marginal q. Using our framework, we document a considerable identification gain that results from analyzing the investment, saving, and debt equations jointly. Further, the measurement error in Tobins q can reconcile the discrepancy with the theories if, and only if, Tobin’s q is a noisy proxy for marginal q. If a researcher maintains that Tobins q is a moderately accurate proxy for marginal q, then we show that a more elaborate theory or specification must be considered.
# TABLE OF CONTENTS

ACKNOWLEDGEMENT ........................................ iii

ABSTRACT .................................................. iv

LIST OF TABLES ........................................ vi

LIST OF ILLUSTRATIONS ................................ vii

CHAPTER 1: Shareholder Recovery and Leverage .................. 1
  1.1 Introduction ......................................... 1
  1.2 Bankruptcy Law in the United States .................... 7
  1.3 Model ............................................. 9
  1.4 Estimation ........................................ 15
  1.5 Empirical Results .................................. 21
  1.6 Robustness ....................................... 25
  1.7 Conclusion ....................................... 35

CHAPTER 2: Measurement Error in Multiple Equations ............ 50
  2.1 Introduction ....................................... 50
  2.2 Data Generation and Assumptions ...................... 53
  2.3 Identification ..................................... 58
  2.4 Numerical Example ................................ 63
  2.5 Estimation and Inference ............................ 65
  2.6 Tobin’s q in Corporate Investment, Saving, and Debt Equations ............................ 72
  2.7 Conclusion ....................................... 81

APPENDIX ................................................. 92

BIBLIOGRAPHY ........................................... 141
LIST OF TABLES

TABLE 1 : Summary Statistics ........................................ 37
TABLE 2 : Aggregate Parameters Values ................................. 38
TABLE 3 : Sensitivity of Moments to Parameters ......................... 39
TABLE 4 : Model Fit .................................................. 40
TABLE 5 : Parameter Estimates ........................................ 40
TABLE 6 : The Effect of Positive Shareholders’ Recovery Rate .......... 41
TABLE 7 : Robustness — Bankruptcy Reform Act of 1978 .............. 42
TABLE 8 : Robustness — Subset Based on Empirical Proxies for Shareholders’ Recovery Rate .................................................. 43
TABLE 9 : Robustness — Firm Heterogeneity ............................... 44
TABLE 10 : Numerical Example .......................................... 84
TABLE 11 : Summary Statistics .......................................... 85
TABLE 12 : Bounds for the Investment, Saving, and Debt Equations for Year 1993 86
TABLE 13 : Bounds for the Investment, Saving, and Debt Equations for Year 1993 87
TABLE 14 : Test of Joint Hypothesis .................................... 89
TABLE 15 : Bounds on the Cash Flow Coefficients in the Investment, Saving, and Debt Equations Using the Full Panel ......................... 91
TABLE 16 : Bounds on the Cash Flow Coefficients in the Investment, Saving, and Debt Equations Using the Full Panel and Accounting for Asset Tangibility ................................. 141
LIST OF ILLUSTRATIONS

FIGURE 1: Illustration of the Trade-off Theory ........................................ 45
FIGURE 2: Identification of $\eta$ and $\alpha$ ............................................. 46
FIGURE 3: Time-Series Variation of Leverage and Default Probabilities .... 47
FIGURE 4: Evolution of Shareholder Recovery Rate and Default Cost in Percentages ................................................................. 48
FIGURE 5: Identification of Time-Series Change in $\eta$ and $\alpha$ .................... 49
FIGURE 6: Identification Regions ............................................................... 83
FIGURE 7: Sensitivity of Partial Identification Region for Cash Flow Coefficients 88
FIGURE 8: Evolution of Identification Region over Time ................................. 90
FIGURE 9: Distribution of Time-Series Length in Quarters ............................... 94
FIGURE 10: Illustration of trade-off theory when upward refinancing is allowed 106
FIGURE 11: Evolution of Identification Region over Time ............................... 140
CHAPTER 1 : Shareholder Recovery and Leverage

1.1. Introduction

In the United States, the absolute priority rule\(^1\) states that shareholders should recover nothing in default unless creditors are paid in full. However, shareholders have received a positive payoff due to a sequence of historical events, notably the Bankruptcy Reform Act of 1978. Between 1970 and 2005, shareholders received a positive payoff in 30.3% of bankruptcy instances (Bharath, Panchapegesan and Werner, 2007). In this paper, I use U.S. data to study the economic consequences of a positive shareholder recovery rate. To that end, using a dynamic model, this paper structurally estimates shareholder recovery rate and conducts counterfactual analysis.

The key insight of the dynamic model is as follows. When shareholders expect to receive a positive payoff in default, they choose to strategically default sooner, which increases borrowing costs, reducing optimal leverage ex-ante. Shareholders would like to commit to zero recovery in default because this would enable them to take higher leverage ex-ante, which generates a greater tax shield benefit. Yet, due to the unique nature of the bankruptcy system in the United States, shareholders sometimes are able to recover some value ex-post, and thus the commitment to zero recovery is not credible. This commitment problem is amplified by allowing default to be costly even when shareholders receive a positive payoff by renegotiating with creditors. The costly default contrasts with model implications of Fan and Sundaresan (2000) and yet is realistic and consistent with empirical findings (Andrade and Kaplan, 1998). As default cost increases, borrowing costs, conditional on leverage, increase, reducing optimal leverage ex-ante. Taken together, this commitment problem can help explain the observed leverage, and thus addresses the “underleverage puzzle,” which states that the “trade-off theory” produces counterfactually high leverage levels when given realistic default costs. (Miller, 1977; Graham, 2000)

\(^1\)See the U.S. Bankruptcy Code §1129(b)(2)(B)(ii))
How much does this commitment problem reduce leverage? I address this question by estimating the structural parameters of my model, targeting leverage and default probabilities. I am able to identify shareholder recovery rate and default cost in the following way. I define shareholder recovery rate as a fraction of remaining firm value before default cost is realized. Accordingly, shareholder recovery increases with shareholder recovery rate although it does not move with respect to changes in default costs. Thus, conditional on leverage, in making a strategic default decision, shareholders consider their recovery rate and yet do not consider default cost. This implies that conditional default probabilities increase with shareholder recovery rate but do not move with respect to changes to default cost. The difference in sensitivities of conditional default probabilities with respect to the two structural parameters helps to separately identify them. Then, I structurally estimate for the representative firm similar to Hennessy and Whited (2005, 2007).

I document dramatic changes over time in the U.S. bankruptcy system. As Hackbarth, Haselmann and Schoenherr (2015) show, the Bankruptcy Reform Act of 1978 increased shareholders’ bargaining power vis-à-vis creditors, and thus shareholder recovery rate increased. In order to test this, similar to Hackbarth et al., I form two subperiods, 1975Q1-1978Q3 and 1981Q2-1984Q4. Between these two subperiods, I allow tax rates to vary in order to account for the other concurrent change: the Economic Recovery Tax Act of 1981. Consistent with Hackbarth et al.’s finding, my structural estimation shows that shareholder recovery rates statistically significantly increased from 0.1% to 29%, whereas default cost statistically insignificantly increased from 19.0% to 21.0%. In response to the change in shareholder recovery rates, firms optimally lowered leverage ex-ante by 32.0%. This shows that allowing a positive shareholder recovery rate better explains the empirically observed leverage than does the “trade-off theory.” Due to lower leverage, default probabilities decreased by 62.5%, and credit spreads decreased by 5.4%. Because firms took less advantage of tax shield benefits, firm values decreased by 5.0% and government tax revenue, defined

---

2In addition, in order to analyze how shareholder recovery rates vary across firms, I structurally estimate shareholder recovery rate for each subset of firms.
as a contingent claim to the future tax revenue, increased by 22.2%. Lastly, lower default probabilities, driven by a positive shareholder recovery, implied less frequent realization of deadweight cost. Thus, the sum of firm values and government tax revenue increased by 4.4%.

After a subsequent series of contractual innovations in the bankruptcy process (Skeel, 2003; Bharath, Panchapagesan and Werner, 2007), shareholders’ bargaining power vis-à-vis creditors steadily decreased, and thus shareholder recovery rate decreased. After accounting for changes in tax rates, my estimates for shareholder recovery rate show the consistent trend: 19.9% between 1985Q1 and 1994Q4; 3.8% between 1995Q1 and 2004Q4; and 0.97% between 2005Q1 and 2016Q4. On the other hand, default cost did not significantly change from one subperiod to the next.

Similar to the existing empirical literature on the capital structure, I also estimate model parameters based on the long sample period, 1970Q1-2016Q4. The implied shareholder recovery rate was 7.1% and the implied default cost was 17.3%. Compared to the counterfactual world where shareholder recovery rate is set to zero, firms’ optimal leverage was 9.4% lower. Due to lower leverage, default probabilities were 8.1% lower, and credit spreads were 1.7% lower. Moreover, firm values were 1.5% lower, government tax revenue was 9.9% greater, and the sum of firm values and government tax revenue was 0.9% greater.

My empirical strategy complements existing literature that relies on natural experiment and direct measurement. Even though results from a natural experiment can be instructive, due to other concurrent changes, it is empirically challenging to tease out the impact of a positive shareholder recovery rate. For example, a seemingly ideal setting for a natural experiment is the Bankruptcy Reform Act of 1978. However, the Economic Recovery Tax Act of 1981 changed tax rates almost simultaneously and thus it is hard to disentangle the impact of shareholder recovery rate from the impact of tax rate. Moreover, it is empirically challenging to use a natural experiment by itself to estimate an unobservable parameter such as shareholders’ expected recovery rate. Direct measurement analysis calculates a
sample average of shareholder recovery rates among bankrupt firms. While instructive, these results may suffer from sample-selection bias because firms with lower shareholder recovery rates default more frequently. This paper bases estimates on a broad cross-section of firms, including both bankrupt and non-bankrupt firms, and thus is immune from the sample-selection bias. Using direct measurement analysis, literature estimates shareholder recovery rate to be between 0.4% and 7.6% (Eberhart, Moore and Roenfeldt, 1990; Franks and Torous, 1989; Betker, 1995; Bharath, Panchapegesan and Werner, 2007). Even before accounting for the sample-selection bias, this paper’s structural estimate of 7.1% for shareholder recovery rate during 1970Q1-2016Q4 is in-line with direct measurement analysis’ estimates.

Admittedly, a structural estimation has limitations. One limitation is that it is hard to account for heterogeneity across firms. This limitation makes it empirically challenging to structurally estimate the sample-selection bias. Firm heterogeneity can arise due to multiple sources, such as heterogeneous shareholder recovery rates or heterogeneous model misspecification. Because the sample-selection bias arises due to heterogeneous shareholder recovery rates but not due to heterogeneous model misspecification, estimating the sample-selection bias requires identifying a portion of heterogeneity that arises only due to the former. Unfortunately, the structural estimation cannot distinguish between these two sources; thus, structural estimation for the sample-selection bias is significantly biased. Due to this limitation, I do not target the result of direct measurement analysis in the structural estimation. Moreover, the limitation implies that researchers might need to revisit the structural estimation of the sample-selection bias (Glover, 2016).

The rest of the paper is structured as follows. Immediately following the introduction is the literature review. Section 1.2 discusses in detail the sequence of events in the United States that allowed shareholders to receive a positive payoff. Section 1.3 develops the model and Section 1.4 discusses the estimation procedure. Section 1.5 presents my empirical

---

3The structural estimation for the representative firm properly addresses heterogeneity thanks to the law of large numbers.
results on model fit, parameter estimates, and economic consequences. Section 1.6 discusses robustness, and Section 1.7 concludes.

**Literature Review** First, there is growing literature on shareholder recovery rate in default. Shareholders can recover non-negative value in default because shareholders can threaten to exercise several options. Credibility of these threats is best illustrated in Eastern Airlines’ bankruptcy case in 1989 (Weiss and Wruck, 1998). Thus, creditors are forced to accept shareholders’ renegotiating terms and this naturally allows shareholders to recoup non-zero residual value in default. Accordingly, using pre-2000 samples on defaulted firms, a number of empirical papers (Eberhart, Moore and Roenfeldt, 1990; Franks and Torous, 1989; Betker, 1995) document that shareholders recover 2.28% to 7.6% of the remaining firm value on average. However, a subsequent series of contractual innovations in the bankruptcy process (Skeel, 2003) decreased shareholder recovery rate ever since. Shareholders recover 0.4% of the remaining firm value on average in 2000Q1-2005Q4 period (Bharath, Panchapagesan and Werner, 2007). Although these results are instructive, their measures could potentially suffer from sample-selection bias in estimating population cross-sectional mean of shareholder recovery rates, whereas my structural estimation is immune from the bias. Moreover, consistent with documented time-series variation, my subperiods analysis (see Figure 4) yields a downward trend in shareholder recovery rate.

The second strand of literature that this paper relates to is on the underleverage puzzle. Using various approaches, a few papers (Altman, 1984; Andrade and Kaplan, 1998) estimate default cost to be between 10% and 20%. However, researchers find that the empirically observed default cost is too low to justify empirically observed leverage (Miller, 1977; Graham, 2000). In response to this concern, Almeida and Philippon (2007), Elkamhi, Ericsson and Parsons (2012), Ju et al. (2005), Bhamra, Kuehn and Strebulaev (2010), and Chen (2010) use various approaches to address the puzzle. More recently, Glover (2016) esti-

---

4These include 1) an option to take risky actions (asset substitution), 2) an option to enter costly Chapter 11, 3) an option to delay the Chapter 11 process if entered and 4) an option not to preserve tax loss carryfowards (for asset sales).
mates population default cost to be much larger (45%) and attributes sample-selection bias as a possible reason behind the large discrepancy between his estimate and other empirical work. In an attempt to address the same puzzle, this paper uses shareholders’ strategic default action driven by a positive shareholder recovery.

Similar to this paper, Morellec, Nikolov and Schurhoff (2012) allow shareholders to receive a positive payoff and obtain a number for shareholder recovery rate. They set liquidation cost to be 46%, assume the liquidation cost to be a bargaining surplus between creditors and shareholders, and assume that shareholders are as equally powerful as creditors are in bargaining. This implies that default is not costly when shareholders and creditors bargain and shareholders recover 23% in default. The gap between Morellec et al.’s estimate and the direct measurement analysis estimate is too large to be reconciled only by the sample-selection bias. This paper allows default to be costly, focuses on the commitment problem driven by shareholder recovery rate, and obtains a shareholder recovery rate that is more in line with the previously documented numbers, between 0.4% and 7.6%.

More recently, by using a model that forces firms to roll over a fixed fraction of debt, Reindl, Stoughton and Zechner (2017) estimate default cost to be 20%. Although Reindl et al.’s estimate is similar to mine, we differ in a few major areas. Most importantly, I allow shareholders to receive a positive payoff and this extension makes the model general enough to capture some types of debt covenants. Absent such an extension, as Reindl et al. argues, debt covenants could have prevented shareholders from strategically defaulting. Next, Reindl et al. argues that Glover’s estimate is significantly upward biased due to a model misspecification in explaining leverage, and motivates the authors’ choice not to match leverage. As discussed in Section 1.6.4, I show that the first-order reason behind Glover’s large estimate is due to Glover’s particular choice of estimation procedure, which is conducted at each firm level. Moreover, I show that the structural estimation procedure, which is used by Hennessy and Whited (2005, 2007), and used in this paper, suffers significantly less from a model misspecification problem and thus validates use of leverage as a
matching moment. Lastly, I allow firms to optimally choose an upward refinancing point.

The third strand of literature that this paper relates to is as follows. Noting the importance of a positive shareholder recovery, Fan and Sundaresan (2000) model strategic interactions between creditors and shareholders and their model is adopted in a number of recent papers (Davydenko and Strebulaev, 2007; Garlappi, Shu and Yan, 2008; Garlappi and Yan, 2011; Morelec, Nikolov and Schurhoff, 2012; Hackbarth, Haselmann and Schoenherr, 2015; Boualam, Gomes and Ward, 2017). Yet, their models typically assume that firms do not incur any default cost in equilibrium, whereas my model allows firms to incur default cost.

Finally, Green (2018) studies how valuable a restrictive debt covenant is in reducing agency costs of debt. As the author focuses on refinancing, he models firms’ default decision as random events. On the contrary, I take firms’ strategic default decision more seriously and study how it impacts firms’ financing. Although I do not explicitly model covenants in my model, a cash-flow-based covenant can be one-to-one matched with shareholder recovery rate and have the similar effect on firms’ optimal leverage ex-ante (see Section 1.3.1 for more discussion). Corbae and D’Erasmo (2017) studies a welfare implication of a policy counterfactual (hybrid version of Chapter 7 and Chapter 11). Corbae et al. compares a policy counterfactual to the world where firms in default optimally choose between Chapter 7, which complies with APR yet comes with high default cost, and Chapter 11, which violates APR yet comes with low default cost. On the contrary, in this paper, firms are not given an option to choose between Chapter 7 and Chapter 11, and this paper studies a shareholder recovery’s impact on the behavior of “average” firms.

1.2. Bankruptcy Law in the United States

In the United States, the bankruptcy code states that creditors should be paid in full before shareholders can receive anything in default. However, in practice, the bankruptcy process is a negotiated agreement involving both creditors and shareholders. Thus, the code merely serves as a guideline for the process rather than a requirement, and thus shareholders can
receive a positive payoff even when creditors are not paid in full. In this section, I briefly discuss the sequence of historical events in the United States that eventually allowed a positive shareholder recovery rate.

Prior to the nineteenth century, the bankruptcy system in the United States was administrative in nature: bankrupt firms were almost always liquidated, its shareholders did not recover any value and managers were let go. Consequently, APR always held and shareholders were never part of the bankruptcy process.

However, in the late nineteenth century, there was a dramatic turn of events due to a series of bankruptcies in the railroad industry. These bankruptcies prompted the courts to intervene and rescue them for the sake of public interest in an effective transportation system. The courts formed equity receivership to run the bankrupt firm. Equity receivership comprised old shareholders, old creditors and old managers. This is important because this was the first time that shareholders became a part of the bankruptcy process. The practice spread to other industries and persisted over time. The Bankruptcy Reform Act of 1978 formally gave more power to shareholders, leading to larger shareholder recovery rate. Although the bankruptcy laws did not significantly change since then, a subsequent series of contractual innovations in the bankruptcy process gradually decreased shareholder recovery rate over time.

It is important to note that a positive shareholder recovery is a byproduct of the courts’ effort to keep its business alive and pay creditors, which is Chapter 11’s stated objective. The bankruptcy process sometimes requires shareholders’ help or consent and thus requires shareholders to be paid off at the expense of the creditors. Consequently, this effort leads to a positive shareholder recovery. In the rest of the document, through the lens of the model, developed in the next section, I test whether a positive shareholder recovery is implied by firm data, and if so, quantify its magnitude that is implied by firm data.

See Skeel (2001) for a more detailed discussion.
1.3. Model

I extend the workhorse dynamic capital structure model (Goldstein, Ju and Leland, 2001) as follows. Upon default, firms lose $\alpha$, shareholders recover $\eta$, and creditors recover the remainder $1 - \eta - \alpha$ fraction of the remaining firm value. If shareholders are subject to a higher tax rate than creditors are, firms have an incentive to issue debt to shield earnings from taxation. Such a tax shield benefit motive is the only reason that firms want to lever up in my model. To stay in a simple time-homogeneous setting, I consider callable debt contracts that are characterized by a perpetual flow of coupon payments. Shareholders of each firm make three types of corporate financing decisions: (1) when to default, (2) when to refinance, and (3) how much debt to issue upon refinancing. Shareholders exercise their default option if earnings drop below a certain earning level, called the default threshold. Shareholders exercise the refinancing option if earnings rise above a certain earning level, called the upward refinancing threshold. These features are shared with numerous other capital structure models.

1.3.1. Setup and Solution

In the model, a firm $i$’s earnings growth depends on aggregate earnings shocks as well as idiosyncratic shocks specific to the firm. Before-tax earning, $X_{i,t}$, evolves according to

$$\frac{dX_{i,t}}{X_{i,t}} = \mu_i dt + b_i \sigma_A dW^A_t + \sigma^F_i dW^F_{i,t}$$

Firm $i$’s expected earnings growth is given by $\mu_i$. $b_i$ is firm $i$’s exposure to the aggregate earnings shocks generated by the Brownian motion $W^A_t$ and $\sigma_A$ is the volatility of aggregate earnings shocks. $\sigma^F_i$ is the volatility generated by the firm-specific Brownian motion $W^F_{i,t}$.

By assumption, $W^F_{i,t}$ is independent of $W^A_t$ for all firms $i$.

---

6This naturally imposes a restriction that $\eta + \alpha <= 1$.

7Contrary to Leland (1994), the model allows shareholders to recover non-zero value. Contrary to Fan and Sundaresan (2000), firms can potentially incur default cost even when shareholders and creditors enter renegotiation. Here, I want to emphasize that the model does not rule out $\alpha = 0$ nor $\eta = 0$. 

9
The model is partial equilibrium, and thus the pricing kernel is exogenously set as:

\[
\frac{d\Lambda_t}{\Lambda_t} = -rdt - \varphi_A dW_t^A
\]

where \( r \) is the risk-free rate and \( \varphi_A \) is the market Sharpe ratio. Under the risk-neutral measure, the earnings process evolves according to:

\[
\frac{dX_{i,t}}{X_{i,t}} = \hat{\mu}_i dt + \sigma_{i,X} d\hat{W}_{i,t}
\]

where \( \hat{W}_{i,t} \) is Brownian motion under the risk-neutral probability measure, \( \hat{\mu}_i = \mu_i - b_i \sigma_A \varphi_A \) and \( \sigma_{i,X} = \sqrt{(b_i \sigma_A)^2 + (\sigma_i^F)^2} \). In order to guarantee the convergence of the expected present value of \( X_{i,t} \), I impose the usual regularity condition \( r - \hat{\mu}_i > 0 \). For notational convenience, I drop \( i \) in the rest of the document.

This paper uses \( \tau_{cd} \equiv 1 - (1 - \tau_c)(1 - \tau_d) \) as an effective tax rate that shareholders pay on the corporate earnings where \( \tau_c \) denotes tax on corporate earnings and \( \tau_d \) denotes tax on equity distributions. \( \tau_{cdi} \equiv \tau_{cd} - \tau_i \) denotes tax shield benefit rate, where \( \tau_i \) denotes tax on interest income.

Now, I describe how debt value and equity value are derived. For given default threshold \( X_D \) and optimal policies (coupon \( C \) and upward refinancing point \( X_U \)), I use the contingent claims approach to solve for debt value \( D(X) \) and equity value \( E(X) \). The most relevant boundary conditions are as follows. (Please see the Appendix for more detail.)

\[
D(X_D) = (1 - \alpha - \eta) \frac{(1 - \tau_{cd}) X_D}{r - \hat{\mu}} \quad (1.1)
\]

\[
E(X_D) = \eta (1 - \tau_{cd}) X_D \frac{1 - \tau_D}{r - \hat{\mu}} \quad (1.2)
\]

The first boundary condition captures that creditors recover \( 1 - \alpha - \eta \) fraction of the remaining unlevered firm value. The second boundary condition captures that shareholders recover \( \eta \) fraction of the remaining unlevered firm value.
The next step is to solve for an optimal coupon $C$, upward refinancing point $X_U$, and default threshold $X_D$. $C$ and $X_U$ are determined at time 0 (initial point or refinancing point) to maximize firm value minus debt issuance cost. Here, debt issuance cost is $\phi_D$ times the debt value.

$$[C, X_U] = \arg \max_{C^*, X_U^*} ((1 - \phi_D)D(X_0; C^*, X_U^*) + E(X_0; C^*, X_U^*))$$ (1.3)

subject to

$$\lim_{X_t \downarrow X_D} E'(X_t) = \frac{\eta(1-\tau_{ed})}{r - \hat{\mu}}$$ (1.4)

$X_D$ is determined based on the above smooth pasting condition, Equation (1.4) (see the heuristic derivation of the smooth pasting condition in Appendix A.3.2).

Because the model is used to explain the data, it is important to discuss what $\eta$ might potentially capture. In the model, shareholders recover only in default. In reality, prior to declaring bankruptcy, some shareholders can potentially enjoy benefits of control rights by, for example, opportunistically refinancing to change covenants (Green, 2018). To the extent that shareholders’ opportunistic behavior makes debt more costly and higher debt cost is internalized, shareholder recovery rate $\eta$ in the model captures such benefits in addition to explicit ex-post recovery value.

When shareholders recover zero amount in default, shareholders might not be able to strategically default at the optimal default threshold due to debt covenants (Reindl, Stoughton and Zechner, 2017). Yet, when shareholders are allowed to receive a positive payoff in default, such debt covenants can be easily described with my current model and shareholders will behave as if shareholders strategically default. More specifically, let us consider a cash-flow-based covenant, which specifies that shareholders must default if the earnings-to-coupon ratio goes below a certain threshold, say $\bar{X}_{DC}$. Because strategic default threshold

\[^8\text{For example, fallen angel firms delay refinancing relative to always-junk firms because loose covenants allow shareholders to transfer wealth from creditors.}\]
$X_{DC}(\eta)$ monotonically increases over $\eta$ (as shown below in Equation (1.5)), I can find unique $\eta$ that sets $X_{DC}(\eta) = \bar{X}_{DC}$. Such an implied $\eta$ increases over $\bar{X}_{DC}$, and thus shows that a more restrictive covenant (or equivalently higher $\bar{X}_{DC}$) corresponds to larger $\eta$. As such, my model is general enough to capture cases with some types of debt covenants.

1.3.2. Key Economic Channels: Commitment Problem

In this subsection, I discuss key economic channels in three steps.

Let me explain the first step. In their decision to default, shareholders weigh the benefits of holding on to their control rights, all future dividends, and recovery value against the costs of honoring debt obligations while the firm is in financial distress. As shareholder recovery rate ($\eta$) increases, the trade-off shifts and leads to earlier exercise of the option to default. This intuition can be seen in a closed form for the normalized default threshold ($X_{DC}$):

$$X_{DC} = \frac{X_D}{C} = \frac{r - \hat{\mu} - \lambda_-}{r} \frac{1}{1 - \lambda_-} \frac{1}{1 - \eta}$$  \hspace{1cm} (1.5)

where $\lambda_-$ is a negative solution to the characteristic equation.

On a related note, $X_{DC}$ does not depend on default cost, $\alpha$, yet depends on $\eta$. The intuition is as follows. I define shareholder recovery rate as a fraction of remaining firm value before default cost is realized. Accordingly, shareholder recovery rate does not move with respect to changes to default cost. Thus, conditional on leverage, in making a strategic default decision, shareholders account for their recovery rate and yet do not account for default cost. This implies that conditional default probabilities increase with shareholder recovery rate but default probabilities do not move with respect to changes to default cost. This intuition is a key mechanism in separately identifying $\alpha$ and $\eta$. Thus, I write this as a proposition below and refer back to it later:

**Proposition 1** Conditional on the leverage, default probabilities increase over shareholder recovery rate ($\eta$). Conditional on the leverage, default probabilities do not change over
default cost (\(\alpha\)).

The second step is, conditional on the leverage, debt becomes more costly. In other words, as \(\eta\) increases, borrowing cost increases because creditors lose \(\eta\) to shareholders and \(X_{DC}\) is determined to maximize the equity value at the expense of the bond value. In the third step, firms internalize higher debt cost and optimally lower leverage ex-ante.

[INSERT FIGURE 1]

The aforementioned three-step intuition can be illustrated graphically as shown in Figure 1. Going from point \(A\) to \(B\) illustrates the first two steps, where firm value decreases due to shareholders’ strategic default action. The last step is illustrated by going from point \(B\) to \(C\) where firms optimally lower leverage ex-ante, and thus firm values increase. Interestingly, as Proposition 7 proves, the increase in firm values from point \(B\) to \(C\) is not sufficient to offset the decrease in firm values from point \(A\) to \(B\).

In the rest of this section, I show why costly default is important in my setting. In order to present the intuition with closed forms, I suppress upward refinancing and study terms for firm value minus debt issuance cost:

\[
(1 - \phi_D)D(X_t) + E(X_t) = \frac{1 - \tau_{cd}}{r - \bar{\mu}}X_t + \frac{\tilde{\tau}_{cdi}C}{r} + Loss\left(\frac{X_t}{C \cdot X_{DC}}\right) < 0 \tag{1.6}
\]

where \(\tilde{\tau}_{cdi} = (1 - \tau_i)(1 - \phi_D) - (1 - \tau_{cd})\) and

\[
Loss/C = -\frac{\tilde{\tau}_{cdi}}{r} - (\alpha + \phi_D(1 - \alpha - \eta))(1 - \tau_{cd})X_{DC} \frac{1 - \tau_{cd}}{r - \bar{\mu}} < 0 \tag{1.7}
\]

Here, \(\frac{\tilde{\tau}_{cdi}C}{r}\) captures the tax shield benefit whereas \(Loss\left(\frac{X_t}{C \cdot X_{DC}}\right)\) captures expected firm-value loss. As \(\eta\) increases, firm values decrease because the expected firm-value loss increases due to larger \(X_{DC}\). Firms’ optimal policy, \(C\), has to decrease to equate marginal cost and marginal benefit. It is important to note that \(\eta\) impacts firms’ capital structure decision mainly through \(X_{DC}\). Thus, if default cost (\(\alpha\)) becomes zero, because \(\phi_D\) is very
small in magnitude, $(\alpha + \phi_D(1 - \alpha - \eta))$ in Equation (1.7) becomes negligible, and thus Loss term in Equation (1.6) becomes insensitive to $\eta$. Consequently, $\eta$’s impact on the optimal leverage significantly decreases.

1.3.3. Leverage, Default Probability and Market Beta

In this subsection, I discuss how $\eta$ and $\alpha$ relate to leverage, default probability, and market beta.

Higher $\alpha$ implies higher firm-value loss conditional on defaults and thus higher expected firm-value loss. This consequently implies lower optimal leverage. Higher $\eta$ implies higher firm-value loss and higher default probability. Taken together, this implies higher expected firm-value loss and consequently implies lower optimal leverage.

Proposition 2 Higher default cost ($\alpha$) and higher shareholder recovery rate ($\eta$) lead to lower leverage.

All proofs are in the Appendix. Let us now discuss how $\alpha$ and $\eta$ relate to default probabilities. As $\alpha$ increases, optimal leverage decreases, and thus default probabilities decrease. Higher $\eta$ implies higher default probabilities and higher firm-value loss conditional on defaults. If leverage decreases only to exactly offset the increase in default probabilities but not in firm-value loss, then marginal cost is larger than marginal benefit and thus it is not optimal. Leverage has to further decrease and this exactly implies that default probabilities decrease over $\eta$.

Proposition 3 Higher default cost ($\alpha$) and higher shareholder recovery rate ($\eta$) lead to lower default probability.

Because default probabilities and leverage change over $\eta$ in the same direction as they do over $\alpha$, I need an extra key economic channel to separately identify $\alpha$ and $\eta$. The structural estimation in this paper is akin to solving the system of two equations for two unknowns. The two unknowns correspond to shareholder recovery rate and default cost. Leverage
specifies one equation in terms of two model parameters and default probability specifies
the other equation in terms of the same two parameters. Figure 2 graphically illustrates
the intuition. It shows locus of \( \alpha \) and \( \eta \) that match a given leverage (solid line) and a given
default probability (dashed line). A necessary condition for \( \alpha \) and \( \eta \) to be point-identified
is that both lines have different slopes.

\[ \text{[INSERT FIGURE 2]} \]

The key driver for the aforementioned necessary condition is \textbf{Proposition 1}. The difference
in sensitivities of conditional default probabilities with respect to the two model parameters
helps to separately identify them in the system of two equations. \textbf{Proposition 4} uses this
intuition to prove the necessary condition.

\textbf{Proposition 4} \textit{Leverage and default probability help to separately identify default cost (\( \alpha \))
and shareholder recovery rate (\( \eta \)).}

Lastly, let us discuss how market betas change over \( \eta \). Upon default, higher \( \eta \) implies
that shareholders recover a higher share of the unlevered firm value. Because unlevered
firm value is less risky than equity, positive probability of receiving higher payout that
is less risky in default implies lower market beta. Moreover, as firms actively lower their
leverage, firms face smaller distress risk and thus market betas further decrease. This idea
is consistent with empirical evidence that is documented in Garlappi, Shu and Yan (2008),
Garlappi and Yan (2011) and Hackbarth, Haselmann and Schoenherr (2015).

\textbf{Proposition 5} \textit{Higher shareholder recovery rate (\( \eta \)) leads to lower market beta.}

1.4. Estimation

This section describes the data, aggregate parameters, estimator, and intuition behind the
estimation method.
1.4.1. Data

Sample

I obtain panel data from CRSP and COMPUSTAT. I omit missing observations and all firms whose primary SIC classification is between 4900 and 4999 or between 6000 and 6999 since the model is inappropriate for regulated or financial firms. I follow Bharath and Shumway (2008) and Gomes, Grotteria and Wachter (2018) to construct quarterly distance-to-default (DD) default probability measures. The sample contains 246,090 firm-quarter observations, number of unique firms is 4,435 and spans from 1970Q1 to 2016Q4. Table 1 shows summary statistics for the panel data set that this paper attempts to match. Section A.4.1 describes data variable definitions.

[INSERT TABLE 1]

Market beta is 1.105 on average, quarterly earnings growth rate is 0.6% on average, and leverage is 0.283 on average. Moments of quarterly DD default probability are important in identifying my key parameters. Thus, I validate DD default probability measures. Sample average for my constructed quarterly default probability is 0.27%. This is similar in magnitude to the realized quarterly bankruptcy frequencies that are reported at 0.27% based on the sample period between 1970 and 2014 (Chava and Jarrow, 2004; Chava, 2014; Alanis, Chava and Kumar, 2015) or 0.28% based on the sample period between 1970 and 2003 (Campbell, Hilscher and Szilagyi, 2008). I also validate my measure against Moody’s Expected Default Frequencies (EDF) measures, which are widely used by financial institutions as a predictor of default probability and are used in several academic papers including Garlappi, Shu and Yan (2008) and Garlappi and Yan (2011). My measures and Moody’s are significantly and positively correlated as rank correlation is 0.75 with p-value=0.00.

9Market beta is not 1 on average because this is an equal-weighted average.
Tax Rates

Following Graham (2000), a few papers (Chen, 2010; Glover, 2016) set $\tau_c = 35\%$, $\tau_d = 12\%$ and $\tau_i = 29.6\%$. However, Graham’s sample period covers only from 1980 to 1994. Because my sample spans from 1970 through 2016, it calls for more up-to-date tax rates. I construct panel data of the most up-to-date tax rates ($\tau_c$, $\tau_i$ and $\tau_d$) by closely following Graham (2000) (see Section A.4.1).

The tax rates used in this paper are $\tau_c = 28.21\%$, $\tau_d = 17.76\%$, and $\tau_i = 32.85\%$. Relative to what were used by a few papers, my $\tau_c$ is lower because it captures periods with low earnings growth and thus implies lower corporate tax rates. My $\tau_d$ is larger because the proportion of long-term capital gains that is taxable increased from 0.4 to 1 after 1987 and my sample captures more of post-1987 than Graham (2000) does. Lastly, my $\tau_i$ is slightly larger because it accounts for the fact that $\tau_i$ is larger in the pre-1988 period. In net, $\tau_{cdi}$ decreased from 13.20% to 10.20%.

