How Quickly Do Firms Readjust Capital Structure?

Peter Iliev
Pennsylvania State University
pgi1@psu.edu

Ivo Welch
Brown University
ivo_welch@brown.edu

November 10, 2009

Abstract

This paper seeks to reconcile the different results from prominent estimators of the speed of adjustment (SOA) of firms' leverage ratios. Previous papers overlooked the simple fact that leverage ratios less than 0% or greater than 100% are not possible. This made some of them find mean reversion, which they mistakenly considered as readjustment. When corrected, the best reconciled estimate for the SOA is not positive: On average, firms do not seem to adjust. Moreover, the data is so plentiful that SOA estimates can be extremely accurate even when the firm-specific target is not known. Finally, our paper suggests both a method of reconciling estimators from prior research and a better way of modeling the underlying leverage ratio process.

JEL Code: G32

As early as 1984, Stewart Myers (Myers (1984)) put forth the two contrasting viewpoints that have become the central issue in capital structure research: should one view the existing capital structure to be more in line with a non-adjustment view (such as his pecking order), or more in line with a (low-friction) optimal tradeoff view? Our paper estimates one parameter that is related to these two views: the speed of readjustment (SOA) of leverage ratios to shocks. Its point estimate allows a reader to judge where on the scale between adjustment and non-adjustment firms lie. Although the SOA has been estimated in many papers, even 25 years after Myers (1984), it still remains in dispute. This view is not just our own. In their survey article, Frank and Goyal (2008) state that "the speed at which [corporate leverage is mean-reverting] is not a settled issue." And Huang and Ritter (2009) opine that it is "perhaps the most important issue in capital structure today."

The two extremes are perfect non-readjustment (an SOA of 0) and perfect readjustment (an SOA of 1). The most prominent published estimates for the SOA in the literature range from 34% (Flannery and Rangan (2006, Table 7)), to 25% (Lemmon, Roberts, and Zender (2008, Table 6)), to 23% (Huang and Ritter (2009, Table 8), to 7-18% (Fama and French (2002, Table 4)), to practically zero (Welch (2004, Table 3)). The economic interpretation of these estimates contrasts starkly. An SOA of 20% suggests a half-life for the influence of a shock of about $\log(0.5)/\log(0.8) \approx 3$ years, somewhat in line with active managerial intervention. An SOA of 10% suggests a half-life of 6.6 years, more in line with the "glacial readjustment" view of Fama and French (2002). An SOA of 4% suggests a half-life of 17 years, in line with the "practically no-readjustment" views of Myers (1984) and Welch (2004). 17 years is longer than the median lifetime of firms on Compustat.

Our paper intentionally does not propose a new estimator. Adding yet another technique with yet another estimate would only confuse the matter even further. Instead, the main goal of our paper is explaining why the already existing methods in this literature have come to such different conclusions. Is it different data sets, or different leverage definitions, or different techniques, or something else altogether?

Perhaps the most important insights of our paper are also the simplest ones.

First, the data contains thousands of firms and many decades—a total of over 100,000 firm-years. We can show (in retrospect not surprisingly) that reasonable estimators are able to produce almost perfect estimates. Yet, the aforementioned large discrepancies in empirical estimates remain even if we use the same data definitions and sample. The main challenge in this literature is not estimation uncertainty, but the need to reconcile SOA estimates that should all be perfectly on target and yet are many standard errors away from one another.

Second, if target leverage ratios are firm-specific constants (or at least not moving in a manner that is correlated with the shocks), then it is not important to know the firm-specific targets when estimating the SOA. After suitable corrections, we already have techniques that can estimate the SOA with practically perfect precision even in the absence of firm-specific target information. Thus, differences in control variables that proxy better or worse for firm-specific leverage target cannot be responsible for the discrepancies in empirical estimates, either.

Third, the leverage ratios used in these papers are by definition feasible only between 0% and 100%.[1] Yet, the properties of even some of the best econometric estimators that earlier papers have employed were derived under a process assumption that is impossible in our context, because our dependent variable is not unlimited. This process is $\tilde{x}_{i,t} = \rho \cdot x_{i,t-1} + \alpha_i + \tilde{\epsilon}_{i,t}$, henceforth SDPP (standard dynamic panel process), with iid normally distributed shocks $\tilde{\epsilon}_{i,t}$. In most papers in the literature, readjustment is considered to be the mean reversion, so these papers report a speed of adjustment that is one minus this rho. Yet, when leverage ratios do not follow this process, these estimators become biased. They attribute the fact that leverage ratios remain inside the feasible domain as mean reversion. Therefore, their SOA estimates are too high. Moreover, when the true SOA is low, leverage ratios reach their domain limits more frequently. This causes a concave mapping from empirical estimates to true values: Low speed of adjustment estimates are relatively more biased.

The first point is also a manifestation of our third point. If leverage ratios had followed the SDPP, (some of) the estimators used in previous papers are excellent and should have produced virtually perfect and thus identical rho estimates. Again, the problem is that the SDPP was never suitable in the leverage ratio context.

Our paper then proposes and evaluates a set of processes for (the underlying true) leverage ratios that do not violate their definitional 0-to-1 range and examines the performance of existing estimators under these processes. It shows how these estimators have produced biased estimates under these processes. Rather than leaving the literature with yet another measure for the speed of adjustment (and thus possibly even more fractious), our paper proposes a method that reconciles different estimators from earlier literature. (We see this as a major contribution of our paper.) After suitable adjustments, we find that the SOA estimate that best reconciles the empirical methods and estimates in the literature is just about zero. (It could also be negative. If we ignore some evidence, we can defend it to be "as high as" 5% per year. This still corresponds

[1] This issue of the importance of domain limits of leverage ratios has also been recently and independently noted by Chang and Dasgupto (2009) in their analysis of tests of the pecking order hypothesis and refinancing, all in the context of book-value based ratios. Their paper naturally complements our own. Similarly, Shyam-Sunder and Myers (1999) mention other forms of mechanical mean reversion. In some sense, the Chang-Dasgupta piece builds more on the Shyam-Sunder-Myers piece than on the papers we examine.

to a half-life for the impact of a shock measured in decades, and roughly as long as the typical lifetime of firms in the Compustat data base.)

However, the average SOA is not necessarily applicable to all firms. We also find some evidence that firms may be heterogeneous. Some of this heterogeneity has already been identified. In particular, Leary and Roberts (2005) are correct that firms that experience extreme shocks (a rate of return of less than –20% or greater than +50%) *and that remain publicly traded* do show leverage ratio readjustments. They unlever. However, we show that their evidence suffers from selection issues. A large fraction of firms with very negative stock returns disappear from the tapes before they report their end-of-year leverage ratios. If we assume that these firms exited because they went bankrupt (with a leverage ratio of 100%), then the average SOA coefficient is *much* lower. We find no economically meaningful readjustment (and definitely not readjustment with the half-life of 3 years visually suggested by Leary and Roberts (2005)) even for firms that have experienced stock rates of return of –50% or lower.

Our paper also shows (again) that adjustment and inertia are too different issues. Our firms are *not* inactive. In fact, the average non-stock-return-caused change in leverage ratios is almost 9% per year. Despite such large year-to-year changes (typically every second or third year), managers do not use the opportunity to readjust towards a target ratio. It is *this* evidence—and not the slow speed of adjustment evidence itself—that casts doubts on explanations for slow SOAs that are based on the costs of issuing debt or equity.

Our paper now proceeds as follow: We describe the data and the estimators in Section 1. We successfully replicate the findings of earlier studies in our own data set, and describe the inference that the reader can draw from the empirical estimates without adopting our own leverage ratio processes. Section 2 explains the problems that arise when shocks (epsilons) are additive and normally distributed, and shows empirical evidence that this is an important concern in our context. It then proposes four alternative processes to model the underlying evolution of leverage ratios, in which leverage ratios cannot exceed the unit domain. Section 3 fits the estimators on these processes. Section 4 determines which process and rho can best reconcile the differing estimates in the literature. Section 5 looks at some other specifications. It shows that moving targets do not matter; that lumpiness does not change the inference; that firms with extreme positive shocks and extreme negative shocks do show just a little readjustment behavior (if we consider survivorship bias); and that multi-year estimation methods come to similar conclusions. And Section 6 adds a critical perspective about what our findings imply, especially for (dynamic) tradeoff theories.

1 Empirical Estimates and Replication of Findings

1.1 Data and Variables

[Table 1 here]

Table 1 explains our data and the construction of the leverage measures used in our paper. Our main data source is Compustat from 1963–2007, supplemented with CRSP stock returns without dividends. Depending on data requirements, the samples used in later analyses have up to 136,450 firm-year observations. The panel is irregular. The average number per firms in a given fiscal year is 3,299, with a range from 453 to 5,804. The average number of years per firm in the sample is 19, although the median is only 16.

We use L as our generic notation for leverage. Our (non-exclusive) emphasis will be on the market-value based financial debt-to-capital ratio, DC. It is the sum of long-term and current debt divided by the sum of long-term and current debt and the market value of equity. Table 1 shows that DC has a pooled mean of 27%, a pooled standard deviation of 25%, and an average cross-sectional standard deviation of 23%. In some papers, leverage ratios were defined differently. The most common correct[2] alternative to DC is LA, the ratio of total liabilities over total assets.[3] Table 1 shows that the LA ratio has a mean of 40%, which suggests more indebtedness than the DC ratio. Some papers calculate leverage ratios using the book value of equity. We denote these with lower-case letters. Table 1 shows that the means of book-value based ratios are similar to their market-value based equivalents, but their heterogeneity is a little lower.

The average leverage ratio in the sample drifted up at a rate of around 0.6% to 1.2% per firm-year, with a cross-sectional standard deviation of around 11-12% for the market-based ratios and 9-10% for the book-based ratios.

The IDC (implied debt-to-capital) ratio is a measure of what would happen to the leverage ratio if the firm did not change the quantity of its debt or equity instruments, the value of debt were to remain constant, and only the stock's rate of return (without dividends) changed the value of the equity. The difference between $\mathsf{IDC}_{t-1,t}$ and DC_{t-1} can be loosely interpreted as the leverage change that is *not* due to stock returns, but due to the intervention of management. Comparing the 12.6% variability in leverage ratio changes with the 8.7% variability sans stock-return induced changes suggests that about 30% of the year-to-year change in capital structure is caused by stock returns. The

[2] The common "leverage" ratio of financial debt divided by total liabilities is not correct, because it considers non-financial liabilities the same as equity. Thus, it considers a company *less* levered when it has *more* non-financial liabilities.

[3] Huang and Ritter (2009) do not use our exact LA definition. They use as measure of liabilities "Total Liabilities + Preferred – Deferred Taxes – Convertibles" and divide by "Total Liabilities + Preferred – Deferred Taxes + MVE." The results are similar when we use their definition instead of our simpler LA definition.

remaining 70% is (almost entirely) due to debt and equity net issuing activity (including dividends). Stock returns are a first order determinant of capital structure changes, but firms are by no means passive or inert.

1.2 The SDPP Process and Leverage Ratio Targets

[Table 2 here]

Table 2 summarizes the methods that have been used in the relevant literature and that we are investigating in our current paper. We shall formulate the models in terms of one minus the speed of adjustment (SOA). This coefficient is named "rho." It is the parameter of primary interest to us. A fairly general model for the evolution of leverage in a dynamic panel is

$$\tilde{\mathsf{L}}_{i,t+1} - \mathsf{L}_{i,t} \quad = \quad (1-\rho) \cdot \left[\mathsf{T}_{i,t+1} - \mathsf{L}_{i,t}\right] + \tilde{\epsilon}_{i,t+1} \quad ,$$

where $\mathsf{T}_{i,t+1}$ is a firm-specific target leverage ratio, possibly itself a function of time and covariates. When rho is zero, the panel loses its dynamic aspect and becomes static. When the panel is dynamic, the true error and next firm-year's independent leverage observation are correlated.

For the most part, we will assume that each firm has its own *constant* target. The firm-specific targets themselves are not of interest to us. We only need them as controls in our quest to estimate rho. We will consider three cases. In the first case, we assume that the researcher knows them perfectly,

$$\tilde{\mathsf{L}}_{i,t+1} \quad = \quad \rho \cdot \mathsf{L}_{i,t-1} + (1-\rho) \cdot \mathsf{T}_i + \tilde{\epsilon}_{i,t} \quad .$$

In the second case, we assume that the researcher does not know anything about the targets. The estimators can then control for them, e.g., by including firm-specific fixed effects or differencing. When each target is a firm-specific unknown constant (t_i'), then we can write this model in the form of the econometric *standard dynamic panel process* (SDPP),

$$\tilde{\mathsf{L}}_{i,t+1} \quad = \quad (1-\rho) \cdot t_i' + \rho \cdot \mathsf{L}_{i,t} + \tilde{\epsilon}_{i,t+1} \quad = \quad \rho \cdot \mathsf{L}_{i,t-1} + t_i + \tilde{\epsilon}_{i,t} \quad , \qquad \text{(SDPP)}$$

where t_i are firm-specific intercepts. (The $1-\rho$ scalar on unspecified intercepts does not change rho estimates.)

In the third case (in Section 5), we allow the unknown target to vary over time, although not in a manner that is correlated with shocks. In real life, none of these scenarios is perfectly accurate. The researcher can proxy for the target with fixed effects and observable covariates, but this probably succeeds in capturing the true target for each firm only with modest accuracy.

1.3 Existing Estimators

We now discuss the estimators that have been proposed to recover the rho.

OLS Estimator: The simplest panel estimator is a pooled OLS regression, in which leverage is regressed on past leverage and usually additional firm-specific variables. These covariates are included to capture the unknown firm-specific hetrogeneous targets. It is easy to verify that with thousands of cross-sectional observations, it also makes no difference whether the model is estimated in differences or in levels." However, OLS has two problems. The first problem is not too important in our context. The assumed orthogonality conditions between residuals and independent values does not hold. When an epsilon is high, the next year's independent variable will then also be high. In small samples, with few firms and time-periods, this can create the well-known "Hurwicz bias" in favor of finding mean reversion (estimated rhos less than 1) even when there is none in the underlying process (true rho of 1). Fortunately, our panel is primarily cross-sectional, with far more firms than years, and there is only one parameter to estimate. Moreover, the Hurwicz bias vanishes asymptotically. We will show that this violation of the orthogonality condition has almost no influence on the estimates in our context. The second problem is serious. The omission of firm-specific targets in the estimation creates a bias in favor of not finding mean-reversion if it is indeed present. We will show that this "omitted variable" bias is severe when firms have firm-specific targets unknown to the researcher.

Fixed-Effects (FE) Estimator: Flannery and Rangan (2006) in effect argue that researcher-supplied target covariates may not capture all the heterogeneity across firms' targets. Thus, they suggest adding fixed effects into the estimator. This takes care of the omitted variables problem. However, the FE estimator properties are applicable under static panel processes, not dynamic ones. Huang and Ritter (2009) noted that the Hurwicz bias reappears, because the intercepts consume a large number of degrees of freedom. Thus, the fixed-effects estimator suffers seriously from the correlation between the error and the next independent variable. Put differently, the target fixed effect assumes the mean of the error realizations.

GMM (BB) Estimators: To allow for fixed effects without suffering the Hurwicz bias, Lemmon, Roberts, and Zender (2008) suggest using the Blundell-Bond (BB) estimator. It is a modified version of the more common Arellano-Bond (AB) GMM procedure. An econometric derivation of this estimator is beyond our scope, but we can provide an intuitive explanation of some of their moment conditions to help readers unfamiliar with this technique (and to help them understand its failure in our context).

The GMM method is easiest to understand by assuming that firm-specific targets are zero $t_i = 0$. The moment conditions are chosen to be cross-correlations, so that these means are irrelevant.[4] The following explanation is therefore valid even when firms have their own targets. Any specific rho estimate implies a specific residual for each firm and year. Thus, the error e can be thought of as a function of ρ, i.e.,

$$e_{i,t}(\rho) \equiv \mathsf{L}_{i,t} - \rho \cdot \mathsf{L}_{i,t-1} \quad .$$

Arellano-Bond find their moment conditions by pointing out that these residuals should be uncorrelated with firms' own two-years-or-more-lagged leverage, $(\mathsf{L}_{i,t-2}, \mathsf{L}_{i,t-3}, ...)$.[5] For example, the expected sum over all firms of, say, $\sum_i e_{i,3}(\rho) \cdot \mathsf{L}_{i,1}$ should be zero. To improve on the intuition, consider an example with five years of level leverage data. Differencing eliminates one year. This leaves four years of residuals, of which two ($e_{i,3}$ and $e_{i,4}$) have at least one two-year lagged L. The three conditions that the AB estimator wants to match are $\sum_i (e_{i,3} \cdot \mathsf{L}_{i,1})=0$, $\sum_i (e_{i,4} \cdot \mathsf{L}_{i,2})=0$, and $\sum_i (e_{i,4} \cdot \mathsf{L}_{i,1})=0$. Thus, the GMM procedure seeks to find the rho that minimizes a *squared* penalty function of

$$w_{\mathrm{M1}} \left[\sum_i (\mathsf{L}_{i,4} - \rho \cdot \mathsf{L}_{i,3}) \cdot \mathsf{L}_{i,1} \right] + w_{\mathrm{M2}} \left[\sum_i (\mathsf{L}_{i,4} - \rho \cdot \mathsf{L}_{i,3}) \cdot \mathsf{L}_{i,2} \right] + w_{\mathrm{M3}} \left[\sum_i (\mathsf{L}_{i,3} - \rho \cdot \mathsf{L}_{i,2}) \cdot \mathsf{L}_{i,1} \right]$$

where w_{M} are weights whose optimal values can be asymptotically derived. The intuition for their selection is that we want to weight moment conditions more when they have more to say about ρ—meaning that small changes in ρ would cause large changes in the particular moment sum. (We would not expect a residual to have much correlation with leverage from 20 years earlier regardless of what the actual rho is.)

Rearranging the penalty function yields

$$\begin{aligned}
& (y - \rho \cdot x)^2 \\
\text{where} \quad y &= w_{\mathrm{M1}} \cdot \sum_i \mathsf{L}_{i,1} \cdot \mathsf{L}_{i,3} + w_{\mathrm{M2}} \cdot \sum_i \mathsf{L}_{i,2} \cdot \mathsf{L}_{i,4} + w_{\mathrm{M3}} \cdot \sum_i \mathsf{L}_{i,1} \cdot \mathsf{L}_{i,4} \\
\text{and} \quad x &= w_{\mathrm{M1}} \cdot \sum_i \mathsf{L}_{i,1} \cdot \mathsf{L}_{i,2} + w_{\mathrm{M2}} \cdot \sum_i \mathsf{L}_{i,2} \cdot \mathsf{L}_{i,3} + w_{\mathrm{M3}} \cdot \sum_i \mathsf{L}_{i,1} \cdot \mathsf{L}_{i,3} \quad .
\end{aligned}$$

[4]The fixed effects estimator imposes orthogonality conditions between these errors and the mean of the independent variables that are not likely to be satisfied. The GMM estimator ignores those (mean) conditions that are not likely to be orthogonal, and focuses only on those correlation conditions that are likely to remain valid (orthogonal). By doing so, the GMM estimator loses a large number of (linear) restrictions. With over 100,000 firm-years, this is not a big problem because the remaining conditions are sufficiently powerful.

[5]This also indicates why the BB estimator fails in our context. Intuitively, if the two-period lagged L is very low, e.g., zero, it is common that the one-period lagged L is also zero. For firm's with such zero leverage, the shock ϵ cannot be negative. It is this feature that is ultimately the problem for this estimator in our context. It was built upon the lack of correlation between ϵ and lagged L, which cannot hold here.

Consequently, one can obtain the GMM rho via a least-squares estimator by running a regression $y = \rho \cdot x + \text{noise}$, in which the sums of these predetermined leverage products are regressed on one another. Intuitively, the GMM rho estimate is higher when the AR(t) cross-products are more similar to the AR(t+1) cross-products. Generalizing to an arbitrary number of years, Arellano-Bond provide $(T-1) \cdot (T-2)/2$ such moment conditions. It is not uncommon (and Lemmon, Roberts, and Zender (2008) do so) to use only moment conditions in which e is close (in time) to L. After all, the optimal weight on moment conditions with more than a few years distance between residual and leverage is small anyway.

The Blundell-Bond "SYSTEM" statistic innovates over Arellano-Bond by adding another $T-2$ moment conditions: each residual should also not be correlated with lagged *changes* in leverage. For example, $e_{i,3}(\rho)$ should also not be correlated with $\mathsf{L}_{i,t-2} - \mathsf{L}_{i,t-1}$. BB prove that their GMM estimator is asymptotically efficient as $N \to \infty$ under the SDPP with independent normally distributed errors.