1.4.2. Aggregate Parameters

Because I assume that aggregate variables do not change over time, I calibrate aggregate variables using the longest sample period available: 1947 through 2016. Table 2 summarizes calibrated values for aggregate parameters and the corresponding data sources. For quarterly aggregate earnings growth volatility ($\sigma_A$), I use log growth rates based on the quarterly aggregate earnings data from National Income and Product Accounts (NIPA) Table 1.14. For real risk-free rate ($r$), I subtract realized inflation rate from the nominal three-month Treasury bill rate.10 Lastly, I use quarterly market Sharpe ratio and debt issue cost ($\phi_D$) that are reported in Chen (2010) and Altinkilic and Hansen (2000) respectively.

[INSERT TABLE 2]

10Using the expected inflation rate (Bansal, Kiku and Yaron, 2012) yields very similar value for $r$. 

17
1.4.3. Estimator

I estimate the model using a two-step process. I first set $\mu$ to the sample mean of earnings growths, which is 0.60%. Then, I estimate the remaining four parameters by using the simulated method of moments (SMM), which chooses parameter estimates that minimize the distance between moments generated by the model and their data analogs. The following subsection defines matched moments and explains how they identify the four parameters. The four parameters are $\eta$ and $\alpha$, which characterize the default process, and $b$ and $\sigma^F$, which characterize earning. Appendix A.1 contains additional details on the SMM estimator.

1.4.4. Identification and Selection of Moments, and Heterogeneity

In a structural estimation, proper moment selection is crucial to identify a unique set of model parameters that make the model fit the data as closely as possible. To that end, moments’ predicted values need to be differently sensitive to the model parameters, and there should be a sufficient number of moments. I match eight moments to identify the four parameters.

Before defining the moments, I address the issue of firm heterogeneity. Model parameters vary across firms, and it is undoubtedly important to account for cross-sectional distribution of model parameters as shown by Glover (2016). However, it is empirically challenging to estimate cross-sectional distribution especially when the model is misspecified (see Section 1.6.4). Similar to Hennessy and Whited (2005, 2007), the structural estimation used in this paper addresses firm heterogeneity in the data by removing firm fixed effects and estimates for the representative firm. Consequently, this allows me to match time-series variation of moments.

Now, I define the moments. The eight matched moments are the mean of market beta, the variance of earnings growth, leverage’s three moments (mean, variance and skewness) and
default probability’s three moments (mean, variance and skewness).\textsuperscript{11} One additional candidate for moments, in identifying \( \eta \), are statistical moments of shareholder recovery rates among defaulted firms, which are documented by Eberhart, Moore and Roenfeldt (1990), Franks and Torous (1989), Betker (1995), and Bharath, Panchapegesan and Werner (2007). In matching these moments, it is crucial to address sample-selection bias. Unfortunately, the magnitude of sample-selection bias is very sensitive to cross-sectional distribution of \( \eta \). Because it is empirically challenging to estimate cross-sectional distribution of \( \eta \) (see Section 1.6.4), I decided not to match those moments.

Next, in order to explain how the identification works, I discuss how each parameter is identified by the aforementioned eight moments. Table 3 tabulates how much each moment changes over parameters and supports the description below. Each moment depends on all model parameters, but I explain the moments that are the most important for identifying each parameter.

[INSERT TABLE 3]

Time-series mean, volatility and skewness of leverage and the same statistical moments of default probabilities help to identify \( \eta \). This is illustrated by Figure 3. The figure illustrates two firms that face the same sequence of earnings (top panel). Two firms have the identical model parameter numbers except for \( \eta \). The middle panel illustrates that \( \eta = 0 \) firm has larger leverage than \( \eta \neq 0 \) firm on average (consistent with Proposition 2). Lower target leverage, driven by larger \( \eta \), makes leverage less sensitive to sequence of subsequent earnings growth shocks and thus decreases time-series volatility of leverage.

Moreover, larger \( \eta \) incentivizes firms to upward refinance less frequently. Because upward refinancing makes it more probable to default, larger expected firm-value loss, driven by larger \( \eta \), incentivizes firms to upward refinance less frequently. Similar to debt issuance

\textsuperscript{11}Another relevant moment to match is the credit spread. Yet, I decided not to match the credit spread due to its data limitation. Firms’ debt frequently consists of heterogeneous instruments, and market prices for most of these are less readily available than aforementioned data. Nonetheless, I study its sensitivity to key parameters.
cost (see Leary and Roberts (2005) for empirical support), shareholder recovery rate makes firms' leverage more persistent over time. Consequently, leverage decreases on average and becomes less volatile. As firms reduce their target leverage and stay below its target leverage longer, time-series distribution becomes more positively skewed. Although default probabilities are similarly related to \( \eta \) as leverage is related to \( \eta \), default probabilities are not as sensitive to \( \eta \) as leverage is due to the opposing force from shareholders’ strategic default action (see Table 3). As illustrated in Figure 3, \( \eta = 0 \) firm upward refinances at time 84 whereas \( \eta \neq 0 \) firm does not upward refinance until its earnings reach a higher threshold at time 154. Because \( \eta = 0 \) firm increases its leverage earlier, a series of negative shocks between time 90 and 140 keep its leverage much higher and more volatile than \( \eta \neq 0 \) firm’s. This leads to more significant jumps in default probability for \( \eta = 0 \) (bottom panel).

In addition, market beta helps to identify \( \eta \). \( \eta \) is negatively related to market beta. This relation is consistent with empirical findings reported in Garlappi, Shu and Yan (2008), Garlappi and Yan (2011), and Hackbarth, Haselmann and Schoenherr (2015).

Similar to \( \eta \), time-series mean, volatility and skewness of leverage and the same statistical moments of default probabilities help to identify \( \alpha \). Higher \( \alpha \) makes leverage lower on average, less volatile and more positively skewed and thus default probability lower on average, less volatile and more positively skewed.

Most importantly, I discuss how \( \alpha \) and \( \eta \) are separately identified. Conditional on leverage, shareholders’ strategic default implies that default probabilities increase over \( \eta \). Due to the commitment problem, however, optimal leverage decreases and, consequently, default probabilities decrease over \( \eta \). Thus, default probabilities are less negatively related to \( \eta \) than they would be without the opposing force driven by shareholders’ strategic default. However, because shareholders’ strategic default action does not depend on \( \alpha \), default probabilities are much more negatively related to \( \alpha \). I illustrate this by calculating the implied slopes of two curves shown in Figure 2. Using numbers reported in Table 3, the slope of solid
blue curve is \( -\frac{\partial E(Lev)}{\partial \eta}/\partial \alpha \) = \(-0.317 - 0.005\) = 0.53, whereas the slope of dashed red curve is \( -\frac{\partial E(DP)}{\partial \eta}/\partial \alpha \) = \(-0.081 - 0.530\) = 0.15. Two curves have different slopes, and thus satisfy a necessary condition for \( \alpha \) and \( \eta \) to be point-identified.

Lastly, let us discuss how the remaining two parameters are identified. Larger \( b \) implies higher exposure to the systematic risk. This naturally translates to larger mean of market beta. Larger \( b \) also implies lower risk-neutral earnings growth rate, which implies lower equity value. This translates to larger mean, higher volatility and smaller skewness of leverage. Larger \( b \) also implies higher volatility of earnings growth rate and thus implies larger mean and higher volatility of default probability. \( \sigma_F \) is naturally identified by the earnings growth rate volatility. Moreover, larger \( \sigma_F \) translates to higher volatility of default probability and implies higher volatility of equity value, and thus higher volatility of leverage. Lastly, as \( \sigma_F \) increases, the probability of reaching the default threshold during the next period increases and thus the mean of default probability increases.

1.5. Empirical Results

In this section, I present main structural estimation results and discuss their implications.

1.5.1. Main Results

Table 4 summarizes model fit. The first and the second rows show data moments and standard errors, respectively. The third row shows model-implied moments. The last two rows show difference between data and model-implied moments and t-statistics.

As shown, all the moments are matched well as none of the differences between data and model-implied moments are statistically significantly different from zero. Especially, I want to highlight two main matching moments. Data sample mean of leverage is 0.283, whereas the model counterpart is 0.283. The difference between the data leverage and model-implied leverage is statistically insignificant. Data sample mean of quarterly default probability is
0.3%, whereas the model counterpart is 0.4%. Again, the difference between the data default probability and model-implied default probability is statistically insignificant.

Table 5 summarizes the parameter estimates. Shareholders’ recovery rate ($\eta$) is estimated to be 7.1% and $\alpha$ is estimated to be 17.3%. Most interestingly, $\eta$ is statistically different from zero, thus a natural null hypothesis that $\eta = 0$ is rejected at 1% significance level. This number is in line with the empirically observed counterpart, between 0.4% and 7.6% (Eberhart, Moore and Roenfeldt, 1990; Franks and Torous, 1989; Betker, 1995; Bharath, Panchaipesan and Werner, 2007), and thus strongly validates my estimates (see Section 1.6.4 for more careful validation of my estimates). Another interpretation of results is that average firms expect to enter the Chapter 11 in default and expect shareholders to recover a non-negative amount. If average firms expect to enter the Chapter 7 in default, then the implied $\eta$ should have been 0, but this is statistically significantly rejected. This is consistent with an empirical observation that most publicly listed firms that are declaring bankruptcy file for the Chapter 11 rather than the Chapter 7. Moreover, default cost ($\alpha$) is expected to be statistically significantly positive even when shareholders and creditors are expected to renegotiate in default. Lastly, if $\eta = 0$ is imposed in the structural estimation, $\alpha$ is estimated to be 22.7%. This illustrates how allowing a positive shareholder recovery helps to obtain default cost, which is more in line with the empirically observed counterpart, between 10% and 20% (Altman, 1984; Andrade and Kaplan, 1998), and thus strongly validates my estimates.

[INSERT TABLE 5]

1.5.2. Credit Spread, Firm Value, and Government Tax Revenue

In this section, I discuss how $\eta$ qualitatively relates to credit spreads, firm values and government tax revenue. Then, I quantify such relations in the subsequent sections.

\footnote{According to www.bankruptcydata.com, more than 90% of U.S. public firms file under Chapter 11.}
\footnote{Morellec, Nikolov and Schurhoff (2008, 2012) cite Gilson, John and Lang (1990) to argue that $\alpha$ is small (0%-5%) when shareholders and creditors renegotiate. However, Gilson’s measure does not include indirect cost and thus is not an appropriate measure in my context.}
Credit Spread

Conditional on leverage, larger $\eta$ can be thought of as wealth transfer from creditors to shareholders. As this is disadvantageous to creditors, credit spreads increase. However, as firms lower leverage in response, their default risk decreases, and thus credit spreads decrease. Such a commitment problem is strong enough that credit spreads decrease over $\eta$ in net.

**Proposition 6** Higher shareholder recovery rate ($\eta$) leads to lower credit spread.

In the literature, there is no empirical consensus on how $\eta$ impacts credit spreads. Davydenko and Strebulaev (2007) find the relation to be positive yet economically small in magnitude. Based on cross-country data, Davydenko and Franks (2008) do not find any positive correlation between $\eta$ and credit spreads. Hackbarth, Haselmann and Schoenherr (2015) find that credit spreads increased after $\eta$ supposedly increased due to the Bankruptcy Reform Act of 1978. Yet, I argue that the increase in credit spreads was not due to the increase in $\eta$ but due to the concurrent decrease in personal income tax rates ($\tau_i$). Section 1.6.1 lists more detail on quantitative analysis of the Bankruptcy Reform Act of 1978.

Firm Value and Government Tax Revenue

Higher $\eta$ leads to lower default probabilities and thus lower expected firm-value loss. Simultaneously, lower leverage and less frequent refinancing, driven by higher $\eta$, decrease the tax shield benefit. As default probabilities are small in magnitude, the latter channel more than offsets the former channel. In net, firm values decrease over $\eta$.

**Proposition 7** Higher shareholder recovery rate ($\eta$) leads to lower firm value.

As $\eta$ increases, firms decrease their leverage and upward refinance less often. Both of these lead to less usage of the tax shield benefit and thus the government collects more taxes. In order to quantify how much government tax revenue increases, I assume that government
tax revenue is a contingent claim to the future tax revenue (see the Appendix for the derivation).

**Proposition 8** Higher shareholder recovery rate ($\eta$) leads to larger government tax revenue.

Let us make one more assumption to study $\eta$’s impact on the entire economy: the entire economy consists of firms and the government. In the model, there are two sources of deadweight cost, default cost and debt issuance cost. Larger $\eta$ leads to smaller default probabilities and thus less frequent realizations of default cost. Moreover, larger $\eta$ leads to less frequent realization of debt issuance cost as larger $\eta$ makes refinancing less frequent. Taken together, larger $\eta$ implies less frequent realization of deadweight cost, and consequently larger value for the entire economy (see Section A.3.6 for the mathematical derivation).

1.5.3. The Effect of Positive Shareholders’ Recovery Rate

Column (1) of Table 6 summarizes the counterfactual world when shareholders recover zero amount in default ($\eta = 0$). Under column (2), I allow shareholders to recover non-zero amount in default, yet I force firms to keep the same optimal policies (coupon, $C$, and refinancing point, $X_U$) as under (1). This exercise helps to quantify how much expected firm-value loss increases, conditional on the firms’ optimal policies. Upon default, firms lose 

$$\alpha \frac{(1-\tau_{cd})X_D}{r-\hat{\mu}}$$

where $\frac{(1-\tau_{cd})X_D}{r-\hat{\mu}}$ is the unlevered firm value in default. Firm-value loss increases as $X_D$ increases from 0.077 to 0.083. Simultaneously, default probabilities increase from 0.388% to 0.436%. Taken together, expected firm-value loss increases by 21.3% even when default cost $\alpha$ does not change. Lastly, consistent with a few papers such as Davydenko and Strebulaev (2007), credit spreads increase. Now, under (3), I allow firms to internalize higher borrowing cost and to re-choose their optimal policies. Higher borrowing costs force firms to borrow less, and thus default probabilities decrease, market betas decrease and credit spreads decrease.

In sum, as we allow for a positive shareholder recovery (i.e. comparing column (1) and
(3)), leverages decrease by 9.4% and default probabilities decrease by 8.1%. Market betas decrease by 20.5% and this is qualitatively consistent with Garlappi, Shu and Yan (2008), Garlappi and Yan (2011), and Hackbarth, Haselmann and Schoenherr (2015)’s empirical finding. Moreover, credit spreads decrease by 1.7% (see Appendix A.5 for the discussion on its magnitude).

Interestingly, firm values decrease by 1.5% as firms lose tax shield benefit. Positive $\eta$ decreases the net leverage benefit (defined as a difference between levered firm value and unlevered firm value) by 14.2%. As firms take less advantage of tax shield benefits, the government collects more taxes. That amounts to a 9.9% increase in government tax revenue. Lastly, the sum of firm values and government tax revenue increases by 0.9%.

1.6. Robustness

This section reports how estimates for shareholder recovery rate changed over time and over different subsets of firms. I discuss how firm heterogeneity might affect estimates for shareholder recovery rate. Lastly, I discuss how estimates would change when I use a different assumption in the model.

1.6.1. Bankruptcy Reform Act of 1978

The Bankruptcy Reform Act of 1978 (BRA) is the most important act that shaped the nature of the modern U.S. bankruptcy system (see Appendix A.6 for more institutional details). Through the lens of my model, I test how much this act changed default cost, $\alpha$, and shareholder recovery rate, $\eta$.

Similar to Hackbarth, Haselmann and Schoenherr (2015), I construct two subperiods: 1975Q1-1978Q3 and 1981Q2-1984Q4. A period between 1978Q4-1981Q1 is removed because, as Hackbarth argues, the market was still learning of BRA’s true impact. In order to focus on the impact that BRA had on $\eta$ and $\alpha$, I assume that only $\eta$ and $\alpha$ changed over these two subperiods and assume that the other model parameters did not change. In order
to account for shifts in firms’ optimal decisions driven by changes in tax rates, I allow tax rates to vary across these two periods. More specifically, in each subperiod, I set the tax rate to the panel-wide average of firm-quarter tax rates. For the pre-event subperiod, tax shield benefit rate ($\tau_{\text{cdi}}^{\text{pre}}$) is set to 11.28\%. For the post-event subperiod, tax shield benefit rate ($\tau_{\text{cdi}}^{\text{post}}$) is set to 18.95\%. $\tau_{\text{cdi}}$ changed over these two subperiods because the Economic Recovery Tax Act of 1981 significantly decreased the personal tax rate on interest income ($\tau_i$).\(^{14}\)

Similar to the main structural estimation (Section 1.4.3), I first estimate $\mu$ by using the entire sample. Then, I structurally estimate six parameters by matching sixteen moments. Six parameters include $b$, $\sigma^F$, $\eta$ for the pre-event subperiod ($\eta^{\text{pre}}$), $\eta$ for the post-event subperiod ($\eta^{\text{post}}$), $\alpha$ for the pre-event subperiod ($\alpha^{\text{pre}}$), and $\alpha$ for post-event subperiod ($\alpha^{\text{post}}$). Sixteen moments include eight moments (mean of market beta, variance of earnings growth, three moments of leverage and three moments of default probability) from the pre-event subperiod and the same eight moments from the post-event subperiod.

Table 7 summarizes parameter estimates. Consistent with the literature’s qualitative argument, $\eta$ statistically significantly increased (t-statistics is 11.4). Decrease in market betas, leverages and default probabilities combined with the concurrent decrease in tax rates have contributed to a significant increase in estimated $\eta$. More interestingly, the current paper quantifies such an increase: $\eta$ increased from 0.1\% to 29.0\%. Moreover, $\eta^{\text{pre}} = 0.1\%$ is consistent with Hackbarth et al.’s argument that the impact of BRA was unclear leading up to 1978. In addition, $\alpha$ increased, although not statistically significantly, from 19.0\% to 20.1\%.

When only $\eta$ changed from 0.1\% to 29.0\%, leverages decrease by 32.0\%, market betas de-

\(^{14}\)One caveat to note here is that data on corporate marginal tax rates ($\tau_c$) are not available for pre-1980 and $\tau_c$ used for firms in the pre-event subperiod are imputed as described in Section A.4.1. However, I do not believe that this imputation causes an increase in $\tau_{\text{cdi}}$ as the Economic Recovery Tax Act of 1981 targeted only individual income tax rates.
crease by 17.5% and credit spreads decrease by 5.4%. A slight increase in $\alpha$ further decreases leverage, market betas and credit spreads. Yet, the concurrent change in tax rates mutes the aforementioned changes. In net, leverages decrease by 19.5%, market betas decrease by 12.5% and credit spreads increase by 4.0%. Here, I want to highlight that an empirically observed increase in credit spreads, which Hackbarth, Haselmann and Schoenherr (2015) documents, is not due to the change in the bankruptcy code but rather due to the change in the tax code. Lastly, the rise in $\eta$ decreases firm values by 5.0%, increases government tax revenue by 22.2%, and increases the sum of firm value and tax revenue by 4.4%. Yet, after accounting for all the changes, including tax rates, firm values increase by 0.5%, government tax revenue decreases by 14.7%, and the sum decreases by 4.8%.

1.6.2. Evolution of Shareholders’ Recovery Rate

Even though bankruptcy law has not significantly changed since BRA was passed, Skeel (2003) conjectured that contractual innovations in the bankruptcy process steadily decreased shareholder recovery rate. In support for Skeel’s conjecture, Bharath, Panchapegesan and Werner (2007) documents an empirical evidence: among firms that defaulted between 2000 and 2005, shareholders only recovered 0.4% on average. Bharath et al. attributes such a time-series decline in shareholder recovery rate to contractual innovations, such as debtor-in-possession financing and key employee retention plans.

In order to test whether this is reflected in firm data, I structurally estimate shareholder recovery rate and other model parameters for more recent periods. I first divided post-1985 era into three subperiods: 1985Q1-1994Q4, 1995Q1-2004Q4, and 2005Q1-2016Q4. I estimate model parameters for each subperiod independently from the others. Figure 4 graphically illustrates subperiod results. For completeness, the figure also illustrates pre-1985 estimates, which were discussed in Section 1.6.1. Consistent with Skeel’s conjecture and Bharath et al.’s empirical finding, shareholder recovery rate steadily decreased over time. Shareholders’ recovery rate decreased from 19.9% during 1985Q1-1994Q4 to 0.97% during 2005Q1-2016Q4. On the contrary, default costs slightly increased, yet the increase
is not statistically significant.

The time-series changes in shareholder recovery rates and default costs are driven jointly by various moving parts. Keeping other parameters constant, the time-series decrease in tax shield benefit rate implies time-series decrease in leverage and time-series decrease in default probability over 1985Q1-2016Q4. Absent the change in tax shield benefit rate, leverage and default probabilities would have actually increased, under which case Figure 5 illustrates how $\alpha$ and $\eta$ are identified. As leverages and default probabilities increase, solid line (locus of $\alpha$ and $\eta$ match a leverage) shifts downward and dashed line (locus of $\alpha$ and $\eta$ that match a default probability) shifts downward. As shown, $\eta$ significantly decreases yet $\alpha$ slightly increases. The time-series increase in volatility of default probabilities and leverages further help to identify significant decrease in $\eta$ and modest increase in $\alpha$.

Finally, this result helps to alleviate a possible concern that my structural estimates might be picking up other alternative economic factors that influence leverage. Those alternative factors are, but certainly not limited to, business cycle variation (Chen, 2010) or agency costs (Morelec, Nikolov and Schurhoff, 2012). However, there is no clear explanation on why either of these alternative factors shows the time-series trend that is shown in Figure 4, and thus puts more weight on my economic story: shareholder recovery rate.

1.6.3. Empirical Proxies for Shareholders’ Recovery Rate

The results so far quantify how much shareholders expect to recover in default for the representative firm. Now, I explore how these values vary over firms with different characteristics. Based on empirical proxies for shareholder recovery rate, discussed below, I construct a subset of firms and estimate model parameters for each subset independently from others. Then, I conduct counterfactual analysis in each subset to quantify economic impacts of a positive shareholder recovery rate.

I first discuss empirical proxies for shareholder recovery rate that I use to construct subsets. Due to lack of guidance on proxies’ validity, the literature uses a wide range of measures.
Unfortunately, in many cases, these empirical measures simultaneously proxy other unob-
servable firm characteristics, and thus its validities can be ambiguous. This subsection
studies two commonly used proxies, firm size and intangible assets. I use total asset (Com-
pustat: AT) to measure firm size. I use two separate measures to proxy intangible assets:
normalized R&D expense (Compustat: XRD/AT) and Intangibility proposed by Peters and
Taylor (2016).

For a given empirical proxy, I form two subsets. In order to make sure that firms do not
move from one subset to the other over time, I perform the following procedure. At each
quarter, I calculate the proxy’s cross-sectional median, and I temporarily allocate a firm
with proxy value greater than the median to High-subset and a firm with proxy value smaller
than the median to Low-subset. This generates a time-series of subsets for a given firm.
Then, I allocate the firm’s entire time-series data to one subset that the firm spends the
most time in.

Across different subsets, I allow tax rates to vary but keep other aggregate variables con-
stant. Table 8’s first panel summarizes tax shield benefit rates and other matching moments.
For example, low-R&D firms’ tax shield benefit rate is much larger than high-R&D firms’
because high-R&D firms have higher expenses.

Next, I structurally estimate model parameters for each subset independently. Table 8’s
middle panel summarizes estimates for default cost ($\alpha$) and shareholder recovery rate ($\eta$)
for different subsets. First, I find that estimates for $\eta$ increase over firm size. This result is
consistent with the literature’s use of $\eta$ as a positive proxy. Citing more frequent occurrences
of a positive shareholder recovery rate\footnote{Please see Weiss (1990), Betker (1995), and Franks and Torous (1994) for more detail.} in larger firms, Garlappi, Shu and Yan (2008)
Garlappi and Yan (2011), and Hackbarth, Haselmann and Schoenherr (2015) use firm size
as a positive proxy for $\eta$. They argue that small firms usually have a higher concentration
of bond ownership. So, close monitoring by concentrated creditors severely decreases $\eta$.

Second, although $\eta$ increases over R&D, the increase is not statistically significantly different
from 0, and thus casts doubt on this literature’s (Garlappi, Shu and Yan, 2008; Garlappi and Yan, 2011; Hackbarth, Haselmann and Schoenherr, 2015)’s use of R&D as a negative proxy for \( \eta \). They use it as a negative proxy because firms with high R&D are more likely to face liquidity shortages (Opler and Titman, 1994) during financial crises, thus are more likely to forgo intangible investment opportunities that shareholders value (Lyandres and Zhdanov, 2013). Firms’ urgent need for liquidity effectively acts as cash-flow-based covenants, and thus high intangibility puts shareholders at a disadvantage \( \text{vis-à-vis} \) creditors and implies low \( \eta \). However, \( \eta \) can increase over R&D because some R&D investments are more valuable under shareholders’ possession, which increases shareholders’ bargaining power \( \text{vis-à-vis} \) creditors. Due to these offsetting forces, R&D might not be a good empirical proxy for \( \eta \) as evident in my estimates. A similar result holds for the other empirical proxy: intangibility.

Third, the results show that both R&D and intangibility are strong positive proxies for \( \alpha \). Consistent with some findings (Reindl, Stoughton and Zechner, 2017), this variable captures how non-transferable a firm’s asset might be in default and thus is positively related to \( \alpha \).

Lastly, interestingly, the estimates imply that firm size is a strong positive proxy for \( \alpha \) and this is inconsistent with some previous findings (Reindl, Stoughton and Zechner, 2017). If the same tax rates were used for both small and big firms, as was done in other studies, then small firms’ \( \alpha \) would have been larger than big firms’ because small firms’ leverage is smaller than big firms’. Yet, as shown in Table 8, big firms face much larger tax shield benefit rates (almost 3 times) than small firms do. Thus, through the lens of my model, it implies that big firms have to face larger \( \alpha \) than small firms do. This finding illustrates the importance of using appropriate tax rates in estimating \( \alpha \) and \( \eta \) for each subset.

Using these estimates, I do a counterfactual analysis for each subset of firms. Allowing \( \eta \) to be positive can have different implications on each subset because different subsets have different \( \alpha \), \( \eta \) and tax shield benefit rates. Table 8’s last panel summarizes such counterfactual analysis results.

[INSERT TABLE 8]
For example, I focus on firms sorted by R&D. $\eta$ for high-R&D firms is larger than that for low-R&D firms and thus allowing $\eta$ to be positive should have larger economic consequences on high-R&D firms. High-R&D firms have larger $\alpha$ and thus reinforces the commitment problem that positive $\eta$ plays. Thus, high-R&D’s leverages and default probabilities decrease significantly more (17.1% and 24.0%, respectively) than low-R&D’s (7.3% and 4.4%, respectively). However, these do not necessarily translate to lower dollar amount of tax shield benefit for high-R&D firms. Because high-R&D firms face a lower tax shield benefit rate than low-R&D firms, high-R&D firms face lower firm value loss than low-R&D firms despite that high-R&D firms reduce their leverage more. Accordingly, the percentage increase in government tax collection is larger for low-R&D firms than it is for high-R&D firms.

1.6.4. Firm Heterogeneity: Monte Carlo Simulations

As emphasized by Glover (2016), cross-sectional distribution of model parameters is important to consider, especially in estimating default cost and shareholder recovery rate. Although I agree with its importance, it is empirically challenging to estimate the cross-sectional distribution. Below, I first discuss how model misspecification at the firm level makes it empirically challenging to estimate the cross-sectional distribution. Then, I illustrate that estimating for the representative firms does not suffer from model misspecification problem. Lastly, I quantify sample-selection bias and validate bias-adjusted estimates.

To quantitatively analyze the aforementioned three points, it is the most ideal to know population cross-sectional distribution of model parameters. Thus, the most suitable way to analyze this is through a Monte Carlo simulation. In order to illustrate how my analysis is sensitive to different data-generating process (DGP), I create two simulated panel data (see Appendix A.2 for details). Both DGPs are calibrated to resemble data on a few aspects, including population cross-sectional mean of model parameters, which are set almost equal to those reported in Table 5. In both simulated data set, firm heterogeneity arises due to heterogeneous model misspecification, heterogeneous model parameters or different realiza-
tions of earnings. The only difference between these two DGPs is as follows. I create the first simulated panel data set by randomly drawing $\alpha$ and $\eta$ from truncated normal PDF. I create the second simulated panel data set by randomly drawing $\alpha$ and $\eta$ from truncated exponential PDF. Panel A in Table 9 shows population cross-sectional mean of shareholder recovery rate for both DGPs.

**Firm-Level Estimation**

Glover (2016) estimates the population default cost to be 45% even though default cost among population conditional on defaults is 25%, and Glover attributed the large discrepancy to sample-selection bias. Even though I qualitatively agree with existence of sample-selection bias, I want to illustrate that Glover’s particular choice of estimation method might have significantly upward biased its estimate of the sample-selection bias.

As long as the estimation procedure cannot distinguish between different sources of heterogeneity, estimated cross-sectional distribution of model parameters are biased by model misspecification. In order to illustrate my point, I study firm-level estimation, used by Glover, which cannot distinguish between different sources of heterogeneity. Due to small time-series data, the law of large numbers cannot help to “fix” the model misspecification problem, and consequently firm-level estimates are biased. Firm-level estimates are upward biased because both model-implied functions for leverage and default probability are convex functions in default cost, $\alpha$, and shareholder recovery rate, $\eta$. Consequently, cross-sectional average of firm-level estimates are significantly upward biased. Panel B.1 in Table 9 reports estimation bias for both set of simulated panel data.

[INSERT TABLE 9]

**Numerical Validation of Structural Estimation**

Similar to Hennessy and Whited (2005, 2007), this paper structurally estimates for the representative firm. This subsection numerically shows that the structural estimation procedure
used in this paper properly uncovers population cross-sectional mean of model parameters. The results are summarized in Panel B.2 in Table 9. As shown, the structural estimation procedure’s estimate biases are small for both simulated panel data, yet not zero due to finite-sample bias. Thanks to the law of large numbers, estimating for the representative firm always yields lower bias than firm-level estimates as long as the model is nonlinear in model parameters.

Sample-Selection Bias and Validation of Estimates

Many earlier papers attempt to estimate $\eta$ and $\alpha$ by examining defaulted firms. Thus, it seems natural to check my estimates against realized counterparts documented in those papers. However, as noted by Glover (2016), sample-selection bias can be large because firms with small $\eta$ and/or small $\alpha$ tend to default more frequently. Thus, the sample average of $\eta$ ($\alpha$) conditional on defaults can be smaller than the unconditional sample average of $\eta$ ($\alpha$).

I quantify sample-selection bias and report results in Panel C in Table 9. Population cross-sectional mean of $\alpha$ and $\eta$ is almost identical for both simulated data. Yet, cross-sectional distributions of $\alpha$ and $\eta$, which follow the normal distribution, have smaller mass on small values than those that follow exponential distribution. Thus, the magnitude of sample-selection bias should be smaller under the first specification. Consistent with the intuition, sample-selection bias is 5.5% for $\alpha$ and 2.3% for $\eta$ using the first simulated data set, whereas sample-selection bias is 8.3% for $\alpha$ and 3.3% for $\eta$ using the second simulated data set. As illustrated, the magnitude of sample-selection bias heavily depends on $\eta$ and $\alpha$’s cross-sectional distribution.

Nonetheless, I use these bias-adjusted numbers to validate my estimates. Among defaulted firms, direct measurement literature estimate shareholder recovery rate after default cost is realized. Thus, the comparable number is the sample average of $\frac{\eta}{1-\alpha}$. The value is 5.4% under the first simulated data and 4.2% under the second simulated data. Both numbers are in line with empirically observed counterpart between 0.4% and 7.6%. Bias-adjusted
default cost is 11.9% under the first simulated data and 8.4% under the second simulated data. Both numbers are in line with the empirically observed counterpart between 10% and 20%. This external validation exercise strongly validates my results. Lastly, I check if the parameter estimates imply a reasonable value for creditors’ recovery rate. According to Moody’s Ultimate Recovery Database announcement in April 2017, the median recovery for corporate bonds was 36% between 1987 and 2016. The model counterparts\textsuperscript{16} are 25.4% using the first simulated data and 27.9% using the second simulated data. The discrepancy in creditors’ recovery rate could arise due to different sample period.

1.6.5. Uncertainty in Shareholders’ Recovery Rate

In this subsection, I study how uncertainty in shareholder recovery rate could impact firms’ optimal leverage ex-ante.

I assume that shareholder recovery rate, \( \eta \), is drawn once, immediately after firms decide on its initial leverage. One reasonable conjecture is that uncertainty in \( \eta \) decreases the commitment problem played by positive \( \eta \) because its power seems to have subsided due to its uncertainty. However, the opposite can happen for the following reason. Expected firm-value loss is a convex function in \( \eta \) (because \( X_{DC} \) is proportional to \( \frac{1}{1-\eta} \) as shown in Equation (1.5)). Thus, the average of high \( \eta \)'s optimal leverage and low \( \eta \)'s is smaller than medium \( \eta \)'s leverage. On a related note, as long as \( \eta = 1 \) event happens with some positive probability, expected firm-value loss becomes infinity (again, because \( X_{DC} \) is proportional to \( \frac{1}{1-\eta} \) as shown in Equation (1.5)) and firms optimally choose zero leverage ex-ante. Thus, introducing uncertainty in \( \eta \) can allow us to match the empirically observed leverage even with lower magnitude of \( \eta \).

What does this mean for my \( \eta \) estimate? If uncertainty in \( \eta \) truly exists in the real world, \textsuperscript{16}I define creditors’ recovery rate as

\[
\frac{(1 - \alpha' - \eta')(D(X_D) + E(X_D))}{D(X_0)}
\]

where the numerator represents the creditor’s realized recovery value and the denominator represents what creditors are owed.

34
because the current model does not account for uncertainty, $\eta$ reported in Table 5 is an upper bound for the population cross-sectional mean of $\eta$. Although this is a very interesting extension, quantitative analysis of the role of uncertainty in $\eta$ is beyond the scope of this paper and I will leave this for later study.

1.7. Conclusion

According to the absolute priority rule (APR), shareholders should recover nothing in default unless creditors are paid in full. However, in practice, shareholders do receive a positive payoff in default even if creditors are not paid in full. In this paper, I develop a dynamic tradeoff model to examine the importance of a positive shareholder recovery rate. Consistent with existing empirical findings, I allow default to be costly even when shareholders recover a positive amount as a renegotiation outcome with creditors. A positive recovery makes shareholders more willing to default, which increases borrowing costs. In response, firms optimally reduce leverage ex-ante. This channel helps to match distributions of both leverage and default probability.

A structural estimation of the model yields a default cost of 17.3% and a shareholder recovery rate of 7.1%. Counterfactual analysis reveals that allowing a positive shareholder recovery rate decreases the leverage by 9.4%. Consequently, default probabilities decrease by 8.1%, market betas decrease by 20.5% and credit spreads decrease by 1.7%. As firms lose the tax shield benefit, firm values decrease by 1.5% and government tax revenue increases by 9.9%. Lastly, lower default probability, driven by a positive shareholder recovery, implies less frequent realization of deadweight cost. Thus, the sum of firm values and government tax revenue increases by 0.9%. Even though this paper does not do complete welfare analysis, these findings can still be used to shed some light on an important bankruptcy policy question as this paper highlights its consequences. Additionally, subperiod analysis reveals that shareholder recovery rate increased immediately after the Bankruptcy Reform Act was passed in 1978, and a shareholder recovery rate has steadily decreased ever since. Consistent with the empirical literature, my subset estimates show that firm size is a good
positive proxy for shareholder recovery rate.
Table 1: Summary Statistics

This table reports the summary statistics for my sample of firm data. The sample contains 246,090 firm-quarter observations, number of unique firms is 4,435 and spans from 1970Q1 to 2016Q4. Market beta is calculated based on a rolling window of 24 months of monthly returns. Earnings growth is \( \tilde{e}_{t,t+1} = \log\left(\frac{\sum_{j=0}^{K} OIADPQ_{t,t+1-j}}{\sum_{j=0}^{K} OIADPQ_{t,t-j}} - 1\right) \) where \( K \) is set to 8 and \( OIADPQ \) is operating income after depreciation. Default probabilities are constructed by following distance-to-default model (Bharath and Shumway, 2008). Leverage is defined as

\[
\frac{DLTTQ+DLCQ}{DLTTQ+DLCQ+ME}
\]

where \( DLTTQ, DLCQ \) and \( ME \) are long-term debt, short-term debt and market equity, respectively.

<table>
<thead>
<tr>
<th></th>
<th>Market beta</th>
<th>Earnings Growth</th>
<th>Default Probability</th>
<th>Leverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1.105</td>
<td>0.0060</td>
<td>0.0027</td>
<td>0.283</td>
</tr>
<tr>
<td>Median</td>
<td>1.050</td>
<td>0.0144</td>
<td>0.0000</td>
<td>0.231</td>
</tr>
<tr>
<td>Standard dev</td>
<td>0.858</td>
<td>0.3018</td>
<td>0.0437</td>
<td>0.225</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.763</td>
<td>-0.9793</td>
<td>19.2656</td>
<td>0.815</td>
</tr>
<tr>
<td>Minimum</td>
<td>-12.771</td>
<td>-8.3713</td>
<td>0.0000</td>
<td>0.000</td>
</tr>
<tr>
<td>Max</td>
<td>15.278</td>
<td>8.0986</td>
<td>1.0000</td>
<td>0.999</td>
</tr>
<tr>
<td>Number of obs</td>
<td>246,090</td>
<td>246,090</td>
<td>246,090</td>
<td>246,090</td>
</tr>
</tbody>
</table>
Table 2: Aggregate Parameters Values

This table reports values used for aggregate parameters and their data sources. Quarterly aggregate earnings growth volatility ($\sigma_A$) and quarterly real risk-free rate ($r$) are calibrated based on the sample period from 1947Q1 through 2016Q4. A value for market Sharpe ratio is obtained from Chen (2010) and a value for proportional debt issuance cost is obtained from Altinkilic and Hansen (2000).

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_A$ Quarterly aggr earnings growth vol</td>
<td>0.052</td>
<td>NIPA</td>
</tr>
<tr>
<td>$r$ Quarterly real risk-free rate</td>
<td>0.0016</td>
<td>FRED</td>
</tr>
<tr>
<td>$\varphi_A$ Quarterly market Sharpe ratio</td>
<td>0.165</td>
<td>Chen (2010)</td>
</tr>
<tr>
<td>$\phi_D$ Prop’ debt issuance cost</td>
<td>0.015</td>
<td>Altinkilic and Hansen (2000)</td>
</tr>
</tbody>
</table>
Table 3: Sensitivity of Moments to Parameters

This table shows the local sensitivity of model-implied moments (in columns) with respect to model parameters (in rows). To make the sensitivities comparable across parameters and moments, the sensitivities are normalized as \( \frac{\partial \text{moment}}{\partial \text{parameter}} \times \frac{\text{parameter's standard error}}{\text{moments' standard error}} \). Local sensitivities are calculated around the estimates that are reported in Table 5. From left to right, moments are mean of market beta \((E(\beta))\), variance of earnings growth \((\text{var}(EG))\), mean of leverage \((E(Lev))\), variance of leverage \((\text{var}(Lev))\), skewness of leverage \((\text{skew}(Lev))\), mean of default probability \((E(DP))\), variance of default probability \((\text{var}(DP))\) and skewness of default probability \((\text{skew}(DP))\). Parameter definitions are as follows. \(b\) is earnings growth beta, \(\sigma^F\) is volatility of firm-specific earnings growth shock, \(\eta\) is shareholder recovery rate and \(\alpha\) is default cost.

<table>
<thead>
<tr>
<th></th>
<th>(E(\beta))</th>
<th>(\text{var}(EG))</th>
<th>(E(Lev))</th>
<th>(\text{var}(Lev))</th>
<th>(\text{skew}(Lev))</th>
<th>(E(DP))</th>
<th>(\text{var}(DP))</th>
<th>(\text{skew}(DP))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(b)</td>
<td>0.837</td>
<td>0.015</td>
<td>0.275</td>
<td>0.312</td>
<td>-0.204</td>
<td>0.235</td>
<td>0.202</td>
<td>-0.088</td>
</tr>
<tr>
<td>(\sigma^F)</td>
<td>0.003</td>
<td>0.215</td>
<td>-0.006</td>
<td>0.280</td>
<td>0.154</td>
<td>0.280</td>
<td>0.251</td>
<td>-0.107</td>
</tr>
<tr>
<td>(\eta)</td>
<td>-0.208</td>
<td>0.000</td>
<td>-0.317</td>
<td>-0.655</td>
<td>0.027</td>
<td>-0.081</td>
<td>-0.072</td>
<td>0.030</td>
</tr>
<tr>
<td>(\alpha)</td>
<td>-0.276</td>
<td>0.001</td>
<td>-0.603</td>
<td>-0.733</td>
<td>0.465</td>
<td>-0.530</td>
<td>-0.452</td>
<td>0.196</td>
</tr>
</tbody>
</table>
Table 4: Model Fit

This table shows how well the model fits the eight moments targeted in the SMM estimation. The first and the second rows show data moments and standard errors respectively. The third row shows model-implied moments. The last two rows show the difference between data and model-implied moments and t-statistics. From left to right, moments are mean of market beta ($E(\beta)$), variance of earnings growth ($\text{var}(EG)$), mean of leverage ($E(Lev)$), variance of leverage ($\text{var}(Lev)$), skewness of leverage ($\text{skew}(Lev)$), mean of default probability ($E(DP)$), variance of default probability ($\text{var}(DP)$) and skewness of default probability ($\text{skew}(DP)$).

<table>
<thead>
<tr>
<th></th>
<th>$E(\beta)$</th>
<th>$\text{var}(EG)$</th>
<th>$E(Lev)$</th>
<th>$\text{var}(Lev)$</th>
<th>$\text{skew}(Lev)$</th>
<th>$E(DP)$</th>
<th>$\text{var}(DP)$</th>
<th>$\text{skew}(DP)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>1.105</td>
<td>0.078</td>
<td>0.283</td>
<td>0.019</td>
<td>0.451</td>
<td>0.003</td>
<td>0.002</td>
<td>16.869</td>
</tr>
<tr>
<td>Std. Err.</td>
<td>(0.036)</td>
<td>(0.010)</td>
<td>(0.012)</td>
<td>(0.003)</td>
<td>(0.155)</td>
<td>(0.001)</td>
<td>(0.000)</td>
<td>(3.960)</td>
</tr>
<tr>
<td>Model</td>
<td>1.100</td>
<td>0.060</td>
<td>0.283</td>
<td>0.018</td>
<td>0.736</td>
<td>0.004</td>
<td>0.002</td>
<td>16.068</td>
</tr>
<tr>
<td>Difference</td>
<td>0.005</td>
<td>0.018</td>
<td>-0.001</td>
<td>0.001</td>
<td>-0.285</td>
<td>-0.001</td>
<td>0.000</td>
<td>0.801</td>
</tr>
<tr>
<td>t-stat</td>
<td>0.151</td>
<td>1.786</td>
<td>-0.041</td>
<td>0.251</td>
<td>-1.839</td>
<td>-1.662</td>
<td>-0.689</td>
<td>0.202</td>
</tr>
</tbody>
</table>

Table 5: Parameter Estimates

This table reports the model’s parameter estimates from the simulated method of moments (SMM). Here, I cluster by industries to account for apparent correlation between firms in the same industry. I use 17 industry definitions from Kenneth French’s website. This clustering strategy also accounts for time-series autocorrelation within firms. This is more conservative than clustering by firms. Parameter definitions are as follows. $b$ is earnings growth beta, $\sigma^F$ is volatility of firm-specific earnings growth shock, $\eta$ is shareholder recovery rate and $\alpha$ is default cost.

<table>
<thead>
<tr>
<th></th>
<th>$b$</th>
<th>$\sigma^F$</th>
<th>$\eta$</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>0.816</td>
<td>0.244</td>
<td>0.071</td>
<td>0.173</td>
</tr>
<tr>
<td>Standard Error</td>
<td>(0.019)</td>
<td>(0.002)</td>
<td>(0.009)</td>
<td>(0.010)</td>
</tr>
</tbody>
</table>
Table 6: The Effect of Positive Shareholders’ Recovery Rate

This table illustrates the effect of a positive shareholder recovery rate. The first column reports values for the counterfactual world where shareholders are expected to recover nothing in default. The second column allows shareholders to recover non-zero amount in default yet forces firms to keep the same optimal policies (coupon, $C$, and refinancing point, $X_U$) as under the first column. This exercise helps to quantify how much expected firm-value loss increases, conditional on the firms’ optimal policies. The third column summarizes the data-matched world where shareholders are expected to recover 7.1% in default. The first three rows are coupon, default threshold and upward refinancing boundary, all scaled by initial earnings level. The fourth row shows panel-wide average of leverage. The fifth row shows panel-wide average of default probabilities. The sixth row shows panel-wide average of market betas. The last row shows panel-wide average of credit spreads.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>η = 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>η = 7.1%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>η = 7.1%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coupon ($C$)</td>
<td>1.145</td>
<td>1.145</td>
<td>1.006</td>
</tr>
<tr>
<td>Upward Refinancing Point ($X_U$)</td>
<td>3.882</td>
<td>3.882</td>
<td>3.918</td>
</tr>
<tr>
<td>Default Threshold ($X_D$)</td>
<td>0.077</td>
<td>0.083</td>
<td>0.073</td>
</tr>
<tr>
<td>Leverage</td>
<td>0.314</td>
<td>0.306</td>
<td>0.285</td>
</tr>
<tr>
<td>Default Probability (%)</td>
<td>0.388</td>
<td>0.436</td>
<td>0.356</td>
</tr>
<tr>
<td>Market-beta</td>
<td>1.386</td>
<td>1.132</td>
<td>1.102</td>
</tr>
<tr>
<td>Credit Spread (BP)</td>
<td>194</td>
<td>201</td>
<td>191</td>
</tr>
</tbody>
</table>
Table 7: Robustness — Bankruptcy Reform Act of 1978

This table reports the model’s parameter estimates from the simulated method of moments (SMM). These estimates help to test how the Bankruptcy Reform Act of 1978 changed shareholder recovery rate ($\eta$) and default cost ($\alpha$). Similarly to the baseline parameter estimates, I cluster by industries to account for apparent correlation between firms in the same industry. I use 17 industry definitions from Kenneth French’s website. This clustering strategy also accounts for time-series autocorrelation within firms. This is more conservative than clustering by firms. Parameter definitions are as follows. $b$ is earnings growth beta, $\sigma^F$ is volatility of firm-specific earnings growth shock, $\eta^{pre}$ is $\eta$ for the pre-event subperiod (1975Q1-1978Q3) and $\eta^{post}$ is $\eta$ for the post-event subperiod (1981Q2-1984Q4), $\alpha^{pre}$ is $\alpha$ for the pre-event subperiod (1975Q1-1978Q3) and $\alpha^{post}$ is $\alpha$ for the post-event subperiod (1981Q2-1984Q4).

<table>
<thead>
<tr>
<th></th>
<th>$b$</th>
<th>$\sigma^F$</th>
<th>$\eta^{pre}$</th>
<th>$\eta^{post}$</th>
<th>$\alpha^{pre}$</th>
<th>$\alpha^{post}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>0.812</td>
<td>0.220</td>
<td>0.001</td>
<td>0.290</td>
<td>0.190</td>
<td>0.210</td>
</tr>
<tr>
<td>Standard Error</td>
<td>(0.012)</td>
<td>(0.002)</td>
<td>(0.019)</td>
<td>(0.017)</td>
<td>(0.008)</td>
<td>(0.032)</td>
</tr>
</tbody>
</table>
Table 8: Robustness — Subset Based on Empirical Proxies for Shareholders’ Recovery Rate

This table reports results for subset analysis. For each proxy (size, R&D, or intangibility), I form two subsets. At each quarter, I calculate the proxy’s cross-sectional median, and I temporarily allocate a firm with proxy value greater than the median to High-subset and a firm with proxy value smaller than the median to Low-subset. This generates a time-series of subsets for a given firm. Then, I allocate the firm’s entire time-series data to one subset that the firm spends the most time in. The first panel reports summary statistics for each subset. The second panel reports estimates for shareholder recovery rate ($\eta$) and default cost ($\alpha$) and standard errors (in parentheses) for each subset. The last panel quantifies economic consequences of allowing shareholders to recover a positive amount in default. It reports percent changes on leverage, default probability, firm value and government tax revenue.

<table>
<thead>
<tr>
<th></th>
<th>Size</th>
<th>R&amp;D</th>
<th>Intangibility</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Small</td>
<td>Big</td>
<td>Low</td>
</tr>
<tr>
<td>Summary Statistics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tax Shield Benefit Rate</td>
<td>0.048</td>
<td>0.132</td>
<td>0.112</td>
</tr>
<tr>
<td>Leverage</td>
<td>0.273</td>
<td>0.287</td>
<td>0.240</td>
</tr>
<tr>
<td>Default Prob</td>
<td>0.201</td>
<td>0.291</td>
<td>0.142</td>
</tr>
<tr>
<td>Estimates</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shareholders’ Recovery Rate ($\eta$)</td>
<td>0.015</td>
<td>0.122</td>
<td>0.071</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.009)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>Default Cost ($\alpha$)</td>
<td>0.105</td>
<td>0.242</td>
<td>0.131</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.028)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>Economic Consequences: $\eta = 0 \Rightarrow \eta \neq 0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$%\Delta$ Leverage</td>
<td>−2.6</td>
<td>−14.6</td>
<td>−7.3</td>
</tr>
<tr>
<td>$%\Delta$ Default Prob</td>
<td>−4.3</td>
<td>−12.2</td>
<td>−4.4</td>
</tr>
<tr>
<td>$%\Delta$ Firm Value</td>
<td>−0.1</td>
<td>−4.0</td>
<td>−2.0</td>
</tr>
<tr>
<td>$%\Delta$ Tax Revenue</td>
<td>0.6</td>
<td>79</td>
<td>146.4</td>
</tr>
</tbody>
</table>
Table 9: Robustness — Firm Heterogeneity

This table illustrates how firm-level heterogeneity can impact structural estimates. For illustration, I simulate two sets of panel data sets that feature firm heterogeneity due to variety of sources: heterogeneous model parameters ($\alpha$, $\eta$), heterogeneous model misspecification and idiosyncratic shocks. The first simulated data set assumes truncated normal distribution for model parameters and the table’s first column summarizes such results. The second simulated data set assumes truncated exponential distribution for cross-sectional distribution of model parameters and the table’s second column summarizes such results (see Appendix Section A.2 for more detail). Panel A summarizes population cross-sectional mean of heterogeneous model parameters. Panel B reports estimates for population cross-sectional mean using different structural estimation procedures. Panel C reports conditional mean in default, and quantifies sample-selection bias.

<table>
<thead>
<tr>
<th></th>
<th>Truncated Normal</th>
<th>Truncated Exponential</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Panel A. Population Cross-sectional Mean of Heterogeneous Model Parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_i$</td>
<td>17.4%</td>
<td>17.0%</td>
</tr>
<tr>
<td>$\eta_i$</td>
<td>7.1%</td>
<td>7.0%</td>
</tr>
<tr>
<td><strong>Panel B. Estimates Using Different Estimation Procedures</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>estimation bias</td>
<td>estimation bias</td>
</tr>
<tr>
<td><strong>B.1 Firm-Level Structural Estimation (Glover (2016))</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>19.6%</td>
<td>2.2%</td>
</tr>
<tr>
<td>$\eta$</td>
<td>21.1%</td>
<td>14.0%</td>
</tr>
<tr>
<td><strong>B.2 Structural Estimation Used in the Current Paper</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>17.7%</td>
<td>0.3%</td>
</tr>
<tr>
<td>$\eta$</td>
<td>7.4%</td>
<td>0.3%</td>
</tr>
<tr>
<td><strong>Panel C. Conditional Mean Upon Default</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>selection bias</td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>11.9%</td>
<td>5.5%</td>
</tr>
<tr>
<td>$\eta$</td>
<td>4.8%</td>
<td>2.3%</td>
</tr>
</tbody>
</table>
Figure 1: Illustration of the Trade-off Theory

This figure illustrates the Trade-off theory. Solid line illustrates the Trade-off theory when $\eta = 0$. Dashed line illustrates the Trade-off theory when $\eta \neq 0$. $A \rightarrow B$ illustrates an economic intuition that the firm value decreases due to shareholders’ strategic default action. $B \rightarrow C$ illustrates that firms actively de-lever to optimize firm value minus debt issuance cost and illustrates a commitment problem.
This figure illustrates how \( \eta \) and \( \alpha \) are separately identified. Solid line is the locus of \( \alpha \) and \( \eta \) that match a given leverage. Dashed line is the locus of \( \alpha \) and \( \eta \) that match a given default probability.
This figure illustrates how shareholder recovery rate ($\eta$) impacts time-series variation of leverage and default probabilities. Two firms have the same model parameters except for $\eta$: blue firm’s $\eta$ is zero, whereas red firm’s $\eta$ is non-zero. Both firms face the same sequence of earnings that are shown in the top panel. Middle panel shows time-series of leverage for both firms. Blue firm’s leverage is larger than red firm’s on average. Moreover, blue firm upward refinances earlier at time 84 than red firm does. Consequently, blue firm’s leverage is more volatile and more positively skewed. Similar patterns are observed in the sequence of default probabilities (bottom panel).
This figure illustrates how structural estimates for shareholder recovery rate (upper panel) and default cost (lower panel) change over time. Red dashed line illustrates estimates for the entire sample: 1970Q1-2016Q4: $\hat{\eta} = 7.1\% (0.9\%)$ and $\hat{\alpha} = 17.3\% (1.0\%)$. Black solid line (estimates) along with gray region (95% confidence interval) illustrate estimates for each subperiod. $\hat{\eta} = 0.1\% (1.9\%)$ and $\hat{\alpha} = 19.0\% (0.8\%)$ for 1975Q1-1978Q3. $\hat{\eta} = 29.0\% (1.7\%)$ and $\hat{\alpha} = 21.0\% (3.2\%)$ for 1981Q2-1984Q4. $\hat{\eta} = 19.9\% (1.1\%)$ and $\hat{\alpha} = 14.0\% (1.3\%)$ for 1985Q1-1994Q4. $\hat{\eta} = 3.8\% (1.0\%)$ and $\hat{\alpha} = 15.3\% (1.1\%)$ for 1995Q1-2004Q4. Lastly, $\hat{\eta} = 0.97\% (0.3\%)$ and $\hat{\alpha} = 17.1\% (0.8\%)$ for 2005Q1-2016Q4.
This figure illustrates how time-series change in leverage and default probabilities help to identify time-series change in $\eta$ and $\alpha$. As it moves from 1985Q1-1994Q4 to 2005Q1-2016Q4, solid line (locus of $\alpha$ and $\eta$ that match a leverage) shifts downward and dashed line (locus of $\alpha$ and $\eta$ that match a default probability) shifts downward. Intersection of thick solid line and thick dashed line is a structural estimate for 1985Q1-1994Q4. Intersection of thin solid line and think dashed line is a structural estimate for 2005Q1-2016Q4. As shown, as it moves from 1985Q1-1994Q4 to 2005Q1-2016Q4, $\eta$ significantly decreases yet $\alpha$ slightly increases.
2.1. 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
2.2. Data Generation and Assumptions

We assume that the data is generated as follows.

**Assumption A$_1$**

**Data Generation:** (i) Let \((X', W, Y')'\) be a random vector with a finite variance. (ii) Let a structural system generate the random variables \(\eta, \varepsilon, U, X, W,\) and \(Y\) such that

\[
Y' = X'\beta + U\delta + \eta' \quad \text{and} \quad W = U + \varepsilon \tag{2.1}
\]

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

A$_1$ decomposes the proxy \(W\) into the “signal” component\(^1\) \(U\) and the “noise” or error \(\varepsilon\).

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

\[
Y_j = X'\beta_j + U\delta_j + \eta_j. \tag{2.2}
\]

\(X\) 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 \(\text{Cov}[\eta', \varepsilon', X] = 0\), 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 2.3.1). We do not require such exclusion restrictions. Here, the challenge in identifying \((\delta, \beta)\) is due to \(U\) being unobserved and possibly correlated with \(X\). In particular, we maintain two

\(^1\)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 \(U \equiv V\psi\).
standard assumptions about the other unobservables $\eta$ and $\varepsilon$. First, the “disturbance” $\eta$ is uncorrelated with $(X',U)'$.

**Assumption A$_2$ 1 Uncorrelated Disturbance:** $\text{Cov} [\eta, (X',U)'] = 0$.

Second, the measurement error $\varepsilon$ is uncorrelated with $(X',U,\eta)$.

**Assumption A$_3$ 1 Uncorrelated measurement error:** $\text{Cov} [\varepsilon, (X',U,\eta)'] = 0$.

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$^2$ $(\beta, \delta)$. We do not require these stronger assumptions$^3$.

Instead, we impose the uncorrelation assumptions A$_2$-A$_3$ and study partially identifying $\delta$ and $\beta$. DiTraglia and Garcia-Jimeno (2017) relax A$_2$ to allow $X$ (or its instrument) to be endogenous and, similarly to this paper’s joint equation analysis, they advocate analyzing 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 $\delta$ or of a component of $\beta$. Klepper and Leamer (1984) and Bollinger (2003) characterize the sharp identification regions for $\delta$ and $\beta$ under A$_1$-A$_3$. Chalak and Kim (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

$^2$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.
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.

2.2.1. Notation

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

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

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

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

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

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

the population coefficient of determination (R-squared) from a regression\(^4\) of $A$ on $B$.

---

\(^4\)If $\sigma^2_B$ is singular, we set $R^2_{A,B} = R^2_{A,B_o}$ where $B_o$ is a maximal linearly independent subset of $B$. Further, if either $\sigma^2_A = 0$ or $\sigma^2_B = 0$ then we set $r_{A,B} = 0$ and $R^2_{A,B} = 0$. 

2.2.2. Linear Projection

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

$$b_{Y,X} = \beta + b_{W,X}\delta.$$  \hspace{1cm} (2.3)

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:

$$\tilde{Y}' = \tilde{U}\delta + \tilde{\eta}' \hspace{1cm} \text{and} \hspace{1cm} \tilde{W} = \tilde{U} + \tilde{\varepsilon}$$  \hspace{1cm} (2.4)

to study identifying $\delta$. The identification region for $\beta$ then obtains using equation (2.3).

2.2.3. 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 $\sigma_{\varepsilon}^2 = 0$ by imposing an upper bound $\kappa$ on the net-of-$X$ “noise to signal ratio.”

**Assumption A$_4$ 1** Bounded Net-of-X Noise to Signal Ratio: $\sigma_{\varepsilon}^2 \leq \kappa \sigma_U^2$ where $0 \leq \kappa$.

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 $U$, $\sigma_{\varepsilon}^2 \leq \sigma_U^2$. Given $A_1$-$A_3$, $A_4$ equivalently
imposes a lower bound \( \frac{1}{1+\kappa} \) on \( \rho \), the net-of-X “signal to total variance ratio”:

\[
\frac{1}{1+\kappa} \leq \rho \equiv \frac{\sigma^2_\theta}{\sigma^2_\varepsilon + \sigma^2_\varepsilon}.
\]

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

Consider the coefficient of determination \( R^2_{Y_j,\tilde{U}} \equiv 1 - \frac{\sigma^2_{\eta_j}}{\sigma^2_{\tilde{Y}_j}} \) in the \( \tilde{Y}_j \) equation from display (2.4). By \( A_1-A_3 \) and since \( W \) measures \( U \) with error, Lemma B.4.1 in the Online Appendix gives that \( R^2_{Y_j,\tilde{W}} \leq R^2_{Y_j,\tilde{U}} \). The second auxiliary assumption controls the fit of the model by imposing a bound \( \tau_j \) on how large can \( R^2_{Y_j,\tilde{U}} \) be.

**Assumption A_5 1** Bounded Net-of-X Coefficient of Determination: \( R^2_{Y_j,\tilde{U}} \leq \tau_j \) where \( 0 < \tau_j \) and \( R^2_{Y_j,\tilde{W}} \leq \tau_j \leq 1 \) for \( j = 1, \ldots, p \).

Since \( R^2_{A,(X',B)'} = \frac{\sigma^2_A}{\sigma^2_A} (R^2_{A,B} - 1) + 1 \), \( A_5 \) equivalently imposes an upper bound \( \tau^*_j \equiv \frac{\sigma^2_{\eta_j}}{\sigma^2_{\tilde{Y}_j}} (\tau_j - 1) + 1 \) on the coefficient of determination \( R^2_{Y_j,(X',U)'} \equiv 1 - \frac{\sigma^2_{\eta_j}}{\sigma^2_{\tilde{Y}_j}} \) in the \( Y_j \) equation from display (2.2). We let \( \tau \equiv (\tau_1, \ldots, \tau_p)' \) and \( \tau^* \equiv (\tau^*_1, \ldots, \tau^*_p)' \).

The third auxiliary assumption weakens the assumption that \( \sigma^2_\eta \) is diagonal by specifying the sign of the correlation \( r_{\eta_j,\eta_h} \) among the cross-equation disturbances, if at all.

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

\( A_6 \) encodes the sign restrictions (if any) imposed in \( A_6 \) on the \( \frac{1}{2}p(p-1) \) off-diagonal elements of \( \sigma^2_\eta \). For example, \( (\xi_{jh}, \xi_{jh}) = (-1, 0) \) encodes that \( r_{\eta_j,\eta_h} \leq 0 \) whereas if \( A_6 \) does not restrict the sign of \( r_{\eta_j,\eta_h} \) then we set \( (\xi_{jh}, \xi_{jh}) = (-1, 1) \). We collect these restrictions in the matrix \( \mathbf{c} = (\xi, \overline{\xi}) \) where \( \xi = (\xi_{12}, \ldots, \xi_{(p-1)p})' \) and \( \overline{\xi} = (\overline{\xi}_{12}, \ldots, \overline{\xi}_{(p-1)p})' \). For example, when \( \sigma^2_\eta \) is assumed to be diagonal, we set \( \mathbf{c} = 0 \).
Online Appendix A extends A₆ to A'₆, which sets \( c_{jh} \leq r_{\eta_j, \eta_h} \leq \tau_{jh} \) with \(-1 \leq c_{jh} \leq \tau_{jh} \leq 1\).

In particular, A'₆ may restrict the sign and/or magnitude of the correlation \( r_{\eta_j, \eta_h} \). While A'₆ is conceptually similar to A₆, the expression for the identification region under A₁-A₆ is more complex. To ease the exposition, we report these results in the Online Appendix.

Here and in the empirical analysis in Section 6, we focus on specifying the sign of \( r_{\eta_j, \eta_h} \), if at all, which can be more salient in empirical work and is sometimes more easily inferred from economic theory.

As we show in Section 3, whereas A₄ directly restricts the net-of-X signal to total variance ratio \( \rho \) (i.e. the extent of the measurement error), A₅ and A₆ indirectly restrict \( \rho \). We vary \( \kappa, \tau, \) and \( c \) in A₄-A₆ to conduct a sensitivity analysis that weakens the no measurement error assumption \( \kappa = 0 \) (or \( R_{Y_j,W}^2 = \tau_j \) in A₃), controls the fit of the model (\( R_{Y_j,W}^2 \leq \tau_j \)), and weakens the assumption that \( \sigma_\eta^2 \) is diagonal (\( c = 0 \)). Conversely, we study for what configuration of \( (\kappa, \tau, c) \) does the identification region admit a plausible value or range e.g. for a component of \( \delta \) or \( \beta \). To keep the exposition concise, we impose A₄-A₆ throughout and obtain the results when A₄, A₅, or A₆ is not binding as a special case in which \( \kappa \to +\infty \), \( \tau = (1, \ldots, 1)' \), or \( c \) is such that \( (c_{jh}, \tau_{jh}) = (-1, 1) \) for all \( j < h \).

2.3. Identification

We study identifying \( \delta \), and consequently \( \beta = b_{Y,X} - b_{W,X} \delta \), under A₁-A₃ 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 A₄-A₆ on the identification regions for \( \delta \) and \( \beta \).

2.3.1. Characterization Theorem

From Theorem 2.3.1, under A₁-A₃, the moments in \( Var[(\bar{Y}', \bar{W})'] \) can be expressed as

\[
\sigma_W^2 = \sigma_U^2 + \sigma_\varepsilon^2, \quad \sigma_{W, \bar{Y}} = \sigma_{W, \bar{U}} \delta = \sigma_U^2 \delta, \quad \text{and } \sigma_Y^2 = \delta' \sigma_U^2 \delta + \sigma_\eta^2.
\]
Dividing $\sigma_{W,Y}$ by $\sigma^2_W$, we obtain that

$$b_{\bar{Y},\bar{W}} = \rho \delta$$

where $\rho \equiv \frac{\sigma^2_U}{\sigma^2_W} = \frac{\sigma^2_U}{\sigma^2_W + \sigma^2_\varepsilon}$. \hfill (2.5)

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

$$\sigma^{-2}_W \sigma^2_Y = \delta' \rho \delta + \sigma^{-2}_W \sigma^2_\eta,$$

where we have that

$$\Gamma \equiv \sigma^{-2}_W \sigma^2_\eta$$

is positive semi-definite (denoted by $0 \preceq \Gamma$). \hfill (2.7)

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

$$\Gamma_{jh} \equiv \sigma^{-2}_W \sigma_{\eta_j,\eta_h} = \sigma^{-2}_W \sigma_{\bar{Y}_j,\bar{Y}_h} - \delta_j \rho \delta_h.$$ \hfill (2.8)

As we show in Corollary 2.3.2, the system of (in)equalities (2.3,2.5,2.6,2.7) 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 $\frac{\sigma^2_Y}{\sigma^2_W} (1 - \tau_j) \leq \Gamma_{jj}$, and $A_6$ may specify the (weak) sign of $\Gamma_{jh}$.

When $U$ and $X$ are not perfectly collinear, i.e. $\rho \neq 0$, Theorem 2.3.1 uses equations (2.3,2.5,2.6) to express $\delta$, $\beta$, and $\Gamma$ as functions $D$, $B$, and $G$ of $\rho$. This mapping enables characterizing the identification region for $(\rho, \delta, \beta, \Gamma)$ in terms of restrictions on $\rho$ only and facilitates a sensitivity analysis that studies the consequences of deviating from the “no measurement error” assumption $\rho = 1$. 

59
Theorem 2.3.1 Assume $A_1$-$A_3$ and let $\text{Var}[X', U']$ be nonsingular so that $0 < \rho$. Then

$$
\delta = D(\rho) \equiv \frac{1}{\rho} b_{Y_j \tilde{W}}', \\
\beta = B(\rho) \equiv b_{Y_i X} - b_{W_i X} \frac{1}{\rho} b_{Y_j \tilde{W}}', \text{ and} \\
\Gamma = G(\rho) \equiv \sigma_{\tilde{W}}^2 \sigma_Y^2 - b_{Y_j \tilde{W}}' \frac{1}{\rho} b_{Y_j \tilde{W}}'.
$$

Theorem 2.3.1 reveals how if there is no measurement error ($\rho = 1$) then $(\delta, \beta, \Gamma)$ is point identified. Further, even when $\rho < 1$, $b_{Y_j \tilde{W}} = 0$ if and only if $(\delta_j, \beta_j, \Gamma_{jh}) = (0, b_{Y_i X}, \sigma_{\tilde{W}}^2 \sigma_Y^2)$. Similarly, if the $l^{th}$ element $b_{W_i X, l}$ of $b_{W_i X}$ is 0 then $\beta_l = b_{Y_i X, l}$.

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

2.3.2. 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 $(k, \tau, c)$ value). Corollary 2.3.2 states the general result. We then discuss several special cases.