Long Difference (LD) Estimator: Huang and Ritter (2009) use the "long difference instrumental variables" estimator, suggested by Hahn, Hausman, and Kuersteiner (2007) (HHK). The LD approach suggests running an OLS regression on the longest differences that are available. For example, with 20 years of data per firm, it would estimate

$$(\mathsf{L}_{i,20} - \mathsf{L}_{i,2}) \quad = \quad \rho \cdot (\mathsf{L}_{i,19} - \mathsf{L}_{i,1}) \quad + \quad e1_{i,\text{Estimation Stage}}, \quad .$$

However, $(\mathsf{L}_{19} - \mathsf{L}_1)$ is measured with error. Even though the average true epsilon error becomes smaller as the difference increases, higher ϵ errors can still cause higher L_{19} values. This error-in-variables problem means that the estimate of rho would be biased towards 0 if simple OLS were used. The lagged difference in leverage is still not orthogonal to the (differenced) error. To reduce this problem, $(\mathsf{L}_{19} - \mathsf{L}_1)$ is replaced with its fitted value from an IV regression, which (given a rho estimate) is essentially

$$(\mathsf{L}_{i,19} - \mathsf{L}_{i,1}) = g1 \cdot \mathsf{L}_{i,1} + g2 \cdot (\mathsf{L}_{i,3} - \rho \cdot \mathsf{L}_{i,2}) + g3 \cdot (\mathsf{L}_{i,4} - \rho \cdot \mathsf{L}_{i,3}) + \ldots + e2_{i,\text{IV Stage}} \quad .$$

This is then iterated a few times. Although HHK are quite concerned with choosing a good rho starting value, ultimately this matters little as long as the two equations are iterated more than a few times. Given an arbitrary starting value for rho between 0 and 1, iterating at least a half a dozen times yields the unique rho estimate. HHK prove that their LD estimator is asymptotically efficient as $N \to \infty$ under the SDPP with independent normally distributed errors.

Intuitively, as with the GMM estimators, the quality of the LD estimates in small samples depends on the strength of its moment restrictions (instruments) that are not ignored. Because the estimator requires a long time series, and because the data set is irregular, Huang and Ritter (2009) consider different length estimators (4, 8, 18 and 28 years). These trade off accuracy against data selection in different ways. Our replications focus on their overlapping 8-year estimates, but we also confirmed that the results are robust to different choices. (For example, in Table 3, the estimates typically varied by about 1%, with the largest discrepancy being 4%.)

Welch (2004) (W) "Implied Target" Estimator: This method is quite different from the tests above. It is *not* based on the SDPP model. Instead, it is based on a specific economic identification. As already noted on Page 4, the implied leverage ratio, IDC, that comes about in the absence of corporate capital structure activity is $\mathsf{IDC}_{t,t+1} \equiv D_t/[D_t + (1 + r_{t,t+1}) \cdot E_t]$. The W method is a competitive OLS regression of

$$\widetilde{\mathsf{DC}}_{i,t+1} \quad = \quad a + b \cdot \mathsf{DC}_{i,t} + \rho \cdot \mathsf{IDC}_{i,t,t+1} + \tilde{e}_{i,t} \quad .$$

The lagged DC is intended to capture the most recent firm's own target (somewhat similar to a fixed effect), while the constant is intended to capture a common target. The coefficient ρ measures the degree to which only stock-return-caused noise is or is not undone.

In our current paper, we implement the estimation with a simple pooled regression (clustered standard errors) rather than with the Fama-Macbeth-style method used in the original paper.[6] The rho estimates are almost identical. The W estimator nests the simple pooled OLS estimator. In Table 3 below, we also report the rho estimate from a simple extension that allows for fixed effects. This can nest the FE estimator.

Not being an SDPP approach, the W estimator cannot be used in other contexts—not even when leverage is defined by book value. It also has the disadvantage that its properties are not known. Moreover, it is not clear whether the past leverage ($\mathsf{DC}_{i,t}$) term in the regression can be successful in controlling for firm-specific targets, as interpreted by Welch (2004). Thus, one wonders how this estimator performs if each firm has its own constant target.

1.3.1 What Shock?

The estimators are non-chalant in spelling out what shocks they are considering. This deserves some clarification.

Changes to leverage ratios have to come either from changes to debt or changes to equity. In turn, these can come from changes in the quantity or the value of claims.

[6] There was really no justification for using the Fama-Macbeth method in Welch (2004). It served only to show that standard errors remain small even when one sacrifices power. The standard errors in Welch (2004) were wider (around 1.4%) than those reported in our current paper (around 0.4%).

As for changes to the quantity of debt and equity claims, it is difficult to interpret the meaning of what a "shock" is and what consequent managerial readjustment is—after all, these *are* usually "shocks" caused by the capital structure actions of the managers *themselves*.[7]

Adjustment by managers to manager-caused leverage ratio changes are arguably less interesting than shocks that are mostly imposed on the firm from the outside. These more interesting shocks would be to the value of the firm, which would manifest themselves primarily in the equity portion. (There could also be shocks to the risk of the claims.) Such shocks are *relatively*[8] more "exogenous." The W estimator considers only these shocks. Thus, strictly speaking, the W method does not even ask the same question as the other estimators: Instead of investigating the question of how firms respond to *any* kind of shock, it investigates the response only to a specific kind of shock—leverage ratio changes associated with stock returns.

But the big advantage of focusing only on stock-return related shocks is that it is a lot easier to learn the response to them. In this case, the researcher has ex-post knowledge of the exact amount of the shock that stock-returns would have caused in the absence of managerial response. Ignoring this information would be equivalent to ignoring information about the true disturbances in an OLS regression if they were available. (Such knowledge would allow one to pin down OLS coefficients without error.) In other words, one can answer the question of how firms respond to stock-return caused shocks with a lot more accuracy than one can answer the question of how firms respond to arbitrary shocks of any kind.

---

[7] We are not entertaining separate adjustment speeds for stock return caused shocks and other shocks. This is beyond the scope of our current paper and left to future work. Briefly, the tests could further be generalized if one could identify more exogenous shocks (as opposed to debt or equity changes explicitly caused by firms in order to adjust their leverage ratios). For example, one could measure (unusual) debt coming due in one year (disclosed by firms in their financial footnotes), which would allow measuring how firms respond to this kind of shock. However, firms would likely have already planned for them in their capital structure. In this case, "no reaction" may not mean "no adjustment." That is, it might not be surprising if firms were not reacting to these anticipated shocks.

[8] The identification comes at a cost of ignoring one form of reverse causality: if debt issues *cause* negative stock returns, then the no-readjustment leverage ratio would go up, and W would mistakenly attribute managers' intentional levering up to the no-readjustment hypothesis. The same holds if equity issues *cause* positive stock returns. Equity issue stock price announcement evidence suggests the opposite, however.

In addition, because the identification is so specific (based an a quantitatively exact ratio on own stock returns), it is both less sensitive to this reverse-causality problem—unless the stock return and net issuing activity happen in exactly the right proportion, IDC is unlikely to pick it up strongly—and more difficult to correct. A correction would be difficult because any stock return other than the actually observed exact one should not work under the functional identification. Unlike common IV techniques in linear models, in W even a constant added to the stock return would destroy the usefulness of the IDC identification.

(Association rather than reverse causality is not a problem due to the way this evidence is interpreted.)

In sum, if the question is how managers respond to the specific shock of stock returns, then the W estimator has an "unfair" advantage similar to knowledge of the true OLS epsilons. It uses information ignored by the other methods. Of course, given our large sample and the fact that in one variation we allow the researcher to know the firm targets perfectly, other methods should be extremely accurate, too.

1.4 Replication of Findings

[Table 3 here]

Table 3 shows the results of the estimators applied to the empirical data. The coefficient that is to be interpreted as the influence of past leverage is in bold. For the OLS, FE, BB, and LD methods, this is simply the coefficient on lagged leverage (DC). For the W estimators, with its interest in the role of shocks from stock returns, this is the coefficient on IDC. Different sections in Table 3 consider different definitions of leverage. Table 4 repeats Table 3 in the 40,708 firm-years that can be used in all estimations. This sample is mostly limited by the LD estimator's need for long time series. In such a "constant sample," differences in the firm-years that are used by different estimators cannot be responsible for differences in estimates.

[Table 4 here]

The first observation from both tables is that it does not matter greatly which definition of leverage is used. The rho estimates are remarkably similar, with differences caused by leverage definitions of less than 0.02. The often heated disagreements about whether book-value or market-value based leverage ratios are more appropriate seems irrelevant in this context.

The second observation is that the estimation methods themselves are responsible for the ordering of inference. The W estimate for rho is always highest, followed by OLS, BB, LD, and FE. As in the literature, the range of estimates is uncomfortably large, with FE seemingly suggesting a rho of less than 0.7, LD suggesting 0.78, BB suggesting 0.85, OLS suggesting 0.9, and W suggesting 1.

Comparing Table 3 and 4 shows that for the DC leverage ratio estimates, the OLS estimate decreases by 0.1%, the FE estimate decreases by 1.6%, the BB estimate increases by 1.0%, the LD estimate decreases by 0.04%, and the W estimate increases by 1.2%. Thus, all in all, Table 4 estimates are similar to those in Table 3. Differences in the sample cannot explain the discrepancies in estimates.

Our estimates are also similar to the coefficient estimates reported in our predecessors. Lemmon, Roberts, and Zender (2008, Table 6) explain book-leverage and report a coefficient of 87% in an OLS estimation (a), 64% in an FE estimation (d), and 78% in their own BB estimation. This is just 4% lower than the rho estimate for the book-value based *dc* that we obtain in our sample. Huang and Ritter (2009, Table 8) even include a similar summary table, although it is based on reporting of the earlier papers and not based on an independent replication. They report rhos of between 93% (dividend paying firms) and 85% (non-dividend paying firms) for the OLS method in Fama and French (2002); 91.7% for the OLS method in Kayhan and Titman (2007); 64.5% for the fixed-effects method in Flannery and Rangan (2006); 67.8% for the fixed-effects method in Antoniou, Guney, and Paudyal (2008); and 76.8% for their own long-difference method.

Half-Lives: The common way to gain intuition into the meaning of these SOA estimates is to translate them into "half-lives." (1-ρ) is the expected percentage by which the gap between the past leverage and the target closes in one period. Half-life is the time that it takes a firm to adjust back one-half the distance to its target leverage after a one unit shock to the error term ($\epsilon_{i,t}$). For an AR(1) process, half-life is $\log(0.5)/\log(\rho)$. The discrepancies in these methods' rho estimates cause stark contrasts in interpretation. The fixed-effect estimate of 0.66 suggests a half-life of less than 2 years, while the W estimate of 1 suggests a half-life of infinity. Half-life is particularly sensitive when rho is close to 1. Rhos of 0.8 or lower suggest half-lives of around 3 years, in line with the limited trade-off based view of capital structure in Flannery and Rangan (2006), Huang and Ritter (2009), and Lemmon, Roberts, and Zender (2008). (If projects appear with much higher frequency, and if they need to be financed quickly, even this adjustment seems slow, however.) Rhos above 0.9 suggest half-lives of 6 years, more in line with the "glacial readjustment" view Fama and French (2002). Rhos above 0.95 suggest half-lives in excess of 13.5 years, in line with the "practically no-readjustment" view of Myers (1984) and Welch (2004). 13.5 years is only slightly lower than the 16-year median life of firms in our sample.

1.5 Discrepancies and SDPP Inference

The estimated standard errors of estimates in Tables 4 and 3 are very low. This should not be surprising. The observation count is thousands of firms multiplied by dozens of years, noise is modest (leverage ratios can only change so much), and there is really only one parameter of interest. Therefore, the estimation of the standard errors of these rhos is easy—in fact, if one simply considered them to be economically close to zero, it would not seem like a big mistake. More precisely, some simple unreported simulations of the SDPP with normally distributed ϵ errors confirm that the BB and LD estimators are not only asymptotically unbiased, but also unbiased in our sample; adn these simulations show that their asymptotic standard errors are right on the mark, too. Put simply, they have excellent properties in a sample as large as our's.

Therefore we can view the BB and LD estimators as "SDPP diagnostics" for one another. Consequently, we can draw one conclusion from their uncomfortably large disagreements:

Proposition: We can reject with extremely high confidence the hypothesis that leverage ratios follow the SDPP with iid normal errors.

The reason is that we would simply not expect to observe distances between the rho estimates that are as large as those observed. Even the smallest difference—from the 0.847 and 0.772 for the BB and LD estimators in the full sample—is still 15 standard deviations.

It is further unlikely that the key problem would be correlations among unknown normally distributed errors (ϵ's) that we have not corrected for—again, there are so many observations that we believe that the rho point estimates are very accurate even under reasonably high cross-correlations. The failure is more likely in the functional process specification itself, and/or in the symmetry and zero assumption of the errors.

2 Leverage Processes

In this section, we first assess the consequences of real-world relevance of the domain limits of leverage ratio, and then propose some processes that cannot exceed them.

2.1 Problems Due To The Additive Normal Shocks

The process under which the BB and LD panel estimators (and, to a lesser extent, the FE estimator) were derived and shown to be asymptotically efficient is the SDPP. In our context, it is

$$\widetilde{\mathsf{DC}}_{i,t+1} \quad = \quad \rho \cdot \mathsf{DC}_{i,t} + t_i + \tilde{\epsilon}_{i,t+1}$$

with independent normally distributed epsilon errors. As already noted in the introduction, this is intrinsically flawed. Assuming additive independent normal shocks is incorrect, because it is impossible in the leverage ratio context. Leverage ratios have to be between 0 and 1. Thus, when the initial leverage ratio is 100%, a shock to leverage can only be negative. Fortunately, few firms have very high leverage. When the initial leverage ratio is 0%, a shock to leverage can only be positive. Unfortunately, this is not rare. Figure 1 extends the information in Table 1. It shows that firm-years with low leverage are quite common. Although mean leverage in the sample is 27%, zero is the mode of the distribution. With a cross-sectional standard deviation of about 25%, our variable of interest is left-truncated at just about –1 standard deviation. And as Table 1 showed, the average year-to-year change in leverage ratios, limited by the domain, is a remarkably high 12%, of which 8.7% is due to managerial intervention.[9]

[Figure 1 here]

[9]Note that one option is to abandon the definition of leverage that has been used in the literature. For example, the profession could explain leverage that counts cash as negative debt or the (far more

The fact that leverage ratios are limited in their domain causes predictable biases that can be understood even without sophisticated econometrics. A firm with a leverage ratio of 0% simply cannot experience a shock that further reduces its leverage ratio. It follows that it is logically not possible to expect standard estimators, built for the SDPP process with iid errors, to report a coefficient estimate of rho of 1 even if managers never readjust (the true rho is 1).[10]

> **Proposition:** Even dynamic panel estimators that are unbiased under the SDPP with normal errors are predictably biased towards finding positive speed of adjustment in the leverage ratio context. They attribute the inability of the leverage ratio to exceed its domain of [0,1] as mean-reversion.

That is, even if managers do not take any measures to readjust, it is quite possible that the previously used estimators would detect evidence of mean reversion (a rho less than 1).

The obvious next question is whether hitting the domain limit is merely an intellectual curiosity or a real issue. When we simulate SDPP processes without truncating leverage ratios, the number of firm-years with impossible leverage ratios is large. Figure 2 shows that even when the true rho is 0, normally distributed random shocks of 12.5% per year (the average change in the data) would still push just under one-fifth of firm-years into the infeasible region under the SDPP. When rho increases, there are increasingly many leverage ratios that are not just exceeding the feasible range but that are just absurdly high. When rho is 1, this figure rises to more than one quarter. The range of observed leverage ratios under different rhos is similarly damning. With a rho of 1.0, the 95 percentile range of observed leverage ratios reaches from –72% (unreasonable) to +112% (far from the 27% DC average, but still close to the feasible range). With a rho of 1.1, this range increases to –350% to +400%. With a rho of 1.2, it increases to –2,500% to +2,500%. With this 1.2 rho, even the 80 percentile range reaches from –233% to +260%. **[Figure 2 here]**

We thus conclude that assuming an unconstrained leverage process as a model for observed leverage ratios is not just wrong conceptually, but also difficult to defend as a reasonable empirical approximation that is only violated in rare firm-years.

---

volatile) interest coverage ratios. However, this would change the meaning of the inference. More important to our own paper, it would defeat its purpose. We want to reconcile the existing findings on the existing leverage measures, and not propose new methods to assess the SOA.

[10]There are well-known biases of the OLS estimator when variables are artificially limited even when the model is not dynamic. For example, assume that the true relation is $y = 1 \cdot x + \epsilon$, both ϵ and x are unit-noise, and the sample is asymptotically large. If the y variable is winsorized at 0, then the OLS estimate is 0.5. If the x variable is winsorized at 0, the coefficient is 1.47. If both x and y are winsorized, the coefficient is 0.91—a bias of 0.09. If instead of winsorizing, observations less than 0 are not observed ("selected"), the equivalent coefficient estimates are 0.53, 1.00, and 0.66. Of course, the same biases apply in dynamic panels, with the twist that any one limited leverage data point (usually) enters both the x and y series, in addition to being subject to the confounding Hurwicz bias.

The appendix describes two more potential violations of the SDPP assumption in our context that could also mechanically cause mean-reversion even where there is none. (However, the most important seems to be the domain restriction, which is why we focus our attention primarily on this issue.)

2.2 A Stance

To understand the properties of the estimators better, especially in quantitative terms, we must now take a stance on what a reasonable alternative process for leverage ratios could look like.[11] We want to stay largely within the same conceptual framework that lead to the SPP. Loosely, this means that we presume that leverage follows

$$\mathsf{L}_{\text{Next Year}} \quad = \quad \rho \cdot \mathsf{L}_{\text{No Readjustment}}(\cdot) + (1-\rho) \cdot \mathsf{L}_{\text{Active Readjustment}}(\cdot)$$

with some noise to be specified. The first term captures what happens to the leverage ratio when the manager does not readjust. The second term captures what happens to the leverage ratio when the manager does readjust. The dots in the arguments indicate that both the no-readjustment or the active-readjustment leverage ratios can change or be a function of covariates, such as lagged leverage. The parameter of interest is still the coefficient rho. A rho of 1 is perfect non-adjustment. As already noted, our paper focuses mostly on modeling DC leverage, and active readjustment is considered to be the moving towards a firm-specific target, so we can rewrite this as

$$\widetilde{\mathsf{DC}}_{\text{Next Year}} \quad = \quad \rho \cdot \mathsf{DC}_{\text{No Readjustment}}(\cdot) + (1-\rho) \cdot \mathsf{T}(\cdot) \quad .$$

The SDPP is one example of a process that fits into this framework.

2.3 Limiting The Dependent's Domain in Additive Shock Processes

Our first two processes seek to mimick the SDPP closely. This is important *because* the GMM and LD estimators (and to some extent the FE estimator) were derived under this SDPP process. Their goal is to recover the true rho coefficient. Consequently, our first proposals are processes that modify the SDPP to fit our specific leverage ratio context.

We now introduce a "truncation" mechanism that keeps simulated leverage ratios within the feasible domain (the unit interval). The most obvious possible truncation

[11] We consider only statistical rather than fully structural models. They should be viewed as processes that are locally robust to alternatives, in the same sense in which OLS assumptions are never fully satisfied. The alternative of estimates from structural models are much more difficult to interpret intuitively outside their own contexts (which by necessity usually impose a set of strong economic and functional assumptions).

methods that retain most of the functional process specification of the SDPP are the following:

1. We can assume that firms' leverage ratios "bounce back" as if the 0 and 1 boundaries were mirrors. For example, a –10% shock to a firm with a leverage ratio of 3% would result in a leverage ratio of 7%. In this process, almost no firms would end up at the borders. This is not realistic, so we will not use this modification.

2. We can assume that when a firm's leverage ratio exceeds the unit range, its leverage ratio is simply winsorized. Thus, on the next draw of a (normally distributed) shock, if the firm is at a border and rho is 1, then it has an even chance of remaining there and an even chance of returning to an interior leverage ratio.

3. We can assume that the underlying leverage ratio is not the observed leverage ratio. The underlying leverage ratio can take on any value, but the observed leverage ratio is only equal to the underlying leverage ratio when the latter is between 0 and 1. Otherwise, observed leverage is the corresponding border leverage ratio. Firms are more likely to remain at this border than they are in method 2.[12]

4. We can assume that once a firm's leverage ratio reaches a border, it remains there forever. Unfortunately, such a process yields the non-sensible result that with reasonable year-to-year variability, too many firms would end up in the absorbing state. For example, assuming a 12.5% year-to-year change in leverage ratios, even with a true rho of 0, 58% of all leverage ratios should have ended up with zero leverage in our final year 2007.[13]

The first methods are more favorable to the hypothesis of no-readjustment. They imply that leverage ratios have more reverting behavior even when the true rho is 1. This makes some (possibly observed) empirical reversion behavior achievable even with very high rhos. *We adopt the third method as our way of truncating leverage ratios,* although the second method yields very similar results.[14]

[12]In a sense, this converts a vice of our process into a virtue. Because we allow for additive process errors, our simulated process can enter the negative domain. In this case, while in the negative domain, we assume that the reported leverage is zero. This allows firms to maintain zero observed leverage for some time—just as they do in the data. Figure 2, which showed the fraction of firm-years that firms would report 0% or 100% leverage ratios was actually computed under this third truncation method.

[13]That is, if rho were 1, 65% should end up with zero leverage. To match the year-to-year empirical variability in leverage ratios would require even higher true variability than 12.5% in this process. Thus, the reported probabilities of being in an absorbing state are still much too low. For contrast, the equivalent frequency of being at a leverage ratio of 0 in the previous truncation method are 35% (28%) if rho were 1 (0). These are reasonable frequencies.