Corollary 2.3.2 Under the conditions of Theorem 2.3.1 and $A_4$-$A_6$, for $j, h = 1, ..., p$ with $j < h$, $(\rho, \delta, \beta, \Gamma)$ is partially identified in the sharp set

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

$$
\epsilon_{jh} \leq \text{sgn}(G_{jh}(r)) \leq \overline{\epsilon}_{jh} \text{ for } j, h = 1, ..., p \text{ and } j < h \}. \tag{5}
$$

For $a \in \mathbb{R}$, define the sign function: $\text{sgn}(a) = -1$ if $a < 0$, $\text{sgn}(a) = 0$ if $a = 0$, and $\text{sgn}(a) = 1$ if $a > 0$.\(^6\)
Further, $\rho, \delta, \beta,$ and $\Gamma$ are partially identified in the sharp sets

$$\mathcal{R}^{k,\tau,c} = [R^2_{W,Y}, 1] \cap \left[ \frac{1}{1 + \kappa}, 1 \right] \cap_{j=1}^{p} \left( \frac{1}{r_j} R^2_{W,Y_j}, 1 \right) \bigcap_{j,h=1 \atop j < h}^{p} \mathcal{R}^{c}_{jh},$$

where

$$\mathcal{R}^{c}_{jh} = \begin{cases} 
\frac{b_{\tilde{y}_j, \tilde{w}} b_{\tilde{y}_h, \tilde{w}}}{a^\tau_{\omega} a^\tau_{\omega_2}} & \text{if } (\xi_{jh}, \eta_{jh}) = (0, 0) \text{ and } a^\tau_{\omega_2} \sigma_{\tilde{y}_j, \tilde{y}_h} \neq 0 \\
(-\infty, \frac{b_{\tilde{y}_j, \tilde{w}} b_{\tilde{y}_h, \tilde{w}}}{a^\tau_{\omega} a^\tau_{\omega_2}}) & \text{if } (\xi_{jh}, \eta_{jh}) \in \{(-1, 0), (0, 1)\} \text{ and } \text{sgn}(a^\tau_{\omega_2} \sigma_{\tilde{y}_j, \tilde{y}_h}) \notin [\xi_{jh}, \eta_{jh}] \\
\left[ \frac{b_{\tilde{y}_j, \tilde{w}} b_{\tilde{y}_h, \tilde{w}}}{a^\tau_{\omega} a^\tau_{\omega_2}}, \infty \right) & \text{if } (\xi_{jh}, \eta_{jh}) \in \{(-1, 0), (0, 1)\} \text{ and } \text{sgn}(a^\tau_{\omega_2} \sigma_{\tilde{y}_j, \tilde{y}_h}) \in [\xi_{jh}, \eta_{jh}] \setminus \{0\} \\
\emptyset & \text{if } (\xi_{jh}, \eta_{jh}) \neq (-1, 1), - \text{sgn}(b_{\tilde{y}_j, \tilde{w}} b_{\tilde{y}_h, \tilde{w}}) \notin [\xi_{jh}, \eta_{jh}], \text{ and } a^\tau_{\omega_2} \sigma_{\tilde{y}_j, \tilde{y}_h} = 0 \\
(-\infty, \infty) & \text{otherwise} 
\end{cases}$$

$$\mathcal{D}^{k,\tau,c} = \{ D(r) : r \in \mathcal{R}^{k,\tau,c} \}, \mathcal{B}^{k,\tau,c} = \{ B(r) : r \in \mathcal{R}^{k,\tau,c} \}, \text{ and } \mathcal{G}^{k,\tau,c} = \{ G(r) : r \in \mathcal{R}^{k,\tau,c} \}.$$
Next, consider imposing $A_6$. To illustrate how restricting the sign of the off-diagonal elements in $\sigma^2_\eta$ can help identify $\delta$ and $\beta$, 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 $(\delta_j, \beta_j)' = b_{Y_j,W|Y_h}(Y_h,X')$. Indeed, Corollary 2.3.2 shows that, even when $A_4$-$A_5$ are not binding, if $\sigma_{\eta_j,\eta_h} = 0$ (i.e. $(\varepsilon_{jh}, \delta_{jh}) = (0,0)$) then, provided $\sigma_{Y_h}^{-2} \neq 0$, $\rho = \frac{\alpha_{Y_j,W} \alpha_{Y_h,W} b_{Y_j,W|Y_h}}{\sigma_{Y_j,W} \sigma_{Y_h,Y_h}}$ is point identified. When $b_{Y_j,W|Y_h}$ exists and is nonzero, we can express $\rho = \frac{b_{Y_j,W|Y_h}}{b_{Y_j,W|Y_h}}$ as the ratio of the regression and IV regression estimands. It follows from the mappings in Theorem 2.3.1 that the full vector of system coefficients $(\rho, \delta, \beta, \Gamma)$ is point identified, with $\delta_j = b_{Y_j,W|Y_h}$ and $\beta_j = b_{Y_j,X} - b_{W,X} b_{Y_j,W|Y_h}$ as obtains via the IV regression $b_{Y_j,(W,X')|(Y_h,X')}$.
What if $\sigma_{\eta_j,\eta_h} = 0$ fails? Corollary 2.3.2 answers this question by deriving the identification regions for $\rho$, $\delta$, and $\beta$ under weaker restriction in $A_6$ on the sign of $\Gamma_{jh} \equiv \sigma^{-2}_W \sigma_{\eta_j,\eta_h}$. First, if the identification region $G_{jh}^{k,\tau}$ identifies the sign of $\Gamma_{jh}$ when $A_6$ is not binding (i.e. when $(\zeta_{jh}, \tau_h) = (-1, 1)$ for all $j < h$) then imposing the (correct) sign restriction on $\Gamma_{jh}$ in $A_6$ is uninformative about $\rho$, $\delta$, and $\beta$. Otherwise, restricting the sign of $\Gamma_{jh}$ in $A_6$ can rule out a region of $R_{c,\tau}$. Specifically, recall that

$G_{jh}^{k,\tau} = \{ \sigma^{-2}_W \sigma_{\bar{y}_j,\bar{y}_h} - b_{\bar{y}_j,\bar{y}_h} b_{\bar{y}_h,\bar{y}_r} : r \in R_{c,\tau} \}$.

Thus, provided $\sigma^{-2}_W \sigma_{\bar{y}_j,\bar{y}_h}$ is nonzero, $0 \in R_{c,\tau}$ if and only if

$$\frac{b_{\bar{y}_j,\bar{y}_h} b_{\bar{y}_h,\bar{y}_r}}{\sigma^{-2}_W \sigma_{\bar{y}_j,\bar{y}_h}} \in int(R_{c,\tau}).$$

Corollary 2.3.2 demonstrates how restricting the sign of $\Gamma_{jh}$ can rule out elements of $R_{c,\tau}$ that are either smaller or larger than $\frac{b_{\bar{y}_j,\bar{y}_h} b_{\bar{y}_h,\bar{y}_r}}{\sigma^{-2}_W \sigma_{\bar{y}_j,\bar{y}_h}}$, as encoded in $R_{c,\tau}$. In turn, this can tighten the identification regions for $\delta$ and $\beta$. Last, if Corollary 2.3.2 yields $R_{c,\tau} = \emptyset$ 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 $\delta_j$ or $\beta_{jl}$ may tighten the identification region of $\rho$, and therefore of $\delta$, $\beta$, and $\Gamma$ using Theorem 2.3.1’s mappings. We do not pursue this here; instead, we focus on the auxiliary assumptions $A_4$-$A_6$ which do not directly restrict $\delta$ or $\beta$.

2.4. 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:

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

If $\sigma^{-2}_W \sigma_{\bar{y}_j,\bar{y}_h} = 0$ then restricting the sign of $\Gamma_{jh}$ is either contradictory or uninformative about $\rho$, depending on the sign of $b_{\bar{y}_j,\bar{y}_h} b_{\bar{y}_h,\bar{y}_r}$, as encoded in $R_{c,\tau}$. 

63
where \( \eta_2 = (\eta_2, \eta_3)' \), \( X_2 = (X_1, X_2)' \), \( \eta = (\eta_1, \eta_2, \eta_3)' \), and \( Y_3 = (Y_1, Y_2, Y_3)' \). We let \( \eta_2, U, \varepsilon, \) and \( \eta \) be jointly independent and normally distributed with mean 0 so that A2 and A3 hold. We allow the components of \( \eta_2 \) (respectively \( \eta \)) to be correlated. It follows that \((X', W, Y')\) is normally distributed and we can analytically express the identification regions for \( \rho \), \( \delta \), and \( \beta \) in terms of the elements of \( \text{Var}[(\eta'_2, U, \varepsilon, \eta)'\)]\). In this example, we set the equation coefficients to

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

and the variances of \( \eta_2, U, \varepsilon, \) and \( \eta \) to

\[
\sigma^2_\varepsilon = 3, \quad \sigma^2_U = 5, \quad \sigma^2_{\eta_2} = \begin{bmatrix} 1 & 0.14 \\ 0.14 & 1 \end{bmatrix}, \quad \text{and} \quad \sigma^2_\eta = \begin{bmatrix} 1.1 & -0.31 & 0.63 \\ -0.31 & 1.99 & -0.59 \\ 0.63 & -0.59 & 2.25 \end{bmatrix}.
\]

We obtain that \( \rho = 0.53 \) and thus any restriction \( 0.89 = \frac{\sigma^2_\varepsilon}{\sigma^2_\varepsilon} \leq \kappa \) in A4 is valid. Further, we obtain that \( R^2_{W, Y_1} = 0.31, R^2_{W, Y_2} = 0.34, R^2_{W, Y_3} = 0.27, \) and \( R^2_{W, \tilde{Y}} = 0.44 \).

Using a grid search, we approximate 4 types of identification regions, illustrated in Figure 1. The first is the single-equation identification regions \( S_j \) that consider each \( Y_j \) equation separately. The second is the joint-equations region \( J \) that considers the \( Y \) equations jointly. \( S_j \) and \( J \) obtain under A1-A3 only (i.e. when \( \kappa = \infty, \tau = (1, 1, 1)' \), and \((\underline{c}_j, \overline{c}_j) = (-1, 1)
\) for all \( j < h \)). The third identification region is the joint-equations bounds \( J^{\kappa, \tau} \) that obtains under A1-A5, with \( \kappa = 1 \) and \( \tau = (0.7, 0.7, 0.7)' \). The fourth region \( J^{\kappa, \tau, c} \) obtains under A1-A6, with \( \kappa \) and \( \tau \) as in \( J^{\kappa, \tau} \), where \( c \) imposes the (correct) sign restrictions \( r_{\eta_1, \eta_2} \leq 0, r_{\eta_1, \eta_3} \geq 0, \) \( r_{\eta_2, \eta_3} \leq 0 \). Figure 1 illustrates these regions by plotting their two dimensional projections onto the \((\rho, \delta_j), (\rho, \beta_j)\), and \((\rho, \beta_j)\) spaces for \( j = 1, 2, 3 \). The plus sign denotes the true parameter values. Further, the asterisk corresponds to the regression estimand.
and the cross sign corresponds to the identification region $\mathcal{J}^{\kappa,\tau,c^*}$ (the IV regression estimand) where $c^*$ incorrectly sets $\sigma_{\eta_1,\eta_2} = 0$ and leaves $\sigma_{\eta_1,\eta_3}$ and $\sigma_{\eta_2,\eta_3}$ 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 $\mathcal{J}_{\kappa,\tau}$. The second darkest region corresponds to the joint-equations region $\mathcal{J}_{\kappa,\tau,c}$ 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 $\mathcal{J}^{\kappa,\tau,c}$.

Table 10 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 $\rho$, depending on $j$. The third column reports the joint-equations bounds $\mathcal{J}^{\kappa,\tau,c}$ 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 $\mathcal{J}^{\kappa,\tau,c^*}$ where $c^*$ incorrectly assumes that $\sigma_{\eta_1,\eta_2} = 0$, with $\sigma_{\eta_1,\eta_3}$ and $\sigma_{\eta_2,\eta_3}$ unrestricted. The last column reports the regression estimand $b_{Y,(W,X)'}$ which would point identify $\delta$ and $\beta$ if there is no measurement error in $W$. Table 10 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 $\mathcal{J}^{\kappa,\tau,c} \subseteq \mathcal{S}_1^{\kappa,\tau} \times ... \times \mathcal{S}_p^{\kappa,\tau}$ which become tighter as stricter valid restrictions on $\kappa$, $\tau$, and/or $c$ are imposed.

2.5. Estimation and Inference

For inference, we implement a procedure that delivers $1 - \alpha$ (e.g. 50% or 95%) confidence regions for each of the partially identified parameters $\rho$, $\delta_j$, $\beta_{jl}$, and $\Gamma_{jh}$ for $j, h = 1, ..., p$ and $l = 1, ..., k$. The procedure consists of three steps. First, we express each of the bounds
where\(^9\) vec(\(\sigma_Y^2\)) collects the \(\frac{1}{2}p(p+1)\) variance and covariance elements of \(\sigma_Y^2\). Further, we construct an estimator \(\hat{\pi}\) for \(\pi\) and give conditions under which \(\hat{\pi}\) is \(\sqrt{n}\) consistent and asymptotically normally distributed. Second, we employ results on intersection bounds to construct a \(1 - \alpha\) confidence region \(\text{CR}_{1-\alpha}\) for the parameter \(\rho\) that is partially identified in \(\mathcal{R}^{\kappa,\tau,c}\) for any \((\kappa,\tau,c)\) configuration. The last step uses the mappings, given in Theorem 2.3.1, that express \(\delta_j, \beta_{jl}\), and \(\Gamma_{jh}\) as functions of \((\pi,\rho)\) to construct \(1 - \alpha\) confidence regions for the partially identified parameters \(\delta_j, \beta_{jl}\), and \(\Gamma_{jh}\).

### 2.5.1. Estimation of \(\pi\)

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

\[
\hat{\pi} \equiv (\text{vec}(b_{Y,(W,X)'}')', b_{W,(Y,X)'}', b_{W,(Y_1,X_1)'}', \ldots, b_{W,(Y_p,X_p)'}', \text{vec}(b_{Y,X}')', b_{W,X}', \sigma_W^{-2} \text{vec}(\sigma_Y^2)')',
\]

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

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

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 \(\pi\) is \(\sqrt{n}\) consistent and

---

\(^7\)An alternative would express the bounds in Corollary 2.3.2 as a function of \(\text{Var}([1,Y',W,X]')\) and constructs an estimator for these moments.

\(^8\)Throughout this discussion, we assume that \(\sigma_Y^2\) is nonsingular. Otherwise, we drop the redundant \(Y\) elements from \((Y,X')'\) in \(b_{W,(Y,X)'}\) and \(R_W^{\kappa,\tau,c}\).

\(^9\)Let \(A \equiv [A_1, \ldots, A_q]_{m \times 1}\). Then \(\text{vec}(A) \equiv (A_1', \ldots, A_q')'\). Further, if \(q = m\) and \(A\) is symmetric then we let \(\text{vec}(A) \equiv [A_{11}, \ldots, A_{mm}, A_{12}, \ldots, A_{1m}, \ldots, A_{(m-1)1}, \ldots, A_{(m-1)m}']'\) collect the diagonal and upper-diagonal elements of \(A\).
asymptotically normally distributed. For this, let $\mu_A^2 = E(AA')$ and define the square block-diagonal matrix $Q$:

$$Q \equiv \text{diag}\{ I \otimes \mu^2_{1,W,X'}, \mu^2_{2,1,W,X'}, \ldots, \mu^2_{p(p+1),1,W,X'} \},$$

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

**Theorem 2.5.1** Assume $A_1 (i)$ and that $Q$ is nonsingular. Suppose further that:

(i) $\frac{1}{n} \sum_{i=1}^{n} (1, Y_i', W_i, X_i')' (1, Y_i', W_i, X_i') n^{-1/2} \rightarrow N(0, \Sigma)$ where $\Sigma$ is nonsingular. Suppose further that:

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

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

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

**2.5.2. Inference on $\rho$**

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

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

67
which may depend on a vector of nuisance parameters \( \lambda \in \mathbb{R}^{2T \times 1} \) (\( T \equiv \frac{1}{2}p(p-1) \)), a function of \( \pi \):

\[
\lambda_{2T \times 1} = (\sigma_{\widetilde{W}}^{-2}\sigma_{\widetilde{Y}_1,\widetilde{Y}_2}, \sigma_{\widetilde{W}}^{-2}\sigma_{\widetilde{Y}_1,\widetilde{Y}_3}, ..., \sigma_{\widetilde{W}}^{-2}\sigma_{\widetilde{Y}_{p-1},\widetilde{Y}_p}, \hat{b}_{\widetilde{Y}_1,\widetilde{W}}\hat{b}_{\widetilde{Y}_2,\widetilde{W}}, ..., \hat{b}_{\widetilde{Y}_{p-1},\widetilde{W}}\hat{b}_{\widetilde{Y}_p,\widetilde{W}}).
\]

Further, for a given \( \lambda \), each of the bounds \( \rho_v^l(\lambda) \) and \( \rho_v^u(\lambda) \) can be expressed as a function of \( \pi \). For example, in the numerical example in Section 4, the identification region \( \mathcal{R}_{12}^{\kappa,\tau,c} \) under A1-A6 (with \( \Gamma_{12} \leq 0, \Gamma_{13} \geq 0 \), and \( \Gamma_{23} \leq 0 \)) for \( \rho \) is

\[
\mathcal{R}_{12}^{\kappa,\tau,c}(\lambda) = \cap_{v=1}^8 [\rho_v^l(\lambda), \rho_v^u(\lambda)]
\]

\[
= [R_{\widetilde{W},\widetilde{Y}}^2, 1] \cap \left[ \frac{1}{\tau_1} R_{\widetilde{W},\widetilde{Y}_1}, 1 \right] \cap \left[ \frac{1}{\tau_2} R_{\widetilde{W},\widetilde{Y}_2}, 1 \right] \cap \left[ \frac{1}{\tau_3} R_{\widetilde{W},\widetilde{Y}_3}, 1 \right]
\]

\[
\cap (-\infty, \frac{b_{\widetilde{Y}_1,\widetilde{W}}b_{\widetilde{Y}_2,\widetilde{W}}}{\sigma_{\widetilde{W}}^{-2}\sigma_{\widetilde{Y}_1,\widetilde{Y}_2}}, \infty) \cap (-\infty, \frac{b_{\widetilde{Y}_2,\widetilde{W}}b_{\widetilde{Y}_3,\widetilde{W}}}{\sigma_{\widetilde{W}}^{-2}\sigma_{\widetilde{Y}_2,\widetilde{Y}_3}}, \infty)
\]

where the last three intersected regions \( \mathcal{R}_{12}^{\kappa,\tau,c}(\lambda) \), \( \mathcal{R}_{13}^{\kappa,\tau,c}(\lambda) \), and \( \mathcal{R}_{23}^{\kappa,\tau,c}(\lambda) \) in \( \mathcal{R}_{12}^{\kappa,\tau,c}(\lambda) \) obtain from Corollary 2.3.2 based on the signs of the nuisance parameters (here \( T = 3 \))

\[
\lambda_{2T \times 1} = (\sigma_{\widetilde{W}}^{-2}\sigma_{\widetilde{Y}_1,\widetilde{Y}_2}, \sigma_{\widetilde{W}}^{-2}\sigma_{\widetilde{Y}_1,\widetilde{Y}_3}, ..., \sigma_{\widetilde{W}}^{-2}\sigma_{\widetilde{Y}_{p-1},\widetilde{Y}_p}, \hat{b}_{\widetilde{Y}_1,\widetilde{W}}\hat{b}_{\widetilde{Y}_2,\widetilde{W}}, ..., \hat{b}_{\widetilde{Y}_{p-1},\widetilde{W}}\hat{b}_{\widetilde{Y}_p,\widetilde{W}}).
\]

Thus, \( \lambda \) determines whether each \( \mathcal{R}_{jh}^{\kappa,\tau,c}(\lambda) = \emptyset, (-\infty, \infty), (-\infty, \frac{b_{\widetilde{Y}_j,\widetilde{W}}b_{\widetilde{Y}_h,\widetilde{W}}}{\sigma_{\widetilde{W}}^{-2}\sigma_{\widetilde{Y}_j,\widetilde{Y}_h}}, \infty) \), or \( \frac{b_{\widetilde{Y}_j,\widetilde{W}}b_{\widetilde{Y}_h,\widetilde{W}}}{\sigma_{\widetilde{W}}^{-2}\sigma_{\widetilde{Y}_j,\widetilde{Y}_h}}, \infty \).

**Known Nuisance Parameters**

First, suppose that the nuisance parameter \( \lambda \) is known (or that A6 is not binding and \( \lambda \) is irrelevant). As discussed in Manski and Pepper (2009) and Chernozhukov, Lee, and Rosen (2013), the sample analog estimator \( \hat{\mathcal{R}}_{\kappa,\tau,c}^{\kappa,\tau,c}(\lambda) \equiv \cap_{v=1}^M [\hat{\rho}_v^l(\lambda), \hat{\rho}_v^u(\lambda)] \) tends to be biased “inward” in finite samples, leading to estimates that are narrower on average than \( \mathcal{R}_{\kappa,\tau,c}(\lambda) \). Further, the sampling error may vary with \( v \), across the intersected regions \( \mathcal{R}_v(\lambda) \), which complicates the inference on \( \mathcal{R}_{\kappa,\tau,c}(\lambda) \). To overcome these difficulties, we follow Chernozhukov, Lee, and Rosen (2013) and use the “precision-corrected” estimators
for \( \rho'_v(\lambda) \) and \( \rho^u_v(\lambda) \), \( v \in \mathcal{V} \equiv \{ 1, ..., M \} \) in order to construct estimators for \( \rho'_0(\lambda) \) and \( \rho^u_0(\lambda) \) as follows:

\[
\hat{\rho}'_0(\lambda; 1-\alpha_{21}) \equiv \sup \{ \hat{\rho}'_v(\lambda) - c_{1-\alpha_{21}}(\lambda)se'_v(\lambda) \} \quad \text{and} \quad \hat{\rho}^u_0(\lambda; 1-\alpha_{21}) \equiv \inf \{ \hat{\rho}^u_v(\lambda) + c^u_{1-\alpha_{21}}(\lambda)se^u_v(\lambda) \}
\]

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

\[
\Pr[\hat{\rho}'_0(\lambda; 1-\alpha_{21}) \leq \rho'_0(\lambda)] \geq 1 - \alpha_{21} - o(1) \quad \text{and} \quad \Pr[\rho^u_0(\lambda; 1-\alpha_{21})] \leq \hat{\rho}^u_0(\lambda; 1-\alpha_{21})] \geq 1 - \alpha_{21} - o(1).
\]

In particular, setting \( \alpha_{21} = \frac{1}{2} \) yields half-median-unbiased estimators \( \hat{\rho}'_0(\lambda; \frac{1}{2}) \) and \( \hat{\rho}^u_0(\lambda; \frac{1}{2}) \).

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

\[
CI_{1-\alpha_{21}}^R(\lambda) \equiv [\hat{\rho}'_0(\lambda; 1-\frac{\alpha_{21}}{2}), \hat{\rho}^u_0(\lambda; 1-\frac{\alpha_{21}}{2})] \quad \text{such that} \quad \liminf_{n \rightarrow \infty} \Pr[R^{\kappa,\tau,c}(\lambda) \subseteq CI_{1-\alpha_{21}}^R(\lambda)] \geq 1 - \alpha_{21}.
\]

\( CI_{1-\alpha_{21}}^R(\lambda) \) is a valid, but conservative, confidence region for \( \rho \in R^{\kappa,\tau,c}(\lambda) \). To conduct inference on \( \rho \) directly, we invert a test statistic that combines the lower and upper bounds. This yields an asymptotically valid \( 1 - \alpha_{21} \) (e.g. 95%) confidence regions \( CI_1^\rho(\lambda; 1-\alpha_{21}) \) for the parameter \( \rho \) that is partially identified in \( R^{\kappa,\tau,c}(\lambda) \):

\[
\liminf_{n \rightarrow \infty} \Pr[\rho \in CI_1^\rho(\lambda; 1-\alpha_{21})] \geq 1 - \alpha_{21}.
\]

In particular, we apply the results in\(^\text{10}\) 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 \( c^t_{1-\alpha_{21}}(\lambda) \) and \( c^u_{1-\alpha_{21}}(\lambda) \) and construct the bias-adjusted estimates \( \hat{\rho}'_0(\lambda; 1-\alpha_{21}) \) and \( \hat{\rho}^u_0(\lambda; 1-\alpha_{21}) \) and the confidence region \( CI_{1-\alpha_{21}}^\rho(\lambda) \), we implement their algorithm 1. For brevity, we describe the details of the algorithm in Online Appendix B.1.

\(^{10}\)See also Chernozhukov, Kim, Lee, and Rosen (2015).
Estimated Nuisance Parameters

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

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

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

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

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

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

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

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

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 $c = 0$ then one rejects (under $A_1$-$A_5$) that $\sigma^2_\theta$ is diagonal. Otherwise, imposing tighter restrictions on $(\kappa, \tau, c)$ 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 $\lambda$. For example, if the sign of $\sigma^{-2}_W \sigma_{\tilde{y}_j, \tilde{y}_h}$ is imprecisely estimated then forming the union over
$\Lambda_{1-\alpha_{22}}$ may effectively mute the impact of the $\sigma_{\eta_j,\eta_h}$ restriction in $A_6$ on $CR_{1-\alpha_2}^\theta$.

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

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

Each identification region for $\delta_j$, $\beta_{jl}$, and $\Gamma_{jh}$ for $j,h = 1,...,p$, $j < h$, and $l = 1,...,k$ in Corollary 2.3.2 is of the form

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

where $R^{k,\tau,c}$ is the identification region for $\rho$ under any given $(\kappa,\tau,c)$ configuration, and $H(\cdot;r)$ is a function of $\pi$, given in Theorem 2.3.1. For example,

$$D_{j}^{k,\tau,c} = \{1/r b_{Y_j,\tilde{W}} : r \in R^{k,\tau,c}\}.$$

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:

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

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

71
by applying Proposition 3 of Chernozhukov, Rigobon, and Stoker (2010) to form the union:

\[ CR_{1-\alpha_1-\alpha_2} = \bigcup_{r \in CR_{1-\alpha_1}} CR_{1-\alpha_2}^\phi (r). \]

In the empirical application, we report the confidence regions \( CR_{0.5}^\phi \) and \( CR_{0.95}^\phi \) for \( \delta_j \) and \( \beta_{jl} \) (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.

2.6. 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\(^\text{11}\)). 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 econo-

\(^{11}\)This result assumes quadratic investment adjustment costs, constant return to scale, perfect competition, and an efficient financial market (see Hayashi, 1982).
metric method in Erickson and Whited (2002), which uses higher order moments\textsuperscript{12} 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.

\subsection*{2.6.1. 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\textsuperscript{13} between 1970 to 1994. Erickson and Whited (2002) strengthen A\textsubscript{2}-A\textsubscript{3} to require $\varepsilon$, $\eta$, and $(X', U)'$ to be jointly independent.

\textsuperscript{12}Erickson and Whited (2002) strengthen A\textsubscript{2}-A\textsubscript{3} to require $\varepsilon$, $\eta$, and $(X', U)'$ to be jointly independent.

\textsuperscript{13}Specifically, we apply 4 firm filters: INDFMT=INDL (industrial), DATAFMT=STD (standardized data reporting), POPSRC=D (domestic (North American)), and CONSOL=C (consolidated).
2017. We remove financial firms (Standard Industrial Classification (SIC) code 6000 to 6999) and regulated firms (SIC code 4900 to 4999). To exclude small firms, we delete observations in which a firm has at most $2 million in real total assets (COMPUSTAT item: AT) or $5 million in real capital (COMPUSTAT item: PPEGT) at either the end or the beginning of a time period. 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\(^\text{14}\) them by the firm’s total assets\(^\text{15}\). 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\(^\text{16}\) 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\(^\text{17}\) 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\(^\text{18}\). Section 6.6 considers including asset tangibility \(X_3\) in

---

\(^{14}\)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)).
\(X\), 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 for the panel variables.

2.6.2. 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 \(\rho\), \(\delta_j\), and \(\beta_{jl}\) when \(\kappa = \infty\) and \(\tau = (1,1,1)'\), with \(A_4\) and \(A_5\) not binding. Column 1 reports the results corresponding to the single equation bounds \(S_{\kappa,\tau}^{j}\). This yields different identification regions for \(\rho\) 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 \(J_{\kappa,\tau}\). 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 \((0.181,0.223)\) 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 \(J_{\kappa,\tau}\) contains 0. For example, the 95% confidence region for the effect \(\beta_{11}\) of a $1 increase in cash flow on investment is \((-0.397,0.278)\). Column 3 reports the results for \(J_{\kappa,\tau,c}\) when \(A_6\) sets \(c\) such that the investment and saving disturbances are negatively correlated, the investment and debt disturbances are positively correlated, and the saving
and debt disturbances are negatively correlated. This yields comparable confidence regions to $J^{\kappa,\tau}$. In this case, the identification gain from imposing $A_6$ (with the $c$ configuration above) is offset by the decrease in the precision of the estimates. Column 4 reports the (IV-type) results under $J^{\kappa,\tau,c^*}$ when $A_6$ sets $c^* = 0$ so that the variance matrix of the disturbances is diagonal. For instance, $c^* = 0$ 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 $\rho$ (note that the 50% confidence region for $\rho$ is empty). Last, column 5 reports the results from the regression estimator which would be consistent if there is no measurement error in Tobin’s q. The regression estimates for $\delta_1$, $\delta_2$, and $\delta_3$ are possibly attenuated relative to the bounds in $J^{\kappa,\tau}$, $J^{\kappa,\tau,c}$, and $J^{\kappa,\tau,c^*}$. 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 $\kappa^*$ and $\tau^*$ 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 $S^{\kappa,\tau}_j$, $J^{\kappa,\tau}$, and $J^{\kappa,\tau,c}$ (with $c$ 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 $J^{\kappa,\tau,c^*}$, when $A_6$ assumes that $\sigma^2_\eta$ 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

\[ R^2_{Y_1,W}, R^2_{Y_2,W}, \text{and } R^2_{Y_3,W} \text{ are estimated to be } 3.66\%, 0.9\%, \text{ and } 3.56\% \text{ respectively.} \]
incompatible with the assumption that the cross-equation disturbances are uncorrelated.

2.6.3. Sensitivity Analysis

If $\kappa = 0$ then there is no measurement error and $(\delta, \beta, \Gamma)$ is point identified. Next, we study the sensitivity of the identification regions for the cash flow coefficients $\beta_{j1}$ for $j = 1, 2, 3$ to deviations from $\kappa = 0$. For this, we set $\tau = (1, 1, 1)'$ in $A_5$ and directly control the extent of the measurement error by varying $\kappa$ in $A_4$. Using the sample from the middle year 1993, Figure 2 plots the 50% and 95% confidence regions for the partially identified $\beta_{11}, \beta_{21}$ and $\beta_{31}$ as $\kappa$ ranges from 0 to $\infty$ (or equivalently as $\kappa^*$ ranges from 1 to $R^2_{W,X}$). It plots the regions under $A_1$-$A_4$ when each equation is analyzed separately ($S^*_j$), the three equations are analyzed jointly ($J^*$), and the three equations are analyzed jointly and $A_6$ is imposed ($J^{\kappa,c}$ with $c$ set as in Tables 3 and 4). The 95% confidence region for the effect $\beta_{11}, \beta_{21}$, or $\beta_{31}$ 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, $\beta_{11}$ and $\beta_{21}$ are each estimated to be significantly positive and $\beta_{31}$ to be negative.

Table 5 studies the joint consequences of measurement error on the identification of the coefficients $(\beta_{11}, \beta_{21}, \beta_{31})$ on cash flow in the three equations. This safeguards against maintaining empirical conclusions about $\beta_{11}, \beta_{21}$, and $\beta_{31}$ that rest 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 $\kappa$, and the corresponding largest $\kappa^*$, such that a null hypotheses about $(\beta_{11}, \beta_{21}, \beta_{31})$ is not rejected. In particular, if the reliability ratio of the proxy exceeds $\kappa^*$, the reported threshold value of $\kappa^*$, 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 $\kappa^*$, 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 $R_{W,U}^2 \geq 25.7\%$, 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.

2.6.4. 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 $\beta_{j1}$ (the coefficients on cash flow) that is partially identified in $B_{j1}$, $B_{j1}^c$, or $B_{k,\tau,c}^\kappa$, $j = 1, 2, 3$, for each cross section in our sample (years 1970 to 2017). The first column reports the joint equations bounds for $\beta_{j1} \in B_{j1}$ 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 $\beta_{j1} \in B_{j1}^c$ under $A_1$-$A_3$ and the $A_6$ diagonal restriction $c^* = 0$. We reject this specification in 19 of the 48 years at the 96% level (i.e. $CR_{0.96}^\theta = \emptyset$ and hence $CR_{0.95}^\theta = \emptyset$). When nonempty, the confidence regions for $\beta_{j1} \in B_{j1}^c$ yield mixed results across different years. The third column reports the bounds for $\beta_{j1} \in B_{j1}^{k,\tau,c}$ under $A_1$-$A_6$ where, for each year, we set $\kappa$ and $\tau$ such that the estimated $\kappa^*$ and $\tau^*$ are 0.5 and (0.9, 0.9, 0.9). Setting $\kappa^* = 0.5$ assumes that Tobin’s q is a moderately accurate proxy for marginal q. Further, as before, 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.

2.6.5. 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 $\pi$ 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

---

\(^{20}\)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.
improve substantially over the single equation bounds. Specifically, for the specification under A1-A3 without fixed effects, the 95% confidence regions for the effect of cash flow \( \beta_{j1} \in B_{j1} \) falls in the positive range for investment and in the negative range for debt. Imposing the A6 sign restrictions encoded in \( c \) tightens the bounds further (\( \beta_{j1} \in B_{j1}^c \)) 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 A6 (\( \beta_{j1} \in B_{j1}^c \)). We also report the bounds when A4-A5 set \( \kappa \) and \( \tau \) such that \( \kappa^* \) and \( \tau^* \) are estimated to be 0.5 and (0.9, 0.9, 0.9)' respectively. This yields 95% confidence regions that are close to the regression estimates, whereby cash flow is estimated to affect investment and saving positively and debt negatively. Last, 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 A1-A3 but remains significantly positive under the sign restrictions in A6.

### 2.6.6. 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 \( \beta_{j1} \in B_{j1} \) 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 A1-A6 (\( \beta_{j1} \in B_{j1}^c \)), reported in the second column, in 5 (as opposed to 19) of the 48 years. Further, the bounds under A1-A6 (\( \beta_{j1} \in B_{j1}^{c,\tau,c} \)) in the last column remain close to the regression estimates and
this specification is now rejected in 11 (as opposed to 5) years. Last, Table 7 in the Online Appendix replicates the panel data analysis in Table 6 after augmenting $X$ with $X_3$ and, here too, the results share similar features to those in Table 6. To summarize the differences, for the specification without fixed effects, the 95% confidence region for the effect of cash flow on investment 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 $\beta_{j1} \in B_{j1}^{\kappa,\tau,c}$ 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.