[14]For example, the second variant (truncation) raises the rho estimates in Process I by 0.03 for the LD, FE, and BB estimator. (The poorly performing OLS estimator is raised from 0.71 to 0.89.) In Process II, it raises the estimate for LD from 0.82 to 0.83; for FE from 0.90 to 0.94; for BB from 0.93 to 1.0; and

We admit that these truncation methods are ad-hoc, but we are unaware of better alternatives that retain the general SDPP specification. (We have no guidance from previous papers, because these did not spell out feasible alternatives.) For a different approach, we will also offer alternative ways to model leverage ratios.

2.4 Process I: Additive Shocks To Leverage and a Constant No-Adjustment Ratio

Our Process I's change vis-a-vis the SDPP is that instead of assuming independent normal errors, we now assume that leverage ratios suffer the (third method of) winsorizations just discussed,

$$\widetilde{\mathsf{DC}}_{i,t+1} \quad = \quad \rho \cdot \mathsf{DC}_{i,t} + (1-\rho) \cdot \mathsf{T}_i + \tilde{\epsilon}_{i,t+1} \quad , \qquad \text{(Process I)}$$

Actual leverage follows the SDPP process, but *observed* leverage is limited to the unit domain.

There is one other issue of importance (and which also applies to earlier papers, except W). The no-readjustment hypothesis in Process I ironically requires active capital structure management. This is best explained by example. Consider a firm with \$1 in debt and \$3 in equity. It has a leverage ratio of 25%. Assume that it experiences a stock return of 10%. This changes its equity value to \$3.3, and its firm-value to \$4.3. To keep its debt ratio, this firm must change its debt from \$1 to \$1.075 in debt, and its equity from \$3.30 to \$3.225. This means that it has to issue \$0.075 in debt and use it to repurchase \$0.075 in equity under the hypothesis of zero adjustment. These issuing and repurchasing amounts change for different initial leverage ratios and stock returns. To use the P I process, it is this adjustment activity *exactly* that must be considered "non-adjustment."

---

does not change the W estimate. In Process III, truncation is so rare that it makes no difference.) In Process IV, it lowers the BB from 0.92 to 0.91 and the W from 1.11 to 1.06 (with the LD and FE estimators staying the same).

2.5 Process II: Additive Shocks And Stock-Return Dependent No-Adjustment Ratio

Our Process I is exactly like Process II, except that it changes the assumption of what the no-readjustment leverage ratio is. Absent managerial issuing/repurchasing activity, if a stock return perturbs the leverage ratio, it impacts the equity component in the leverage ratio. The impact on leverage ratios is therefore specific, non-linear, and depends on the initial leverage. The new no-readjustment benchmark is

$$
\begin{aligned}
\mathsf{IDC}_{i,t,t+1} &= \frac{D_{i,t}}{D_{i,t}+E_{i,t+1}} = \frac{D_{i,t}}{D_{i,t}+(1+r_{i,t,t+1})\cdot E_{i,t}} = \frac{\frac{D_{i,t}}{D_{i,t}+E_{i,t}}}{\frac{D_{i,t}+E_{i,t}}{D_{i,t}+E_{i,t}}+r_{i,t,t+1}\cdot\frac{E_{i,t}}{D_{i,t}+E_{i,t}}} \\
&= \frac{\mathsf{DC}_{i,t}}{1+r_{i,t,t+1}\cdot(1-\mathsf{DC}_{i,t})}\ ,
\end{aligned}
$$

where $r_{i,t,t+1}$ is the contemporaneous rate of return (without dividends). For example, a firm that has a zero leverage ratio and experiences an exogenous shock that doubles its equity value will still have a zero leverage ratio. A firm that has a 50% leverage ratio and experiences a shock that doubles its equity value will experience a leverage ratio decline from 50% to 33%.

The change in the no-readjustment debt ratio leads to Process II,

$$
\begin{aligned}
\mathsf{IDC}_{i,t+1} &= \frac{\mathsf{DC}_{i,t}}{1+r_{i,t,t+1}\cdot(1-\mathsf{DC}_{i,t})} \\
\widetilde{\mathsf{DC}}_{i,t+1} &= \rho\cdot\mathsf{IDC}_{i,t+1} + (1-\rho)\cdot\mathsf{T}_i+\tilde{\epsilon}_{i,t+1}\ .
\end{aligned}
\qquad \text{(Process II)}
$$

In our process simulations below, we will use firms' own stock returns—so we do not draw them randomly. The random draw is only on $\tilde{\epsilon}$, just as the process suggests. Implementation-wise, we continue to shock leverage ratios by $\tilde{\epsilon}$ as in Process I, but then also apply firms' observed stock returns to determine their next leverage ratios. This means that a good calibration for the variance of the additive shocks in Process II is smaller than its equivalent in Process I, because stock returns can explain some of the variability in leverage ratios.

The plausibility of Process I vis-a-vis Process II rests on the question of what capital structure is in the absence of readjustment. The first process (which is closer to the SDPP) requires that the leverage ratio remains constant, even if stock returns are not zero. The second process requires that the leverage ratio changes in a very specific way, consistent with lack of managerial capital structure activity. It is left to the reader to judge whether the active issuing and repurchasing behavior in Process I appropriately reflects the "no-readjustment" hypothesis better than the no-issuing behavior in Process II. Our paper investigates the estimators under both processes.

2.6 Processes III and IV: Modeling Debt and Equity Changes

Our next processes propose an entirely different approach to modeling the evolution of leverage ratios.

A more natural way to deal with the intrinsic non-linearity of leverage ratios may be to model debt and equity changes themselves, and only then compute the ratio. Without loss of generality, corporate leverage can be written to evolve as

$$\begin{aligned}
\widetilde{\mathsf{DC}}_{i,t+1} &= \frac{D_{i,t+1}}{D_{i,t+1}+E_{i,t+1}} = \frac{D_{i,t}\cdot(1+\tilde{\nu}_{i,t,t+1})}{D_{i,t}\cdot(1+\tilde{\nu}_{i,t,t+1})+(1+\tilde{\eta}'_{i,t,t+1})\cdot E_{i,t}} \\
&= \frac{\frac{D_{i,t}}{D_{i,t}+E_{i,t}}}{\frac{D_{i,t}}{D_{i,t}+E_{i,t}}+\frac{1+\tilde{\eta}'_{i,t,t+1}}{1+\tilde{\nu}_{i,t,t+1}}\cdot\frac{E_{i,t}}{D_{i,t}+E_{i,t}}}
\end{aligned}$$

where $\eta'_{i,t,t+1}\in[-1,\infty]$ is the fractional change in equity and $\nu_{i,t,t+1}\in[-1,\infty]$ is the fractional change in debt. We can further decompose $\tilde{\eta}'$ into a shock due to stock returns ($r_{i,t,t+1}$) and a shock due to everything else ($\tilde{\eta}$), i.e., $(1+\tilde{\eta}')=(1+r)\cdot(1+\tilde{\eta})$.

The hypothesis of no-readjustment can then be viewed as the statement that $\tilde{\nu}$ and $\tilde{\eta}$ are not dependent on the firm's current leverage ratio. Unfortunately, we have no theoretical guidance that tells us the joint distribution of the value shocks to debt and equity and the stock return under the no-readjustment hypothesis. Thus, we will not identify the shocks parametrically. Instead, we randomly draw some other firm-year and adopt the sampled (percent) debt and equity changes (excluding stock returns) as our shocks. It is important for the meaning of this test that this match is chosen without regard to the match's leverage ratio. We combine these two shocks, $\tilde{\eta},\tilde{\nu}$ with the firm's known existing stock returns, and compute the firm's next leverage ratio.

The rest of this process is still similar. We can adopt the additive process from earlier sections,

$$\begin{aligned}
\widetilde{\mathsf{SDC}}_{i,t+1} &= \frac{\mathsf{DC}_{i,t}\cdot(1+\tilde{\nu}_{i,t,t+1})}{\mathsf{DC}_{i,t}\cdot(1+\tilde{\nu}_{i,t,t+1})+(1+\tilde{\eta}_{i,t,t+1})\cdot(1+r_{i,t,t+1})\cdot(1-\mathsf{DC}_{i,t})} \\
\widetilde{\mathsf{DC}}_{i,t+1} &= \rho\cdot\widetilde{\mathsf{SDC}}_{i,t+1} + (1-\rho)\cdot\mathsf{T}_{i,t+1}
\end{aligned}$$

(Process III)

For rhos between 0 and 1, the evolution of leverage ratios is guaranteed to be in the feasible domain. For rhos above 1, it can happen that the weighted average exceeds the domain. In this case, we impose the same winsorization method as in the other two processes.

We entertain a second similar process, in which the target itself can vary in an iid fashion from year to year,

$$\widetilde{\mathsf{SDC}}_{i,t+1} = \frac{\mathsf{DC}_{i,t} \cdot (1 + \tilde{v}_{i,t,t+1})}{\mathsf{DC}_{i,t} \cdot (1 + \tilde{v}_{i,t,t+1}) + (1 + \tilde{\eta}_{i,t,t+1}) \cdot (1 + r_{i,t,t+1}) \cdot (1 - \mathsf{DC}_{i,t})}$$

$$\widetilde{\mathsf{DC}}_{i,t+1} = \rho \cdot \widetilde{\mathsf{SDC}}_{i,t+1} + (1 - \rho) \cdot \mathsf{T}_{i,t+1} \qquad \text{(Process IV)}$$

and $\tilde{\mathsf{T}}_{i,t+1} = \mathsf{T}_{i,t+1} + \tilde{u}_i$, where $\tilde{u}_i$ is normal noise with standard deviation of 9.5%. The rationale is that in the earlier processes, the variability of shocks was independent of rho. In contrast, in Process III, rho also changes the impact of the (observed) shocks on leverage ratio changes. This is because all shocks are in the first term. For example, when $\rho = 0$, then Process III would imply no annual changes in leverage. Thus, under Process III, the estimators might choose to try to match the annual time-series volatility of leverage ratios. This is not the case in Process IV.[15]

These processes are not without drawbacks, either. The biggest problem is one of realistic leverage ratio dynamics for firms with border leverage ratios. We face a conceptual issue: it is not clear how we should interpret the fact that about one in five firms with a zero debt ratio acquires a positive debt ratio in the following year (i.e., the firm issues debt). On the one hand, such changes can legitimately be seen as evidence of active adjustment behavior towards some positive leverage ratio target. On the other hand, if one believes in the presence of shocks, a positive change in leverage is the only possible shock. If we knew how frequently firms would deviate "randomly" under the hypothesis of no readjustment, our tests could become more powerful.

Note also that in Processes I and II, it is possible for firms to move away from the border even if rho is 1 (when they experience a large shock). This is the case in Process IV, too, but not in Process III. Recall that we use the empirically observed first leverage ratio for each firm, this is not the case. If a firm starts out with zero leverage, no percent shock to debt or equity can move it. Further, if a firm has a target ratio of zero leverage, it remains in this absorbing state when rho is greater than 1 and can never exit it. Thus, Process III could benefit especially from a reasonable conjecture about random managerial behavior that could occasionally move a firm with a leverage ratio

[15] Simulations suggest that for rhos between 0 and 1, the standard deviation of annual leverage changes is now between 10.5 (for a true rho of 0.5) and 12.5% (for a true rho of either 0 and 1). In the most relevant domain of rho between 0.8 and 1, the standard deviation of annual leverage changes is now above 12%. Sidenote: in Process IV, after the $\tilde{u}$ shock, the target can be outside the feasible domain for this time period. Thus, we occasionally need to apply the same truncation method as that in the first two processes.

of zero *under the hypothesis of no-readjustment.* However, we are reluctant to impose our own priors on such reasonable disturbances in such a case. This not only favors the adjustment hypothesis over the no-readjustment hypothesis, but also represents a shortcoming in the ability of this process to describe the empirical data.

2.7 Process Comparison

In sum, we do not believe that there is an *a priori* single best candidate process to model the evolution of capital structures. Consequently, we entertain four different processes in the remainder of our paper:

Process I is a simple adaptation of the SDPP assumed in the econometric literature. It differs only by truncating observed (but not underlying) leverage ratios to lie in the unit interval.

Process II is like Process I but changes the assumption of the capital structure under no-readjustment to one that does not require issuing or repurchasing to counteract the influence of stock returns.

Process III makes debt and equity experience (non-parametrically) correlated geometric shocks, and only then computes a leverage ratio. (This process requires no truncation for interior rhos.)

Process IV is like Process III, but adds noise to the target.

The first two are adaptations of the conventional SDPP models that led to the estimators used in the earlier literature. The third and fourth seem more natural. However, neither the additive-shock processes [with its necessary alteration of the domain limit] nor the debt-and-equity processes are perfect descriptions of reality. After all, they are just models. Fortunately, we can evaluate the relative performances of these processes empirically (in Section 4).

The reader may object that other processes could produce different properties for the estimators. This is an appropriate concern. However, our approach is the same as that of every other paper. All econometric estimators and estimates are justified by and interpretable only under their process assumptions—the BB and LD processes were derived under an assumed SDPP process, too. Clearly, this SDPP process does not fit our context. We are unaware of better processes for our context that have been proposed elsewhere. Thus, our assumed processes seem to be the best current way to assess the quality of the existing SOA estimators. Moreover, as just noted, we can even evaluate the relative performances of our processes.

2.8 Summary of Our Simulation Protocol

[Table 5 here]

Table 5 summarizes our simulation protocols. *It is important to remain as close to the empirically observed leverage data as possible.* After all, we are interested in assessing the magnitude of any biases, not just their presence.

- We adopt the exact irregular panel dimension of the Compustat data set.
- Each firm starts with the leverage ratio that it has in the empirical data.
- When a firm disappears from the Compustat data, it also disappears from the simulated data set.
- We draw random values only for the ϵ's that the simulation processes call for. In Process II (the process with additive shocks and stock-return perturbed no-readjustment leverage ratio) and Process III (the debt-and-equity process), each firm also retains its actual stock returns.
- We need to select a target that each firm would want to pursue. We simply assume that firms start out with the correct target ratio one period before our simulations and experience one shock. We have confirmed that the results are similar if we assume that each firm's first leverage ratio was also its target.[16]
- The assumed standard deviations for our shocks are as follows:

  Process I (the additive shock process with constant no-readjustment ratios): We use the standard deviation of shocks of 12.5% (close to 12.6% from Table 1). This would be suitable especially if the true model rho is 1, because it is the standard deviation of the series $\mathsf{DC}_{i,t} - \mathsf{DC}_{i,t-1}$. Variations based on rho estimates from 0.7 to 1.1, yielding $\mathsf{DC}_{i,t} - \rho \cdot \mathsf{DC}_{i,t-1}$ do not greatly alter the shocks' standard deviations. For example, a rho of 0.9 would suggest a shock of 12.2% per year.

  Process II (the additive shock process with stock-return perturbed no-readjustment ratios): We use a standard deviation of shocks of 8.7%. This is the standard deviation of the $\mathsf{DC}_{i,t} - \mathsf{IDC}_{i,t-1,t}$ series from Table 1.

  Processes III and IV (the debt-and-equity process): We are sampling another firm from the panel. The random match is drawn without regard to the firm's or the match's own capital structure. In Process IV, unlike in the earlier processes, to preserve similar year-to-year variability, the target is additionally subject to a yearly iid error of 9.5%.

[16] The only reason why we do not draw a random target for each firm is that we want some congruence between the original leverage ratio and the firm's target. The exact method of choosing a constant firm-specific target choice does not greatly influence the power of the techniques to identify the AR1 coefficient.

In processes I, II, and IV, we are keeping the noise distribution of annual leverage ratio roughly constant when we vary simulating rhos.[17]

- We simulate the capital structure processes and use the procedures from Table 2 to estimate rho with different techniques. Our inference is based on 500 simulations of each process, with each simulation exploring values of true rho from 0.0 to 1.4.

The plots have lines to show the best estimates of the true rho as a function of each estimator's actual empirical rho estimate. This is easier to understand than it is to explain.

3 Fitting Estimators On Processes

3.1 Process I: Additive Shocks To Leverage and a Constant No-Adjustment Ratio

[Figure 3 here]

Figure 3 plots the performance of the first four estimators if target leverage follows Process I. Observed leverage cannot exceed 0 or 100% (via the third truncation method, as explained on page 16). The figure shows the following:

- The OLS estimate is severely upward biased when the true rho is not 1. This is because it fails to pick up that firms have distinct own capital structure targets. It is not a very informative estimator. (Estimating the OLS or FE in changes instead of in levels yields practically identical results.) OLS without target controls is best ignored.
- The FE estimate is severely downward biased. The constants are correlated with the residuals. This bias was also observed in Huang and Ritter (2009, Figure 5). However, the FE estimator has a good slope, and thus a bias-adjusted version can make a good estimator. After such a bias-adjustment, the 0.68 empirical rho estimate corresponds to a best true rho estimate of 0.86.
- The BB estimator performs well. It has a downward bias of 0.04. Thus, it suggests a true speed of adjustment estimate of 0.89.
- The LD estimator performs well. It has a downward bias of 0.03. Thus, it suggests a true speed of adjustment estimate of 0.80.

[17] In Process IV, this is roughly true for rhos below 1.1. Note that we could vary the noise with different rhos, too, but then it would be difficult to distinguish between the influence of rho and the influence of changing residual variance estimates. The variance estimate changes very little if we compute $\mathsf{DC}_{i,t} - a \cdot \mathsf{DC}_{i,t-1}$, where a is any number between 0.8 and 1.2; and our inference stays virtually the same.

- All inference functions are concave. The graphs are more concave when they approach the unit-root. This is primarily because high rhos lead to many more observed leverage ratios that would exceed the feasible leverage ratio domain.[18] Not shown in this figure (but obvious in later figures) is the fact that these functions produce a second set of rhos that match the empirical estimates. Here this happens only for rhos that are implausibly large (greater than 2), so we can exclude these estimates on *a priori* grounds.

The best true rho estimates of between 0.80 and 0.89 suggest *half*-lives of 3 to 6 years, which includes readjustment to changes that managers have caused themselves and are now undoing.

[Figure 4 here]

Figure 4 assumes that the researcher has access to covariates that perfectly identify the target leverage ratio, T. This makes a difference only to the previously badly biased and practically useless OLS estimate. Knowing the target perfectly compensates for the omitted variables problem, i.e., the fact that different firms have different targets. Comparing the figures shows that none of the other estimates are improved by knowledge of the target. The best point estimates for the true rho that correspond to the empirically observed measures remain exactly the same. The standard error is only mildly perturbed, too. In retrospect, this is not surprising. After all, we have over 130,000 firm-years for our estimation. This provides for almost perfect accuracy to infer rho, even in the absence of knowledge about the target.

> **Proposition:** With the exception of the OLS estimator, knowledge of firms' target leverage ratios does not substantially improve the inference. The Compustat data set is large enough to produce accurate inference even without this knowledge.

The question of what the best SOA estimate is can thus be answered even without good covariates for firms' target leverage ratios. Agonizing over controls for such covariates is not important under the assumed process (in which the shocks do not correlate with a time-changing target).

[18] There is a secondary cause. No known time-series estimator and none of these panel estimators are robust when rho is greater than one. This is due to so-called "super-convergence." The OLS bias that increases as the true rho increases from 0 to 1 actually decreases again after rho increases beyond 1. Both BB and LD have desirable properties in the SDPP only when rho is less than 1.

3.2 Process II: Additive Shocks And Stock-Return Dependent No-Adjustment Ratio

Our second process changes the leverage ratio that comes about when firms do nothing. This no-adjustment change in leverage ratio assumption is now a leverage ratio perturbed by firms' actual stock returns. The standard error of shocks is now lower, too. Everything else stays the same. The process continues to assume that shocks are additive and that observed leverage cannot exceed 0 or 100% (via the third truncation method, as explained on page 16).

[Figure 5 here]

Figure 5 shows that the influence of stock returns worsens the concavity of all estimators. This concavity is so strong that many estimators lose their monotonicity over the graphed domain. A single estimate is then compatible with two different and potentially plausible underlying true rhos. It would be tempting to select the lower value, but these higher true rhos are still reasonable enough that one cannot dismiss them out of hand. Suggesting that (some) firms have true rhos above 1 over a few decades is not absurd. There is even a theoretical basis for this. In the theory of market-timing in Baker and Wurgler (2002), firms issue more equity after their stock prices have gone up, in effect amplifying exogenous stock return shocks.

- The empirical LD estimate of 0.77 is consistent with either a true rho of 0.82 or a true rho of 1.13.

  Comparing the inference from Process I (rho of 0.80) with that from this Process II (rho of 0.82) shows that the change in what the no-readjustment leverage ratio means has caused only a modest 0.02 increase on the (*lower*) LD estimate.

- The empirical FE estimate of 0.68 is consistent with either a true rho of 0.94 or a true rho of 1.16.

  Comparing the bias-adjusted inference from Process I (rho of 0.86) with that from this Process II (rho of 0.94) shows that the change in what the no-readjustment leverage is has caused an economically meaningful 0.08 increase on the (*lower*) estimate.

- The empirical BB estimate of 0.85 suffers from a different but modest problem. No assumption on the true rho can reach a mean BB estimate this high in our simulations. The closest the process simulation can get to matching the empirical value occurs when the true rho is 1.0. Fortunately, if the true rho is at 1.0, then we are not only at the peak of the BB estimate, but we also observe BB estimates that are close to 0.85 with good frequency. Thus, we can adopt 1.0 as our unique BB estimate.

Comparing the bias-adjusted inference from Process I (rho of 0.89) with that from this Process II (rho of 1.00) shows that the change in the no-readjustment leverage assumption has caused an economically meaningful 0.11 increase.

- The W estimator is less concave and therefore is monotonic over the graphed domain. Its empirical estimate of 1.0 translates into a best estimate of the true rho of 1.04.

In sum, two methods, BB and W provide unique matches over the considered domain, and these yield the same true rho estimates of 1.0. Two methods are not determinate. Their lower rho matches are 0.82 (LD) and 0.94 (FE), their upper rho matches are 1.13 (LD) and 1.16 (FE). Even at their lower rho matches, the evidence suggests that earlier papers had rho estimates that were much too low (their SOA estimates were much too high).

[Figure 6 here]

Figure 6 shows that under Process II, too, knowing the target is unimportant, except for correcting the omitted variables problem suffered by OLS. Again, we have enough observations that estimation uncertainty has become irrelevant.

3.3 Processes III and IV: Modeling Debt and Equity Changes

Our third and fourth processes model the shocks to firms' debt and equity itself and only then compute the leverage ratio.

[Figure 7 here]

Figure 7 shows the performance of the estimators in Process III. As in the previous figures, the inference functions of estimators other than W is highly concave and non-monotonic over the considered domain.

- The FE estimator suggests either 0.89 or 1.13 as best true rho estimates.
- The BB estimator suggests either 0.87 or 1.05 (ignoring the unreasonable 0.35 value).[19]
- The LD estimator suggests either 0.86 or 1.07.
- The W estimator is monotonic and suggests an average of 1.04. It is again the only estimator that can be deemed robust with respect to heterogeneity in rho.

The true rho estimates for the different methods are now closer to one another, both for the true rhos below 1 and for the true rhos above 1. In particular, the LD and BB

[19] Note that this estimator has strange behavior when far away from the estimates that fit its empirical value (i.e., when the true rho is less than 0.6). Further investigation showed that this is not due to our particular process, but due to the fact that the variance of the shock is very small when rho is very small. If we assume a similarly small error variance in the additive-shock processes, we get the same strange behavior. When the noise is 0, the BB estimator is undefined.

estimators, with their rather different empirical estimates, now map into more true rho estimates. Thus, reconciling the different estimates should be easier under Process III than under the first two processes.

[Figure 8 here]

Figure 8 shows the performance of the estimators in Process IV. This narrows the estimates, but has little effect otherwise. Still, having seen the estimates under this process has assured us that the best true rho estimates for Process III are not driven just by the desire of the estimators to match the year-to-year variability in leverage ratios.

4 Reconciling the Evidence

We now try to reconcile the estimates that one obtains from different methods. This also allows us to judge the relative quality of the four processes.[20]

4.1 A Single (Homogeneous) Best Rho Estimate

Consider the following thought experiment. Each arbitrary true rho has a unique expected estimate of rho for each of the estimators under a given process. We can measure the degree to which this true rho fits one particular estimator by how far its empirical value is from its expected (i.e., average simulated) value. This can be done using the T-statistic that tests for equality of the simulated average rho and the empirical rho.

To reconcile the estimates, we want to find the true rho estimate (call it μ) that minimizes the differences between the empirical and the expected (i.e., average simulated) rhos. We consider each estimator equally important. Thus, our objective function is the equal-weighted[21] mean squared error T-statistic for the difference between the empirically observed rho estimate and the average simulated rho value.

$$\min_{\mu} \text{ Penalty}(\mu) = \sqrt{\frac{\sum_{\text{E}}[T_{\text{E}}(\mu)]^2}{\text{Number of E's}}}$$

$$\text{where } T(\mu) \equiv \frac{\rho_{\text{sim}}(\mu) - \rho_{\text{empirical}}}{\text{Stderr}(\rho_{\text{sim}}(\mu))}$$

[20]It also alleviates another concern. In Process III, a different rho implied a different year-to-year volatility of leverage ratios. If estimators under Process III just tried to match the volatility and thus produced poor fitting rho estimates, then we should observe a poor fit of Process III's best true rho estimates with one another.

[21]A better way would be to weight the estimators according to their estimated standard errors or slopes at the estimated values. An estimator with a steeper slope and lower standard error seems more accurate than one with a shallower slope.

where $\rho_{\text{sim}}(\mu)$ is the estimate when the rho is mu, the best parameter (to be optimized). The standard error of each estimator is obtained from the simulated (true expected) distribution. The value of the penalty function can loosely be interpreted as an average T-statistic.

Naturally, the best fitting true rho depends on both the underlying processes and the estimators that are supposed to fit their empirical values. Table 6 considers all four processes, and tries to reconcile the FE, LD, and BB estimates. (We are omitting the W estimator, because we want to compare Process I to the other two processes.) Under Process I, the optimization suggests that the single best true rho is about 0.83. However, it fits very poorly. The penalty function—our sort of T-statistic for fit—at the optimum is 12. The smaller table underneath these figures shows that all simulation averages (i.e., expected values) under a true rho of 0.83 would be more than 5 standard deviations from their empirically observed rho estimates. Process II fits better than Process I. Under its best true rho estimate, 0.937, the average penalty is a lower 5.26. Process III fits best. Under its best true rho estimate, 1.066, the penalty is only 3.43. Process IV does not fit as well, but offers similar inference about rho.[22] Note also that the W estimate fits quite nicely under Processes III and IV, even though it was not included in the objective function. **[Table 6 here]**

Under the additional assumption that the penalty function is approximately normally distributed at its optimum, the inverse Hessian is an approximation of the asymptotic standard error. For our purposes, we prefer to view it as a generic sensitivity measure of for our infered estimates, rather than as formal standard errors. Table 6 shows that the rho estimates are fairly accurate.

4.2 Concavity of Inference and Heterogeneity in SOA

Now, recall that all inference functions in Figures 3-8 were not linear, but concave. This means that it is possible that the empirical estimates have an additional downward bias if firms' true rhos are heterogeneous. This is easiest to explain with an example—e.g., via the LD inference function in Figure 5. If all firms had a true rho of 1.0, the function suggests that we would expect to observe an LD estimate of around 0.97. Now contrast this to a case in which half the firms have a true rho coefficient of 0.7 and the other half have a true rho coefficient of 1.3. Therefore the true average rho coefficient of firms is again 1. However, the LD estimator would now likely return a rho estimate of 0.7 for the first half and 0.9 for the second half. The overall-average adjusted LD estimate would therefore be only about 0.8 in the second (heterogeneous) case and not 0.97

[22]Not shown here, Process IV also has a second *local* minimum at a rho of about 0.9. Multiple minima occur because these estimators are non-monotonic. If we add the more monotonic W to the objective function (as we do in Table 8), this lower local minimum disappears.

as in the first (homogeneous) case.[23] This also suggests that the W estimator has an advantage relative to the other methods: its approximate linearity makes it relatively more robust to firm-specific heterogeneity in rho.

Given the concavity of the inference functions and the reasonable prior that some firms may adjust while others may not, it makes sense to allow the true rho to be either homogeneous or heterogeneous. Thus, our reconciliation objective function now generalizes to

$$\min_{\mu,\sigma} \text{Penalty}(\mu,\sigma) = \sqrt{\frac{\sum_{\text{E}}[T_{\text{E}}(\mu,\sigma)]^2}{\text{Number of E's}}}$$

$$\text{where } T(\mu,\sigma) \equiv \frac{\rho_{\text{sim}}(\mu,\sigma) - \rho_{\text{empirical}}}{\text{Stderr}(\rho_{\text{sim}}(\mu,\sigma))}$$

where $\rho_{\text{sim}}(\mu,\sigma)$ is the estimate when the true rho is drawn from a normal distribution with a mean μ and a standard deviation σ, which are now the two parameters to be determined.

[Table 7 here]

Table 7 shows the best true rho mean and rho spread estimates under the four processes if we try to reconcile only the FE, LD, and BB estimates. Allowing for heterogeneity in adjustment speed does not alter the Process I estimates—it remains poor. However, Processes II through IV all suggest substantial heterogeneity across firms: the standard deviation of rho is around 11%. This suggests that a significant fraction of firms may have adjustment speeds that are economically meaningfully positive. (Note also that our heterogeneity is unspecific. With instruments for specific heterogeneity in rho, one could probably improve the estimates.) The three latter processes now offer virtually identical inference. Their best average rhos are between 0.95 and 0.99. The standard errors (from the Hessians) are high enough that we cannot reject the hypothesis that the average adjustment speed is zero. Finally, Process III fits best. Its penalty is a remarkably low 2.07. (This magnitude can also be directly compared to the penalty in Table 6.) This is better than we had hoped for, given the hundreds of thousands of observations and the consequent high accuracy of the estimators.

Neither Table 6 nor Table 7 considered W in their penalty functions. It should be more difficult to reconcile four estimators rather than three estimators, so the penalty function at the optimum is not comparable. It should be higher. We also know that the W estimator fit poorly in Table 7; that it is less susceptible to multiple optimums than the other estimates (e.g., Figure 7); and that it requires the additional assumption

[23] This is not exactly correct, but meant to be illustrative. Our argument is analogous to one that if half of all observations have an OLS slope of 1, and the other half also have an OLS slope of 1, then the slope for all observations should be 1, too. Yet this is really the case only under additional assumptions. In the example, if the observations are ($x_1 = 0, y_1 = 0$) and ($x_2 = 1, y_2 = 1$) in one half of the sample, and ($x_3 = 10, y_3 = -10$) and ($x_4 = 11, x_4 = -9$), the slope with all observations included is not 1, but negative.

that the shocks are stock-return caused. Table 8 shows that Process III remains the best process when we also care about the W estimate. The best fitting estimate of the underlying true rho average is then about 1.06, with fairly low estimated heterogeneity. With so little heterogeneity, and the good fit of W in Table 6, it is not surprising that the estimate for the true rho mean is also similar to that in Table 6. Although this estimate is uncomfortably above 1 (suggesting a negative speed of adjustment), the inverse Hessian (pseudo standard error) of around 0.02 suggests that a speed of adjustment estimate of just about 0 is reasonable. **[Table 8 here]**

In sum, Process III produces the best fit for all estimators. Under this process, the best true rho estimate that reconciles the estimates reported in previous publications is around or just above 1.

5 Alternatives and Robustness

The estimations so far have examined the estimators proposed in the papers referenced above. We now briefly explore some variations of the underlying processes and the estimators.

5.1 Time-Varying Slowly Moving Targets

[Figure 9 here]

Figure 9 assumes a process similar to Process II, but one in which firms' own targets are moving. We assume that the change in the target is normally distributed with mean zero and 5% standard deviation. (The targets are not truncated.)

The figure shows that a randomly changing target makes little difference. After all, a randomly changing target does not alter the inference when the true rho is very high anyway. Given that the empirical estimates earlier suggested a high rho, having a randomly changing target is not important. The only noteworthy finding is that the W estimator performs best. This is also not surprising, because it takes the leverage target to be the firm's most recent own leverage ratio. This can handle time-changing leverage ratios better than the fixed-effects intercepts.

In sum, our inference is robust to underlying targets that move *randomly*. We will discuss on Page 38 the case in which time-varying targets would not be innocuous: if the optimal leverage ratio target moves inversely with stock returns in a manner similar to the functional ratio form of leverage under no-readjustment, then optimal adjustment theories have implications that are practically the same as those of non-adjustment theories. In this case, SOA tests cannot—and may not need to—distinguish between them.

Nevertheless, it is important to emphasize, however, that the W test is very sharp: not reported, even just a little noise added to the IDC measure drastically changes the estimated coefficient. Thus, the fact that the IDC coefficient is around 1 is very informative. The W test suggests not only that the target moves in the same direction as implied by the stock-return (for which one can easily create a tradeoff theory), but it also suggests that this theory would have to predict that the target moves **exactly** in the same amount as implied by a stock return movement.

5.2 Simple Lumpiness

A simple version of fixed adjustment costs can be tested by assuming that observed leverage ratios would not change until actual leverage reaches a distance of at least, say, 5% from the observed leverage. For example, an actual leverage ratio sequence of (15%, 12%, 13%, 8%, 11%, 20%) would became an observed leverage ratio sequence of (15%, 15%, 15%, 8%, 8%, 20%). Recall from Figure 3 that under Process I, we inferred an estimate of 0.71 for the OLS estimator (with a very flat inference function), 0.80 for LD, 0.86 for FD, and 0.89 for BB. The only inference change in our simulations when we impose lumpiness is that the OLS estimate declines from 0.71 to 0.70. All other estimators remain at the same point estimates. *Such lumpiness does not affect the estimates*, and there is also no evidence that leverage ratio changes are lumpy in this simple form. Firms are not only very active in their capital structure policy, but regularly active.[24]

[24]Leary and Roberts (2005, page 2601) define activity as a dummy that represents a leverage ratio change greater than a 5% (or 3% or 7%) cutoff. They state that "perhaps the most striking result is that in 72% of the quarters in our sample no adjustment occurs. That is, a majority of the time firms are inactive with respect to their capital structures." This is however not particularly unusual. If leverage ratio changes are perfectly normally distributed with a standard deviation of 5% per quarter, then we would see no adjustment activity in an even greater fraction (85%) of all quarters. The median time between activity would be 5 quarters, the mean time would be 6.5 quarters, and the max time would be 44 quarters. These are all similar to the numbers that Leary and Roberts report.

Their more important evidence is however about duration. Interestingly, they define optimal adjustment with frictions as a *positive* auto-correlation of changes, not a negative one. Although such a positive auto-correlation seems counterintuitive from the perspective of readjustment, they show that this pattern can be consistent with high fixed costs and convex variable costs under certain parameters. (DTTs are indeed flexible!) We do not examine their evidence further, because it relies on quarterly book-value-based figures, and thus is twice removed from the main variable in our study.

5.3 Extreme Shocks and Asymmetry

In their critique of some of the findings in Welch (2004), Leary and Roberts (2005) look at the subset of firms that experience more extreme shocks. This is a natural place to look for adjustment behavior, even if it is not the average firm that Welch (2004) and our own paper are focusing on. They find that firms respond more to large shock-caused leverage increases (large negative stock returns) than they respond to large shock-caused leverage decreases.

Unfortunately, their test can only be easily done with the W estimator. First, the common dynamic panel estimators are not designed to conditionally ignore firm-years within a panel's timeseries. Second, we need an *a priori* identification for the shock. We do not know in advance whether a firm has a leverage ratio above its target or below its target, and individual firm-specific target estimates are likely to be quite noisy (unlike the single common adjustment parameter estimate). Fortunately, the W estimator proved to be relatively accurate and robust, so having only this one estimator is not a problem.

Figure 4 in Leary and Roberts (2005) suggests a fairly speedy SOA for firms experiencing large shocks (visually roughly a half-life of around 2–3 years). We do not replicate their exact method, but prefer to adapt the standard W method to their hypothesis.[25] Each firm-year, we divide the sample into firms that have had a stock return less than a cutoff C and firms that have had a stock return greater than C. The stock return itself is the same that is used for calculation of IDC. Figure 11 shows the coefficient estimates on IDC as a function of C. (We do not calibrate the estimated rhos to true rhos.) The figure confirms the main directional finding of Leary and Roberts (2005): Firms having experienced strong negative shocks indeed show more readjustment. Firms with –66% rates of return (about the 5th percentile) had average rhos of about 85%, corresponding to half-lives of about 4.3 years—somewhere between active and glacial. Firms with –35% rates of return (about the 20th percentile) had rhos of about 90%, corresponding to half-lives of 6.6 years. Firms having experienced strong positive shocks show some readjustment, too. Firms with +88% rates of return (about the 90th percentile) had rhos of about 95%, corresponding to half-lives of 13.5 years—an

[25] The Leary and Roberts (2005) paper uses ex-post data rather liberally. Their sample is limited to large firms that have experienced a 1-sd shock. First, they define the return shocks themselves based on the ex-post rates of return distribution. That is, they compute each firm's rate of return over the entire sample, and then select firms which had a rate of return above or below one standard deviation. Second, they require a firm to have *five* years of data *after* the shock has occurred in order to enter their graphs. This leaves fewer than 25% of the sample. In contrast, our selection is ex-ante. We also experimented with multi-period variations of our own method. The inference looks quite similar.

economically glacial readjustment speed.[26] In sum, our SOA estimates are considerably lower, but qualitatively similar to their's.

Our next question is how much of this (slower) readjustment rate could be due to survivorship bias. A firm that experiences a –66% rate of return and fails to unlever could be more likely to go bankrupt or delist, exit Compustat, not report an end-of-year leverage ratio, and thus not enter the regression. The dashed red line shows that about one third of firms that have had a –66% rate of return during their fiscal year (from CRSP) disappear before they report an end-of-year leverage ratio. The lower the rate of return, the more likely the firm is to disappear. This suggests that disappearance is not benign.

One way to assess the potential significance of this bias is to assume that firms with a negative rate of return during a fiscal year in which they disappeared went "bankrupt" with a leverage ratio of 100% at the end of the year. *This is an assumption.* Within this year of disappearance, we compute the stock return from CRSP including the delisting rate of return. In the figure, the thinner red line includes these bankrupt observations. This line shows that IDC coefficient is 5% to 10% higher than that of the survivors only. Firms that have experienced a rate of return of –52% (about the 10th percentile) now have half-lives of 13.5 years and not 6.2 years.

In sum, the data suggests that the market disproportionally weeds out firms that do not lower their leverage ratios after extreme negative shocks. Although we can confirm the Leary and Roberts (2005) finding that firms that experience a strong negative shock readjust towards a less levered capital structure *if they remain publicly traded*, the average such firm may have adjusted much less. It may thus have been the market and not the manager that "readjusted" publicly traded firms' leverage ratio. In any case, the fraction of the sample that engages in *active* and quick readjustment remains very small.

5.4 5-Year Leverage Ratio Changes

Finally, we examine estimators that examine multi-year leverage ratio changes. This can also help inference if the adjustment processes have higher-order AR terms. Figure 10 repeats the estimations with overlapping 5-year intervals under Process II. Beyond a 5-year rho of 1, most econometric estimators decline. The fixed effects estimator is now a weak test for true rhos below about 0.8, showing almost no slope. Interpreting an empirical value of 0.0252 as evidence of a true 5-year rho of 0.88 is uncomforting, even if it corresponds in our simulations to a best *1-year* rho estimate of 0.88 (or 1.27).

[26] The figure also shows that our main findings are robust when we eliminate firm-years with extreme stock-returns. For example, when the stock return is greater than –50%, the blue line estimate remains above 1. When the stock return is less than 100%, the red line remains just a smidgen below 1.

The BB, W, and LD estimators have better slopes. Their empirical estimates are quite low, but their translations into true rhos suggests a different picture. Even their lower estimates suggest one-year adjustment coefficients (rho) above 90%, 96%, and 99%, respectively.

Not reported, we also considered some longer-memory processes driving the 1-year leverage ratios. Our reported conclusions are robust with respect to some reasonable alternative processes. For example, we increased the AR(1) parameter by 5% and introduced an AR(2) parameter of –7%. The half-life of shocks under the AR(2) process is similar to that of the AR(1) process.[27] We then estimated the BB and LD methods on these processes. There was no *differential* impact of the omitted second-order lag term on the biases of the BB and LD estimators. Thus, a longer-term memory process is not likely to be capable of explaining the discrepancy in estimates.

6 Interpretation and A Critical View of Dynamic Tradeoff Theory

Replication and reconciliation of the empirical evidence from earlier papers suggests that firms *on average* do not readjust—or are glacially slow to readjust—their capital structures. This is not inconsistent with the view that *some* firms do adjust. Some of our reconciliations even delivered estimates for the degree of heterogeneity across firms. Specific heterogeneity has also been specifically identified in earlier papers. For example, Kisgen (2006) finds that firms near a change in credit ratings issue less debt. (Of course, leverage ratio increases and debt issuances are not one-to-one.) Roberts and Sufi (2009) find that firms issue less debt when they violate a covenant. Leary and Roberts (2005) find that firms that experience a dramatic leverage increase in one year due to large negative stock returns are a little more inclined to readjust their leverage ratios—even though our point estimates suggest *much* slower reversion than their's (and especially when we take survival bias into account). In sum, the evidence is not that no firm adjusts, but that firms on average do not actively readjust.

[27]The BB and LD estimators are not designed for or easily adaptable to higher-order AR processes. Fortunately, extending the FE estimator to include two lags is easy. A sample run suggested that the AR(1) term was about 5% too low (rising from 0.68 to 0.73), and the AR(2) terms was negative (–7%) and statistically significant.

6.1 Dynamic Tradeoff Theories and Speed of Adjustment Evidence

The SOA evidence that our paper reexamined is often viewed as having the potential to confirm or reject dynamic tradeoff theories (DTTs). It is correct that our evidence can reject a subset of these theories, in which frictions are small, the benefits to an optimal leverage ratio are large, and managers optimize firm value. However, it is not correct that our evidence of extremely slow adjustment rejects these theories. In fact, we cannot think of *any* SOA evidence that could by itself reject DTTs.