2.7. 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 COMPU-STAT, 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 6: Identification Regions

Identification regions $S_j$ (light) for $j = 1, 2, 3$, $J$, $J^{\kappa, \tau}$, and $J^{\kappa, \tau, c}$ (dark) for $\kappa = 1$, $\tau = (0.7, 0.7, 0.7)'$ and $c$ set to the (correct) sign restrictions $r_{\eta_1, \eta_2} \leq 0$, $r_{\eta_1, \eta_3} \geq 0$, $r_{\eta_2, \eta_3} \leq 0$. The plus, asterisk, and cross signs correspond to the true parameter values, $b_{Y,(W,X)'},$ and $J^{\kappa, \tau, c^*}$ respectively, where $c^*$ incorrectly sets $\sigma_{\eta_1, \eta_2} = 0$ (with $\sigma_{\eta_1, \eta_3}$ and $\sigma_{\eta_2, \eta_3}$ unrestricted).
<table>
<thead>
<tr>
<th>DGP</th>
<th>$S_{\kappa,\tau}^\rho$</th>
<th>$\mathcal{I}_{\kappa,\tau}^\rho$</th>
<th>$\mathcal{J}_{\kappa,\tau,\epsilon}^\rho$</th>
<th>$\mathcal{J}_{\kappa,\tau,\epsilon}^{\ast}$</th>
<th>$b_{Y, (W, X)^\prime}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa \to \infty$ and $\tau = (1, 1)'$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.527</td>
<td>[0.315, 1]</td>
<td>[0.441, 1]</td>
<td>[0.441, 0.604]</td>
<td>0.604</td>
</tr>
<tr>
<td>$\delta_1$</td>
<td>0.700</td>
<td>[0.369, 1.71]</td>
<td>[0.369, 0.835]</td>
<td>[0.610, 0.835]</td>
<td>0.610 0.369</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>1</td>
<td>[0.551, 1.316]</td>
<td>[0.871, 1.316]</td>
<td>[0.871, 1.086]</td>
<td>1.086 1.316</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>0.700</td>
<td>[0.543, 0.811]</td>
<td>[0.655, 0.811]</td>
<td>[0.655, 0.730]</td>
<td>0.730 0.811</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.527</td>
<td>[0.342, 1]</td>
<td>[0.441, 1]</td>
<td>[0.441, 0.604]</td>
<td>0.604</td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>1.050</td>
<td>[0.553, 1.618]</td>
<td>[0.553, 1.253]</td>
<td>[0.915, 1.253]</td>
<td>0.915 0.553</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>0.850</td>
<td>[0.308, 1.324]</td>
<td>[0.656, 1.324]</td>
<td>[0.656, 0.979]</td>
<td>0.979 1.324</td>
</tr>
<tr>
<td>$\beta_{22}$</td>
<td>0.950</td>
<td>[0.761, 1.16]</td>
<td>[0.882, 1.16]</td>
<td>[0.882, 0.995]</td>
<td>0.995 1.16</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.527</td>
<td>[0.269, 1]</td>
<td>[0.441, 1]</td>
<td>[0.441, 0.604]</td>
<td>0.604</td>
</tr>
<tr>
<td>$\delta_3$</td>
<td>1.100</td>
<td>[0.334, 1.479]</td>
<td>[0.945, 1.479]</td>
<td>[0.945, 1.203]</td>
<td>1.203 1.479</td>
</tr>
<tr>
<td>$\beta_{31}$</td>
<td>0.950</td>
<td>[0.761, 1.116]</td>
<td>[1.146, 1.116]</td>
<td>[1.146, 0.995]</td>
<td>0.995 1.116</td>
</tr>
<tr>
<td>$\beta_{32}$</td>
<td>1.200</td>
<td>[0.932, 1.333]</td>
<td>[1.146, 1.333]</td>
<td>[1.146, 1.236]</td>
<td>1.236 1.333</td>
</tr>
<tr>
<td>$\kappa = 1$ and $\tau = (0.7, 0.7, 0.7)'$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.527</td>
<td>[0.500, 1]</td>
<td>[0.500, 1]</td>
<td>[0.500, 0.604]</td>
<td>0.604</td>
</tr>
<tr>
<td>$\delta_1$</td>
<td>0.700</td>
<td>[0.369, 0.737]</td>
<td>[0.369, 0.737]</td>
<td>[0.610, 0.737]</td>
<td>0.610 0.369</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>1</td>
<td>[0.965, 1.316]</td>
<td>[0.965, 1.316]</td>
<td>[0.965, 1.086]</td>
<td>1.086 1.316</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>0.700</td>
<td>[0.688, 0.811]</td>
<td>[0.688, 0.811]</td>
<td>[0.688, 0.730]</td>
<td>0.730 0.811</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.527</td>
<td>[0.500, 1]</td>
<td>[0.500, 1]</td>
<td>[0.500, 0.604]</td>
<td>0.604</td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>1.050</td>
<td>[0.553, 1.106]</td>
<td>[0.553, 1.106]</td>
<td>[0.915, 1.106]</td>
<td>0.915 0.553</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>0.850</td>
<td>[0.797, 1.324]</td>
<td>[0.797, 1.324]</td>
<td>[0.797, 0.997]</td>
<td>0.997 1.324</td>
</tr>
<tr>
<td>$\beta_{22}$</td>
<td>0.950</td>
<td>[0.931, 1.116]</td>
<td>[0.931, 1.116]</td>
<td>[0.931, 0.995]</td>
<td>0.995 1.116</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.527</td>
<td>[0.500, 1]</td>
<td>[0.500, 1]</td>
<td>[0.500, 0.604]</td>
<td>0.604</td>
</tr>
<tr>
<td>$\delta_3$</td>
<td>1.100</td>
<td>[0.334, 1.479]</td>
<td>[0.945, 1.479]</td>
<td>[0.945, 1.203]</td>
<td>1.203 1.479</td>
</tr>
<tr>
<td>$\beta_{31}$</td>
<td>0.950</td>
<td>[1.058, 1.479]</td>
<td>[1.058, 1.479]</td>
<td>[1.058, 1.203]</td>
<td>1.203 1.479</td>
</tr>
<tr>
<td>$\beta_{32}$</td>
<td>1.200</td>
<td>[1.185, 1.333]</td>
<td>[1.185, 1.333]</td>
<td>[1.185, 1.236]</td>
<td>1.236 1.333</td>
</tr>
</tbody>
</table>

This table reports population identification regions and point estimands. $\frac{\sigma^2}{\hat{\epsilon}} = 0.89$ and $R_{\hat{W}, \hat{Y}}^2 = 0.44$. $c^\ast$ correctly sets $\left(c^\ast_{12}, c^\ast_{12}\right) = (-1, 0)$ and $\left(c^\ast_{13}, c^\ast_{13}\right) = (0, 1)$ whereas $c$ incorrectly sets $\left(c_{12}, c_{12}\right) = (0, 0)$ and $\left(c_{13}, c_{13}\right) = (c_{23}, c_{23}) = (-1, 1)$. 

84
Table 11: Summary Statistics

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.

<table>
<thead>
<tr>
<th></th>
<th>Investment</th>
<th>Saving</th>
<th>Debt</th>
<th>Tobin’s Q</th>
<th>Cash Flow</th>
<th>Firm Size</th>
<th>Tangibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>0.084</td>
<td>0.009</td>
<td>0.261</td>
<td>1.633</td>
<td>0.071</td>
<td>5.327</td>
<td>0.354</td>
</tr>
<tr>
<td>std dev</td>
<td>0.098</td>
<td>0.106</td>
<td>0.214</td>
<td>1.147</td>
<td>0.137</td>
<td>1.949</td>
<td>0.236</td>
</tr>
<tr>
<td>min</td>
<td>0.002</td>
<td>-0.298</td>
<td>0.000</td>
<td>0.526</td>
<td>-0.536</td>
<td>0.391</td>
<td>0.022</td>
</tr>
<tr>
<td>max</td>
<td>0.593</td>
<td>0.541</td>
<td>1.016</td>
<td>7.364</td>
<td>0.390</td>
<td>10.216</td>
<td>0.925</td>
</tr>
</tbody>
</table>
Table 12: Bounds for the Investment, Saving, and Debt Equations for Year 1993

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Investment Equation ($I^\rho_j$)</th>
<th>Saving Equation ($J^{\delta_1}_{1j}$)</th>
<th>Debt Equation ($J^{\beta_{12}}_{1j}$)</th>
<th>$b_{Y,(W,X')}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho$</td>
<td>[0.037, 1]</td>
<td>[0.076, 1]</td>
<td>[0.069, 1]</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.019, 1)</td>
<td>(0.055, 1)</td>
<td>(0.051, 1)</td>
<td>(0.050, 0.244)</td>
</tr>
<tr>
<td>$\delta_1$</td>
<td>[0.016, 1.136]</td>
<td>[0.016, 0.360]</td>
<td>[0.016, 0.409]</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>(0.012, 1.382)</td>
<td>(0.012, 0.438)</td>
<td>(0.012, 0.497)</td>
<td>(0.013, 0.022)</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>[-1.308, 0.242]</td>
<td>[-0.243, 0.242]</td>
<td>[-0.309, 0.242]</td>
<td>0.228</td>
</tr>
<tr>
<td></td>
<td>(-1.800, 0.278)</td>
<td>(-0.397, 0.278)</td>
<td>(-0.484, 0.278)</td>
<td>(0.190, 0.266)</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>[-0.010, 0.052]</td>
<td>[-0.010, 0.009]</td>
<td>[-0.010, 0.012]</td>
<td>-0.009</td>
</tr>
<tr>
<td></td>
<td>(-0.012, 0.073)</td>
<td>(-0.012, 0.016)</td>
<td>(-0.012, 0.020)</td>
<td>(-0.011, -0.007)</td>
</tr>
<tr>
<td>$\rho$</td>
<td>[0.009, 1]</td>
<td>[0.076, 1]</td>
<td>[0.069, 1]</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.000, 1)</td>
<td>(0.055, 1)</td>
<td>(0.051, 1)</td>
<td>(0.050, 0.244)</td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>[0.007, $\infty$]</td>
<td>[0.007, 0.213]</td>
<td>[0.007, 0.242]</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>(0.002, $\infty$)</td>
<td>(0.002, 0.309)</td>
<td>(0.002, 0.351)</td>
<td>(0.004, 0.015)</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>[-$\infty$, 0.181]</td>
<td>[-0.115, 0.181]</td>
<td>[-0.154, 0.181]</td>
<td>0.165</td>
</tr>
<tr>
<td></td>
<td>(-$\infty$, 0.223)</td>
<td>(-0.270, 0.223)</td>
<td>(-0.328, 0.223)</td>
<td>(0.120, 0.209)</td>
</tr>
<tr>
<td>$\beta_{22}$</td>
<td>[-0.003, $\infty$]</td>
<td>[-0.003, 0.009]</td>
<td>[-0.003, 0.010]</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>(-0.004, $\infty$)</td>
<td>(-0.004, 0.014)</td>
<td>(-0.004, 0.016)</td>
<td>(-0.004, -0.000)</td>
</tr>
<tr>
<td>$\rho$</td>
<td>[0.035, 1]</td>
<td>[0.076, 1]</td>
<td>[0.069, 1]</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.024, 1)</td>
<td>(0.055, 1)</td>
<td>(0.051, 1)</td>
<td>(0.050, 0.244)</td>
</tr>
<tr>
<td>$\delta_3$</td>
<td>[-1.641, -0.033]</td>
<td>[-0.697, -0.033]</td>
<td>[-0.791, -0.033]</td>
<td>-0.035</td>
</tr>
<tr>
<td></td>
<td>(-1.873, -0.028)</td>
<td>(-0.796, -0.028)</td>
<td>(-0.903, -0.028)</td>
<td>(-0.041, -0.030)</td>
</tr>
<tr>
<td>$\beta_{31}$</td>
<td>[-0.311, 1.994]</td>
<td>[-0.311, 0.654]</td>
<td>[-0.311, 0.787]</td>
<td>-0.287</td>
</tr>
<tr>
<td></td>
<td>(-0.372, 2.728)</td>
<td>(-0.372, 0.961)</td>
<td>(-0.372, 1.136)</td>
<td>(-0.352, -0.221)</td>
</tr>
<tr>
<td>$\beta_{32}$</td>
<td>[-0.078, 0.012]</td>
<td>[-0.026, 0.012]</td>
<td>[-0.031, 0.012]</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td>(-0.107, 0.016)</td>
<td>(-0.038, 0.016)</td>
<td>(-0.045, 0.016)</td>
<td>(0.008, 0.015)</td>
</tr>
</tbody>
</table>

The sample size is 3,454 observations. $Y_1$, $Y_2$, and $Y_3$ denote Investment, Saving, and Debt respectively and $X = [\text{Cash Flow, Firm Size}]$. $c$ sets $(c_{12}, c_{13}) = (-1, 0)$ and $(c_{13}, c_{23}) = (0, 1)$ whereas $c^* = 0.50\%$ and 95% confidence regions are in brackets and parentheses respectively.
Table 13: Bounds for the Investment, Saving, and Debt Equations for Year 1993

<table>
<thead>
<tr>
<th></th>
<th>$\mathcal{S}^{\kappa,\tau}$</th>
<th>$\mathcal{J}^{\kappa,\tau}$</th>
<th>$\mathcal{J}^{\kappa,\tau,\epsilon}$</th>
<th>$b_{Y(W,X)^*}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho$</td>
<td>[0.486, 1]</td>
<td>[0.486, 1]</td>
<td>[0.486, 1]</td>
<td>-</td>
</tr>
<tr>
<td>$\delta_1$</td>
<td>[0.016, 0.040]</td>
<td>[0.016, 0.040]</td>
<td>[0.016, 0.040]</td>
<td>-</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>[0.188, 0.242]</td>
<td>[0.188, 0.242]</td>
<td>[0.188, 0.242]</td>
<td>-</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>[-0.010, -0.008]</td>
<td>[-0.010, -0.008]</td>
<td>[-0.010, -0.008]</td>
<td>-</td>
</tr>
<tr>
<td>$\rho$</td>
<td>[0.486, 1]</td>
<td>[0.486, 1]</td>
<td>[0.486, 1]</td>
<td>-</td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>[0.007, 0.024]</td>
<td>[0.007, 0.024]</td>
<td>[0.007, 0.024]</td>
<td>-</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>[0.134, 0.181]</td>
<td>[0.134, 0.181]</td>
<td>[0.134, 0.181]</td>
<td>-</td>
</tr>
<tr>
<td>$\beta_{22}$</td>
<td>[-0.003, -0.001]</td>
<td>[-0.003, -0.001]</td>
<td>[-0.003, -0.001]</td>
<td>-</td>
</tr>
<tr>
<td>$\rho$</td>
<td>[0.486, 1]</td>
<td>[0.486, 1]</td>
<td>[0.486, 1]</td>
<td>-</td>
</tr>
<tr>
<td>$\delta_3$</td>
<td>[-0.077, -0.033]</td>
<td>[-0.077, -0.033]</td>
<td>[-0.077, -0.033]</td>
<td>-</td>
</tr>
<tr>
<td>$\beta_{31}$</td>
<td>[-0.311, -0.212]</td>
<td>[-0.311, -0.212]</td>
<td>[-0.311, -0.212]</td>
<td>-</td>
</tr>
<tr>
<td>$\beta_{32}$</td>
<td>[0.008, 0.012]</td>
<td>[0.008, 0.012]</td>
<td>[0.008, 0.012]</td>
<td>-</td>
</tr>
</tbody>
</table>

The sample size is 3,454 observations. $Y_1$, $Y_2$, and $Y_3$ denote Investment, Saving, and Debt respectively and $X = \text{[Cash Flow, Firm Size]}$. $c$ sets $(c_{12}, \tau_{12}) = (c_{23}, \tau_{23}) = (-1, 0)$ and $(c_{13}, \tau_{13}) = (0, 1)$ whereas $c^* = 0$. 50% and 95% confidence regions are in brackets and parentheses respectively.
50% (dark) and 95% (light) confidence regions for the partially identified coefficients $\beta_{j1}$ 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 $S^\kappa$, $J^\kappa$, and $J^{\kappa, c}$ when $\kappa \in [0, \infty)$ and $c$ 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 $\kappa$ (largest $\kappa^*$) 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.

Figure 7: Sensitivity of Partial Identification Region for Cash Flow Coefficients
Table 14: Test of Joint Hypothesis

Joint test for the possible signs of the components of \((\beta_{11}, \beta_{21}, \beta_{31})\) under A1-A4 at the 5% level for year 1993. \(\kappa^*\) is the largest \(\kappa^*\) such that \(H_0\) is not rejected. \(\kappa\) is the smallest \(\kappa\) such that \(H_0\) is not rejected.

| \(H_0\) | \(\beta_{11}\) | - | - | - | - | + | + | + | + |
| \(\beta_{21}\) | - | - | + | + | - | - | + | + | + |
| \(\beta_{31}\) | + | - | + | - | + | - | + | - |
| \(\kappa^*\) | 0.157 | 0.157 | 0.177 | 0.177 | 0.197 | 0.197 | 0.257 | 1 |
| \(\kappa\) | 6.483 | 6.483 | 5.486 | 5.486 | 4.723 | 4.723 | 3.230 | 0 |
50% (dark) and 95% (light) confidence regions for $\beta_{j1}$ (cash flow) for $j = 1, 2, 3$ (investment, saving, and debt) from year 1970 to 2017. We consider the regions $B_{j1}$, $B_{j1}^{c^*}$, and $B_{j1}^{\kappa, \tau, c}$ where $c^* = 0$, $\kappa$ and $\tau$ are such that $\hat{\kappa}^* = 0.5$ and $\hat{\tau}^* = (0.9, 0.9, 0.9)'$, and $c$ is such that $(\xi_{12}, \xi_{13}) = (-1, 0)$ and $(\xi_{23}, \xi_{13}) = (0, 1)$. The shaded vertical bars indicate years in which the maintained assumptions are rejected at the 96% level.
<table>
<thead>
<tr>
<th>( S_{j}^{\kappa,\tau} )</th>
<th>( F_{j}^{\kappa,\tau} )</th>
<th>( F_{j}^{\kappa,\tau, c} )</th>
<th>( F_{j}^{\kappa,\tau, c^*} )</th>
<th>( Y_{1}(W, X') )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_{11} )</td>
<td>([-0.242, 0.197])</td>
<td>([0.096, 0.197])</td>
<td>([0.193, 0.197])</td>
<td>-</td>
</tr>
<tr>
<td>([-0.258, 0.198])</td>
<td>([0.092, 0.198])</td>
<td>([0.192, 0.198])</td>
<td>-</td>
<td>((0.187, 0.205))</td>
</tr>
<tr>
<td>([0.115, 0.124])</td>
<td>([-0.016, 0.126])</td>
<td>([0.119, 0.126])</td>
<td>-</td>
<td>((0.116, 0.131))</td>
</tr>
<tr>
<td>([-0.124, 0.453])</td>
<td>([-0.352, -0.087])</td>
<td>([-0.352, -0.345])</td>
<td>-</td>
<td>-0.351</td>
</tr>
<tr>
<td>([-0.355, 0.485])</td>
<td>([-0.355, -0.076])</td>
<td>([-0.355, -0.341])</td>
<td>-</td>
<td>((-0.369, -0.332))</td>
</tr>
</tbody>
</table>

Results without fixed effects for \( \kappa = \infty \) and \( \tau = (1, 1, 1)' \)\)

| \( \beta_{11} \) | \([0.189, 0.197]\) | \([0.189, 0.197]\) | \([0.193, 0.197]\) | - | 0.196 |
| \((0.187, 0.198)\) | \((0.187, 0.198)\) | \((0.192, 0.198)\) | - | \((0.187, 0.205)\) |
| \([0.115, 0.124]\) | \([0.115, 0.124]\) | \([0.120, 0.124]\) | - | 0.124 |
| \([-0.352, -0.333]\) | \([-0.352, -0.333]\) | \([-0.352, -0.345]\) | - | -0.351 |
| \([-0.355, -0.330]\) | \([-0.355, -0.330]\) | \([-0.355, -0.341]\) | - | \((-0.369, -0.332)\) |

Results without fixed effects for \( \kappa^* = 0.5 \) and \( \tau^* = (0.9, 0.9, 0.9)' \)\)

| \( \beta_{11} \) | \([-0.627, 0.130]\) | \([-0.493, 0.130]\) | \([0.017, 0.130]\) | - | 0.129 |
| \((-0.632, 0.131)\) | \((-0.497, 0.131)\) | \((0.016, 0.131)\) | - | \((0.122, 0.137)\) |
| \([-2.504, 0.172]\) | \([-0.248, 0.172]\) | \([0.096, 0.172]\) | - | 0.172 |
| \((-2.542, 0.173)\) | \((-0.255, 0.173)\) | \((0.094, 0.173)\) | - | \((0.161, 0.182)\) |
| \([-0.367, 17.982]\) | \([-0.367, -0.270]\) | \([-0.367, -0.349]\) | - | -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 \( \kappa = \infty \) and \( \tau = (1, 1, 1)' \)\)

| \( \beta_{11} \) | \([0.084, 0.130]\) | \([0.084, 0.130]\) | \([0.084, 0.130]\) | 0.129 |
| \((0.083, 0.131)\) | \((0.083, 0.131)\) | \((0.083, 0.131)\) | - | \((0.122, 0.137)\) |
| \([0.141, 0.172]\) | \([0.141, 0.172]\) | \([0.141, 0.172]\) | - | 0.172 |
| \((0.139, 0.173)\) | \((0.139, 0.173)\) | \((0.139, 0.173)\) | - | \((0.161, 0.182)\) |
| \([-0.367, -0.359]\) | \([-0.367, -0.359]\) | \([-0.367, -0.359]\) | - | -0.366 |
| \([-0.368, -0.357]\) | \([-0.368, -0.357]\) | \([-0.368, -0.357]\) | - | \((-0.381, -0.351)\) |

Results with year and firm fixed effects for \( \kappa^* = 0.5 \) and \( \tau^* = (0.9, 0.9, 0.9)' \)\)

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 \( X = \text{[Cash Flow, Firm Size]} \). When year fixed effects are included, \( X \) also includes year indicator variables. When firm fixed effects are included, the equations' variables undergo a within transformation. \( c \) sets \((c_{12}, c_{13}) = (-1, 0)\) and \((c_{23}, c_{33}) = (0, 1)\) whereas \( c^* = 0 \). Robust standard errors for \( \pi \) are clustered by firm. 50% and 95% confidence regions are in brackets and parentheses respectively.
A.1. Estimation Procedure

The objective here is to estimate parameters: $b$, $\sigma^F$, $\eta$ and $\alpha$.

First of all, why do simulation at all? Don’t I have everything in closed-forms? I do have closed-forms for firm value, equity value and debt value. But I do not have closed-forms for my moments because moments are path-dependent and a simulated sample is an unbalanced panel. Thus, I rely on simulations to generate model counterparts.

In order to address firm heterogeneity in the data, I account for firm fixed effects in calculating the higher-order moments. More specifically, let us assume that firm $i$’s data at time $t$ is $d_{it}$. I convert $d_{it}$ to $\tilde{d}_{it}$ by accounting for firm fixed effects as follows:

$$
\tilde{d}_{it} = d_{it} - \frac{1}{T_i} \sum_{t=1}^{T_i} d_{it} + \frac{1}{\sum_{i=1}^{N} T_i} \sum_{i=1}^{N} \sum_{t=1}^{T_i} d_{it}
$$

Using the above, I construct $8 \times 1$ data moments vector $M$. Similarly, for parameter $\theta$, for $s$-th simulated collection of earnings sample path, I calculate the model-implied moments $\mathcal{M}_s(\theta)$. Similar to the data counterpart, I account for firm fixed effects in the simulated data. Then, I estimate $\theta$ by minimizing SMM-weight weighted distance between data moments and model-implied moments:

$$
\hat{\theta} = \arg \min_{\theta} \left( M - \frac{1}{S} \sum_{s=1}^{S} \mathcal{M}_s(\theta) \right)' W \left( M - \frac{1}{S} \sum_{s=1}^{S} \mathcal{M}_s(\theta) \right)
$$

Here, $W$ is covariance matrix of data-moments after accounting for time-series and intra-industry dependence:

$$
W = \left( \frac{1}{\sum_{i=1}^{N} P_i} \sum_{i=1}^{N} [u_i u_i'] \right)^{-1}
$$

Here, I cluster by industries to account for apparent correlation between firms in the same industry. I use 17 industry definitions from Kenneth French’s website. This clustering strategy also accounts for time-series autocorrelation within firms. This is more conservative than clustering by firms.
where $u_i$ is an $8 \times P_i$ matrix of influence functions. Here, $N$ is the number of industries and $P_i$ is the sample size for the industry $i$.

In calculating standard errors, I correct standard errors for the sampling variability in initially estimating $\mu$. To that end, I update $W$ as follows (Newey and McFadden, 1994):

$$ \tilde{W} = \left( \frac{1}{\sum_{i=1}^{N} P_i} \sum_{i=1}^{N} \left[ \left( u_i(\hat{\mu}) - \frac{\partial u_i(\mu)}{\partial \mu} u_i^\mu \right) \left( u_i(\hat{\mu}) - \frac{\partial u_i(\mu)}{\partial \mu} u_i^\mu \right) \right] \right)^{-1} $$

where $u_i(\hat{\mu})$ is an influence function for 8 moments for given $\hat{\mu}$ and $u_i^\mu$ is an influence function for the earnings growth mean. Then, the standard errors for parameter estimates are given by:

$$ \sqrt{\sum_{i=1}^{N} P_i (\hat{\theta} - \theta_0) \to N \left( 0, \left( \frac{1}{S} \right) \left( (H_0)' \tilde{W} H_0 \right)^{-1} \right) } $$

where $H_0 = E \left[ \frac{\partial M_i(\theta)}{\theta} \right]$.

I first simulate $S = 10$ time-series of the aggregate earnings growth. For each time series of the aggregate earnings growth, I simulate 4,435 firm-specific sample paths as there are 4,435 unique firms in my panel data set. In each simulation, I generate a sample path of 50+$T_i$ quarters long earnings $X_{i,t}$. I discard the first 50 quarters of simulated earnings to reduce solutions’ dependence on $X_{i,t}$ at time $t = 0$. I set $T_i$ to actual time-series length of firm $i$ in order to simulate the unbalanced panel as shown in Figure 9. This small time-series sample bias is important because default probability is significantly sensitive to how long $T_i$ is. For example, for firms whose earnings growth was hit by negative shock, they would not have had enough time to recover under the shorter time, and thus their time-series sample average of default probability would have been larger compared to what would have been if they had been given a longer time.
A.2. Simulated Panel Data

Using the models that are described in Section 1.3, I simulate panel data of leverage and default probability and moderately “taint” the simulated data with a model misspecification. I simulate panel data of leverage and default probability for 4,435 firms to mimic the true number of unique firms in the data. For a sequence of earnings level, \( \{X_{i,t}\} \), firm \( i \)'s observable leverage and observable default probabilities at time \( t \) are:

\[
l(\alpha_i, \eta_i; X_{i,t}) + \epsilon^l_i
\]
\[
d(\alpha_i, \eta_i; X_{i,t})
\]

where \( l \) and \( d \) are aforementioned functions of leverage and default probability, respectively, in terms of \( \alpha_i \) and \( \eta_i \). I randomly draw firm fixed effects \( \epsilon^l_i \) from normal distribution \( \epsilon^l_i \sim \mathcal{N}(0\%, 4\%^2) \) in order to simulate a model misspecification. I use two different cross-sectional distributions of \( \alpha_i \) and \( \eta_i \).

First, I randomly draw \( \eta_i \) and \( \alpha_i \) from truncated normal distributions: \( \eta_i \sim \mathcal{T}\mathcal{N}(7.1\%, 4.2\%^2) \) and \( \alpha_i \sim \mathcal{T}\mathcal{N}(17.4\%, 10.4\%^2) \), respectively. Here, a model misspecification accounts for 13.8% of the total cross-sectional variation for leverage. Second, I randomly draw \( \eta_i \) and \( \alpha_i \) from truncated exponential PDF, \( \lambda_{\alpha} \exp(-\lambda_{\alpha}\alpha_i) \) and \( \lambda_{\eta} \exp(-\lambda_{\eta}\eta_i) \), respectively. \( \lambda_{\alpha} \) and
are chosen to set the cross-sectional mean of $\eta_i$ and $\alpha_i$ equal to 17.0% and 7.0%. Here, a
model misspecification accounts for 3.6% of the total cross-sectional variation for leverage.

Note that the above formulation captures different sources of firm heterogeneity: default
cost, $\alpha_i$, shareholder recovery rate, $\eta_i$, $\epsilon_i^l$, or realized sequence of $X_{i,t}$.

A.3. Mathematical Appendix

A.3.1. Solution

For an arbitrary value for $X_D$, $X_U$ and $C$, I first derive the debt value. Debt is a contingent
claim to an after-tax interest payment. Thus, debt value $D(X)$ satisfies the following ODE:

$$\frac{1}{2} \sigma_X X^2 D'' + \hat{\mu} XD' + (1 - \tau_i) C = rD$$

Boundary conditions are

$$D(X_D) = (1 - \alpha - \eta) \frac{(1 - \tau_{cd}) X_D}{r - \hat{\mu}}$$

$$D(X_U) = D(X_0)$$

The first boundary condition captures that creditors recover only $1 - \alpha - \eta$ fraction of the
remaining unlevered firm value and the second boundary condition captures that creditors
receive the par-value if the debt gets called at the refinancing point. Closed form solution
for debt value is:

$$D(X_t) = \frac{(1 - \tau_i) C}{r} + A_1 X_t^{\lambda^+} + A_2 X_t^{\lambda^-}$$
\[
\lambda_{\pm} = \left(\frac{1}{2} - \frac{\hat{\mu}}{\sigma_X^2}\right) \pm \sqrt{\left(\frac{1}{2} - \frac{\hat{\mu}}{\sigma_X^2}\right)^2 + \frac{2r}{\sigma_X^2}}
\]

Similarly, for an arbitrary value for \(X_D, X_U\) and \(C\), equity value is:

\[
\begin{align*}
E(X_t) &= \sup_{\tau^D} \mathbb{E}^Q \left[ \int_0^{\tau^D} e^{-rs}(1 - \tau_{cd})(X_t - C) ds + e^{-r\tau^D} \cdot E(X_D) \right]
\end{align*}
\]

where \(\tau^D \equiv \inf\{t : X_t \leq X_D\}\).

Here, it is important to note that the above tries to maximize equity value for given coupon amount \(C\). This implies that “optimal” default decision \(X_D\) is made without internalizing the default decision’s impact on cost of debt and leverage. For example, if the default decision was made after internalizing its decision’s impact on cost of debt, the optimal default decision is not to default at all, i.e., \(X_D = \infty\). As firms never choose to default, this effectively makes expected firm-value loss zero and thus firms choose to max out their leverage to enjoy the tax shield benefit. However, this is possible only when shareholders commit to constantly supplying cash by issuing equity even when firms’ earnings are significantly low. This is economically unfeasible and unrealistic and thus I make an assumption that “optimal” default decision was made without regard to its impact on cost of debt and leverage.

Again, following the contingent claims approach, equity value \(E(X)\) satisfies the following ODE:

\[
\frac{1}{2} \sigma_X X^2 E'' + \hat{\mu} X E' + (1 - \tau_{cd})(X - C) = rE
\]

\(^2\)Here, \(A_1 < 0\) and \(A_2 < 0\) where

\[
\begin{bmatrix}
A_1 \\
A_2
\end{bmatrix} = \begin{bmatrix}
X^\lambda_D^{-} & X^\lambda_D^+ \\
X^\lambda_U^- - X^\lambda_U^+ & X^\lambda_U^- - X^\lambda_U^-
\end{bmatrix}^{-1} \begin{bmatrix}
(1 - \alpha - \eta)(1 - \tau_{cd})X_D \\
(1 - \alpha - \eta)(1 - \tau_{cd})X_U - \frac{(1 - \tau_{cd})C}{r}
\end{bmatrix}
\]
Boundary conditions are:

\[
E(X_D) = \frac{\eta(1 - \tau_{cd}) X_D}{r - \hat{\mu}}
\]

\[
E(X_U) = [(1 - \phi_D)D(X_U) + E(X_U)] - D(X_0) = \frac{X_U}{X_0}[(1 - \phi_D)D(X_0) + E(X_0)] - D(X_0)
\]

The first boundary condition captures that shareholder recover \(\eta\) fraction of the remaining unlevered firm value and the second boundary condition captures that shareholders receive the firm value minus debt issuance cost and original debt’s par value. The second equality in the second boundary arises due to homogeneity. Analytical solution for \(E(X_t)\) is:

\[
E(X_t) = \frac{1 - \tau_{cd}}{r - \hat{\mu}} X_t - \frac{(1 - \tau_{cd})C}{r} + B_1 X_t^{\lambda_+} + B_2 X_t^{\lambda_-}
\]

where \(B_1\) represents additional benefit for being allowed to upward refinance and \(B_2\) represents additional benefit for being allowed to default.\(^3\)

The last remaining step is to solve for an optimal coupon \(C\), upward refinancing point \(X_U\) and default threshold \(X_D\). \(C\) and \(X_U\) are determined at time 0 (initial point or refinancing point) by solving the following maximization problem:

\[
[C, X_U] = \arg \max_{C^*, X_U^*} \left( E(X_0; C^*, X_U^*) + (1 - \phi_D)D(X_0; C^*, X_U^*) \right)
\]

The equity value at the time of debt issuance is equal to the total firm value. Thus, shareholders’ incentives are aligned with the maximization of the total firm value. Here, \(X_D\) is determined based on the following smooth pasting conditions (see the heuristic derivation

\[^3\]Here, \(B_1 > 0\) and \(B_2 > 0\) where

\[
\begin{bmatrix}
B_1 \\
B_2
\end{bmatrix} = \left[ \begin{bmatrix}
X_{U}^{\lambda_+} - X_{0}^{\lambda_+} & X_{D}^{\lambda_-} - X_{0}^{\lambda_-} \\
X_{U}^{\lambda_-} - X_{0}^{\lambda_-} & X_{D}^{\lambda_+} - X_{0}^{\lambda_+}
\end{bmatrix}^{-1}
\right]
\left[ \begin{bmatrix}
\frac{X_U}{X_0} (1 - \phi) - 1 \left( A_1 X_0^{\lambda_+} + A_2 X_0^{\lambda_-} + \frac{(1 - \tau_{cd})C}{r} \right) + \frac{X_U}{X_0} \left( \frac{(1 - \tau_{cd})C}{r - \hat{\mu}} X_0 - \frac{(1 - \tau_{cd})C}{r} \right) - \left( \frac{X_U}{X_0} - \frac{(1 - \tau_{cd})C}{r} \right)
\end{bmatrix} \right]
\]
of smooth pasting condition in Appendix A.3.2)

\[ \lim_{X_t \downarrow X_D} E'(X_t) = \frac{\eta(1 - \tau_{cd})}{r - \hat{\mu}} \]

A few points are worth noting here. First, \( X_D \) can be smaller than \( C \), i.e., firms are allowed to costlessly issue equity. Second, as emphasized by Bhamra, Kuehn and Strebulaev (2010), due to fluctuations in the earnings and the assumed cost of refinancing, the firm’s actual leverage drifts away from its optimal target. In the model, the firm is at its optimally chosen leverage ratio only at time 0 and subsequent refinancing dates.

Now, I solve for government tax revenue. Following the contingent claims approach, I solve:

\[ \frac{1}{2} \sigma_X X^2 G'' + \hat{\mu} X G' + (\tau_{cd} X - \tau_{cdi} C) = rG \]

I impose the following boundary conditions:

\[ \lim_{X_t \rightarrow X_D} G(X_t) = 0 \]
\[ \lim_{X_t \rightarrow X_U} G(X_t) = \frac{X_U}{X_0} G(X_0) \]

The first boundary condition specifies that the government does not collect any future tax if firms declare bankruptcy. The second boundary condition captures the homogeneity of the problem. Then, the analytical solution for \( G(X_t) \) is:\(^4\)

\[ \begin{bmatrix} G_1 \\ G_2 \end{bmatrix} = \begin{bmatrix} \frac{X_U}{X_0} X_D^{\lambda_+} - X_U^{\lambda_+} & \frac{X_U}{X_0} X_D^{\lambda_-} - X_U^{\lambda_-} \\ \frac{X_U}{X_0} X_D^{\lambda_+} - X_U^{\lambda_+} & \frac{X_U}{X_0} X_D^{\lambda_-} - X_U^{\lambda_-} \end{bmatrix}^{-1} \begin{bmatrix} -\frac{\tau_{cd}}{r - \hat{\mu}} X_D + \frac{\tau_{cdi} C}{r} \\ \left(\frac{X_U}{X_0} - 1\right) \frac{\tau_{cdi} C}{r} \end{bmatrix} \]

\(^4\)Here, \( G_1 \) and \( G_2 \) satisfy:
A.3.2. Smooth Pasting Condition

As a reminder, a function for equity value is

\[ E(X_t) = \frac{1 - \tau_{cd}}{r - \mu} X_t - \frac{(1 - \tau_{cd}) C}{r} + B_1 X_t^{\lambda_+} + B_2 X_t^{\lambda_-} \]

First, because \( X_D \) is chosen to maximize \( E(X) \), we need to have:

\[ B_1'(X_D) = 0 \text{ and } B_2'(X_D) = 0 \]

Second, the value-matching condition specifies that

\[ \frac{1 - \tau_{cd}}{r - \mu} X_D - \frac{(1 - \tau_{cd}) C}{r} + B_1(X_D) X_D^{\lambda_+} + B_2(X_D) X_D^{\lambda_-} = \frac{\eta(1 - \tau_{cd})}{r - \mu} X_D \]

where \( B_1 \) and \( B_2 \) are functions of \( X_D \). Let us take a derivative of both sides with respect to \( X_D \)

\[ \frac{1 - \tau_{cd}}{r - \mu} + B_1'(X_D) X_D^{\lambda_+} + B_1(X_D) \lambda_+ X_D^{\lambda_+ - 1} + B_2'(X_D) X_D^{\lambda_-} + B_2(X_D) \lambda_- X_D^{\lambda_- - 1} = \frac{\eta(1 - \tau_{cd})}{r - \mu} \]

Substituting \( B_1'(X_D) = 0 \) and \( B_2'(X_D) = 0 \), we have:

\[ \frac{1 - \tau_{cd}}{r - \mu} + B_1(X_D) \lambda_+ X_D^{\lambda_+ - 1} + B_2(X_D) \lambda_- X_D^{\lambda_- - 1} = \frac{\eta(1 - \tau_{cd})}{r - \mu} \]

Thus, we have:

\[ \lim_{X_t \downarrow X_D} E'(X_t) = \frac{\eta(1 - \tau_{cd})}{r - \mu} \]

and this is exactly the smooth pasting condition.
A.3.3. Firm Characteristics

This section summarizes formulas for each firm characteristic. I define the leverage as follows:

\[
\frac{D(X_t)}{D(X_t) + E(X_t)}
\]

Based on Harrison (1985), I define default probability under physical measure as:

\[
DP(X_t) = \begin{cases} 
\Phi \left( \frac{\log \left( \frac{X_t}{X_D} \right) - (\mu - \frac{\sigma^2}{2})T}{\sigma X \sqrt{T}} \right) + \left( \frac{X_t}{X_D} \right)^{1-2(\mu)/\sigma^2} \Phi \left( \frac{\log \left( \frac{X_t}{X_D} \right) + (\mu - \frac{\sigma^2}{2})T}{\sigma X \sqrt{T}} \right) & \text{if } X_t \geq X_D \\
1 & \text{Otherwise}
\end{cases}
\]

where I set \( T = 1 \).