This is because DTTs are too flexible. The empirical tests to-date have not identified the *net* benefit of capital structure activity *a priori*. Instead, they have in effect calibrated the net benefit of capital structure change as free parameters. This has made the DTT so flexible that it can explain speeds of adjustment from practically instant to practically never. Consequently, DTT theories should be viewed as largely irrelevant to interpreting SOA evidence. It is only in reverse—that the SOA evidence can impose certain restrictions on the wide set of possible DTTs—that the empirical evidence is useful.

This critique also suggests that *if* one could obtain good *a priori* identification of the *net* benefit of capital structure change—instead of allowing it to be a free parameter that is calibrated to/inferred from capital structure change behavior itself—then the DTT theories could become more meaningful to interpreting SOA evidence. Some papers have argued that investment banking fees could proxy for the cost. This is a step in the right direction, but these fees may also cover services beyond just the issue itself and they may understate the cost to the firm itself. And most importantly, these papers did not specify the net benefit *a priori*.

Some papers have adopted the DTT as their NULL hypothesis and declared victory after showing that the SOA evidence can be consistent with their DTT. Although this is reasonable, we would caution against it: The empirical evidence in our paper suggests that a DTT works only when it explains the same SOA as one from a theory in which "managers are random action generators." It is not clear that the DTTs deserve NULL hypothesis status relative to such “random action” hypotheses.

In sum, it is difficult to think of feasible empirical tests that can broadly reject the DTT. Of course, one can write down specific functional forms of the theory that are rejectable. Such models essentially identify the relations through the structure following from their functional assumptions. Unfortunately, they become very quickly very complex and very specific. In the end, rejecting them may just be rejecting their functional form identifications. Other DTT parameterizations with more free parameters can then fairly quickly step into the breach. Absent quantitative *a priori* identification of the costs and benefits, the challenge for DTTs is thus the question of *which broad stylized facts that can feasibly be empirically established would ultimately be inconsistent with the theory?* Otherwise, the DTT is not a theory, but just a conceptual framework (i.e., an optimization paradigm)—not testable, but assumed—that suggests how firms *should* react.

6.2 A Potential Evidence Problem For Friction-Based DTTs

Despite the generality of the DTT, simple versions of the theory seem to fail if we broaden the empirical dimensions on which we measure its success. We should judge a DTT-based explanation not just by its ability to explain slow SOA, but also by its ability to explain the variability of capital structure activity.[28] This deserves elaboration.

A number of papers have shown that the DTT fits the data, because under certain costs of issuing and repurchasing debt and equity, it can explain a slow speed of adjustment. Intuitively, the role of these frictions is to make firms reluctant to change their capital structures. Yet the data suggests that firms are actively changing their capital structures and that this is *not* rare.

Of the 12.5% standard deviation in annual leverage ratio changes, only about 3.8% is not caused by corporate activity. The remaining 8.7% year-to-year variability in leverage ratio changes ($\mathsf{DC}_{i,t} - \mathsf{IDC}_{i,t-1,t}$) is due to active issuing and repurchasing. Although occasional large debt or equity issues are a good part of this 8.7%, many firms that do not readjust are almost constantly active. Managerial activity is zero in only about 7% of firm-years. It is small (less than 1% in absolute terms [0.1 standard deviations in firm-normalized terms]) in about one-third of all firm-years; modest (between 1% and 5%) in another one-third of all firm-years; and large (greater than 5%) in the final third of all firm years. A capital structure change of 5% is economically meaningful by any definition, and it happens on average every three years.[29] Such evidence of active corporate intervention is not new: Fama and French (2002) show that certain common annual equity issuing activity is quite sizeable, and Welch (2004, Table 4) suggests that year-to-year capital structure changes are disproportionally due to firms' debt refinancing activities.

The challenge is to understand this economically large variability in capital structure activity and slow SOAs at the same time—and why the capital structure activity is largely orthogonal to capital structure shocks. Why do managers not use the opportunity to readjust if they are changing their capital structure anyway?

[28] The pecking order may also suggest more passivity than we see. Thus, our paper focuses on the related question of speed of adjustment only, rather than on the broader test of this particular theory.

[29] Another way to see whether managerial activity is limited to rare and large issuing activity is to remove the influence of large changes. If we winsorize active leverage ratio changes at –25% and +25%, the year-to-year variation in active claims management drops from 8.7% only to 7.1% per year. If we winsorize at –10% and +10,% the variation remains at 5% per year.

In sum, our view is not that frictions are unimportant in the real world, or that the DTT is wrong because it cannot explain the SOA empirical evidence, or that it is not possible to build a DTT with more parameters that can explain the evidence. Instead, we caution only that such a DTT may look quite different from current manifestations. The high frictions employed to explain slow SOAs do not seem compatible with the empirical evidence of *active leverage ratio changes*. If anything, in order to explain the evidence of frequent active leverage ratio changes in a DTT context, lower frictions may end up making it easier to fit the data than higher frictions.

From the perspective of the DTT—to explain both slow SOAs and much capital structure activity—what is needed is an as-of-yet unknown factor that [a] greatly moves target ratios from year to year and sufficiently so to induce firms to actually change them frequently; and [b] on average tends to move the target ratio about one-to-one with stock returns. As for [a], if the factor can be empirically measured, it would become the key to a better understanding of capital structure. As for [b], it could be that stock returns mostly reflect growth options that (must) disappear in financial distress, so that when a firm experiences a positive stock return, its optimal debt ratio simultaneously becomes lower. (Of course, this intuition is opposite to the more common intuition in other DTTs, e.g., as in Chen (2008).)

Importantly, it is not enough for such a DTT theory to predict that optimal targets move in the same direction as those induced by stock returns. The optimality theory must literally explain almost precisely the same one-to-one movement as the no-readjustment theory. The reason is the sensitivity of the Welch (2004) test. Table 1 showed that IDC has an annual standard deviation of about 25%. Now add a normally distributed noise term with standard deviation s to the target (IDC) as a standin for the prediction of an optimality theory, and reestimate the Welch regression. Starting with the benchmark case in Table 3, the estimated IDC regression coefficients change as follows

s(noise)	IDR	ADR	constant
0.00	1.003	−0.075	0.029
0.01	0.990	−0.062	0.029
0.02	0.952	−0.025	0.030
0.03	0.894	0.0310	0.030

s(noise)	IDR	ADR	constant
0.04	0.823	0.100	0.031
0.05	0.744	0.175	0.032
0.10	0.421	0.489	0.035
0.20	0.154	0.748	0.038

(Recall that the IDC estimation standard errors are negligable.) This shows that if a DTT theory predicts an optimized target that is just 3-5% off from the "pseudo-target" implied by stock-return changes, this test will not provide estimates that are as high as those observed in the data.

Of course, a successful "active optimization DTT" that suggests such similar behavior as the "no-readjustment" theory would itself raise not only the question of how one would distinguish the two, but also raise the question of why it was useful above and beyond the more simple theory.

Similar to looking for a novel factor that drives changes in the optimal capital structure, one could define alternative theories which have specific predictions linked to variables that are not claimed by common tradeoff theories. Good potential candidate would be theories in which managerial identity (e.g., as in Bertrand and Schoar (2003)) or uncontrolled managerial preferences matter. (For example, ceteris paribus, many managers may like an equity-heavy capital structure more than a debt-heavy capital structure, even if the latter maximizes firm value.)

7 Conclusion

Our paper has pointed out that discrepancies in SOA estimates in the literature, given the almost asymptotic sample, point to violations of these papers' underlying process assumptions. It has identified the relative unimportance of knowing the firm-specific leverage ratio target (*provided* that it does not covary strongly with stock returns). It has identified the importance of the fact that leverage ratios have a limited domain. It has pointed out the potential importance of concavity in the inference mapping and of firm-specific heterogeneity in SOAs. It has quantified how important it is to specify what the leverage ratio under the hypothesis of no readjustment is presumed to be—whether it requires firms to issue and repurchase debt and equity in specific proportions or whether it allows passivity.

Methodologically, our paper has introduced and compared different process assumptions for leverage ratio changes, and it has shown how one should reconcile different estimators and processes.

Our paper has found that a novel process can describe the data better than those used to derive existing estimators. In this process, one models joint debt and equity processes, rather than a process on leverage ratios themselves. In fact, this dual variable process seems good enough that it can (almost) reconcile the different estimates provided by the methods in the literature—a very challenging task. To reconcile the FE, LD, and BB estimators, our best estimate of rho has a mean of about 96% (an average half-life of greater than 15 years), with about 10% heterogeneity. To reconcile the FE, LD, BB, and W estimators requires a negative speed of adjustment, as suggested by the theory of Baker and Wurgler (2002). We could confirm the Leary and Roberts (2005) finding that firms experiencing extreme negative shocks (say, –75% or worse) do lower their leverage ratios—but the magnitude is lower than they suggest. Much of this readjustment can simply be due to the fact that firms that do not lower their leverage ratios after a large negative stock return disappear from the sample (probably because they went bankrupt). Randomly moving capital structure targets and lumpiness in capital structure adjustments do not create inference problems.

The natural next steps will be more difficult. First, one could model heterogeneity in SOAs better. In an IV approach, extreme stock returns, other variables in particular years, or firm characteristics could help identify rho. This would require the development of novel econometric techniques. Second, one could try to derive the exact algebraic conditions under which the estimators diverge. This would aid in understanding the class of true processes that can potentially reconcile the observed differences in process estimates better than we have. (This is not an easy task. Even the papers in the econometrics literature itself struggle with understanding the differences in estimators—and this is under the assumption of normally distributed errors and only one simple AR(1) term!) Third, we could benefit from a better understanding of exactly what we should consider as "shocks." It would be especially helpful if we could identify specifically what kind of random shocks one would expect under the no-readjustment hypothesis for firms with very low leverage ratios. About one in five such firms departs the zero-leverage state in any given year—should this be viewed as adjustment (mean reversion) or can this be viewed as random? Fourth, we could improve on the statistical models of the evolution of debt and equity. For example, firms may choose to smooth short-term needs with debt and long-term needs with equity.

But the most interesting empirical regularity that begs for an explanation is also the simplest: what theory and factors consistent with almost no *readjustment* can also explain the dramatic year-to-year *changes* in firms' leverage ratios?

References

Antoniou, A., Y. Guney, and K. Paudyal, 2008, "The Determinants of Capital Structure: Capital Market Oriented versus Bank Oriented Institutions," Journal of Financial and Quantitative Analysis, 43, 59–92.

Baker, M., and J. Wurgler, 2002, "Market Timing and Capital Structure," The Journal of Finance, 57, 1–32.

Bertrand, M., and A. Schoar, 2003, "Managing With Style: The Effect of Managers on Firm Policies," Quarterly Journal of Economics, 118, 1169–1208.

Blundell, R., and S. Bond, 1998, "Initial conditions and moment restrictions in dynamic panel data models.," Journal of Econometrics, 87, 115–143.

Chang, X., and S. Dasgupto, 2009, "Target Behavior and Financing: How Conclusive is the Evidence," The Journal of Finance, 64, forthcoming.

Chen, H., 2008, "Macroeconomic Conditions and the Puzzles of Credit Spreads and Capital Structure," Discussion paper, MIT.

Fama, E. F., and K. R. French, 2002, "Testing Trade-Off and Pecking Order Predictions about Dividends and Debt," The Review of Financial Studies, 15, 1–33.

Flannery, M. J., and K. P. Rangan, 2006, "Partial adjustment toward target capital structures," Journal of Financial Economics, 79, 469–506.

Frank, M. Z., and V. K. Goyal, 2008, "Trade-off and Pecking Order Theories of Debt," in B. Espen Eckbo (ed.), Handbook of Corporate Finance: Empirical Corporate Finance, vol. 2, . chap. 12, Elsevier.

Hahn, J., J. Hausman, and G. Kuersteiner, 2007, "Long Difference Instrumental Variables Estimation for Dynamic Panel Models with Fixed Effects.," Journal of Econometrics, 140, 574–617.

Huang, R., and J. R. Ritter, 2009, "Testing Theories of Capital Structure and Estimating the Speed of Adjustment," .

Kayhan, A., and S. Titman, 2007, "Firms' Histories and Their Capital Structure," Journal of Financial Economics, 83, 1–32.

Kisgen, D. J., 2006, "Credit Ratings and Capital Structure," The Journal of Finance, 61, 1035–1072.

Leary, M. T., and M. R. Roberts, 2005, "Do Firms Rebalance Their Capital Structures?," The Journal of Finance, 60, 2575–2619.

Lemmon, M. L., M. R. Roberts, and J. F. Zender, 2008, "Back to the Beginning: Persistence and the Cross-Section of Corporate Capital Structure," .

Myers, S. C., 1984, "The Capital Structure Puzzle," The Journal of Finance, 39, 575–592.

Nickell, S., 1981, "Biases in Dynamic Models with Fixed Effects," Econometrica, 49, 1417–1426.

Roberts, M. R., and A. Sufi, 2009, "Control Rights and Capital Structure: An Empirical Investigation," The Journal of Finance, p. forthcoming.

Shyam-Sunder, L., and S. C. Myers, 1999, "Testing Static Tradeoff Against Pecking Order Models of Capital Structure," Journal of Financial Economics, 51, 219–244.

Welch, I., 2004, "Capital Structure and Stock Returns," Journal of Political Economy, 112, 106–131.

Table 1: Measures of Leverage Ratios and Sample Characteristics

		Levels			Changes		
Variable	Explanation	Mean(i,t)	Sdv(i,t)	$\overline{\text{Sdv(t)}}$	Mean(i,t)	Sdv(i,t)	$\overline{\text{Sdv(t)}}$
DC	Financial Debt/Financial Capital, Market Value $\frac{\text{DLTT+DLC}}{\text{DLTT+DLC+CSHO} \cdot \text{PRCC\_F}}$	27%	25%	23%	1.2%	12.6%	11.4%
LA	Liabilities/Assets, Market Value $\frac{\text{LT+PSTKL–TXDITC–DCVT}}{\text{CSHO} \cdot \text{PRCC\_F+LT+PSTKL–TXDITC–DCVT}}$	40%	25%	24%	1.5%	12.4%	11.2%
dc	Financial Debt/Financial Capital, Book Value $\frac{\text{DLTT+DLC}}{\text{AT}}$	24%	20%	19%	0.6%	9.7%	9.0%
la	Liabilities/Assets, Book Value $\frac{\text{LT+PSTKL–TXDITC–DCVT}}{\text{AT}}$	45%	21%	20%	0.8%	10.2%	9.5%
IDC	Stock-Return Implied DC (and non-stock return changes) $\frac{\text{DLTT+DLC}}{\text{DLTT+DLC+(CSHO} \cdot \text{PRCC\_F)} \cdot (1 + r_{i,t})}$	26%	25%	23%	$(\text{IDC}_{t-1,t} - \text{DC}_{t-1}) =$ –1%	8.7%	8.1%

(continued)

Table 1: Measures of Leverage Ratios and Sample Characteristics (continued)

where Compustat defines variables as

Variable	Explanation	Mean(i,t)	Sdv(i,t)	$\overline{\text{Sdv(t)}}$
DLTT	Long-Term Debt, Total	\$352	\$2,629	\$1,655
DLC	Debt in Current Liabilities, Total	\$122	\$1,847	\$1,122
CSHO	Common Shares Outstanding	\$50	\$244	\$136
PRCC_F	Price Annual Close, Fiscal Year End	\$25	\$748	\$357
LT	Liabilities, Total	\$1,011	\$7,582	\$4,736
PSTKL	Preferred Stock, Liquidating Value	\$13	\$329	\$159
TXDITC	Deferred Taxes and Investment Tax Credit	\$72	\$591	\$406
DCVT	Debt, Convertible	\$19	\$208	\$128
AT	Assets, Total	\$1,632	\$10,239	\$6,677

and

	Mean	Sdv	Min	Median	Max
Number of Firms Per Year	3,299	1,436	453	3,307	5,804
Number of Years Per Firm	19	12	1	16	45

Explanation: In addition to the data from Compustat, we use the rate of return $r_{i,t}$ without dividends from CRSP. We confirmed that using the CRSP market value of equity instead of CSHO·PRCC changes empirical alpha estimates by no more than by 1%. Mean and Sdv are pooled statistics. $\overline{\text{Sdv(t)}}$ is the average cross-sectional standard deviation per year. The observation count includes firm-years from 1963 to 2007 with non-missing DC, excluding financial firms and utilities (SIC codes 6000–6999 and 4900–4949), and prior year total assets of more than \$10 million. This leaves a total of 14,615 firms with 148,464 firm years. The change in IDC is an exception, in that it subtracts lagged DC. (It is the non-stock return caused change in capital structure.)

Table 2: Estimation Procedures

Procedure	Explanation
OLS, Auto (OLS) Lemmon, Roberts, and Zender (2008), Fama and French (2002), etc.	$\mathsf{DC}_{i,t} = c + \rho \cdot \mathsf{DC}_{i,t-1} \;\; [+(1-\rho)\cdot\gamma\cdot \mathsf{T}_{i,t-1}] + \epsilon_{i,t}$ (Note that estimations in which the dependent variable is measured in differences come to virtually identical inferences.)
Firm Fixed Effects (FE) Flannery and Rangan (2006)	$\mathsf{DC}_{i,t} = c_i + (1-\rho)\cdot \mathsf{DC}_{i,t-1} \;\; [+\rho\cdot\gamma\cdot \mathsf{T}_{i,t-1}] + \eta_i + \epsilon_{i,t}$ Similar to "OLS, Auto," but each firm receives its own intercept.
Blundell Bond GMM (BB) Lemmon, Roberts, and Zender (2008)	Stata: **xtabond2 dc L.dc [T], gmm(L.(dc), lag(1 3)) [iv(T)] robust** System GMM estimation (Blundell and Bond (1998)), for estimating dynamic panel data models in the presence of firm fixed effects. Includes variable levels, as well as differences, in the instrument set to address the problem of persistent regressors. (Uses book values and exact GMM stata specification, as noted.)
Long Difference Estimator (LD) Huang and Ritter (2009)	$\mathsf{LA}_{i,t} - \mathsf{LA}_{i,t-8} = \rho\cdot(\mathsf{LA}_{i,t-1} - \mathsf{LA}_{i,t-9}) \;\; [+\delta\cdot(\mathsf{T}_{i,t-1} - \mathsf{T}_{i,t-9})] + \epsilon_{i,t} - \epsilon_{i,t-8}$ Uses iterative two-stage least squares instruments ($\mathsf{LA}_{i,t-1} - \mathsf{LA}_{i,t-9}$) with $\mathsf{LA}_{i,t-9}$. Obtains initial values for $\tilde{\rho}$ and $\tilde{\delta}$. Uses the residuals $\mathsf{LA}_{i,t-1} - \tilde{\rho}\cdot\mathsf{LA}_{i,t-2} - \tilde{\delta}\cdot\mathsf{T}_{i,t-2}, \ldots, \mathsf{LA}_{i,t-7} - \tilde{\rho}\cdot\mathsf{LA}_{i,t-8} - \tilde{\delta}\cdot\mathsf{T}_{i,t-8}$ as additional instruments. (HR perform 3 iterations overall.) This estimator was suggested by Hahn, Hausman, and Kuersteiner (2007).
Implied Target (W) Welch (2004)	$\mathsf{DC}_{i,t} = c + \rho\cdot\mathsf{IDC}_{i,t} + (1-\rho)\cdot\mathsf{DC}_{i,t-1} \;\; [+\gamma\cdot\mathsf{T}_{i,t-1}] + \epsilon_{i,t}$ where $\mathsf{IDC}_{i,t}$ is the implied financial debt to capital ratio (leverage), which is the past leverage adjusted for the change in the stock price. One can also extend this method to use fixed effects, $\mathsf{DC}_{i,t} = c_i + \rho\cdot\mathsf{IDC}_{i,t} + (1-\rho)\cdot\mathsf{DC}_{i,t-1} \;\; [+\gamma\cdot\mathsf{T}_{i,t-1}] + \epsilon_{i,t}$

Explanation: Variables are described in Table 1. ($\mathsf{DC}_{i,t}$ is financial debt to capital ratio, our main measure for leverage.) $\mathsf{T}_{i,t-1}$ is a proxy for the firm target, usually instrumented by observable variables. $\epsilon_{i,t}$ is a random disturbance term. ρ is the autocoefficient, our main parameter of interest, which is also 1- SOA.

Table 3: Estimates With Full Sample

Measure ↓	Method →	**OLS**	**FE**	**BB**	**LD**	**W**	**WFE**
Market DC	Lagged Leverage (DC)	**0.895**	**0.681**	**0.847**	**0.772**	–0.075	–0.191
		(0.002)	(0.004)	(0.005)	(0.004)	(0.004)	(0.004)
	Implied Debt Ratio (IDC)					**1.003**	**0.971**
						(0.004)	(0.004)
	Constant	0.040	many	0.053		0.029	many
		(0.001)		(0.001)			
	Observations	132,412	132,412	132,412	55,967	128,943	128,943
	R-Squared	0.76	0.81	.	.	0.88	0.90
Market LA	Lagged Leverage (LA)	**0.886**	**0.660**	**0.828**	**0.779**	–0.026	–0.122
		(0.002)	(0.004)	(0.005)	(0.004)	(0.003)	(0.003)
	Implied Debt Ratio (ILA)					**0.969**	**0.942**
						(0.003)	(0.003)
	Constant	0.147	0.147	0.082		0.035	0.082
		(0.002)	(0.002)	(0.002)		(0.001)	(0.001)
	Observations	120,827	120,827	120,827	44,783	117,626	117,626
	R-Squared	0.77	0.82	.	.	0.92	0.93
Book *dc*	Lagged Leverage (*dc*)	**0.893**	**0.677**	**0.829**	**0.772**		
		(0.002)	(0.005)	(0.008)	(0.004)		
	Constant	0.032	many	0.047			
		(0.001)		(0.002)			
	Observations	136,450	136,450	136,450	58,345		
	R-Squared	0.77	0.81	.	.		
Book *la*	Lagged Leverage (*la*)	**0.888**	**0.683**	**0.864**	**0.789**		
		(0.002)	(0.005)	(0.009)	(0.004)		
	Constant	0.0582	many	0.0688			
		(0.0009)		(0.0040)			
	Observations	119,919	119,919	119,919	43,694		
	R-Squared	0.77	0.82	.	.		

Explanation: The variables are defined in Table 1. The estimation techniques are defined in Table 2. The dependent variable is always the same as the first independent variable (lagged) (except in the LD column, where it is a change.) The table reports results for two measures of market leverage and two measures of book leverage. Standard errors are reported in brackets, and adjusted to be robust to heteroscedasticity and clustered within firms.

Interpretation: The estimates seem uncomfortably far from one another.

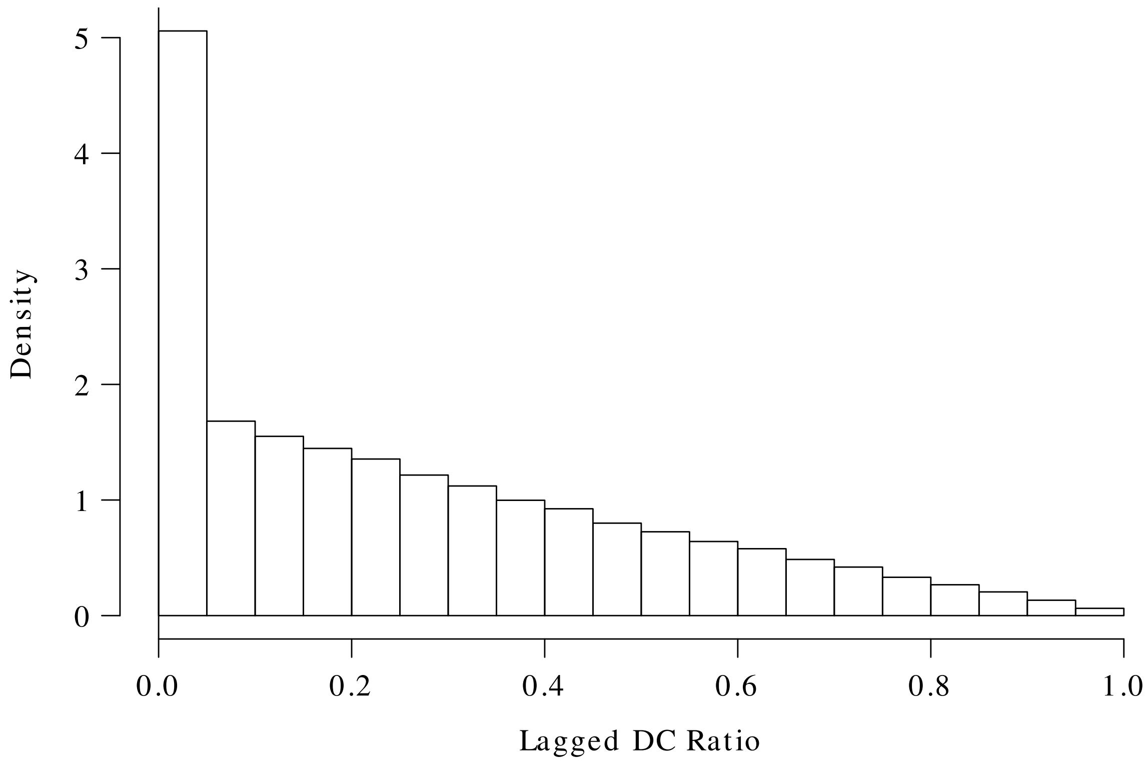
Table 4: Estimates With Constant Sample (40,708 Firm-Years)

Measure ↓	Method →	OLS	FE	BB	LD	W
Market DC	Lagged Leverage (DC)	**0.894**	**0.665**	**0.857**	**0.768**	-0.081
		(0.0027)	(0.0066)	(0.0070)	(0.0039)	(0.0075)
	Implied Debt Ratio (IDC)					**1.015**
						(0.0076)
	Constant	0.027	many	0.037		0.025
		(0.0008)		(0.0019)		(0.0006)
	R-Squared	0.79	0.83	.	.	0.88
Market LA	Lagged Leverage (LA)	**0.902**	**0.673**	**0.851**	**0.775**	-0.020
		(0.0023)	(0.0063)	(0.0072)	(0.0035)	(0.0052)
	Implied Debt Ratio (IDC)					**0.969**
						(0.0050)
	Constant	0.129	many	0.059		0.032
		(0.0025)		(0.0029)		(0.0007)
	R-Squared	0.80	0.84	.	.	0.92
Book *dc*	Lagged Leverage (*dc*)	**0.903**	**0.677**	**0.850**	**0.771**	
		(0.0029)	(0.0077)	(0.0095)	(0.0043)	
	Constant	0.024	many	0.035		
		(0.0007)		(0.0022)		
	R-Squared	0.79	0.83	.	.	
Book *la*	Lagged Leverage (*la*)	**0.921**	**0.730**	**0.908**	**0.787**	
		(0.0026)	(0.0072)	(0.0097)	(0.0040)	
	Constant	0.041	many	0.046		
		(0.0012)		(0.0044)		
	R-Squared	0.81	0.85	.	.	

Explanation: This table repeats Table 3 but restricts the sample to observations that are used in all methods. The principal reason why the number of firm-years is so low is that the LD estimator requires long sample time-series.

Interpretation: The estimates remain uncomfortably far from one another. Sample differences are not the reason.

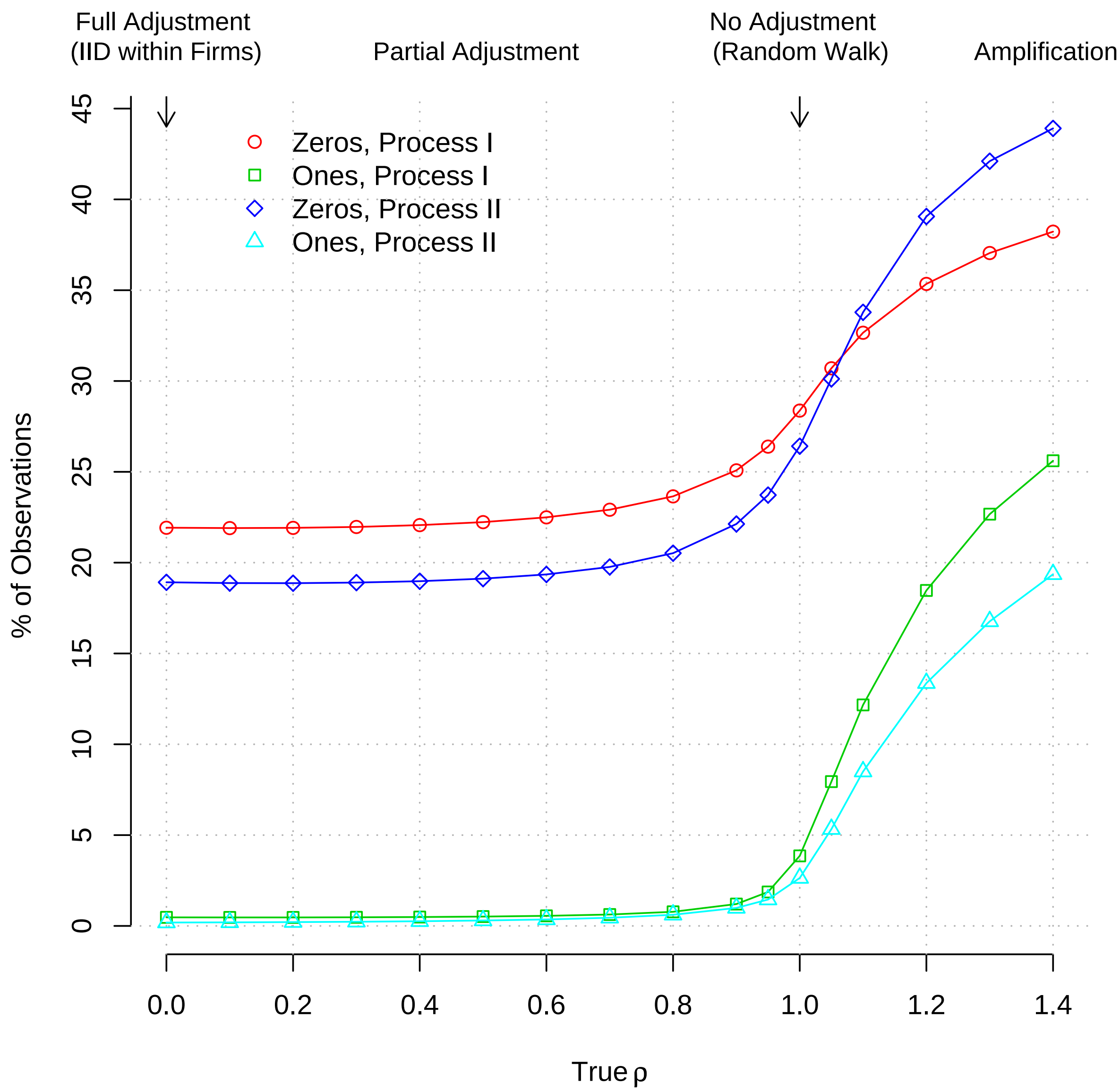
Figure 1: Histogram of Debt-to-Capital Ratios (DC)



Explanation: This figure plots a histogram of (lagged) leverage ratios. This is typically an independent variable in our regressions predicting its own future value. The figure shows that the distribution is skewed to the right, and there is a large number of firms with leverage ratios less than 10%. With a mean of 27% and a standard deviation of 25%, the truncation is about one standard deviation from the mean. Moreover, not shown in this figure, the year-to-year standard deviation of changes in DC in the sample is 12.5%.

Interpretation: With an additive error process, firms are likely to run especially into the lower domain limit quite frequently.

Figure 2: Percentage of Leverage Ratios Below 0% and Above 100%



Explanation: This figure shows the percentage of firm-years in which an unconstrained AR(1) process produce infeasible leverage ratios. The simulated true leverage processes are Process I, i.e., $\mathrm{DC}_{i,t+1} = \rho \cdot \mathrm{DC}_{i,t} + t_i + u_{i,t+1}$ or Process II, i.e., $\mathrm{DC}_{i,t+1} = \rho \cdot \mathrm{IDC}_{i,t,t+1} + t_i + u_{i,t+1}$.

Interpretation: When rho is high, firms would run into the domain limit imposed by additive error processes much more frequently.

Table 5: Summary of Simulation Protocol

- The dimension of the firm-year matrix is the same in the simulations as in the actual data. That is, the sample is equivalent in dimension to that in Table 3. In the typical run, this means that we have 132,412 firm-years with at least one lagged leverage ratio. A few firm-years do not have the necessary stock return data to simulate the process in the additive and feasible shocks process.
- Each firm is assumed to start out with its actually observed first target leverage ratio.
- The target leverage ratio is presumed to be one random draw away from the firm's actual first leverage ratio. (Thus, targets are heterogeneous.)
- Each simulation is a set of draws from the error distribution, one for each firm-year. Each error draw (simulation) is combined with the full range of underlying rho parameters that we plot. (Each method is then estimated on these data.)
- In the additive shock specifications, we assume that its true leverage ratio is unobserved and only the (possibly truncated at the feasible border) leverage ratio is observed. We continue adding shocks to the true leverage ratio.
- Simulated Shocks (Errors):

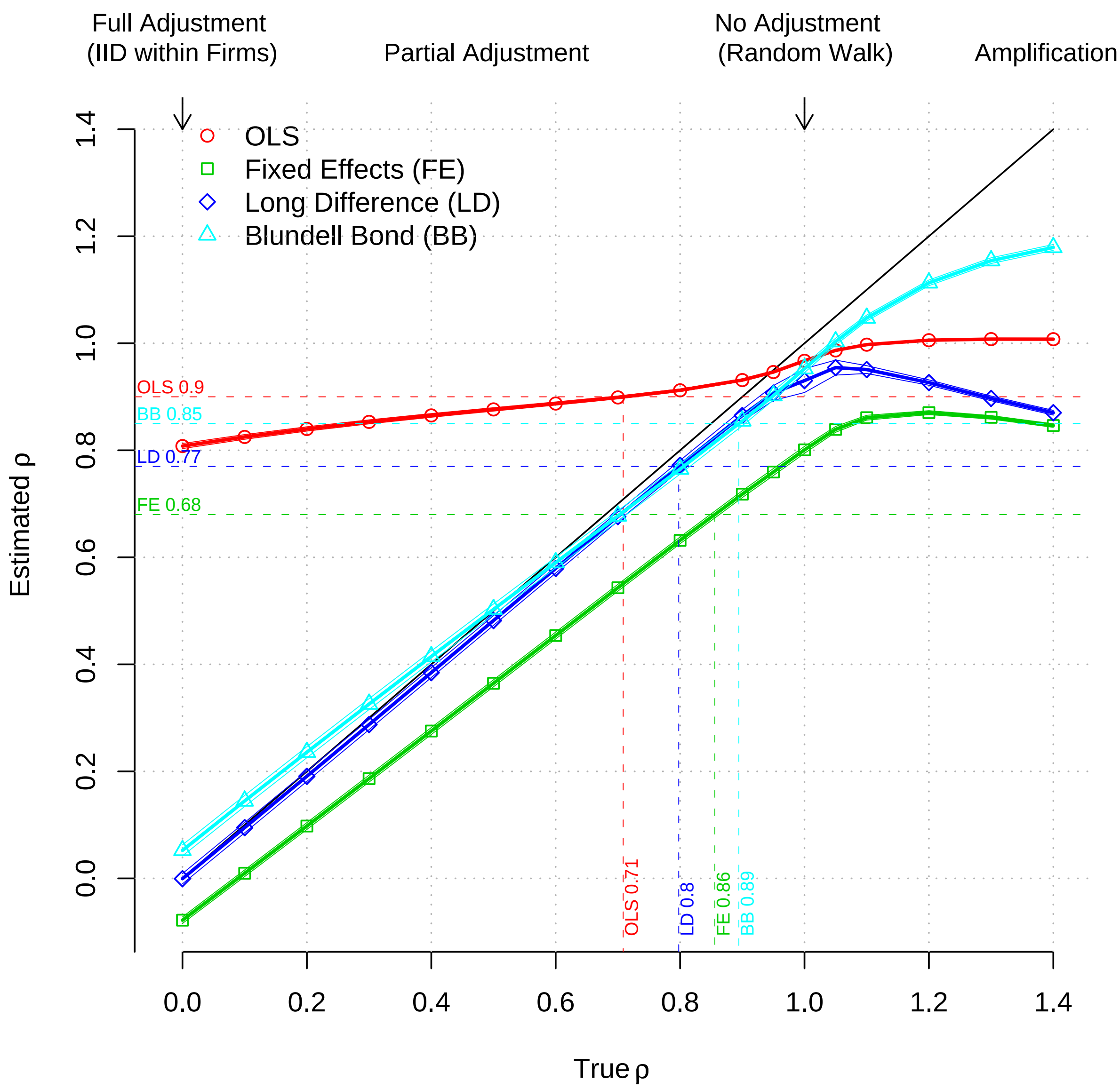
Additive Shock: In the additive shock specification, $\widetilde{\mathsf{DC}}_{i,t+1} = (1-\rho)\cdot \mathsf{T}_i + \rho \cdot \mathsf{DC}_{i,t} + \tilde{u}_{i,t+1}$, errors are drawn from a normal distribution with a standard deviation of 12.5%.[30]

Additive Shocks Plus Stock Return Shocks: In the additive shock specification with stock-return induced changes in equity values, $\widetilde{\mathsf{DC}}_{i,t+1} = (1-\rho)\cdot \mathsf{T}_i + \rho \cdot \mathsf{IDC}_{i,t,t+1} + \tilde{u}_{i,t+1}$, where $\mathsf{IDC}_{i,t,t+1} = \mathsf{DC}_{i,t}/[1 + r_{i,t,t+1}\cdot(1-\mathsf{DC}_{i,t})]$, where r is without dividends. Errors are drawn from a normal distribution with a standard deviation of 8.7%. Stock returns are as observed in the data, and also change the equity market-capitalization.

Feasible Shocks: In the feasible shock specification, $\widetilde{\mathsf{DC}}_{i,t+1} = \rho \cdot \mathsf{SDC}_{i,t,t+1} + (1-\rho)\cdot \mathsf{T}_i$, where $\mathsf{SDC}_{i,tt+1} = [\mathsf{DC}_{i,t}\cdot(1+\tilde{v}_{i,t,t+1})]/[\mathsf{DC}_{i,t}\cdot(1+\tilde{v}_{i,t,t+1}) + (1+\tilde{\eta}_{i,t,t+1})\cdot(1+r_{t,t+1})\cdot(1-\mathsf{DC}_{i,t})]$ and $\tilde{v}_{i,t,t+1}$ is the percent change in the value of debt, $r_{i,t,t+1}$ is the realized firm stock return returns from CRSP, and $\tilde{\eta}_{i,t,t+1}$ is the percent change in the value of equity that is not due to stock returns. Stock returns are retained from the firm's actual history in the data. The two shocks, $\tilde{\eta}$ and $\tilde{v}$, are drawn as pairs from the actual empirical distribution (i.e., a randomly sampled different firm-year). This preserves the correlation structure of non-stock related debt and equity changes. Thus, the hypotheses tested is that these changes are dependent or independent of the firm's current leverage ratio.

In this simulation, the target is again one-shock away from the firm's initial leverage. However, here it is not additive/reversible. Instead $\mathsf{T}_i = \frac{\mathsf{DC}_{i,0}\cdot(1+\tilde{\eta}_{i,t-1,t})\cdot(1+r_{i,t-1,t})}{\mathsf{DC}_{i,0}\cdot(1+\tilde{\eta}_{i,t-1,t})\cdot(1+r_{i,t-1,t})+(1+\tilde{v}_{i,t-1,t})\cdot(1-\mathsf{DC}_{i,0})}$, where $\tilde{v}_{i,t-1,t}$, $\tilde{r}_{i,t-1,t}$ and $\tilde{\eta}_{i,t-1,t}$ are drawn as a vector from the complete sample distributions of observed triplets. (Note: this is the only place where a stock return is drawn rather than retained.)

Figure 3: Process I with Unknown (Firm-Specific) Target



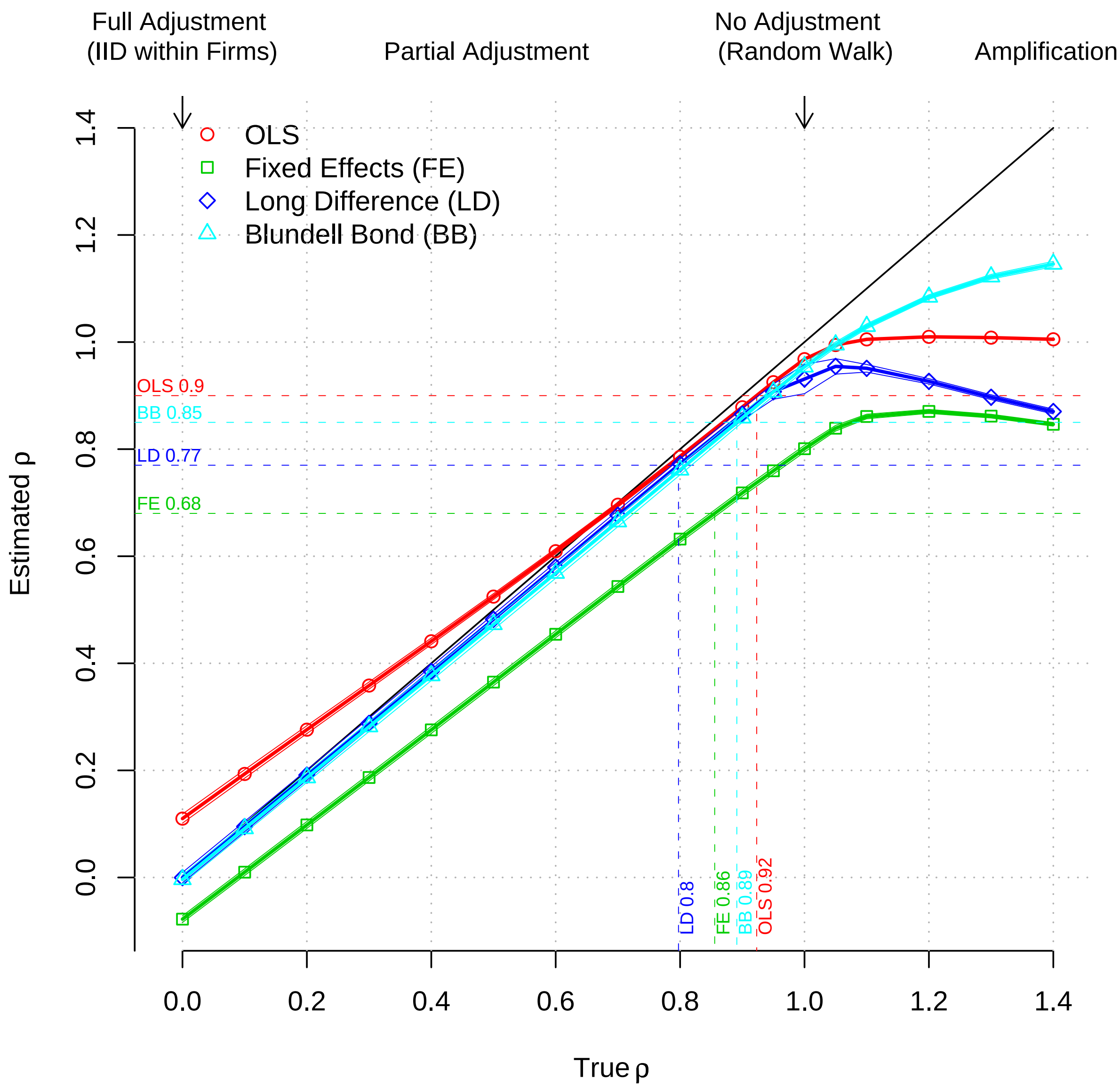
Explanation: Variables are described in Table 1. Estimators are described in Table 2. The simulation process is described in Table 5. In brief, the simulated process is $\widetilde{\text{DC}}_{i,t} = \rho \cdot \text{DC}_{i,t-1} + (1-\rho) \cdot T_i + \tilde{\epsilon}_{i,t}$.