Next, I discuss the formula for the market beta. The equity return is:

\[
dR_t = \frac{dE(X_t) + (1 - \tau_{cd})(X_t - C)dt}{E(X_t)} \\
= \left( (1 - \tau_{cd})(X_t - C) \right) + \frac{E'(X_t)X_t}{E(X_t)} \mu + \frac{1}{2} \frac{E''(X_t)X_t^2}{E(X_t)} \sigma^2_X + \frac{E'(X_t)X_t}{E(X_t)} (b \sigma_A dW^A_t + \sigma_F dW^F_t)
\]

Let \( x_t^A \) be a log of aggregate earnings \( X_t^A \). Then,

\[
dx_t^A = \sigma_A dW^A_t
\]

Using this, a term for market beta is:

\[
\text{Market beta} = \frac{1}{dt} \mathbb{E}_t [dx_t^A dR_t] / \frac{1}{dt} \text{var}_t(dx_t^A) = \frac{E'(X_t)X_t}{E(X_t)} b
\]
Lastly, I define credit spread as:

\[ \frac{C}{D(X_t)} \]

### A.3.4. Proof

This subsection lists all of the proofs for all of the propositions when upward refinancing is suppressed. I allow upward refinancing in the simulation and numerically show that the same intuition still carries through.

\[ D(X_t) = (1 - \tau_i) \frac{C}{r} + \tilde{A}_2 \left( \frac{X_t}{C \cdot X_{DC}} \right)^{\lambda_c} \]

where

\[ \tilde{A}_2 / C = -\frac{1 - \tau_i}{r} + (1 - \alpha - \eta) \frac{(1 - \tau_{cd}) X_{DC}}{r - \hat{\mu}} < 0 \]

On the contrary, as \( \eta \) increases, \( E(X_t) \) increases because shareholders gain \( \eta \) and \( X_{DC} \) is determined to maximize \( E(X_t) \). Similarly, this increase gets magnified by larger \( \left( \frac{X_t}{C \cdot X_{DC}} \right)^{\lambda_c} \).

\[ E(X_t) = \frac{1 - \tau_{cd}}{r - \hat{\mu}} X_t - \frac{(1 - \tau_{cd}) C}{r} + \tilde{B}_2 \left( \frac{X_t}{C \cdot X_{DC}} \right)^{\lambda_c} \]

\[ \tilde{B}_2 / C = \frac{1 - \tau_{cd}}{r} + (\eta - 1) \frac{(1 - \tau_{cd}) X_{DC}}{r - \hat{\mu}} > 0 \]

**Proof** of Proposition 2: I can approximately write the leverage as:

\[ lev = \frac{D(X_t)}{E(X_t) + D(X_t)} \approx \frac{(1 - \tau_i) C}{r + \frac{(1 - \tau_{cd}) X_t + \tau_{cd} C}{r}} \]

where the second approximate equality exists because default probability is typically very
small. Then, I can write partial derivative terms as:

\[
\frac{\partial \text{lev}}{\partial \eta} = \frac{\partial \text{lev}}{\partial C} \frac{\partial C}{\partial \eta} \\
\frac{\partial \text{lev}}{\partial \alpha} = \frac{\partial \text{lev}}{\partial C} \frac{\partial C}{\partial \alpha}
\]

Because \( \frac{\partial \text{lev}}{\partial C} > 0 \), proving \( \frac{\partial \text{lev}}{\partial \eta} < 0 \) and \( \frac{\partial \text{lev}}{\partial \alpha} < 0 \) is equivalent to proving that \( \frac{\partial C}{\partial \eta} < 0 \) and \( \frac{\partial C}{\partial \alpha} < 0 \). So, let us focus on terms for \( C \). The optimization problem to solve for \( C \) is as follows:

\[
C = \arg \max_{C^*} \left\{ \frac{1 - \tau_{ci}}{r - \mu} X_0 + \frac{\tau_{cd} - \phi_D(1 - \tau_i)}{r} C^* + \left( \frac{X_0}{X_{DC}} \right)^{\lambda_+} (1 - \phi_D) A_{2c} + B_{2c} C^{(1 - \lambda_+)} \right\}
\]

where \( A_{2c} = A_2/C \) and \( B_{2c} = B_2/C \). The closed form solution for optimal coupon \( C \) is

\[
C = \left[ \frac{\tau_{cd} - \phi_D(1 - \tau_i)}{r} \right]^{-1/\lambda_-} \cdot \frac{X_0}{X_{DC}} \cdot \left[ \left( \frac{X_0}{X_{DC}} \right)^{\lambda_-} (1 - \phi_D) A_{2c} + B_{2c} \right]^{1/\lambda_-}
\]

(A.1)

Because \( \frac{\partial C}{\partial \alpha} < 0 \) and \( \frac{\partial C}{\partial \eta} < 0 \), I can say that \( C \) decreases over \( \alpha \) and \( \eta \). Intuitively, the denominator of the second term shows that \( C \) decreases as shareholders strategically determine high threshold \( X_{DC} \). High \( X_{DC} \) implies a high default probability thus a high expected firm-value loss and low optimal \( C \). The third term represents the loss of firm value upon bankruptcy adjusted for debt issuance cost. High loss of firm value \((1 - \phi_D) A_{2c} + B_{2c}\) implies low \( C \). ■

**Proof of Proposition 3** As shown in Section A.3.3, for given \( \mu, b \) and \( \sigma \), there is monotonic relation between default probability and \( X_D \) (default threshold). Thus, comparative statistics between default probability and \( \eta \) and \( \alpha \) is equivalent to that between \( X_D \) and \( \alpha \)
and \( \eta \). Using closed forms for \( C \) and \( X_{DC} \), I derive closed-form terms for \( X_D \):

\[
\frac{X_D}{X_0} = \left[ \frac{\tau_{cd} - \phi_D (1 - \tau_i)}{\tau - (1 - \tau_i)(1 - \phi_D) A_{2c} + B_{2c}} \right]^{-1/\lambda_-} \cdot \left[ (1 - \lambda_-)((1 - \phi_D) A_{2c} + B_{2c}) \right]^{1/\lambda_-}
\]

Here, \( X_D \) decreases over \( \alpha \) and \( \eta \) because \( \frac{\partial X_D}{\partial \alpha} < 0 \) and \( \frac{\partial X_D}{\partial \eta} < 0 \). Intuitively, for given \( C \), the rise in \( \eta \) increases both default probability and value loss. Thus, \( C \) has to decrease sufficiently enough to offset high expected firm-value loss driven by an increase in both default probability and value loss. Thus, the decrease in \( C \) more than offsets the increase in \( X_{DC} \). As a result, \( X_D \) decreases over \( \eta \) and so does default probability. ■

**Proof of Proposition 5:** Now, I prove for market beta:

\[
\text{Market beta} = b \frac{E'X}{E}
\]

\[
E'X = \frac{1 - \tau_{cd}}{r - \mu} X + B_2 \lambda_- \left( \frac{X}{C \cdot X_{DC}} \right)^{\lambda_-}
\]

\[
E = \frac{1 - \tau_{cd}}{r - \mu} X - \frac{(1 - \tau_{cd})C}{r} + B_2 \left( \frac{X}{C \cdot X_{DC}} \right)^{\lambda_-}
\]

Thus, I have:

\[
\text{Market beta} = \frac{E'(X_t)X_t}{E(X_t)} b = \left( 1 + \frac{(1 - \tau_{cd})C}{r \cdot E} + (\lambda_- - 1) \frac{B_2}{E} \left( \frac{X_t}{C \cdot X_{DC}} \right)^{\lambda_-} \right) b
\]

Because \( \frac{\partial \text{Market beta}}{\partial \eta} < 0 \), this proves the proposition. Intuitively, conditional on leverage, shareholders’ strategic action first decreases market beta. After allowing firms to de-lever in response, its distress risk decreases and market beta further decreases. One interesting point to note is that as \( X \) gets very close to \( X_D \), the market beta becomes less sensitive to \( \eta \) because equity gets converted to a fixed fraction of unlevered asset value in default. ■

**Proof of Proposition 6:** I set \( \phi_D = 0 \) and show closed-form expression for the credit spread.
\[
\frac{C}{D(X_0)} = \frac{1}{\frac{1-\tau_i}{r} + A_2/C \left( \frac{X_0}{X_D} \right)^{\lambda_+} - 1} \\
= \frac{1-\tau_i}{r} + \frac{-1}{\tau(D) - \eta} \left( 1 + \frac{1-\lambda_+}{1-\lambda_-} \right) \left( 1 + \frac{1-\lambda_+}{1-\lambda_-} \right)
\]

It is immediately clear that the credit spread decreases over \( \eta \). Intuitively, lower default probability caused by the commitment problem more than offsets creditors’ higher value loss (normalized by \( C \)) and results in a lower credit spread. Again, this commitment problem exists because \( \alpha \) is non-zero and thus the credit spread decreases over \( \eta \). If \( \alpha = 0 \), then the above formula clearly tells you that the credit spread does not change over \( \eta \). ■

**Proof of Proposition 7:** Let me prove the proposition by illustrating my points in Figure 1. I need to prove that point A always corresponds to higher firm value than point C does. To that end, I prove this by contradiction. Let us assume otherwise: point C’s firm value is greater than point A’s. Now, pick a point D on the solid curve that has the same coupon rate as point C. Because the solid curve always sits above the dotted curve (\( (1 - \phi_D)D(X_t) + E(X_t) \) always decreases as \( \eta \) increases), point D’s firm value is greater than point C’s. This implies that point D’s firm value is greater than point A’s. This contradicts that point A is the optimal point on the solid curve. This completes the proof.

■

**Proof of Proposition 8:** Government tax revenue \( G(X_t) \) is:

\[
G(X_t) = \frac{\tau_{cd}}{r - \hat{\mu}} X_t - \frac{\tau_{cdi}}{r} C + \tilde{G}_2 \left( \frac{X_t}{C \cdot X_{DC}} \right)^{\lambda_-}
\]

where

\[
\tilde{G}_2/C = \frac{\tau_{cdi}}{r} - \frac{\tau_{cd}}{r - \hat{\mu}} X_{DC}
\]

As \( \eta \) increases, \( C \) decreases and thus government tax revenue \( G(X_t) \) increases and this
completes the proof. On the related note, $\tilde{G}_2$ illustrates an interesting intuition. Upon firms’ default, even though the government loses potential tax revenue on the firm’s future income stream, the government is no longer exploited by corporations.

Proof of Proposition 4: Here, I prove that leverage and $DP$ (default probability) have different sensitivities with respect to $\eta$ and $\alpha$.

$$\frac{\partial DP}{\partial \eta} = \frac{(\partial DP/\partial X_D)\partial X_D/\partial \eta}{(\partial DP/\partial X_D)\partial X_D/\partial \alpha} = \frac{\partial X_D/\partial \eta}{\partial X_D/\partial \alpha}$$

where $\frac{\partial X_D}{\partial \eta}$ and $\frac{\partial X_D}{\partial \alpha}$ are:

$$\frac{\partial X_D}{\partial \eta} = \frac{\partial (C \cdot X_{DC})}{\partial \eta} = C \cdot \frac{\partial X_{DC}}{\partial \eta} + X_{DC} \cdot \frac{\partial C}{\partial \eta}$$

$$\frac{\partial X_D}{\partial \alpha} = \frac{\partial (C \cdot X_{DC})}{\partial \alpha} = C \cdot \frac{\partial X_{DC}}{\partial \alpha} + X_{DC} \cdot \frac{\partial C}{\partial \alpha} = X_{DC} \cdot \frac{\partial C}{\partial \alpha}$$

where the last equality holds because $\frac{\partial X_{DC}}{\partial \alpha} = 0$.

Now, let us think of a leverage.

$$lev = \frac{D(X_0)}{E(X_0) + D(X_0)} \approx \frac{(1-t_\theta)C}{r} \frac{X_0}{r} + \frac{\tau_{\theta} \cdot C}{r}$$

$$\Rightarrow \frac{\partial lev}{\partial \eta} = \frac{(\partial lev/\partial C)\partial C/\partial \eta}{(\partial lev/\partial C)\partial C/\partial \alpha} = \frac{\partial C/\partial \eta}{\partial C/\partial \alpha}$$

Thus,

$$\frac{\partial DP}{\partial \eta} \neq \frac{\partial lev}{\partial \eta}$$

Two points are worth making. First, we point out that $\frac{\partial DP}{\partial \eta} < \frac{\partial lev}{\partial \eta}$ because $X_{DC} \frac{\partial C}{\partial \eta} < \frac{\partial X_D}{\partial \eta} < 0$. Second, when $\frac{\partial X_{DC}}{\eta} = 0$, the above becomes equality and implies that leverage and default probability cannot separately identify $\alpha$ and $\eta$. ■
A.3.5. Upward Refinancing

Even when upward refinancing is allowed, the economic channels, discussed in Section 1.3.2, still hold. As debt becomes more costly, firms internalize higher costs and optimally choose to de-lever and refinance less frequently.

![Figure 10: Illustration of trade-off theory when upward refinancing is allowed](image)

When \( \eta = 0 \) (left panel), firm value is maximum at 80.23 when optimal coupon \( C \) is 0.66 and \( X_U \) is 3.7. However, \( \eta \neq 0 \) (right panel), firm value is maximum at 79.75 when coupon \( C \) is 0.59 and \( X_U = 3.75 \). This clearly illustrates that high \( \eta \) implies lower firm value, lower leverage and less frequent upward refinancing.

A.3.6. The Whole Economy

Here, I examine the case when upward refinancing is suppressed. I show that the value of the entire economy, \( D(X) + E(X) + G(X) \), increases over \( \eta \) as lower default frequencies lead to smaller loss of default cost. This can be easily seen below:

\[
D(X) + E(X) + G(X) = \frac{X}{r - \bar{\mu}} + (A_2 + B_2 + G_2) \left( \frac{X}{C \cdot X_{DC}} \right)^{\lambda -}
\]
where

\[(A_2 + B_2 + G_2)/C = -\left[\alpha(1 - \tau_{cd}) + \tau_{cd}\right] \cdot \frac{X_{DC}}{r - \bar{\mu}}\]

Here, \(\alpha(1 - \tau_{cd})\frac{X_{DC}}{r - \bar{\mu}}\) is the deadweight loss from the firms as a whole whereas \(\tau_{cd}\frac{X_{DC}}{r - \bar{\mu}}\) is the government’s loss of future tax revenue.

A.4. Data Variables

A.4.1. Firm-level Variable Definitions

Variables Excluding Default Probabilities

- Earnings growth: \(\tilde{e}_{i,t+1} = \log\left(\frac{\sum_{j=0}^{K} OIADPQ_{i,t+1-j}}{\sum_{j=0}^{K} OIADPQ_{i,t-j}} - 1\right)\) where \(K\) is set to 8 and \(OIADPQ\) is operating income after depreciation.

- Market beta: calculated based on rolling window of 24 months of monthly returns.

- Leverage: \(\frac{DLTTQ + DLCQ}{DLTTQ + DLCQ + ME}\) where \(DLTTQ\), \(DLCQ\) and \(ME\) are long-term debt, short-term debt and market equity, respectively.

Default Probabilities

Because level of default probability is an important matching moment, it warrants a separate discussion. At large, there are two ways to derive default probability. The first is the Merton distance-to-default model, which is based on Merton (1974). The second is based on the hazard model and is used by several papers including Campbell, Hilscher and Szilagyi (2008). I use the former approach, which is more compatible with the model-implied moments that use Merton-style default probability. Specifically, I follow Bharath and Shumway (2008) to construct default probability, which is found to closely match various corporate
default probability measures. Its definition is

\[
\pi = \Phi \left( -\log \frac{V}{B} - \frac{(\mu_v - \sigma^2_v)}{2} \right)
\]

where \( \Phi \) is a cumulative normal distribution function, \( V \) is the market value of assets, \( B \) is the amount of debt that’s due for that quarter, \( \mu_v \) is the expected asset return and \( \sigma_v \) is the asset return volatility. Because \( V, B, \mu_v \) and \( \sigma_v \) are all unobservable, each of these warrant a separate discussion.

First, let us discuss how I derive \( B \) and \( V \). In order to derive distance-to-default over the next one year, Campbell, Hilscher and Szilagyi (2008) and Vassalou and Xing (2004) assume that short-term debt plus one half long-term debt come due in a year. As Campbell et al. noted, “This convention is a simple way to take account for the fact that long-term debt may not mature until after the horizon of the distance to default calculation.” Extending upon this convention, in order to derive distance-to-default in the next quarter as opposed to next year, Gomes, Grotteria and Wachter (2018) assume that one quarter of Campbell et al.’s comes due within the next quarter. Accordingly, I set \( B \) to \( DLCQ/4 + DLTTQ/8 \).

Due to the lack of data on market value of debt, I use \( B \) to proxy market value of debt and set \( V \) to market value of equity plus \( B \). Second, following Bharath and Shumway (2008), I use monthly equity returns to calculate average equity returns and set it to \( \mu_v \). Third, following Bharath and Shumway (2008), I set \( \sigma_v \) to

\[
\sigma_v = \frac{E}{E+F} \sigma_E + \frac{F}{E+F} (0.05/\sqrt{3} + 0.25 \cdot \sigma_E)
\]

where \( \sigma_E \) is the quarterly volatility of the equity returns and \( 0.05/\sqrt{3} + 0.25 \cdot \sigma_E \) is the quarterly volatility of the bond that is due in a quarter. I estimated \( \sigma_E \) using the daily returns of trailing 3 months.
Tax Rates

First, I augment the sample with panel data of corporate marginal tax rates, which were constructed according to Graham (1996a,b). They provide both before-financing marginal tax rates (MTR) and after-financing MTR. Both measure firms’ MTR by incorporating many features present in the tax code, such as tax-loss carryforwards and carrybacks, the investment tax credit, and the alternative minimum tax. Before-financing MTR are based on taxable income before financing expenses are deducted, whereas after-financing MTR are based on taxable income after financing expenses are deducted. As Graham (1998) argues, by construction, after-financing MTR are endogenously affected by the choice of financing. Because the model treats \( \tau_c \) exogenous of firms’ financing decision, this paper uses before-financing MTR to set corporate earnings tax rates \( \tau_c \).

Second, I closely follow Graham (2000) to construct \( \tau_i \) and \( \tau_d \). As documented in Graham (2000), I set \( \tau_i = 47.4\% \) for 1981 or prior, 40.7\% between 1982 and 1986, 33.1\% for 1987, 28.7\% between 1988 and 1992, and 29.6\% afterwards. Based on these estimates for \( \tau_i \), I estimate \( \tau_d \) as \( [d + (1 - d)g\alpha] \tau_i \). The dividend-payout ratio \( d \) is the firm-quarter-specific dividend distribution divided by trailing twelve-quarters moving average of earnings. Since \( d \) needs to be less than or equal to 1, if \( d \) is greater than 1, I set it to 1. If dividend is missing, I set \( d = 0 \). The proportion of long-term capital gains that is taxable \( (g) \) is 0.4 before 1987 and 1.0 afterwards. I assume that the variable measuring the benefits of deferring capital gains, \( \alpha \), equals 0.25. The long-term capital gains rate, \( g \tau_i \), has a maximum value of 0.28 between 1987 and 1997, 0.2 between 1998 and 2003 (Taxpayer Relief Act of 1997) and 0.15 afterwards (Jobs and Growth Tax Relief Reconciliation Act of 2003).

It is worth noting that \( \tau_c \) is different across firms because firms face different tax-loss carryforwards/carrybacks, the investment tax credit and the alternative minimum tax. \( \tau_d \)

---

5I would like to thank John Graham for sharing panel data of corporate marginal tax rates. [https://faculty.fuqua.duke.edu/~jgraham/taxform.html](https://faculty.fuqua.duke.edu/~jgraham/taxform.html). I impute missing marginal tax rates with the time-series average for each firm if the firm is covered in Graham’s database or panel-wide-average if the firm is not covered at all.
is different across firms because dividend-payout ratios are different. However, for given year, \( \tau_i \) is the same across firms because I assume that marginal investors face the same \( \tau_i \). Also, I assume that \( \tau_c \) and \( \tau_i \) stay constant for all four quarters for any given year (due to data limitation) whereas \( \tau_d \) can potentially change every quarter due to varying dividend-payout ratios.

A.4.2. Aggregate Variables

Variable Definitions

- Aggregate earning: Source: NIPA, Section 1, Table 1.14, Series: Line 8 Net Operating Surplus, Quarterly series from 1947Q1 to 2016Q4

- Consumer price index: Source: FRED, Series: CPIAUCNS (Consumer Price Index for All Urban Consumers: All Items), Monthly series from 1913Jan through 2016Dec

- Nominal risk free rate: Source: FRED, Series: TB3MS (3-Month Treasury Bill: Secondary Market Rate), Monthly series from 1934Jan through 2016Dec

Variable Construction

- Realized Inflation\(=\frac{\text{CPI}(t) - \text{CPI}(t-1)}{\text{CPI}(t-1)}\) where \(\text{CPI}(t)\) is the consumer price index in year-quarter \(t\) computed as the average monthly CPI for that year-quarter.

A.5. Magnitude of Credit Spread

Using the estimates reported in Table 5, as noted in Table 6’s column (5), quarterly credit spreads are 191 bp. These credit spreads are much larger than what is empirically observed and contrast with Morellec, Nikolov and Schurhoff (2012)’s success in matching quarterly credit spreads at 53bp. The main reason for this discrepancy is the value used for earnings
growth volatility, $\sigma^F$. Using $\sigma^F = 0.1380^6$ that is reported in Morellec et al. and keeping everything else equal, my model-implied quarterly credit spreads are 56bp, which is very close to Morellec’s. The large sensitivity of credit spreads with respect to $\sigma^F$ is noted in Morellec et al.’s Table 1.

However, lowering $\sigma^F$ decreases quarterly default probabilities from 0.36% to 0.003%, which is counterfactually small. More importantly, $\hat{\sigma}^F = 0.2444$ is a relatively conservative measure, as implied earnings growth variance is under-matched relative to its data counterpart (see Table 4). More specifically, if $\sigma^F$ is chosen just to match earnings growth moments, then $\hat{\sigma}^F = 0.2786$, which is apparently greater than $\hat{\sigma}^F = 0.2444$. Lastly, the current model features bonds with infinite maturity and thus is not suitable to match data counterparts that have finite maturity. Noting this limitation, it is still interesting to study how much credit spreads change over $\eta$ and thus I continue reporting such results.

### A.6. Bankruptcy Reform Act 1978 (BRA)

First, let us review the literature’s stance on how BRA changed shareholder recovery rate. Hackbarth, Haselmann and Schoenherr (2015) argues that BRA increased shareholder recovery rate due to four specific clauses. First, relative to the old code, BRA added equity as one additional class to confirm a reorganization plan. Second, managers were given a 120-day exclusivity period to propose the plan. Third, if no plan could be agreed upon, a new procedure, called cramdown, allowed firms to continue operating while a buyer was sought. This was considered a costly and time-consuming process and thus acted as a disciplinary tool in negotiations in favor of shareholders. Lastly, firms could now declare bankruptcy even when firms were solvent, thus shareholders can use the threat of bankruptcy as a strategic tool against creditors. Thus, BRA increased shareholder recovery rate.

Now, let us discuss how BRA could have changed default cost. Prior to 1978, as discussed in Section 1.2, an increasing number of firms sought to file under shareholder-friendly Chapter

---

^6^Morellec et al. report annual volatility of earnings growth at 28.86%. Thus, $\sigma^F = \sqrt{0.1443^2 - \langle b\sigma^A \rangle^2} = 0.1380$. 

111
11 rather than Chapter 10. However, applying for Chapter 11 required an expensive hearing (LoPucki and Whitford, 1990). Moreover, as discussed in Section 1.2, the prior bankruptcy laws were considerably ambiguous (Posner, 1997; King, 1979). BRA addressed both of these issues. BRA permitted creditors to take less than full payment, in order to expedite or insure the success of the reorganization.\(^7\) This effectively made it easy to deviate from APR. BRA reduced the ambiguity present in the bankruptcy law by spelling out a number of provisions important in enabling the shareholders to reorganize (Skeel, 2001)\(^8\) and this could have reduced friction in the bankruptcy process. Taken together, this could decrease time spent in bankruptcy and thus reduce lawyers’ fees, which are typically charged by the hour, and opportunity cost caused by delayed investment due to uncertain future and loss of business relationships with customers and suppliers. In other words, BRA could have effectively decreased default cost.\(^9\) However, the above argument goes against one of the notorious bankruptcy cases in the post-BRA era, specifically that of Eastern Airlines, which lost 50% of its value during bankruptcy (Weiss and Wruck, 1998). Again, it is not clear how BRA changed default cost. Obvious way to test my hypothesis is to check how BRA changed time spent in bankruptcy. However, data on bankruptcy cases during pre-BRA are limited and thus makes it hard to compare. This warrants a need for a structural estimation that can identify changes in unobservable firm characteristics.

\(^7\)H.R. Rep No. 595, 95th Cong., 1st Sess. 224 (1978)

\(^8\)The list includes an automatic stay, an exclusive period, the ability to use cash collateral and/or obtain post-petition financing, the ability to assume or reject leases and other executory contracts, the ability to sell assets free and clear of liens, the ability to retain and compensate key employees and the ability to reject or renegotiate labor contracts and pension benefits.

\(^9\)This explanation is consistent with the literature’s use of time spent in bankruptcy as a proxy for default cost (Bris, Welch and Zhu, 2006).
APPENDIX for “Measurement Error in Multiple Equations”

B.1. Restricting the Correlations among the Disturbances

We extend $A_6$ to $A'_6$ which restricts the sign and/or magnitude of the correlation $r_{\eta_j,\eta_h}$ between $\eta_j$ and $\eta_h$.

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

In particular, provided $\sigma^2_{\tilde{Y}_j} \sigma^2_{\tilde{Y}_h} \neq 0$, from the proof of Corollary B.1.1 we have that

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

$A'_6$ may restrict the sign of $r_{\eta_j,\eta_h}$ 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 $r_{\eta_j,\eta_h}$ (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 B.1.1, when $R^2_{\tilde{Y}_j,\tilde{Y}_h} \neq 1$, the discriminant of the quadratic function $M_{jh}(\cdot; c)$ is given by

$$\Delta_{jh}(c) \equiv c^2[R^4_{\tilde{W},(\tilde{Y}_j,\tilde{Y}_h)}(1 - R^2_{\tilde{Y}_j,\tilde{Y}_h})^2 - (1 - c^2)(R^2_{\tilde{W},\tilde{Y}_j} - R^2_{\tilde{W},\tilde{Y}_h})^2],$$
and, when $R_{y_j\gamma h}^2 \neq c^2$, the roots of $M_{jh}(\cdot; c)$ are given by

$$\rho_{jh}^-(c) = \frac{F_{jh}(c) - \Delta_{jh}(c)\frac{1}{2}}{2(R_{y_j\gamma h}^2 - c^2)}$$

and

$$\rho_{jh}^+(c) = \frac{F_{jh}(c) + \Delta_{jh}(c)\frac{1}{2}}{2(R_{y_j\gamma h}^2 - c^2)}$$

where

$$F_{jh}(c) \equiv -R_{W,(\gamma h)}^2(1 - R_{y_j\gamma h}^2) + (1 - c^2)(R_{W,y_j}^2 + R_{W,\gamma h}^2).$$

Corollary B.1.1 uses $S_{jh}(r)$ and $M_{jh}(r)$ to encode the sign and magnitudes restrictions in $A_0'$ and to express the identification region for $(\rho, \delta, \beta, \Gamma)$ under $A_1 - A_0'$.

**Corollary B.1.1** Under the conditions of Theorem 2.3.1, $A_4$, $A_5$, and $A_0'$ for $j, h = 1, ..., p$ with $j < h$, $(\rho, \delta, \beta, \Gamma)$ is partially identified in the sharp set

$$J_{k, \tau, e} \equiv \{(r, D(r), B(r), G(r)) : 0 \leq G(r), \frac{1}{1 + \kappa} \leq r \leq 1, \frac{\sigma_{y_j}^2}{\sigma_{W_j}^2}(1 - \tau_j) \leq G_{jj}(r),$$

$$\text{and } \varsigma_{jh} \leq \frac{G_{jh}(r)}{(G_{jj}(r)G_{hh}(r))^{\frac{1}{2}}} \leq \bar{\varsigma}_{jh} \text{ for } j, h = 1, ..., p \text{ and } j < h \}.$$  

Further, $\rho$ is partially identified in the sharp set

$$R_{k, \tau, e}^{\rho} = [R_{W,\gamma h}^2, 1] \cap \left[ \frac{1}{1 + \kappa}, 1 \right] \cap_{j=1}^{p} \left[ \frac{1}{\tau_j} R_{W, y_j}^2, 1 \right] \cap_{j=h}^{p} \cap_{j<h} R_{\rho, \gamma h}^{\rho},$$

with

$$R_{\rho, jh}^{\rho} = \left\{ \begin{array}{ll}
S_{jh}(r) \leq 0 \text{ and } M_{jh}(r; \varsigma_{jh}) \leq 0 \leq M_{jh}(r; \bar{c}_{jh}) & \text{if } \varsigma_{jh} \leq \bar{c}_{jh} \leq 0 \text{ and } \sigma_{y_j}^2 \sigma_{y_h}^2 \neq 0 \\
\{ S_{jh}(r) \leq 0 \text{ and } M_{jh}(r; \varsigma_{jh}) \leq 0 \} \text{ or } & \text{if } \varsigma_{jh} < 0 < \bar{c}_{jh} \text{ and } \sigma_{y_j}^2 \sigma_{y_h}^2 \neq 0 \\
\{ 0 \leq S_{jh}(r) \text{ and } M_{jh}(r; \bar{c}_{jh}) \leq 0 \} & \text{if } 0 \leq \varsigma_{jh} \leq \bar{c}_{jh} \text{ and } \sigma_{y_j}^2 \sigma_{y_h}^2 \neq 0 \\
0 \leq S_{jh}(r) \text{ and } M_{jh}(r; \bar{c}_{jh}) \leq 0 \leq M_{jh}(r; \varsigma_{jh}) & \text{if } 0 \notin [\varsigma_{jh}, \bar{c}_{jh}] \text{ and } \sigma_{y_j}^2 \sigma_{y_h}^2 = 0.
\end{array} \right\},$$

114
where, provided $\sigma^2_{W_j}, \sigma^2_{Y_h} \neq 0$, we have

$$0 \leq S_{jh}(r) \iff \begin{cases} \frac{r_{W_j,Y_h} r_{W_j,Y_h}}{r_{Y_j,Y_h}} \leq r & \text{when } 0 < r_{Y_j,Y_h}, \\ r_{W_j,Y_h} r_{W_j,Y_h} \leq 0 & \text{when } r_{Y_j,Y_h} = 0, \\ r \leq \frac{r_{W_j,Y_h} r_{W_j,Y_h}}{r_{Y_j,Y_h}} & \text{when } r_{Y_j,Y_h} < 0 \end{cases}$$

and if $R^2_{Y_j,Y_h} = 1$ then $0 \leq M_{jh}(r;c) = (1 - c^2)(r - R^2_{W_j,Y_h})^2$ whereas if $R^2_{Y_j,Y_h} \neq 1$ then

$$0 \leq M_{jh}(r;c) \iff \begin{cases} -\infty < r < \infty & \text{when } 0 < c^2 < 1 - \frac{R^4_{W_j,Y_h} (1 - R^2_{Y_j,Y_h})^2}{(R^2_{W_j,Y_h} - R^4_{W_j,Y_h})^2} \\
\rho_{jh}^-([\rho_{jh}(c), \infty)) \cup \rho_{jh}^+([0, \rho_{jh}(c))] & \text{when } c^2 = 0 < R^2_{Y_j,Y_h} \text{ or } 1 - \frac{R^4_{W_j,Y_h} (1 - R^2_{Y_j,Y_h})^2}{(R^2_{W_j,Y_h} - R^4_{W_j,Y_h})^2} \leq c^2 < R^2_{Y_j,Y_h} \\
\rho_{jh}^- \left( \frac{r_{W_j,Y_h} r_{W_j,Y_h}}{r_{Y_j,Y_h}} \right) \leq r & \text{when } c^2 = R^2_{Y_j,Y_h} \text{ and } R^2_{W_j,Y_h} < R^2_{W_j,Y_h} + R^2_{W_j,Y_h} \\
\rho_{jh}^- \left( \frac{r_{W_j,Y_h} r_{W_j,Y_h}}{r_{Y_j,Y_h}} \right) \leq r & \text{when } c^2 = R^2_{W_j,Y_h} \text{ and } R^2_{W_j,Y_h} = R^2_{W_j,Y_h} + R^2_{W_j,Y_h} \\
\rho_{jh}^- \left( \frac{r_{W_j,Y_h} r_{W_j,Y_h}}{r_{Y_j,Y_h}} \right) \leq r & \text{when } c^2 = R^2_{W_j,Y_h} \text{ and } R^2_{W_j,Y_h} < R^2_{W_j,Y_h} \text{ and } R^2_{W_j,Y_h} < R^2_{W_j,Y_h} \\
r \in [\rho_{jh}(c), \rho_{jh}(c)] & \text{when } R^2_{Y_j,Y_h} < c^2 \end{cases}$$

Last, $\delta$, $\beta$, and $\Gamma$ are partially identified in the sharp sets $D^{k,\tau,c} = \{D(r) : r \in R^{k,\tau,c}\}$, $B^{k,\tau,c} = \{B(r) : r \in R^{k,\tau,c}\}$, and $G^{k,\tau,c} = \{G(r) : r \in R^{k,\tau,c}\}$.