The graph plots the mean estimates for four estimators (OLS, FE, LD, BB) from 500 simulations. The graph also shows their two-sigma ranges, except they are so tight here that they are not visible.

Interpretation: The OLS method is so uninformative that it should be ignored. All inference functions are concave. The FE method has a strong bias, while the BB and LD methods have only a modest bias. At LD's empirical sample estimated rho of 0.77, the best true estimate of 0.85 has a bias of around 8%. The BB estimate has a bias of around 4%.

These estimates seem too far apart to be reconcilable.

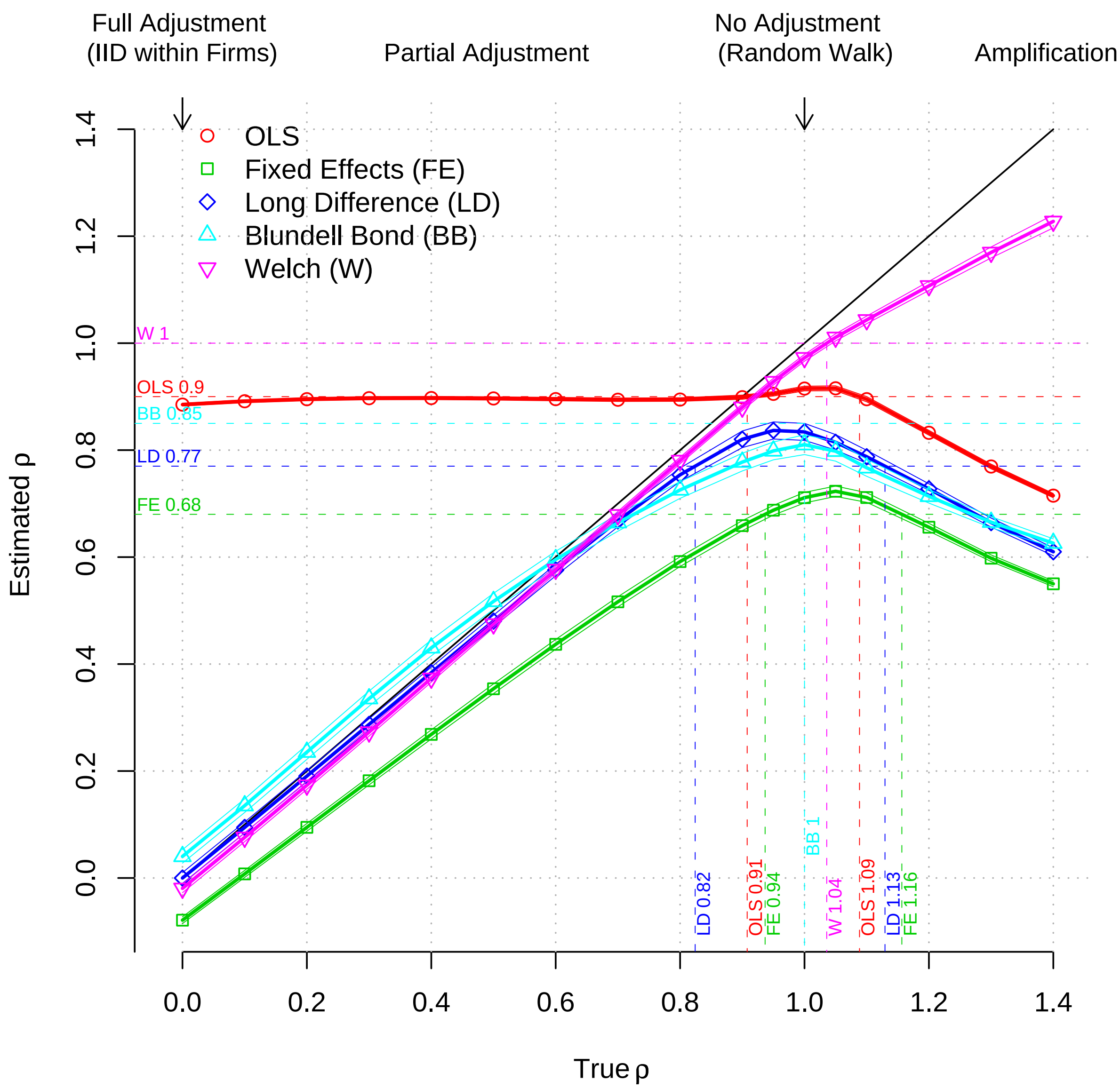
Figure 4: Process I with Known (Firm-Specific) Target



Explanation: This is equivalent to Figure 3, except that the estimators have access to the firm-specific target T_i, which is included as an additional exogenous covariate.

Interpretation: Except in the case of the OLS estimator, knowing the target makes no difference when it comes to estimating an accurate rho coefficient. After all, with 132,412 observations, the estimation accuracy for rho is already almost perfect even if the target is unknown.

Figure 5: Process II with Unknown (Firm-Specific) Target



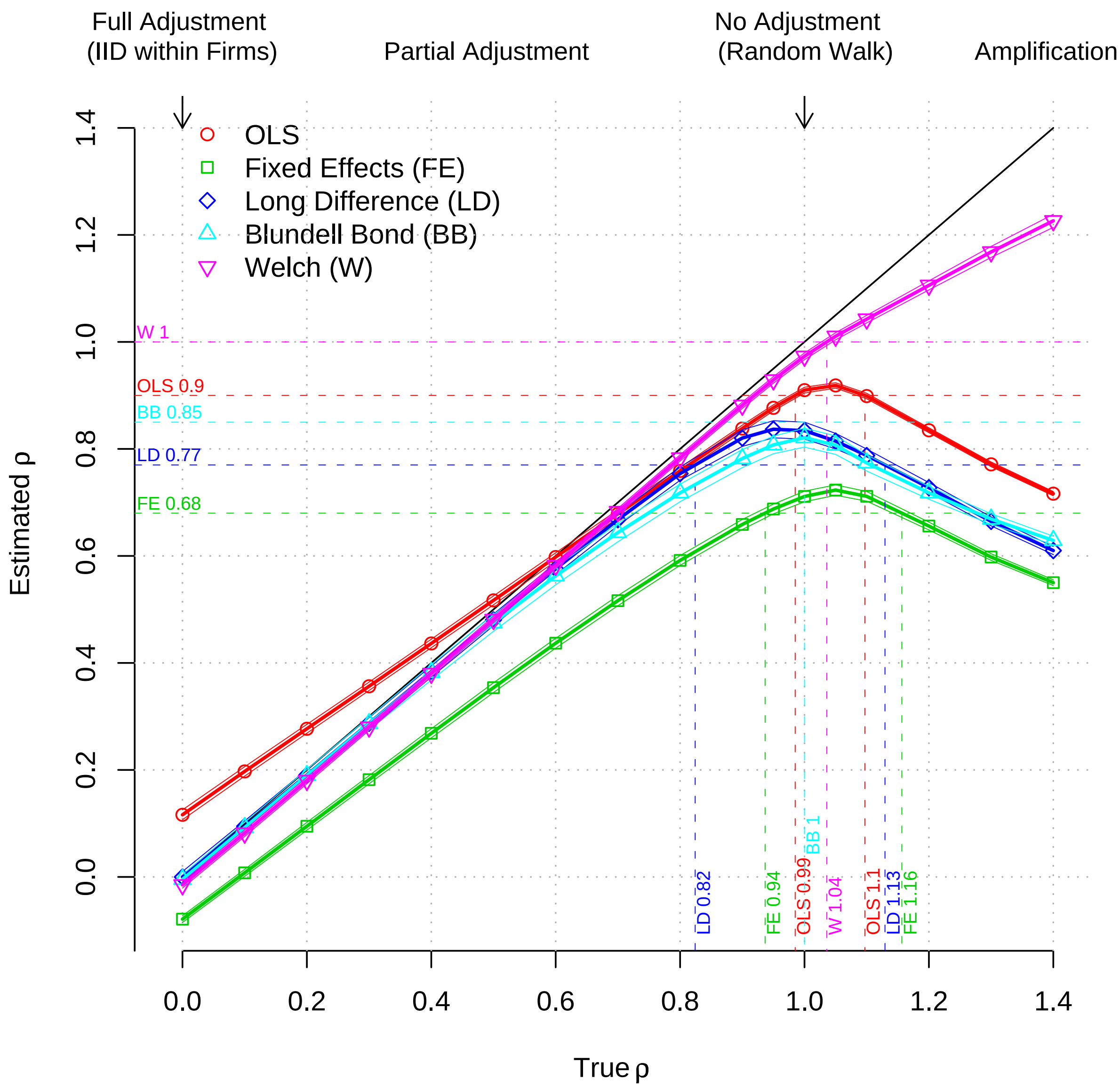
Explanation: Variables are described in Table 1. Estimators are described in Table 2. The simulation process is described in Table 5. In brief, the simulated process is $\widetilde{\mathsf{DC}}_{i,t} = \rho \cdot \mathsf{IDC}_{i,t-1,t} + (1-\rho) \cdot T_i + \tilde{\epsilon}_{i,t}$. Thus, in contrast to Figure 3, the hypothesis of no-adjustment here is not a constant leverage ratio, but one that changes with the firm's stock return (without dividends).

The graph plots the mean estimates for five estimators (OLS, FE, LD, BB, W) from 500 simulations. The graph also shows their two-sigma ranges, except they are so tight here that they are barely visible.

Interpretation: The OLS method is so uninformative that it should be ignored. All inference functions are concave, even to the point where they lose their monotonicity over the graphed domain. The FE method has a strong bias, while the BB and LD methods have only a modest bias for rhos below around 0.8 to 0.9. The W method has less bias and concavity.

These estimates seem too far apart to be reconcilable.

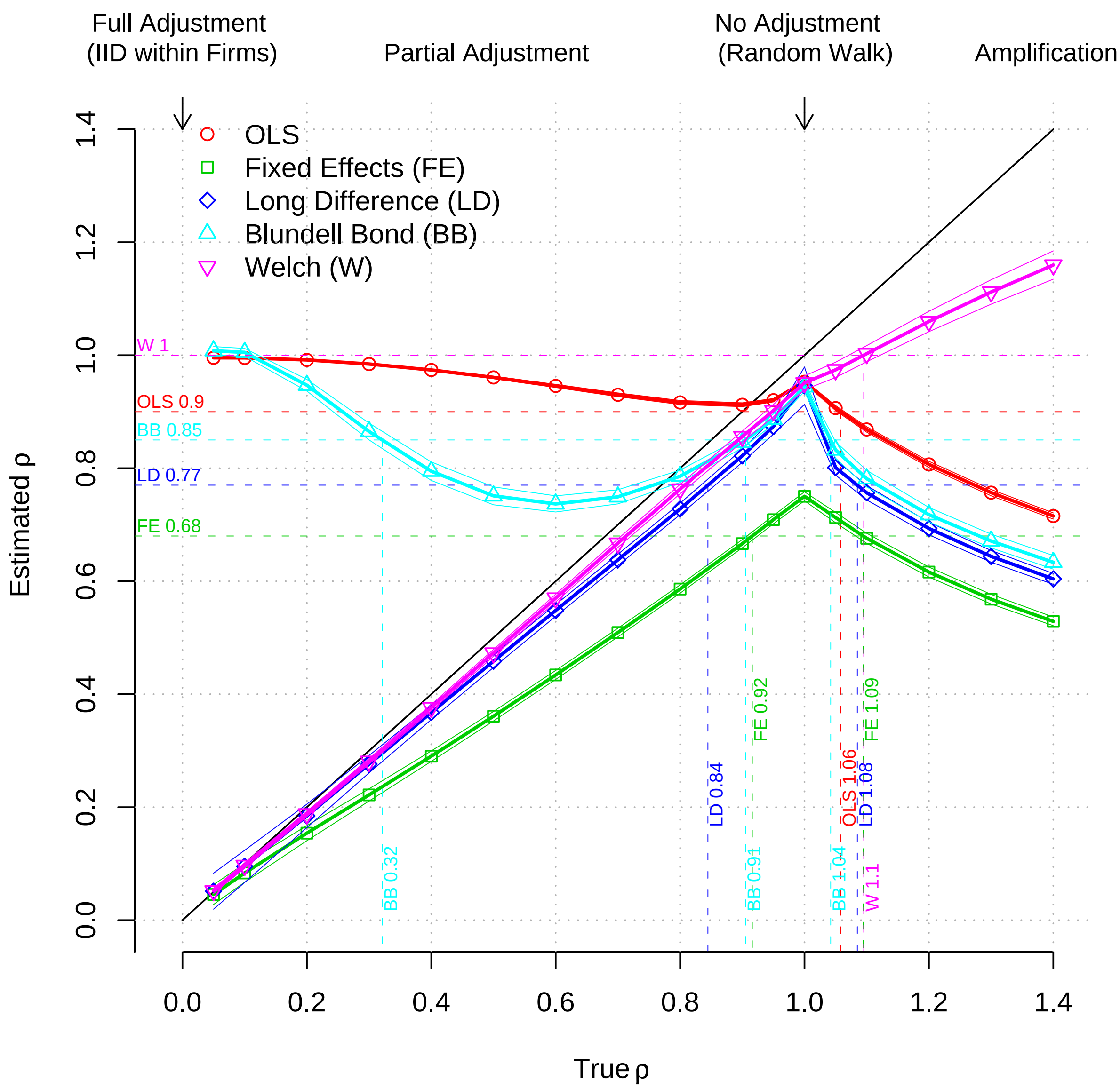
Figure 6: Process II with Known (Firm-Specific) Target



Explanation: This is equivalent to Figure 5, except that the estimations have access to the firm-specific target T_i, which is included as an additional exogenous covariate.

Interpretation: Except in the case of the OLS estimator, knowing the target makes no difference when it comes to estimating an accurate rho coefficient. After all, with 127,556 observations, the estimation accuracy for rho is already almost perfect even if the target is unknown.

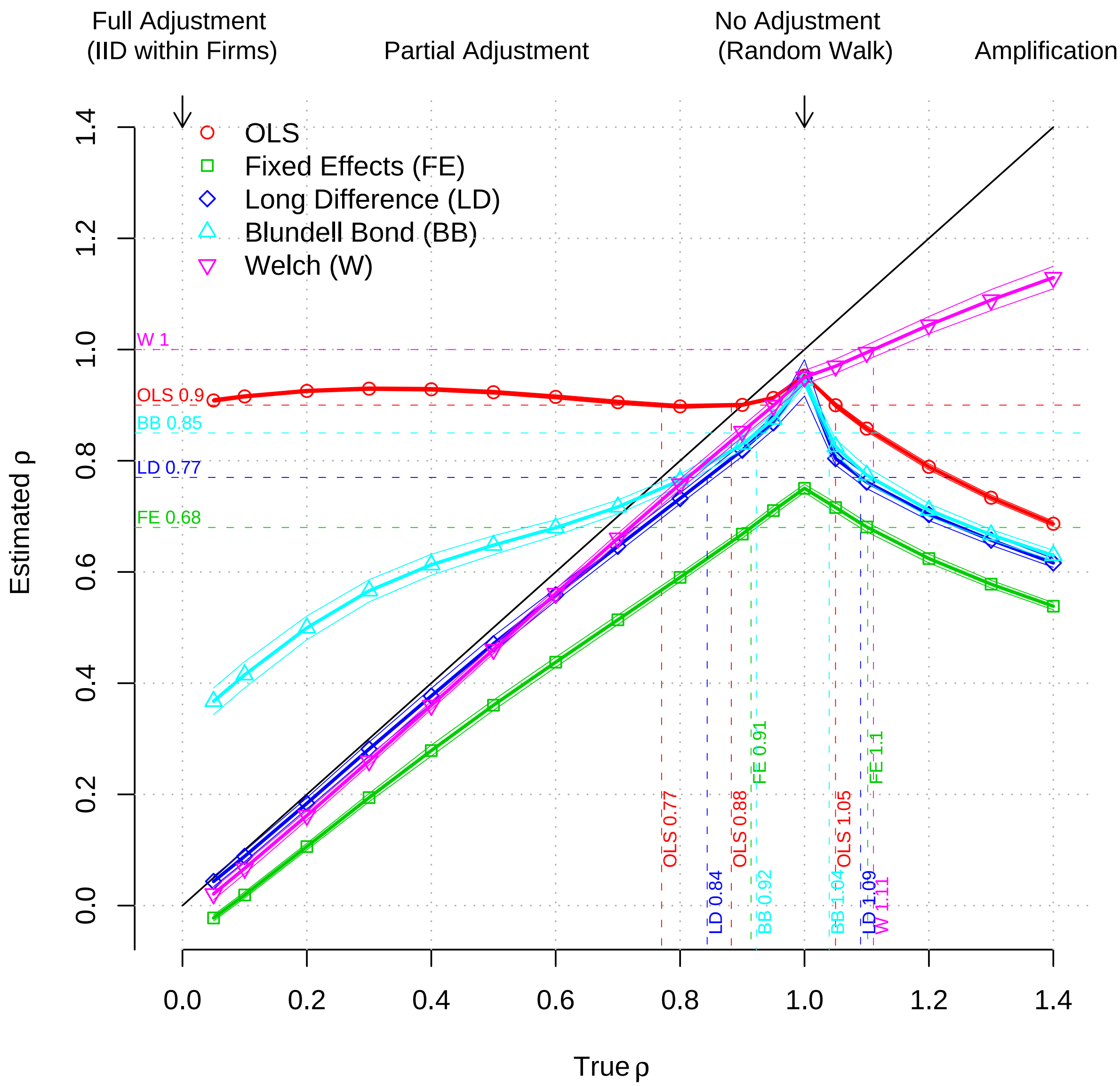
Figure 7: Process III with Unknown (Firm-Specific) Target (-DA)



Explanation: Variables are described in Table 1. Estimators are described in Table 2. The simulation process is described in Table 5. In brief, the simulated process is $\mathrm{DC}_{i,t} = \rho \cdot \mathrm{SDC}_{i,t-1}(\nu_{i,t}, \eta_{i,t}, r_{i,t}) + (1-\rho) \cdot T_i$, where $\mathrm{SDC}_{i,t-1}(\cdot) \equiv D_{i,t-1} \cdot (1+\nu_{t-1,t})/[D_{i,t-1} \cdot (1+\nu_{t-1,t}) + (1+\eta_{t-1,t}) \cdot (1+r_{t-1,t}) \cdot (E_{i,t-1})] \equiv \mathrm{DC}_{i,t-1} \cdot (1+\nu_{t-1,t})/[\mathrm{DC}_{i,t-1} \cdot (1+\nu_{t-1,t}) + (1+\eta_{t-1,t}) \cdot (1+r_{t-1,t}) \cdot (1-\mathrm{DC}_{i,t-1})]$. In contrast to earlier figures, Process III models debt and equity to be joint random draws (from a randomly chosen firm) and only then computes the leverage ratio. If rho is between 0 and 1, leverage ratios are guaranteed to be between 0 and 1, too.

Interpretation: The OLS method is so uninformative that it should be ignored. The BB method is uninformative for low rhos. (This is because it becomes indeterminate when the underlying uncertainty becomes zero.) All inference functions except for the W method are non-monotonic. The FE method has a strong bias, while the BB and LD methods have only a modest bias for rhos below around 0.8 to 0.9. The W method retains monotonicity beyond a rho of 1. These estimates still seem too far apart to be reconcilable, but do seem closer than those in earlier figures.