The bounds in Corollary B.1.1 yield those from Corollary 2.3.2 when $(\xi_{gh}, \xi_{gh})$ is set to $(-1,0)$, $(0,1)$, $(0,0)$, or $(-1,1)$. In particular, when $R^2_{Y_j,Y_h} \neq 0$, the proof of Corollary B.1.1 gives

$$\rho_{jh}^- (0) = \rho_{jh}^+ (0) = \frac{r_{W_j,Y_h} r_{W_j,Y_h}}{r_{Y_j,Y_h}} = \frac{\sigma_{W_j,Y_h} \sigma_{W_j,Y_h}}{\sigma_{W_j,Y_h}^2} \sigma_{Y_j,Y_h},$$

so that $0 \leq M_{jh}(\rho;0) \iff \rho \in (-\infty, \infty)$ and $M_{jh}(\rho;0) \leq 0 \iff \rho = \frac{\sigma_{W_j,Y_h} \sigma_{W_j,Y_h}}{\sigma_{W_j,Y_h}^2} \sigma_{Y_j,Y_h}$. Also,

$$\rho_{jh}^- (-1) = \rho_{jh}^+ (1) = R^2_{W_j,Y_h} \text{ and } \rho_{jh}^- (1) = \rho_{jh}^+ (1) = 0.$$
Thus, when \( R_{\tilde{y}_j, \tilde{y}_h}^2 < 1, M_{j_h}(\rho; 1) = M_{j_h}(\rho; -1) \leq 0 \Leftrightarrow \rho \in (-\infty, 0] \cup [R_{\tilde{W},(\tilde{y}_j, \tilde{y}_h)'}, \infty). \) Since \( R_{\tilde{W},(\tilde{y}_j, \tilde{y}_h)'}^2 \leq R_{\tilde{W},\tilde{y}}^2, \) this magnitude restriction is not binding in \( \mathcal{R}_{j_h}^{k,.c}. \) It follows that Corollary B.1.1 yields the same bound \( \mathcal{R}_{j_h}^{k,.c} \) from Corollary 2.3.2, with \( \mathcal{R}_{j_h}^{c} \) determined by the magnitude restriction encoded in \( M_{j_h}(\rho; 0) \leq 0 \) when \( c_{j_h} = \tilde{c}_{j_h} = 0 \) and by the sign restrictions, if any, encoded in \( S_{j_h}(r) \) otherwise.

B.2. Supplementary Material on Inference

B.2.1. Algorithm for Inference on \( \rho \)

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

\[
g_l^v(\pi; r, \lambda) = (g_{1l}^v(\pi; r, \lambda), \ldots, g_{Ml}^v(\pi; r, \lambda)) \quad \text{where} \quad g_{vl}^v(\pi; r, \lambda) \equiv r - \rho_{vl}(\lambda) \quad \text{for} \quad v = 1, \ldots, M, \quad \text{and} \quad g_u^v(\pi; r, \lambda) = (g_{1u}^v(\pi; r, \lambda), \ldots, g_{Mu}^v(\pi; r, \lambda)) \quad \text{where} \quad g_{vu}^v(\pi; r, \lambda) \equiv \rho_{vu}(\lambda) - r \quad \text{for} \quad v = 1, \ldots, M.
\]

Thus, \( \rho_{vl}(\lambda) = -g_{vl}^l(\pi; 0, \lambda) \) and\(^2\) \( \rho_{vu}(\lambda) = g_{vu}^u(\pi; 0, \lambda). \) Further, we collect all the lower and upper bounds, denoted by \( g_{vl}^c(\pi; r, \lambda) \) for \( v = 1, \ldots, 2M, \) into

\[
g^c(\pi; r, \lambda) = (g_l^l(\pi; r, \lambda)', g_u^u(\pi; r, \lambda)')'.
\]

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 \( g_{vl}^c(\hat{\pi}; r, \lambda) \) 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:

\[
\sqrt{n}(g_{vl}^c(\hat{\pi}; r, \lambda) - g_{vl}^c(\pi; r, \lambda)) \overset{d}{\rightarrow} N(0, \nabla_{\pi}g_{vl}^c(\pi; r, \lambda)\Sigma_{\pi}g_{vl}^c(\pi; r, \lambda)').
\]

\(^2\)We employ \( g_l^l(\pi; 0, \lambda) \) to transform the lower bounds for \( \rho \) into upper bounds for \( -\rho. \) We then use a single algorithm (for an upper bound) when estimating the lower and upper bounds for \( \rho. \)
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 $\ell \in \Lambda_{1-\alpha_2}$, we implement algorithm 1 in Chernozhukov, Lee, and Rosen (2013). To compute, $CI_{1-\alpha_1}(\ell)$, 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:

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

(b) $CI_{1-\alpha}(\ell)$ 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\(^4\) $\hat{h}(v; \ell) = [1(v = 1), ..., 1(v = 2M)] [\nabla_\pi g^c(\hat{\pi}; r, \ell) \hat{\Sigma} \nabla_\pi g^c(\hat{\pi}; r, \ell)]^{\frac{1}{2}}$ and set $se(v; \ell) = \frac{1}{\sqrt{n}} \|\hat{h}(v; \ell)\|$.

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

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

and

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

\(^3\)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.

\(^4\) $\nabla_\pi g^c(\hat{\pi}; r, \ell) \hat{\Sigma} \nabla_\pi g^c(\hat{\pi}; r, \ell)$ may be positive semi-definite and its matrix square root is computed using a singular value decomposition.
5. Compute
\[ c_{\hat{V}^m}(\ell) = (1 - \alpha)\text{-quantile of } \left\{ \sup_{v \in \hat{V}^m} \frac{\hat{h}(v; \ell) Z_s}{\|\hat{h}(v; \ell)\|} \right\}, \quad s = 1, \ldots, S. \]

6. Compute
\[ g_o^m(\hat{\pi}; r, \ell) = \inf_{v \in \hat{V}^m} \left[ g_o^m(\hat{\pi}; r, \ell) + c_{\hat{V}^m}(\ell) se(v; \ell) \right] \]

If \( m = l \) or \( u \) then report
\[ \hat{\rho}_o^l(\ell; 1 - \alpha) = -g_o^l(\hat{\pi}; 0, \ell) \quad \text{or} \quad \hat{\rho}_o^u(\ell; 1 - \alpha) = g_o^u(\hat{\pi}; 0, \ell) \]

Otherwise, if \( m = c \) then report
\[ CI_{1-\alpha}^p(\ell) = \{ r \in (0, 1] : g_o^c(\hat{\pi}; r, \ell) \geq 0 \}. \]

In the single equation bounds or when \( \Lambda_0 \) does not bind, the value \( \ell \) of the nuisance parameters does not affect the bounds. Otherwise, let \( t = 1, \ldots, T \) enumerate the \( T = \frac{1}{2}p(p - 1) \) \((j_t, h_t)\) pairs, \( j_t, h_t = 1, \ldots, p \) with \( j_t < h_t \), that correspond to the first \( T \) components of \( \lambda \). From Corollary 2.3.2, we have that if \( \ell \) is such that \((c_{j_t h_t}, c_{j_t h_t}) \neq (-1, 1), -\text{sgn}(\ell_{T+t}) \notin [c_{j_t h_t}, c_{j_t h_t}]\), and \( \ell_t = 0 \) then \( \mathcal{R}_{j_t h_t}^c(\ell) = \emptyset \). As such, we drop from \( \Lambda_{1-\alpha_{22}} \) the elements that satisfy these conditions since these have no effect on \( CR_{1-\alpha_{22}}^p = \bigcup_{\ell \in \Lambda_{1-\alpha_{22}}} CI_{1-\alpha_{21}}^p(\ell) \). For the remaining components \( \ell \) of \( \Lambda_{1-\alpha_{22}}, \) \( CI_{1-\alpha_{21}}^p(\ell) \) depends only on the signs (negative, zero, or positive) of the elements of the first \( T \) components of \( \ell \). 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 \( \ell \) is represented only once in \( \bigcup_{\ell \in \Lambda_{1-\alpha_{22}}} CI_{1-\alpha_{21}}^p(\ell) \).
B.2.2. Delta Method

Recall that the nuisance parameters \( \lambda \equiv g^\lambda(\pi) \), the vector of lower and upper bounds \( g^c_\pi(\pi; r, \lambda) \) in the intersection bounds algorithm for inference on \( \rho \), and the parameters \( \delta_j, \beta_{jl}, \text{ and } \Gamma_{jh}, j, h = 1, \ldots, p \) and \( l = 1, \ldots, k \), (written in the form \( \theta \equiv H(\pi; \rho) \)) can all be expressed as functions of the estimands

\[
\begin{align*}
\pi' & \equiv (\pi'_1, \pi'_2, \pi'_3, \pi'_4, \pi'_5, \pi'_6, \pi'_7) \\
& \equiv [\text{vec}(b_1^W(Y, X)'Y, X), b_1^W(Y, X)'Y, \ldots, b_p^W(Y, X)'Y, \text{vec}(b_1^W(Y, X)'Y), b_1^W(Y, X)'Y, \ldots, b_p^W(Y, X)'Y, \text{vec}(b_1^W(Y, X)'Y), b_1^W(Y, X)'Y, \ldots, b_p^W(Y, X)'Y].
\end{align*}
\]

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

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

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

Nuisance Parameters

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

\[
\begin{align*}
\lambda = (\lambda_1, \ldots, \lambda_{2T})' = g^\lambda(\pi) \equiv (\sigma_{1W}^{-2}\sigma_{1Y_1Y_2}, \ldots, \sigma_{1W}^{-2}\sigma_{1Y_{p-1}Y_p}, b_{Y_1W}b_{Y_2W}, \ldots, b_{Y_{p-1}W}b_{Y_pW})'.
\end{align*}
\]
It follows that, for \( t = 1, \ldots, T \), the components of \( \nabla_{\pi} g^\lambda(\pi) \) are given by
\[
\nabla_{\pi} g^\lambda_1(\pi) = \begin{bmatrix}
0 \\
1 \times [p(k+1)+(p+k)+p(1+k)+pk+k+p]
\end{bmatrix}
\]
where \( \nu_t \) is the unit vector with 1 in the \( t \)-th position and 0 elsewhere, and for \( t = T+1, \ldots, 2T \)
\[
\nabla_{\pi} g^\lambda(\pi) = \begin{bmatrix}
\nu_{j_t} \otimes \begin{bmatrix} b_{Y_{j_t}, W} & 0 \end{bmatrix}_{1 \times k} + \nu_{h_t} \otimes \begin{bmatrix} b_{Y_{h_t}, W} & 0 \end{bmatrix}_{1 \times k} \\

\end{bmatrix}_{1 \times [(p+k)+(1+k)+pk+k+p+\frac{1}{2}p(p-1)]}
\]

**Lower and Upper Intersection bounds**

Consider the joint equation bounds with \( \lambda = \ell^* \) with \( (\mathcal{E}_{j_{h_t}}, \mathcal{E}_{j_{h_t}}) \in \{(-1,0), (0,1)\} \) and \( sgn(\ell^*_t) \in [\mathcal{E}_{j_{h_t}}, \mathcal{E}_{j_{h_t}}] \setminus \{0\} \) for \( t = 1, \ldots, T \). In this case, we have \( g^c(\pi; r, \lambda) = (g^l(\pi; r, \lambda'), g^u(\pi; r, \lambda')') \) with
\[
g^l(\pi; r, \ell^*) = \begin{bmatrix}
1 - r \\
1 - r \\
1 - r \\
\vdots \\
\inf \\
\inf
\end{bmatrix}
\]
and
\[
g^u(\pi; r, \ell^*) = \begin{bmatrix}
1 - r \\
1 - r \\
1 - r \\
\vdots \\
\inf \\
\inf
\end{bmatrix}
\]
where \( M \equiv 2 + p + T \) (recall \( T \equiv \frac{1}{2}p(p-1) \)) and
\[
R^2_{\bar{W}, \bar{Y}} = b_{\bar{Y}, \bar{W}} b_{\bar{W}, \bar{Y}} = \sum_{h=1}^{p} b_{Y_h,(W,X)'} b_{W,(Y',X)'}, \quad \text{and} \quad R^2_{\bar{W}, \bar{Y}_j} = b_{\bar{Y}_j, \bar{W}} b_{\bar{W}, \bar{Y}_j} = b_{Y_j,(W,X)'} b_{W,(Y',X)'}, \quad \text{for} \quad j = 1, \ldots, 2M \times B.
\]

In this case, the components of \( \nabla_{\pi} g^c(\pi; r, \ell^*) \) are given by
1. For $v = 1$

\[ \nabla_{\pi} g_{1}^{v}(\pi; r, \ell^{*}) = \left[ \sum_{h=1}^{p} \delta_{h} \right] \otimes \begin{bmatrix} -b_{W}(y'; x'; y'_{h}, h) & 0 \end{bmatrix}_{1 \times k} \begin{bmatrix} -b_{Y, \tilde{W}} & 0 \end{bmatrix}_{1 \times k} \]

2. for $v = 2$,

\[ \nabla_{\pi} g_{2}^{v}(\pi; r, \ell^{*}) = 0, \]

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

\[ \nabla_{\pi} g_{v}^{v}(\pi; r, \ell^{*}) = \left[ \nabla_{\pi_{1}} g_{v}^{1}(\pi; r, \ell^{*}) \right]_{1 \times \frac{p}{2}(p-1)} \]

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

\[ \nabla_{\pi} g_{v}^{v}(\pi; r, \ell^{*}) = \left[ \nabla_{\pi_{1}} g_{v}^{1}(\pi; r, \ell^{*}) \right]_{1 \times \frac{p}{2}(p-1)} \]

where

\[ \nabla_{\pi_{1}} g_{v}^{1}(\pi; r, \ell^{*}) = \nabla_{\pi_{1}}(r - \frac{b_{Y_{j}, \tilde{W}_{l}} b_{Y_{ht}, \tilde{W}_{l}}}{\sigma_{W}^{2} \sigma_{Y_{ht}, \tilde{W}_{l}}}) = (\sigma_{W}^{2} \sigma_{Y_{ht}, \tilde{W}_{l}})^{-1} \left\{ \delta_{h} \right\} \otimes \begin{bmatrix} -b_{Y_{ht}, \tilde{W}} \end{bmatrix}_{1 \times k} + \delta_{h} \otimes \begin{bmatrix} -b_{Y_{ht}, \tilde{W}} \end{bmatrix}_{1 \times k} \]

and

\[ \nabla_{\pi_{T}} g_{v}^{1}(\pi; r, \ell^{*}) = \nabla_{\pi_{T}}(r - \frac{b_{Y_{j}, \tilde{W}_{l}} b_{Y_{ht}, \tilde{W}_{l}}}{\sigma_{W}^{2} \sigma_{Y_{ht}, \tilde{W}_{l}}}) = \delta_{t} \otimes \begin{bmatrix} b_{Y_{j}, \tilde{W}_{l}} b_{Y_{ht}, \tilde{W}_{l}} \end{bmatrix}_{1 \times T} = \sigma_{W}^{2} \sigma_{Y_{ht}, \tilde{W}_{l}} \]

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

\[ \nabla_{\pi} g_{v}^{v}(\pi; r, \ell^{*}) = 0 \]

Above, we set $\lambda = \ell^{*}$ where $(c_{j_{ih_{t}}, \tilde{c}_{j_{ih_{t}}}}) \in \{(-1, 0), (0, 1)\}$ and $sgn(\ell_{t}) \in [c_{j_{ih_{t}}, \tilde{c}_{j_{ih_{t}}}} \setminus \{0\}$
for \( t = 1, \ldots, T \equiv \frac{1}{2}p(p - 1) \). More generally, we consider any arbitrary \( \ell \in \Lambda_{1 - \alpha_{22}} \) and define the matrix \( P_{2M \times 2M}^\ell (M \equiv 2 + p + T) \) to operationalize how the nuisance parameters \( \lambda \) determines whether \( \mathcal{R}_{jh}^\ell \) contains an upper or lower bound, if at all, according to Corollary 2.3.2. In particular, for \( \lambda = \ell \), we let

\[
g_c^\ell (\pi; r, \ell) = P_g_c^\ell (\pi; r, \ell^*) \quad \text{and} \quad \nabla g_c^\ell (\pi; r, \ell) = P \nabla g_c^\ell (\pi; r, \ell^*)
\]

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

1. Set \( P = I_{2M \times 2M} \).

2. If \((c_{jh_t}, \tau_{jh_t}) = (0, 0)\) and \( \ell_t \neq 0 \) then change \( P_{M+(2+p+t)} \) to \(-t_{2+p+t}\).

3. If \((c_{jh_t}, \tau_{jh_t}) \in \{(-1, 0), (0, 1)\}\) and \( \text{sgn}(\ell_t) \notin [c_{jh_t}, \tau_{jh_t}] \) then change (a) \( P_{2+p+t} \) to \( t_{M+(2+p+t)} \) and (b) \( P_{M+(2+p+t)} \) to \(-t_{2+p+t}\).

4. If \((c_{jh_t}, \tau_{jh_t}) \in \{(-1, 0), (0, 1)\}\) and \( \text{sgn}(\ell_t) \in [c_{jh_t}, \tau_{jh_t}] \setminus \{0\}\) then keep (a) \( P_{2+p+t} \) as \( t_{2+p+t} \) and (b) \( P_{M+(2+p+t)} \) as \( t_{M+(2+p+t)} \).

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

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

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

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

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

\( \delta_j, \beta_{jt}, \) and \( \Gamma_{jh} \)

We have that \( \delta_j \) is given by:

\[
\delta_j = D_j(\pi; r) \equiv \frac{1}{r} b_{\hat{Y}_j, \hat{W}} \quad \text{for } j = 1, \ldots, p,
\]
where we let $\pi$ enter explicitly in $D_j$. It follows that

$$\nabla_\pi D_j(\pi; r) = \begin{bmatrix} t'_j \otimes \begin{bmatrix} \frac{1}{r} & 0 \\ 1 \times k & 0 \end{bmatrix} \\ 0 \end{bmatrix}.$$ 

Similarly, $\beta_{jl}$ is given by:

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

for $j = 1, \ldots, p$ and $l = 1, \ldots, k$.

It follows that

$$\nabla_\pi B_{jl}(r) = \begin{bmatrix} t'_j \otimes \begin{bmatrix} -\frac{1}{r} b_{W.X,l} & 0 \\ 1 \times (p+k) & 1 \times k \end{bmatrix} \\ 0 \end{bmatrix} - t'_l \otimes \begin{bmatrix} \frac{1}{r} b_{\tilde{Y},\tilde{W}} & 0 \\ 1 \times p & 1 \times p \end{bmatrix}.$$ 

Last, $\Gamma_{jh}$ is given by:

$$\Gamma_{jh} = G_{jh}(\pi; r) \equiv \sigma_{Y_j,Y_h}^2 - b_{\tilde{Y},\tilde{W}} \frac{1}{r} b_{\tilde{Y},\tilde{W}}$$

for $j \leq h$ and $j, h = 1, \ldots, p$.

Letting $t'_{(j,h)}$ take the value 1 at the entry $(j, h)$ corresponding to $\sigma_{Y_j,Y_h}^2$, we have

$$\nabla_\pi G_{jh}(\pi; r) = \begin{bmatrix} t'_j \otimes \begin{bmatrix} -\frac{1}{r} b_{\tilde{Y},\tilde{W}} & 0 \\ 1 \times k & 1 \times k \end{bmatrix} \\ 0 \end{bmatrix} + t'_h \otimes \begin{bmatrix} -\frac{1}{r} b_{\tilde{Y},\tilde{W}} & 0 \\ 1 \times p & 1 \times p \end{bmatrix}.$$ 

**Joint Confidence Regions** We sometimes construct a confidence region for the vector of parameters $\beta_l = (\beta_{1l}, \beta_{2l}, \beta_{3l})'$ associated with the variable $X_l$ in the system of $Y$ equations.

For a given $(\kappa, \tau, c)$ and $\alpha_1$, $\alpha_{21}$ and $\alpha_{22}$, we construct a $1 - \alpha_2$ confidence region $CR_{1-\alpha_2}^\rho$ for $\rho$. For each $r \in CR_{1-\alpha_2}^\rho$, the delta method gives the asymptotic distribution of the plug-in estimator $\hat{B}_1(\pi; r)$ for

$$B_1(\pi; r) = b_{Y.X,l} - b_{W.X,l} \frac{1}{r} b_{\tilde{Y},\tilde{W}}.$$ 

123
Specifically, we have that
\[ \sqrt{n}(B_1(\hat{\pi}; r) - B_1(\pi; r)) \xrightarrow{d} N(0, \Sigma_{B_1}(r)) \]
where \( \Sigma_{B_1}(r) = \nabla_{\pi} B_1(\pi; r) \Sigma \nabla_{\pi} B_1(\pi; r)' \)
and \( \nabla_{\pi} B_1(\pi; r) \) stacks the expressions \( \nabla_{\pi} B_{j1}(\pi; r) \) for \( j = 1, \ldots, p \) derived above. We construct a \( 1 - \alpha \) confidence region \( CR_{1 - \alpha}^{B_1}(r) \) for \( B_1(\pi; r) \), by inverting the following Wald statistic which has an asymptotic \( \chi^2_p \) distribution:
\[
CR_{1 - \alpha}^{B_1}(r) = \{ b_1 \in B_1^*: \sqrt{n}(B_1(\hat{\pi}; r) - b_1)' \Sigma_{B_1}^{-1}(r) \sqrt{n}(B_1(\hat{\pi}; r) - b_1) \leq c_{1 - \alpha} \}
\]
where \( c_{1 - \alpha} \) denotes the \( 1 - \alpha \) quantile of \( \chi^2_p \) and where we search over an initial neighborhood \( B_1^* \). For instance, we let \( B_1^* \) be the cube that contains each of the \( p \) unidimensional 95\% confidence regions:
\[
B_1^* = \{(b_{11}, \ldots, b_{1p}) : B_{j1}(\hat{\pi}; r) - 3se(B_{j1}(\hat{\pi}; r)) \leq b_{j1} \leq B_{j1}(\hat{\pi}; r) + 3se(B_{j1}(\hat{\pi}; r)) \} \text{ for } j = 1, \ldots, p
\]
Last, we construct the confidence region \( CR_{1 - \alpha_1 - \alpha_2}^{\beta_1}(r) \) for \( \beta_1 \) by forming the union:
\[ CR_{1 - \alpha_1 - \alpha_2}^{\beta_1} = \bigcup_{r \in CR_{1 - \alpha_2}^{\beta_1}} CR_{1 - \alpha_1}^{B_1}(r) \]
and use \( CR_{1 - \alpha_1 - \alpha_2}^{\beta_1} \) to form decisions regarding a null hypothesis for \( (\beta_{11}, \ldots, \beta_{p1}) \).

B.3. Extension of the Framework to Panel Data

Consider the unbalanced panel equations with firm fixed effects \( \gamma_i \):
\[
Y_{it}' = \gamma_{i}' + X_{it}' \beta + U_{it} \delta + \eta_{it}' \quad \text{and} \quad W_{it} = U_{it} + \varepsilon_{it} \quad \text{for } i = 1, \ldots, n \text{ and } t \in S_i.
\]
Let the binary indicator $I_{it}$ denote whether the observation $(Y_{it}, X_{it}, W_{it})$ is missing (at random). Let $I_i$ stack $I_{it}$ for $t = 1, ..., T$. Let

$$\sigma_{A_i, B_i} \equiv E(\tilde{A}_i \tilde{B}_i) = E(\sum_{t \in S_i} \tilde{A}_{it} \tilde{B}_{it}) = \sum_{t=1}^{T} E(I_{it} \tilde{A}_{it} \tilde{B}_{it}) = \sum_{t=1}^{T} E(I_{it})E(\tilde{A}_{it} \tilde{B}_{it}).$$

\footnote{The number of time periods $T$ should not be confused with the dimension of the nuisance parameter $\lambda$ in Online Appendix B.}

We assume that the data is missing at random from certain time periods. Specifically, we let $T$ denote 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 $\tilde{A}_{it} \equiv \frac{1}{T_t} \sum_{t \in S_i} A_{it}$ and $\tilde{A}_{i} \equiv A_{it} - \tilde{A}_{it}$, the fixed effect $\gamma_i$ drops out from the $\tilde{Y}_{it}$ equation:

$$\tilde{Y}_{it} = \tilde{X}_{it}' \beta + \tilde{U}_{it} \delta + \tilde{\eta}_{it} + \tilde{\varepsilon}_{it} \quad \text{and} \quad \tilde{W}_{it} = \tilde{U}_{it} + \tilde{\varepsilon}_{it}.$$

Letting $\tilde{A}_{i} \equiv [\tilde{A}'_{i1}, ..., \tilde{A}'_{iT_i}]'$, we obtain the panel analogue of assumption A1:

$$\tilde{Y}_{i} = \tilde{X}_{i}' \beta + \tilde{U}_{i} \delta + \tilde{\eta}_{i} \quad \text{and} \quad \tilde{W}_{i} = \tilde{U}_{i} + \tilde{\varepsilon}_{i} \quad \text{for } i = 1, ..., n.$$

Suppose that A2-A3 hold for this equation. Specifically, let

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

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

$$Cov(\eta_{it}, (\tilde{X}_{is}, \tilde{U}_{is})) = 0 \quad \text{and} \quad Cov(\varepsilon_{it}, (\tilde{X}_{is}, \tilde{U}_{is}, \tilde{\eta}_{is})) = 0 \quad \text{for } i = 1, ..., n \text{ and } t, s \in S_i.$$
In particular, we have $\sigma_{\tilde{X}_i,\tilde{\eta}_i} = 0$ and $\sigma_{\tilde{X}_i,\tilde{\varepsilon}_i} = 0$. Further, let

$$b_{\tilde{A}_i,\tilde{B}_i} = \sigma_{\tilde{B}_i,\tilde{A}_i}^{-2} \text{ and } \epsilon'_{\tilde{A}_i,\tilde{B}_i} = \tilde{A}'_i - \tilde{B}'_i b_{\tilde{A}_i,\tilde{B}_i}.$$ 

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

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

Let $\hat{A}_it \equiv \epsilon_{\tilde{A}_it,\tilde{X}_it}$ and $\hat{A}_i = [\hat{A}'_{1it}, ..., \hat{A}'_{TiT}]'$. By A1-A3, we obtain

$$\tilde{Y}_i = \tilde{U}_i \delta + \tilde{\eta}_i \quad \text{and} \quad \tilde{W}_i = \tilde{U}_i + \tilde{\varepsilon}_i.$$ 

In particular, we have

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

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

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

where

$$\rho = \sigma_{\tilde{W}_i,\tilde{Y}_i}^{-2} \sigma_{\tilde{\eta}_i} = E(\sum_{i \in S_i} \tilde{W}_it \tilde{W}'_it)^{-1} E(\sum_{i \in S_i} \tilde{U}_it \tilde{U}'_it).$$

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 $b_{\tilde{A}_i,\tilde{B}_i}$ and $\epsilon'_{\tilde{A}_i,\tilde{B}_i} = (\epsilon'_{\tilde{A}_{i1},\tilde{B}_{i1}}, ..., \epsilon'_{\tilde{A}_{iT_iT_i},\tilde{B}_{iT_iT_i}})'$ using their plug in sample analogues

$$\hat{b}_{\tilde{A}_i,\tilde{B}_i} = (1/n \sum_{i=1}^{T_i} \tilde{B}'_i \tilde{B}_i)^{-1} (1/n \sum_{i=1}^{T_i} \tilde{B}'_i \tilde{A}_i) \quad \text{and} \quad \hat{\epsilon}'_{\tilde{A}_i,\tilde{B}_i} = \tilde{A}'_i - \tilde{B}'_i b_{\tilde{A}_i,\tilde{B}_i}.$$ 

126
and estimate the asymptotic variance of $\sqrt{n}(\hat{b}_{A_i} - b_{A_i})$ by

$$
\left(\frac{1}{n} \sum_{i=1}^{n} \hat{B}_i - \bar{B}_i\right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} \hat{B}_i - \bar{B}_i\right) \left(\frac{1}{n} \sum_{i=1}^{n} \hat{B}_i - \bar{B}_i\right)^{-1}
$$

Note that the interpretation of $A_4$-$A_6$ applies to the stacked and within-transformed variables. In particular, $A_4$ assumes that

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

For this to hold, it suffices that $E(\hat{e}_{it}^2) \leq \kappa E(\hat{U}_{it}^2)$ for $t = 1, ..., T$. $A_5$ assumes that

$$
R_{\hat{Y}_{jit}, \hat{U}_{it}}^2 = 1 - \frac{\sigma_{\hat{\eta}_{jit}}^2}{\sigma_{\hat{Y}_{jit}}^2} = 1 - \frac{E(\sum_{t \in S_i} \hat{\eta}_{jit}^2)}{E(\sum_{t \in S_i} \hat{Y}_{jit}^2)} = 1 - \frac{\sum_{t=1}^{T} E(I_{it})E(\hat{\eta}_{jit}^2)}{\sum_{t=1}^{T} E(I_{it})E(\hat{Y}_{jit}^2)} \leq \tau_j,
$$

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

$$
\xi_{jh} \leq r_{\hat{\eta}_{jit}, \hat{\eta}_{hit}} = \frac{E(\sum_{t \in S_i} \hat{\eta}_{jit} \hat{\eta}_{hit})}{E(\sum_{t \in S_i} \hat{\eta}_{jit}^2)E(\sum_{t \in S_i} \hat{\eta}_{hit}^2)} \leq \frac{\sum_{t=1}^{T} E(I_{it})E(\hat{\eta}_{jit} \hat{\eta}_{hit})}{\sum_{t=1}^{T} E(I_{it})E(\hat{\eta}_{jit}^2) \sum_{t=1}^{T} E(I_{it})E(\hat{\eta}_{hit}^2)} \leq \tau_{jh},
$$

which holds if one imposes the same sign restriction on $Cov(\hat{\eta}_{jit}, \hat{\eta}_{hit})$ 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 $\bar{A}_{it} = A_{it} - \frac{1}{n} \sum_{i=1}^{n} \sum_{t \in S_i} A_{it}$).

**B.4. Mathematical Proofs**

**Proof of Theorem 2.3.1:** By $A_2$-$A_3$, $Cov[(\eta', \varepsilon)', X] = 0$. Since $Var(X)$ is nonsingular, $A_1$ gives

$$
\beta = b_{Y,X} - b_{W,X} \delta.
$$
A2-A3 also give \( \sigma_{\tilde{U},\tilde{e}} = 0 \) and \( \sigma_{\tilde{U},\tilde{n}} = \sigma_{\tilde{e},\tilde{n}} = 0 \). 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

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

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

\[
b_{\tilde{Y},\tilde{W}} \equiv \sigma_{\tilde{W}}^{-2} \rho \delta \quad \text{and} \quad \sigma_{\tilde{Y}}^{-2} \sigma_{\tilde{Y}}^2 = \delta^2 \rho \delta + \Gamma.
\]

Since \( \rho \neq 0 \), we obtain

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

**Lemma B.4.1** Under the conditions of Theorem 2.3.1, \( R_{\tilde{Y},\tilde{W}}^2 \leq R_{\tilde{Y},\tilde{U}}^2 \).

**Proof of Lemma B.4.1:** If \( \sigma_{\tilde{Y}}^2 = 0 \), set \( R_{\tilde{Y},\tilde{W}}^2 = R_{\tilde{Y},\tilde{U}}^2 = 0 \). If \( 0 < \sigma_{\tilde{Y}}^2 \), we have

\[
R_{\tilde{Y},\tilde{W}}^2 = \frac{\sigma_{\tilde{W}}^2 b_{\tilde{Y},\tilde{W}}}{\sigma_{\tilde{Y}}^2} = \frac{\sigma_{\tilde{W}}^2}{\sigma_{\tilde{Y}}^2} (\delta^2 \rho)^2
\]
\[
R_{\tilde{Y},\tilde{U}}^2 = 1 - \frac{\sigma_{\tilde{n}}^2}{\sigma_{\tilde{Y}}^2} = \frac{1}{\sigma_{\tilde{Y}}^2} (\sigma_{\tilde{Y}}^2 - \sigma_{\tilde{n}}^2) = \frac{1}{\sigma_{\tilde{Y}}^2} \delta^2 \sigma_{\tilde{U}}^2 = \frac{\sigma_{\tilde{W}}^2}{\sigma_{\tilde{Y}}^2} \delta^2 \rho
\]

It follows that

\[
R_{\tilde{Y},\tilde{U}}^2 - R_{\tilde{Y},\tilde{W}}^2 = \frac{\sigma_{\tilde{W}}^2 (\delta^2 \rho - \delta^2 \rho^2)}{\sigma_{\tilde{Y}}^2} = \frac{\sigma_{\tilde{W}}^2}{\sigma_{\tilde{Y}}^2} \rho (1 - \rho) \delta^2 \geq 0.
\]

**Proof of Corollary 2.3.2:** The identification set \( J^{k,\tau,c} \) obtains from A1-A6 and the
\((\text{Var}(\tilde{Y}', \tilde{W}'))\) moments given by (in)equalities \((2.3, 2.5, 2.6, 2.7)\), using the expressions in Theorem 2.3.1. To show that \(\mathcal{J}^{k,r,c}\) is sharp, let \(d = D(r)\), \(b = B(r)\), and \(g = G(r)\). We show that for each \((r, d, b, g) \in \mathcal{J}^{k,r,c}\) 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 \(A_2\)–\(A_3\), \(\text{Cov}[\eta^*, (X', U^*)'] = 0\), \(\text{Cov}[\varepsilon^*, (\eta^*, X', U^*)'] = 0\). Further, \(\frac{\sigma^2_2}{\sigma^2_W} = r\) and thus \(A_4\) holds, \(\sigma^2_{\varepsilon} \leq \kappa \sigma^2_{\tilde{O}^*}\). Last, \(G(r) = \sigma^2_W \sigma^2_{\tilde{Y}'}\) and therefore \(A_5\) holds since, when \(\sigma^2_{\tilde{Y}_j} \neq 0\),

\[
1 - \frac{\sigma^2_{\eta j}}{\sigma^2_{\tilde{Y}_j}} = \frac{\sigma^2_W}{\sigma^2_{\tilde{Y}_j}} (\frac{\sigma^2_{Y_j}}{\sigma^2_W} - G_{jj}(r)) \leq \frac{\sigma^2_W}{\sigma^2_{\tilde{Y}_j}} \left[ \frac{\sigma^2_{Y_j}}{\sigma^2_W} - \frac{\sigma^2_{\tilde{Y}_j}}{\sigma^2_W} (1 - \tau_j) \right] = \tau_j,
\]

and \(A_6\) holds since \(\varepsilon_{jh} \leq \text{sgn}(G_{jh}(r)) \leq \varepsilon_{jh}\).

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 \(\tilde{V} \equiv c_{V,X}\) is nondegenerate and satisfies

\[
\sigma_{\tilde{W},\tilde{V}} = \sqrt{r} \sigma_{\tilde{V}} \sigma_{\tilde{W}} \quad \text{and} \quad \sigma_{\tilde{Y},\tilde{V}} = \frac{1}{\sqrt{r}} \sigma_{\tilde{V}} \sigma_{\tilde{W}} \sigma_{\tilde{Y},\tilde{W}}.
\]

Note that these covariance restrictions are coherent. Specifically,

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

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

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

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

For instance, to construct \(V\), set \(\sigma_{\tilde{V}}\) to some value (e.g. \(\sigma_{\tilde{V}} = 1\)) and let \(\vartheta\) be any ran-
dom variable that is uncorrelated with \((X', W, Y)'\) (e.g. a residual from a regression on \((X', W, Y)\)). When \(\sigma_{(\tilde{W}, \tilde{Y})}'\) 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

\[
\kappa = \frac{1}{\sigma^2} \left[ \sigma^2 - b'_{\tilde{V}, (\tilde{W}, \tilde{Y})}' \sigma^2_{(\tilde{W}, \tilde{Y})}' b_{\tilde{V}, (\tilde{W}, \tilde{Y})}' \right]^{\frac{1}{2}}
\]

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

\[
\tilde{V} = (\tilde{W}, \tilde{Y}) b_{\tilde{V}, (\tilde{W}, \tilde{Y})}' + \kappa \theta.
\]

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 \(U^*, \varepsilon^*,\) and \(\eta^*\) 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
Next, we derive the identification region $\mathcal{R}^{k,r,c}$ for $\rho$. First, we show that $R^2_{W,Y} \leq \rho \leq 1$.
If \( \sigma_{\bar{Y},\bar{W}} = 0 \) then \( R^2_{\bar{W},\bar{Y}} = 0 \leq \rho \leq 1 \). Suppose that \( \sigma_{\bar{Y},\bar{W}} \neq 0 \). Since \( 0 < \rho \) and \( 0 \leq \Gamma \) then for any vector \( x \), we have

\[
0 \leq \rho x' \sigma_{\bar{W}}^{-2} \sigma_{\bar{Y}}^{-2} x - x' b'_{\bar{Y},\bar{W}} b_{\bar{Y},\bar{W}} x.
\]

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

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

Second, by A4, we have \( 1 - \rho = \frac{\sigma_{\bar{W}}^2}{\sigma_{\bar{W}}^2} \leq \frac{\sigma_{\bar{Y}}^2}{\sigma_{\bar{Y}}^2} = \kappa \rho \) and thus \( \rho \in \left[ \frac{1}{1 + \kappa}, 1 \right] \). Third, by A5, we have that for \( j = 1, ..., p \), \( R^2_{\bar{Y}_j,\bar{W}} = (1 - \frac{\sigma_{\bar{Y}_j}^2}{\sigma_{\bar{Y}}^2}) \leq \tau_j \) (recall that if \( \sigma_{\bar{Y}_j}^2 = 0 \) then we set \( R^2_{\bar{Y}_j,\bar{W}} = 0 \)). Multiplying by \( \frac{\sigma_{\bar{Y}_j}^2}{\sigma_{\bar{W}}^2} \) and substituting for \( \Gamma_{j,j} \) we obtain

\[
b_{\bar{Y}_j,\bar{W}}^T \frac{1}{\rho} b_{\bar{Y}_j,\bar{W}} = \frac{\sigma_{\bar{Y}_j}^2}{\sigma_{\bar{W}}^2} - (\sigma_{\bar{W}}^{-2} \sigma_{\bar{Y}_j}^2 - b_{\bar{Y}_j,\bar{W}}^T \frac{1}{\rho} b_{\bar{Y}_j,\bar{W}}) \leq \frac{\sigma_{\bar{Y}_j}^2}{\sigma_{\bar{W}}^2} \]

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

\[
G_{j,h}(\rho) \leq 0 \text{ if and only if } \begin{cases} \frac{b_{\bar{Y}_j,\bar{W}} b_{\bar{Y}_h,\bar{W}}}{\sigma_{\bar{W}}^{-2} \sigma_{\bar{Y}_j,\bar{Y}_h}} \leq \rho \quad \text{when } \sigma_{\bar{W}}^{-2} \sigma_{\bar{Y}_j,\bar{Y}_h} < 0 \\ 0 \leq \frac{b_{\bar{Y}_j,\bar{W}} b_{\bar{Y}_h,\bar{W}}}{\sigma_{\bar{W}}^{-2} \sigma_{\bar{Y}_j,\bar{Y}_h}} \quad \text{when } \sigma_{\bar{W}}^{-2} \sigma_{\bar{Y}_j,\bar{Y}_h} = 0 \\ \rho \leq \frac{b_{\bar{Y}_j,\bar{W}} b_{\bar{Y}_h,\bar{W}}}{\sigma_{\bar{W}}^{-2} \sigma_{\bar{Y}_j,\bar{Y}_h}} \quad \text{when } 0 < \sigma_{\bar{W}}^{-2} \sigma_{\bar{Y}_j,\bar{Y}_h} \end{cases}
\]

Combining the results, we have \( \rho \in \mathcal{R}^{k,\tau,c} = [R^2_{\bar{W},\bar{Y}}, 1] \cap \left[ \frac{1}{1 + \kappa}, 1 \right] \cap_{j=1}^{p} \left[ \frac{1}{\tau_j} R^2_{\bar{Y}_j,\bar{W}}, 1 \right] \cap_{j,h=1}^{p} \mathcal{R}_{j,h}^c \cap_{j<h}^{\rho}. \]

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

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

Further, when $R_{\tilde{W}, \tilde{Y}}^2 \neq 0$, $0 \leq G(R_{\tilde{W}, \tilde{Y}}^2)$. Specifically, $0 < \sigma_{\tilde{W}}^4 R_{\tilde{W}, \tilde{Y}}^2$ and

$$\sigma_{\tilde{W}}^4 R_{\tilde{W}, \tilde{Y}}^2 G(R_{\tilde{W}, \tilde{Y}}^2) = (R_{\tilde{W}, \tilde{Y}}^2 \sigma_{\tilde{W}}^2) \sigma_{\tilde{Y}}^2 = \text{Var}(b'_{\tilde{W}, \tilde{Y}} \tilde{Y}) \sigma_{\tilde{Y}}^2 - \sigma_{\tilde{Y}, \tilde{W}} \sigma_{\tilde{W}, \tilde{Y}} \geq 0$$

since and for any vector $x_p \times 1$, applying the Cauchy–Schwarz inequality gives

$$x' \text{Var}(b'_{\tilde{W}, \tilde{Y}} \tilde{Y}) \sigma_{\tilde{Y}}^2 x - x' \sigma_{\tilde{Y}, \tilde{W}} \sigma_{\tilde{W}, \tilde{Y}} x$$

$$= \text{Var}(b'_{\tilde{W}, \tilde{Y}} \tilde{Y}) \text{Var}(x' \tilde{Y}) - [\text{Cov}(x' \tilde{Y}, \tilde{W})]^2$$

$$= \text{Var}(b'_{\tilde{W}, \tilde{Y}} \tilde{Y}) \text{Var}(x' \tilde{Y}) - [\text{Cov}(x' \tilde{Y}, b'_{\tilde{W}, \tilde{Y}} \tilde{Y})]^2 \geq 0$$

where we make use of $\tilde{W}' = \tilde{Y}' b_{\tilde{W}, \tilde{Y}} + \epsilon_{\tilde{Y}, \tilde{W}}$ and $\text{Cov}(\tilde{Y}, \epsilon_{\tilde{W}, \tilde{Y}}) = 0$ in the last equality.

Then for any $r \in \mathcal{R}^{k, r, e} \subseteq [R_{\tilde{W}, \tilde{Y}}^2, 1]$ there exists $0 \leq \lambda \leq 1$ such that $\frac{1}{r} = \lambda + (1 - \lambda) \frac{1}{R_{\tilde{W}, \tilde{Y}}^2}$ and it follows that

$$0 \leq G(r) = \lambda G(1) + (1 - \lambda) G(R_{\tilde{W}, \tilde{Y}}^2).$$

Clearly, $\frac{1}{1 + r} \leq r \leq 1$. Further, for $j = 1, ..., p$, if $\sigma_{Y_j}^2 = 0$ then we set $\frac{1}{r} R_{\tilde{W}, \tilde{Y}_j}^2 = 0 \leq r$ whereas if $\sigma_{Y_j}^2 \neq 0$ then $\frac{1}{r} b_{Y_j, \tilde{W}_j}^2 \sigma_{Y_j}^2 = \frac{1}{r} R_{\tilde{W}, \tilde{Y}_j}^2 \leq r$ implies that $\frac{\sigma_{Y_j}^2}{\sigma_{\tilde{W}}^2} (1 - \tau_j) \leq \frac{\sigma_{Y_j}^2}{\sigma_{\tilde{W}}^2} - b_{Y_j, \tilde{W}_j}^2 \frac{1}{r} = G_{jj}(r)$. Last, from the expression for $G_{jh}(r)$, we have that $\varepsilon_{jh} \leq \text{sgn}(G_{jh}(r)) \leq \tau_{jh}$ for every $r \in \mathcal{R}^{k, r, e}$ and $j, h = 1, ..., p$ with $j < h$.

The sharp bounds $D^{k, r}, B^{k, r}$, and $G^{k, r}$ for $\delta, \beta$, and $\Gamma$ follow from the mappings $D(\cdot), B(\cdot)$, and $G(\cdot)$ in Theorem 2.3.1.

**Proof of Theorem 2.5.1:** 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 \((\hat{b}_0^0)\) and slope \((\hat{b}_{A,B})\) estimators and the sample residual \((\hat{\epsilon}_{A,B,i})\) by:

\[ \hat{b}_{A,B} = (\hat{b}_0^0, \hat{b}_{A,B})' \equiv \left( \frac{1}{n} \sum_{i=1}^n (1, B'_i)'(1, B'_i) \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n (1, B'_i)'A'_i \right) \quad \text{and} \quad \hat{\epsilon}_{A,B,i} \equiv A'_i - (1, B'_i)\hat{b}_{A,B}. \]

Further, we collect into \( \pi^* \) the following estimands

\[ \pi^* \equiv \left[ \text{vec}(b^*_{Y,(W,X)'})', b^*_{W,(Y,X)'}, b^*_{W,(Y_p,X)'}, \ldots, b^*_{W,(Y_p,X)'}, \text{vec}(b^*_{Y,X})', b^*_{W,X}, \sigma_W^{-2}\text{vec}(\sigma_Y^2) \right], \]

and into \( \hat{\pi} \) the corresponding estimators:

\[ \hat{\pi} \equiv \left[ \text{vec}(\hat{b}_{Y,(W,X)'})', \hat{b}_{W,(Y,X)'}, \hat{b}_{W,(Y_p,X)'}, \ldots, \hat{b}_{W,(Y_p,X)'}, \text{vec}(\hat{b}_{Y,X})', \hat{b}_{W,X}, \hat{\sigma}_W^{-2}\text{vec}(\hat{\sigma}_Y^2) \right]. \]

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

\[ \hat{Q} = \text{diag}\left\{ I_{p \times p} \otimes \hat{\mu}_A^2(1,W,X)' , \hat{\mu}_A^2(1,Y,X)' , \ldots, \hat{\mu}_A^2(1,Y_p,X)' , I_{p \times p} \otimes \hat{\mu}_A^2(1,X)' , \hat{\mu}_A^2(1,X)' , I_{p(p+1) \times p(p+1)} \otimes \hat{\sigma}_W^2 \right\}. \]

and

\[ L = \frac{1}{n} \sum_{i=1}^n \left[ \text{vec}(1, W_i, X'_i)' \epsilon_{Y,(W,X)'}, i, (1, Y_i, X'_i) \epsilon_{W,(Y,X)'}, i, (1, Y_i, X'_i) \epsilon_{Y,X,i}, (1, X'_i) \epsilon_{Y,X,i}, \text{vec}(1, X'_i)' \epsilon_{Y,X,i}, (1, X'_i)' \epsilon_{W,X,i}, \text{vec}(\epsilon_{Y,X,i} \epsilon_{Y,X,i} - \sigma_Y^2)' \right]. \]

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

\[
\sqrt{n}(\hat{\pi}_1 - \pi^*_1) = \hat{Q}_1^{-1} \sqrt{n}L_1 = (\hat{Q}_1^{-1} - \hat{Q}_1^{-1}) \sqrt{n}L_1 + \hat{Q}_1^{-1} \sqrt{n}L_1.
\]

Since \( (i) \) gives \( \hat{Q}_1^{-1} - \hat{Q}_1^{-1} = o_p(1) \) and \( (ii) \) gives \( \sqrt{n}L_1 \to_d N(0, \Sigma^*) \), we obtain that \( \sqrt{n}(\hat{\pi}_1 - \pi^*_1) = \hat{Q}_1^{-1} \sqrt{n}L_1 + o_p(1) \to_d N(0, \Sigma^*) \). Moreover, it follows from \( \hat{\mu}^2_{(1,X')'} \to_p \mu^2_{(1,X')'} \), \( \sqrt{n}(\hat{b}_{Y_j} - \hat{b}_{Y_j}^*) = O_p(1) \), 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 \( j = 1, \ldots, p \) that for any \( j, h = 1, \ldots, p \)

\[
\frac{1}{n} \sum_{i=1}^n \epsilon_{Y_j,X,i}\hat{\epsilon}_{Y_h,X,i} = n^{-\frac{1}{2}} \sum_{i=1}^n (\epsilon_{Y_j,X,i} - (1, X'_i)(\hat{b}_{Y_j} - \hat{b}_{Y_j}^*))((\epsilon_{Y_h,X,i} - (1, X'_i)(\hat{b}_{Y_h} - \hat{b}_{Y_h}^*))
\]

\[
= n^{-\frac{1}{2}} \sum_{i=1}^n \epsilon_{Y_j,X,i}\epsilon_{Y_h,X,i} - \frac{1}{n} \sum_{i=1}^n \epsilon_{Y_j,X,i}(1, X'_i) \sqrt{n}(\hat{b}_{Y_h} - \hat{b}_{Y_h}^*)
\]

\[
- \left[ \frac{1}{n} \sum_{i=1}^n \epsilon_{Y_h,X,i}(1, X'_i) \right] \sqrt{n}(\hat{b}_{Y_j} - \hat{b}_{Y_j}^*) + (\hat{b}_{Y_h} - \hat{b}_{Y_h}^*)' \hat{\mu}^2_{(1,X')} \sqrt{n}(\hat{b}_{Y_j} - \hat{b}_{Y_j}^*)
\]

\[
= n^{-\frac{1}{2}} \sum_{i=1}^n \epsilon_{Y_j,X,i}\epsilon_{Y_h,X,i} + o_p(1).
\]

Similarly, by \( (i) \), we have that

\[
\frac{1}{n} \sum_{i=1}^n \epsilon_{Y_j,X,i}\hat{\epsilon}_{Y_h,X,i} = E(\epsilon_{Y_j,X} \epsilon_{Y_h,X}) + o_p(1) = \sigma_{Y_j,Y_h} + o_p(1)
\]

\[
\frac{1}{n} \sum_{i=1}^n \epsilon_{W,X,i}^2 = \sigma_{W}^2 + o_p(1).
\]

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 \( j, h = 1, \ldots, p \)

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

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

\[
\sqrt{n}(\hat{\pi} - \pi^*) = Q^{-1} \sqrt{n}L + o_p(1) \to_d N(0, \Sigma^*)
\]

and therefore that the subvector \( \sqrt{n}(\hat{\pi} - \pi) \to_d N(0, \Sigma) \).
Proof of Corollary B.1.1: The identification set $\mathcal{J}^{k,\tau,c}$ obtains from $A_1 \cup A_2$ and the $(\text{Var}([\hat{Y}, \hat{W}]))$ the moments given by (in)equalities (2.3.2, 2.6, 2.7), using the expressions in Theorem 2.3.1. The sharpness proof in Corollary 2.3.2 implies that $\mathcal{J}^{k,\tau,c}$ is sharp. Specifically, since $G(r) = \frac{\sigma_y^2 - \sigma_{yj}^2}{\sigma_W^2}$, we have that $\zeta_{jh} \leq r_{\hat{Y}j, \hat{W}h} \leq \bar{c}_{jh}$.

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

$$
\zeta_{jh} \leq \Gamma_{jh} = \frac{G_{jh}(\rho)}{[G_{jj}(\rho)G_{hh}(\rho)]^{1/2}} = \frac{\sigma_W^2 \sigma_{Yj} \bar{Y}_j - b_{\bar{Y}j} \bar{W} \bar{W} \rho b_{\bar{Y}_j} \bar{W}}{(\sigma_W^2 \sigma_{Yj}^2 - \frac{1}{\rho} b_{\bar{Y}j} \bar{W})^{1/2} (\sigma_W^2 \sigma_{Yh}^2 - \frac{1}{\rho} b_{\bar{Y}_h} \bar{W})^{1/2}} \leq \bar{c}_{jh}.
$$

If $\sigma_{Yj}^2 = 0$ or $\sigma_{Yh}^2 = 0$ then $\sigma_{\eta_j}^2 = 0$ or $\sigma_{\eta_h}^2 = 0$ and $\zeta_{jh} \leq \sigma_{\eta_j, \eta_h} \leq \bar{c}_{jh}$ is either incorrect (if $0 \notin [\zeta_{jh}, \bar{c}_{jh}]$) or uninformative about $\rho$ (if $0 \in [\zeta_{jh}, \bar{c}_{jh}]$). Suppose that $\sigma_{Yj}^2 \neq 0$ and $\sigma_{Yh}^2 \neq 0$. Multiplying the numerator and denominator by $0 < \rho \sigma_W^2 - \sigma_{Yj}^2$ gives

$$
\zeta_{jh} \leq \frac{\rho \sigma_{Yj} \bar{Y}_j - \rho \sigma_{Yh} \bar{W} \bar{W} \rho \sigma_{\eta_j, \eta_h}}{(\rho - \sigma_{Yj}^2)^{1/2} (\rho - \sigma_{Yh}^2)^{1/2}} \leq \bar{c}_{jh}.
$$

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

$$
S_{jh}(r) \equiv r \times r_{\bar{Y}j, \bar{Y}h} - r_{\bar{Y}j} r_{\bar{W} \bar{Y}j} - r_{\bar{W} \bar{Y}j} r_{\bar{W} \bar{Y}h}
$$

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

$$
M_{jh}(r; c) \equiv (r \times r_{\bar{Y}j, \bar{Y}h} - r_{\bar{Y}j} r_{\bar{W} \bar{Y}j} - r_{\bar{W} \bar{Y}j} r_{\bar{W} \bar{Y}h})^2 - c^2 (r - \rho \sigma_{Yj}^2 - r - \rho \sigma_{Yh}^2).
$$

By Corollary 2.3.2, we obtain that $\rho \in \mathcal{R}^{k,\tau,c} = [\rho \sigma_{Yj}^2, 1] \cap \{ \frac{1}{\rho \sigma_{Yj}^2}, 1 \} \cap \{ \frac{1}{\rho \sigma_{Yj}^2}, 1 \} \cap \mathcal{R}_{jh}^{c}$. In addition, $\mathcal{R}^{k,\tau,c}$ is sharp since every $r \in \mathcal{R}^{k,\tau,c}$ corresponds to a point $(r, d, b, g) \in \mathcal{J}^{k,\tau,c}$. Specifically, if $r \in \mathcal{R}^{k,\tau,c}$ then $\frac{1}{1 + r} \leq r \leq 1$, $0 \leq G(r)$, and $R_{Yj, \hat{Y}j}^2 \leq \tau_j$ for $j = 1, \ldots, p$ by Corollary 2.3.2. Further, from the sign and magnitude restrictions in $S_{jh}(r)$ and $M_{jh}(r; c)$, we have that $\zeta_{jh} \leq \frac{G_{jh}(r)}{[G_{jj}(\rho)G_{hh}(\rho)]^{1/2}} \leq \bar{c}_{jh}$ for every $r \in \mathcal{R}^{k,\tau,c} \subseteq \mathcal{R}_{jh}^{c}$ and $j, h = 1, \ldots, p$ with $j < h$.
Next, we examine the behavior of $S_{jh}(r)$ and $M_{jh}(r;c)$ when $\sigma_{\tilde{Y}_j}^2 \sigma_{\tilde{Y}_h}^2 \neq 0$. First, we have that

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

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

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

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

$$M_{jh}(r;c) = r^2 R_{\tilde{W}.\tilde{Y}_j}^2 + R_{\tilde{W}.\tilde{Y}_j}^2 R_{\tilde{W}.\tilde{Y}_h}^2 - 2r \times r_{\tilde{Y}_j,\tilde{Y}_h} r_{\tilde{W}.\tilde{Y}_j} r_{\tilde{W}.\tilde{Y}_h}$$

$$- c^2 r^2 + c^2 r(R_{\tilde{W}.\tilde{Y}_j}^2 + R_{\tilde{W}.\tilde{Y}_h}^2) - c^2 R_{\tilde{W}.\tilde{Y}_j}^2 R_{\tilde{W}.\tilde{Y}_h}^2$$

$$= r^2(R_{\tilde{Y}_j,\tilde{Y}_h}^2 - c^2) + r[-2r_{\tilde{Y}_j,\tilde{Y}_h} r_{\tilde{W}.\tilde{Y}_j} r_{\tilde{W}.\tilde{Y}_h} + c^2(R_{\tilde{W}.\tilde{Y}_j}^2 + R_{\tilde{W}.\tilde{Y}_h}^2)] + (1 - c^2)R_{\tilde{W}.\tilde{Y}_j} R_{\tilde{W}.\tilde{Y}_h}^2$$

$$= r^2(R_{\tilde{Y}_j,\tilde{Y}_h}^2 - c^2) + r[R_{\tilde{W}.(\tilde{Y}_j,\tilde{Y}_h)}^2(1 - R_{\tilde{Y}_j,\tilde{Y}_h}^2) - (1 - c^2)(R_{\tilde{W}.\tilde{Y}_j}^2 + R_{\tilde{W}.\tilde{Y}_h}^2)]$$

$$+ (1 - c^2)R_{\tilde{W}.\tilde{Y}_j} R_{\tilde{W}.\tilde{Y}_h}^2,$$

where the last equality makes use of

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

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

$$M_{jh}(r; r_{\tilde{Y}_j,\tilde{Y}_h}) = r[R_{\tilde{W}.(\tilde{Y}_j,\tilde{Y}_h)}^2(1 - R_{\tilde{Y}_j,\tilde{Y}_h}^2) - (1 - R_{\tilde{Y}_j,\tilde{Y}_h}^2)(R_{\tilde{W}.\tilde{Y}_j}^2 + R_{\tilde{W}.\tilde{Y}_h}^2)]$$

$$+ (1 - R_{\tilde{Y}_j,\tilde{Y}_h}^2) R_{\tilde{W}.\tilde{Y}_j} R_{\tilde{W}.\tilde{Y}_h}^2$$

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

Otherwise, if \( c^2 \neq R^2_{W,\tilde{y}_j,\tilde{y}_h} \), the discriminant of \( M_{jh}(\cdot; c) \) is

\[
\Delta_{jh}(c) = [R^2_{W,(\tilde{y}_j,\tilde{y}_h)\tilde{y}_j} - (1 - R^2_{W,\tilde{y}_j,\tilde{y}_h}) - (1 - c^2)(R^2_{W,\tilde{y}_j} + R^2_{W,\tilde{y}_h})]^2 - 4(1 - c^2)(R^2_{W,\tilde{y}_j,\tilde{y}_h} - c^2)R^2_{W,\tilde{y}_j} R^2_{W,\tilde{y}_h} \\
= [R^2_{W,(\tilde{y}_j,\tilde{y}_h)\tilde{y}_j} - (1 - R^2_{W,\tilde{y}_j,\tilde{y}_h}) - (1 - c^2)(R^2_{W,\tilde{y}_j} + R^2_{W,\tilde{y}_h})]^2 \\
- (1 - c^2)4R^2_{W,\tilde{y}_j} R^2_{W,\tilde{y}_j} R^2_{W,\tilde{y}_h} + 4c^2(1 - c^2)R^2_{W,\tilde{y}_j} R^2_{W,\tilde{y}_h} \\
= [R^2_{W,(\tilde{y}_j,\tilde{y}_h)\tilde{y}_j} - (1 - R^2_{W,\tilde{y}_j,\tilde{y}_h}) - (1 - c^2)(R^2_{W,\tilde{y}_j} + R^2_{W,\tilde{y}_h})]^2 \\
- (1 - c^2)[R^2_{W,(\tilde{y}_j,\tilde{y}_h)\tilde{y}_j} - (1 - R^2_{W,\tilde{y}_j,\tilde{y}_h}) - (R^2_{W,\tilde{y}_j} + R^2_{W,\tilde{y}_h})]^2 + 4c^2(1 - c^2)R^2_{W,\tilde{y}_j} R^2_{W,\tilde{y}_h} \\
= c^2[R^4_{W,(\tilde{y}_j,\tilde{y}_h)\tilde{y}_j} - (1 - R^2_{W,\tilde{y}_j,\tilde{y}_h})^2 - (1 - c^2)(R^2_{W,\tilde{y}_j} + R^2_{W,\tilde{y}_h})^2 - 4R^2_{W,\tilde{y}_j} R^2_{W,\tilde{y}_h}] \\
= c^2[R^4_{W,(\tilde{y}_j,\tilde{y}_h)\tilde{y}_j} - (1 - R^2_{W,\tilde{y}_j,\tilde{y}_h})^2 - (1 - c^2)(R^2_{W,\tilde{y}_j} - R^2_{W,\tilde{y}_h})^2].
\]

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

\[
0 < c^2 < 1 - \frac{R^4_{W,(\tilde{y}_j,\tilde{y}_h)\tilde{y}_j} - (1 - R^2_{W,\tilde{y}_j,\tilde{y}_h})^2}{(R^2_{W,\tilde{y}_j} - R^2_{W,\tilde{y}_h})^2}. 
\]

Further, we have that

\[
1 - \frac{R^4_{W,(\tilde{y}_j,\tilde{y}_h)\tilde{y}_j} - (1 - R^2_{W,\tilde{y}_j,\tilde{y}_h})^2}{(R^2_{W,\tilde{y}_j} - R^2_{W,\tilde{y}_h})^2} \leq R^2_{Y_j,\tilde{y}_h} \quad \text{since if } c^2 = R^2_{W,\tilde{y}_j,\tilde{y}_h} \text{ then}
\]

\[
\Delta_{jh}(c) = [R^2_{W,(\tilde{y}_j,\tilde{y}_h)\tilde{y}_j} - (1 - R^2_{W,\tilde{y}_j,\tilde{y}_h}) - (1 - c^2)(R^2_{W,\tilde{y}_j} + R^2_{W,\tilde{y}_h})]^2 - 4(1 - c^2)(R^2_{W,\tilde{y}_j,\tilde{y}_h} - c^2)R^2_{W,\tilde{y}_j} R^2_{W,\tilde{y}_h} \\
= (1 - R^2_{W,\tilde{y}_j,\tilde{y}_h})^2[R^2_{W,(\tilde{y}_j,\tilde{y}_h)\tilde{y}_j} - (R^2_{W,\tilde{y}_j} + R^2_{W,\tilde{y}_h})]^2 \geq 0
\]

138
and if $R_{Y_j,Y_h}^2 = 0$ then

$$1 - \frac{R_{W,(\hat{Y}_j,\hat{Y}_h)}^2(1 - R_{Y_j,\hat{Y}_h}^2)^2}{(R_{W,Y_j}^2 - R_{W,\hat{Y}_h}^2)^2} = 1 - \frac{(R_{W,Y_j}^2 + R_{W,\hat{Y}_h}^2)^2}{(R_{W,Y_j}^2 - R_{W,\hat{Y}_h}^2)^2} \leq 0 = R_{Y_j,\hat{Y}_h}^2.$$ 

It follows that if $0 < c^2 < 1 - R_{W,(\hat{Y}_j,\hat{Y}_h)}^2$ then $c^2 < R_{Y_j,\hat{Y}_h}^2$ and

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

If $c^2 \neq R_{Y_j,\hat{Y}_h}^2$ and $0 \leq \Delta_{jh}(c)$ then define

$$F_{jh}(c) \equiv -R_{W,(\hat{Y}_j,\hat{Y}_h)}^2(1 - R_{Y_j,\hat{Y}_h}^2) + (1 - c^2)(R_{W,Y_j}^2 + R_{W,\hat{Y}_h}^2),$$

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

$$\rho_{jh}^-(c) \equiv \frac{F_{jh}(c) - \Delta_{jh}(c)^{\frac{1}{2}}}{2(R_{Y_j,\hat{Y}_h}^2 - c^2)} \quad \text{and} \quad \rho_{jh}^+(c) \equiv \frac{F_{jh}(c) + \Delta_{jh}(c)^{\frac{1}{2}}}{2(R_{Y_j,\hat{Y}_h}^2 - c^2)}.$$

We then have that

$$0 \leq M_{jh}(r; c) \iff \begin{cases} 
    r \in (-\infty, \rho_{jh}^-(c)] \cup [\rho_{jh}^+(c), \infty) & \text{when } c^2 < R_{Y_j,\hat{Y}_h}^2 \\
    r \in [\rho_{jh}^+(c), \rho_{jh}^-(c)] & \text{when } R_{Y_j,\hat{Y}_h}^2 < c^2.
\end{cases}$$

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

The sharp bounds $D^{k,\tau,e}$, $B^{k,\tau,e}$, and $G^{k,\tau,e}$ follow from the mappings $D(\cdot)$, $B(\cdot)$, and $G(\cdot)$ in Theorem 2.3.1.
50% (dark) and 95% (light) confidence regions for $\beta_{j1}$ (cash flow) for $j = 1, 2, 3$ (investment, saving, and debt) from year 1970 to 2017, when $X$ includes asset tangibility. We consider the regions $B_{j1}, B_{j1}^{c*}$, and $B_{j1}^{\kappa,\tau,c}$ where $c^* = 0$, $\kappa$ and $\tau$ are such that $\hat{\kappa}^* = 0.5$ and $\hat{\tau}^* = (0.9, 0.9, 0.9)'$, and $c$ is such that $(\xi_{12}, \xi_{13}) = (\xi_{23}, \xi_{23}) = (-1, 0)$ and $(\xi_{13}, \xi_{13}) = (0, 1)$. The shaded vertical bars indicate years in which the maintained assumptions are rejected.
Table 16: Bounds on the Cash Flow Coefficients in the Investment, Saving, and Debt Equations Using the Full Panel and Accounting for Asset Tangibility

<table>
<thead>
<tr>
<th></th>
<th>$J^{\kappa,\tau}_{\delta}$</th>
<th>$J^{\kappa,\tau}_{\delta,c}$</th>
<th>$J^{\kappa,\tau,c^*}$</th>
<th>$b_{Y_1(Y_2,Y_3,\bar{Y})}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Results without fixed effects for $\kappa = \infty$ and $\tau = (1, 1, 1)'$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>[-0.134, 0.144]</td>
<td>[-0.028, 0.144]</td>
<td>[0.130, 0.144]</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(-0.142, 0.145)</td>
<td>(-0.033, 0.145)</td>
<td>(0.129, 0.145)</td>
<td></td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>[-0.438, 0.122]</td>
<td>[-0.014, 0.122]</td>
<td>[0.111, 0.122]</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(-0.456, 0.124)</td>
<td>(-0.019, 0.124)</td>
<td>(0.109, 0.124)</td>
<td></td>
</tr>
<tr>
<td>$\beta_{31}$</td>
<td>[-0.419, 1.242]</td>
<td>[-0.419, -0.248]</td>
<td>[-0.419, -0.405]</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(-0.423, 1.294)</td>
<td>(-0.423, -0.242)</td>
<td>(-0.423, -0.402)</td>
<td></td>
</tr>
<tr>
<td>Results without fixed effects for $\kappa^* = 0.5$ and $\tau^* = (0.9, 0.9, 0.9)'$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>[0.127, 0.144]</td>
<td>[0.127, 0.144]</td>
<td>[0.130, 0.144]</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.125, 0.145)</td>
<td>(0.125, 0.145)</td>
<td>(0.129, 0.145)</td>
<td></td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>[0.108, 0.122]</td>
<td>[0.108, 0.122]</td>
<td>[0.111, 0.122]</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.106, 0.124)</td>
<td>(0.106, 0.124)</td>
<td>(0.109, 0.124)</td>
<td></td>
</tr>
<tr>
<td>$\beta_{31}$</td>
<td>[-0.419, -0.401]</td>
<td>[-0.419, -0.401]</td>
<td>[-0.419, -0.405]</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(-0.423, -0.398)</td>
<td>(-0.423, -0.398)</td>
<td>(-0.423, -0.402)</td>
<td></td>
</tr>
<tr>
<td>Results with year and firm fixed effects for $\kappa = \infty$ and $\tau = (1, 1, 1)'$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>[-0.629, 0.130]</td>
<td>[-0.480, 0.130]</td>
<td>[-0.481, 0.130]</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(-0.634, 0.131)</td>
<td>(-0.484, 0.131)</td>
<td>(-0.485, 0.131)</td>
<td></td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>[-2.267, 0.170]</td>
<td>[-0.273, 0.170]</td>
<td>[-0.274, 0.170]</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(-2.300, 0.172)</td>
<td>(-0.279, 0.172)</td>
<td>(-0.280, 0.172)</td>
<td></td>
</tr>
<tr>
<td>$\beta_{31}$</td>
<td>[-0.369, 32.655]</td>
<td>[-0.369, -0.315]</td>
<td>[-0.369, -0.315]</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(-0.370, 35.443)</td>
<td>(-0.370, -0.31)</td>
<td>(-0.370, -0.309)</td>
<td></td>
</tr>
<tr>
<td>Results with year and firm fixed effects for $\kappa^* = 0.5$ and $\tau^* = (0.9, 0.9, 0.9)'$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>[0.083, 0.130]</td>
<td>[0.083, 0.130]</td>
<td>[0.083, 0.130]</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.083, 0.131)</td>
<td>(0.083, 0.131)</td>
<td>(0.083, 0.131)</td>
<td></td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>[0.136, 0.170]</td>
<td>[0.136, 0.170]</td>
<td>[0.136, 0.170]</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.135, 0.172)</td>
<td>(0.135, 0.172)</td>
<td>(0.135, 0.172)</td>
<td></td>
</tr>
<tr>
<td>$\beta_{31}$</td>
<td>[-0.369, -0.364]</td>
<td>[-0.369, -0.364]</td>
<td>[-0.369, -0.364]</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(-0.370, -0.362)</td>
<td>(-0.370, -0.362)</td>
<td>(-0.370, -0.362)</td>
<td></td>
</tr>
</tbody>
</table>

The sample is an unbalanced panel of 161,959 firm-year observations. $Y_1$, $Y_2$, and $Y_3$ denote Investment, Saving, and Debt respectively and $X = [\text{Cash Flow}, \text{Firm Size}, \text{Asset Tangibility}]$. 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. $c$ sets $(c_{12}, c_{13}) = (c_{21}, c_{23}) = (-1, 0)$ and $(c_{13}, c_{12}) = (0, 1)$ whereas $c^* = 0$. Robust standard errors for $\pi$ are clustered by firm. 50% and 95% confidence regions are in brackets and parentheses respectively.
BIBLIOGRAPHY for “Shareholder Recovery and Leverage”


BIBLIOGRAPHY for “Measurement Error in Multiple Equations”