Figure 8: Process IV with Unknown (Firm-Specific *Partly Noisy*) Target



Explanation: This is like Process III, except that the target $\tilde{T}_i$ is random, experiencing iid normal noise of 9.5% per year. This allows rhos between about 0 and 1 to induce similar year-to-year variability in leverage ratios.

Interpretation: The same as Process III: The OLS method is so uninformative that it should be ignored. All inference functions except for the W method are non-monotonic. The FE method has a strong bias, while the BB and LD methods have only a modest bias for rhos below around 0.8 to 1.0. The W method retains monotonicity beyond a rho of 1.
These estimates still seem too far apart to be reconcilable, but do seem closer than those in earlier figures.

Table 6: Multiple Estimators' Best Fitting Homogeneous True Rho:
Reconciling OLS, FE, LD, BB

Process		Best Fitting True Rho	Penalty Function
I	Best Estimate	0.830	12.04
	Inverse Hessian	(0.016)	

Method	Empirical	Fitted	S.E.	Penalty
OLS	0.895	0.917	0.0012	+18.82
FE	0.681	0.658	0.0032	–7.26
LD	0.772	0.800	0.0048	+5.88
BB	0.847	0.792	0.0047	–11.74

Process		Best Fitting True Rho	Penalty Function
II	Best Estimate	0.937	5.26
	Inverse Hessian	(0.023)	

Method	Empirical	Fitted	S.E.	Penalty
OLS	0.895	0.903	0.0021	+4.11
FE	0.681	0.681	0.0047	–0.03
LD	0.772	0.835	0.0082	+7.68
BB	0.847	0.794	0.0089	–5.93
~~W~~	~~1.003~~	~~0.916~~	~~0.0029~~	~~–29.66~~

Process		Best Fitting True Rho	Penalty Function
III	Best Estimate	1.066	3.43
	Inverse Hessian	(0.010)	

Method	Empirical	Fitted	S.E.	Penalty
OLS	0.895	0.893	0.0024	–0.68
FE	0.681	0.700	0.0043	+4.43
LD	0.772	0.785	0.0063	+2.03
BB	0.847	0.813	0.0072	–4.77
~~W~~	~~1.003~~	~~0.983~~	~~0.0070~~	~~–2.89~~

Process		Best Fitting True Rho	Penalty Function
IV	Best Estimate	1.060	4.49
	Inverse Hessian	(0.010)	

Method	Empirical	Fitted	S.E.	Penalty
OLS	0.895	0.891	0.0023	–1.92
FE	0.681	0.709	0.0043	+6.44
LD	0.772	0.794	0.0063	+3.53
BB	0.847	0.813	0.0071	–4.78
~~W~~	~~1.003~~	~~0.976~~	~~0.0070~~	~~–3.81~~

(Explanations on the next page.)

Table 6 continued.
(Multiple Estimators' Best Fitting True Rho: Reconciling FE, LD, BB)

Explanation: Under a given process, any (true) rho produces a measure of fit to its empirical value for each estimator. This measure of fit is the T-statistic which tests for equality of the empirically observed estimate and the average simulated estimate under the given process with the given true rho. We then find the (true) rho that minimizes an objective function that is the square-root of the average squared T-statistic. This is the best reconciliation. The three estimators which are reconciled in this table are FE, LD, and BB. The fit of the W estimator is reported, too, but it is not part of the objective function. Thus, it is struck out. (Note that the minimized penalties [T-stats] do not average to zero, because the penalty function does not have a constant.) The inverse Hessians are a measure of fit of the penalty function around the optimum. For numerical stability, these are calculated from a distance of 0.02 off the maximum. In a normally distributed context, they could be interpreted as pseudo standard errors.

Interpretation: Process I fits very poorly. Process II suggests a rather different estimate than Process III and Process IV. Process III and a rho estimate of 1.066 fit the data best. Process III struggles with BB and FE, although in opposite directions.

Table 7: Multiple Estimators' Best Fitting Heterogeneous True Rho:
Reconciling OLS, FE, LD, BB

Process		Distrib of Rho Mean	Distrib of Rho Sdv	Penalty Function
I	Best Estimates	0.830	0.006	12.03

Method	Empirical	Fitted	S.E.	Penalty
OLS	0.895	0.917	0.001	−18.83
FE	0.681	0.658	0.003	+7.25
LD	0.772	0.800	0.005	−5.88
BB	0.847	0.792	0.005	+11.74

Process		Mean	Sdv	Penalty Function
II	Best Estimates	0.986	0.110	4.12
	Inverse Hessian	(0.023)	(0.020)	

Method	Empirical	Fitted	S.E.	Penalty
OLS	0.895	0.897	0.003	−0.69
FE	0.681	0.688	0.005	−1.34
LD	0.772	0.790	0.008	−2.23
BB	0.847	0.768	0.010	+7.80
~~W~~	~~1.003~~	~~0.941~~	~~0.003~~	~~+18.29~~

Process		Mean	Sdv	Penalty Function
III	Best Estimates	0.986	0.109	2.09
	Inverse Hessian	(0.022)	(0.014)	

Method	Empirical	Fitted	S.E.	Penalty
OLS	0.895	0.897	0.003	−0.92
FE	0.681	0.682	0.005	−0.11
LD	0.772	0.783	0.008	−1.36
BB	0.847	0.814	0.009	+3.85
~~W~~	~~1.003~~	~~0.900~~	~~0.006~~	~~+17.14~~

Process		Mean	Sdv	Penalty Function
IV	Best Estimates	0.950	0.110	2.63
	Inverse Hessian	(0.035)	(0.015)	

Method	Empirical	Fitted	S.E.	Penalty
OLS	0.895	0.899	0.002	−1.91
FE	0.681	0.688	0.005	−1.57
LD	0.772	0.802	0.007	−4.08
BB	0.847	0.822	0.007	+3.38
~~W~~	~~1.003~~	~~0.876~~	~~0.006~~	~~+22.18~~

(Explanations on the next page.)

Table 7 continued:
(Multiple Estimators' Best Fitting Heterogeneous True Rho: Reconciling FE, LD, BB)

Explanation: This table is like Table 6, except that the true rho is allowed to be normally distributed. Thus, its standard deviation is a second parameter to be optimized over. This rho standard deviation is a measure of firm heterogeneity. (Because the heterogeneity estimate is so close to 0 in Process I, we could not compute a reliable Hessian for Process I.)

Interpretation: Process I fits very poorly. The other three processes agree that the best adjustment coefficient is about 0.95 to 0.98 with a heterogeneity of about 0.11. The three processes II through IV fit quite well. However, looking at the unused (and struck out) W estimates suggests that these three processes are far from being able to explain the W estimate. The best process is Process III, although it struggles (modestly) with the BB estimate.

Table 8: Multiple Estimators' Best Fitting Heterogeneous True Rho:
Reconciling OLS, FE, LD, BB, and W

Process		Dist of Rho Mean	Sdv	Penalty Function
II	Estimate	1.056	0.08	4.38
	Inverse Hessian	(0.019)	(0.022)	

Method	Empirical	Fitted	S.E.	Penalty
OLS	0.895	0.893	0.003	+0.64
FE	0.681	0.698	0.005	−3.61
LD	0.772	0.784	0.007	−1.80
BB	0.847	0.767	0.009	+8.90
W	1.003	1.002	0.003	+0.27

Process		Dist of Rho Mean	Sdv	Penalty Function
III	Estimate	1.066	0.039	3.30
	Inverse Hessian	(0.022)	(0.014)	

Method	Empirical	Fitted	S.E.	Penalty
OLS	0.895	0.891	0.002	+1.58
FE	0.681	0.695	0.004	−3.18
LD	0.772	0.780	0.007	−1.19
BB	0.847	0.809	0.007	+5.35
W	1.003	0.979	0.007	+3.42

Process		Dist of Rho Mean	Sdv	Penalty Function
IV	Estimate	1.059	0.030	4.22
	Inverse Hessian	(0.012-0.039)	(0.025-0.084)	

Method	Empirical	Fitted	S.E.	Penalty
OLS	0.895	0.890	0.003	+1.90
FE	0.681	0.705	0.005	−5.34
LD	0.772	0.792	0.007	−2.83
BB	0.847	0.812	0.008	+4.58
W	1.003	0.974	0.007	+4.16

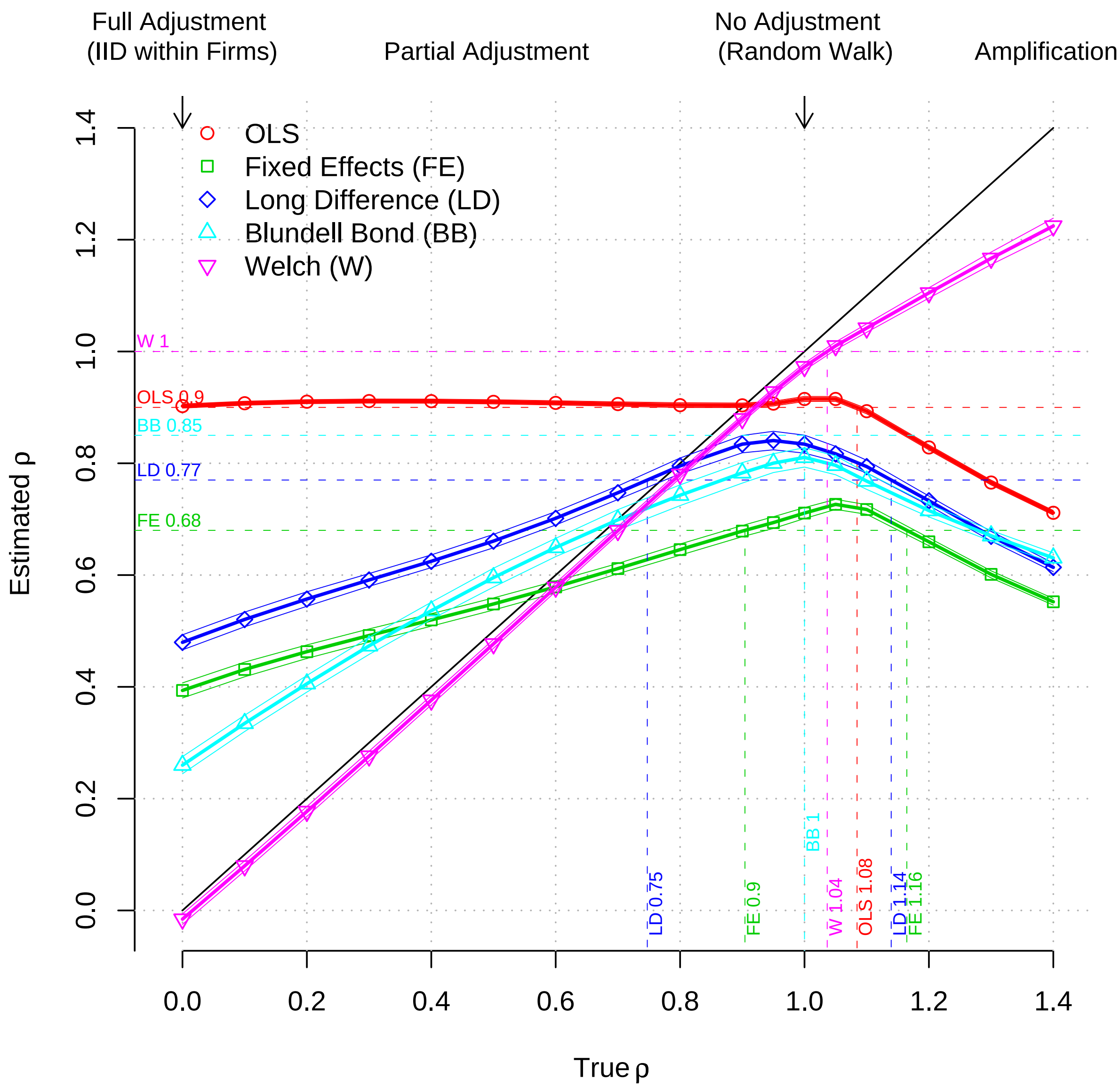
(Explanations on the next page.)

Table 8 continued:
(Multiple Estimators' Best Fitting Heterogeneous True Rho: Reconciling FE, LD, BB)

Explanation: This table is like Table 7, except that the W estimator also enters the objective function. Thus, Process I cannot be fit, because it has no stock return information. For Process IV, the cross-derivative in different directions from the optimum is too variable, which makes it difficult to obtain much precision in the inverse Hessian. Thus, we indicate the range of plausible estimates that our calculations suggested instead of exact amounts.

Interpretation: The three processes agree that the best adjustment coefficient is about 1.06 with a heterogeneity of about 0.02. The best process is III.

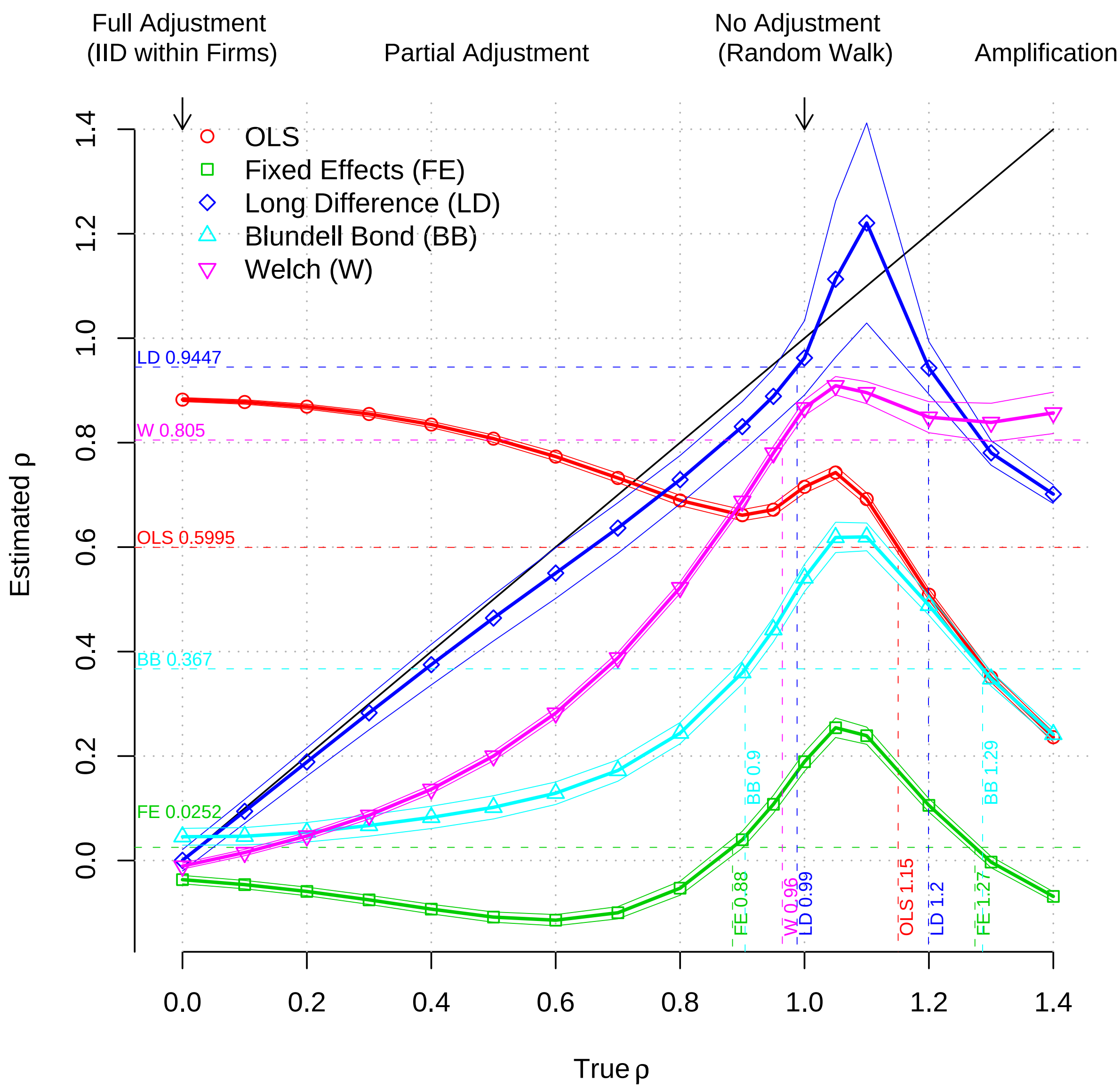
Figure 9: Process II with Changing Unknown (Firm-Specific) Target



Explanation: This figure is the same as Figure 5, except that the unknown firm-specific target changes each period according to the process $T_{i,t} = T_{i,t-1} + \mu_{i,t}$, where $\mu_{i,t}$ are drawn from a normal distribution with a standard deviation of 5%. The graph plots the means of the 100 rho estimates from the four estimators (OLS, FE, LD, BB) that were explained in Table 3. The graph also shows their two-sigma ranges.

Interpretation: Because the W process picks the recent DC as a target control, it works best.

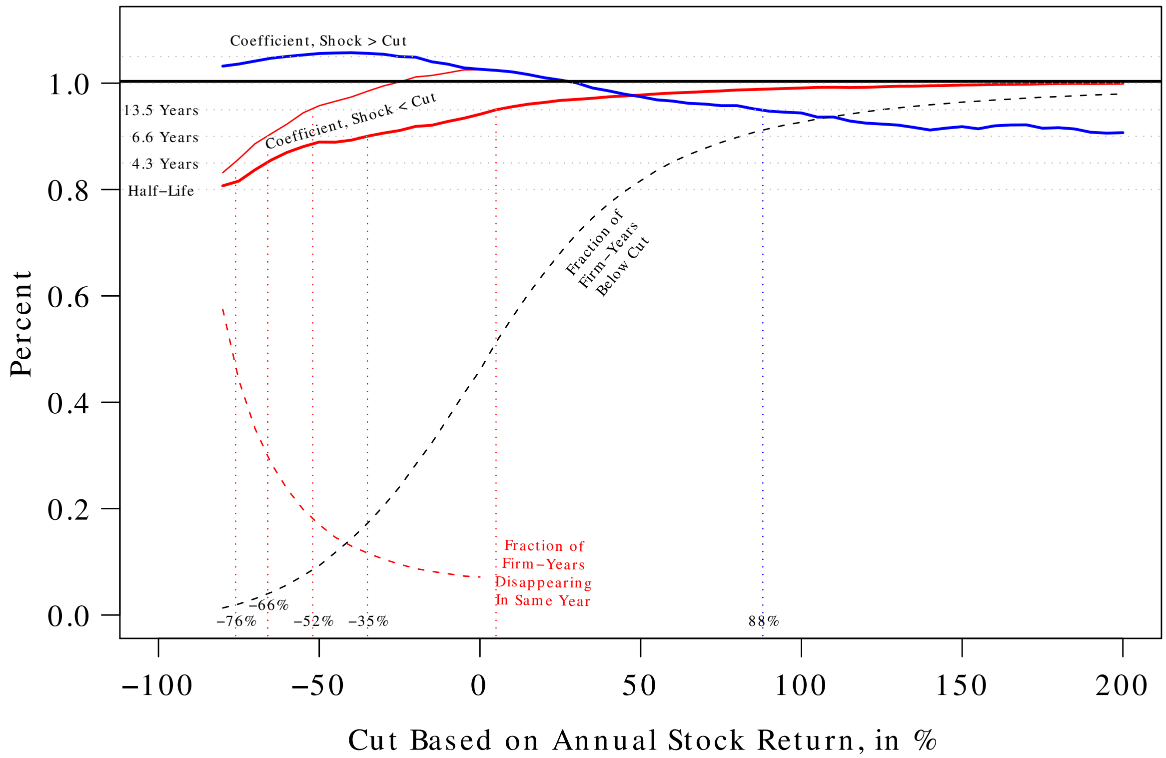
Figure 10: Process II with Unknown (Firm-Specific) Target and 5-Year Estimation



Explanation: This is the equivalent of Figure 5, except that estimations and simulations are carried out with 5-year overlapping intervals. (The dip beyond a true rho of 1 is now due to both super-convergence and truncation.) Note that the units on the x-axis are the 1-year true rhos, not the 5-year rhos.

Interpretation: OLS is useless. FE is uncomfortable—it would require interpreting an empirical coefficient of 2.5% as a true coefficient of 88%. BB, W, and LD offer more reasonable estimates, with the smallest being BB at 0.9, W being 0.96, and LD being 0.99.

Figure 11: Empirical Coefficient Estimates Under Asymmetric and Extreme Stock Return Shocks



Explanation: Observations (firm-years) are first divided into two groups based on their stock return shock that goes into the computation of IDC. The cut-off level that decides on the group membership is on the x-axis. The Welch rho estimate is then computed for both sets, and the uncalibrated coefficient estimate for IDC of each group is graphed. The thinner solid red function assumes that firms that disappear from Compustat after a negative rate of return went "bankrupt" with a leverage ratio of 1.

Interpretation: The figure shows that firms with very negative stock returns shock did readjust. For example, in the 25% of firm-years in which firms experienced a stock return of −35% or lower, the average adjustment coefficient was 0.9. This is weaker but broadly consistent with the finding of Leary and Roberts (2005). However, in about 10% of these firm-years, no end-of-year leverage ratio was observed. If the end-of-year leverage ratio is imputed to be 100% for disappearing firms, the average adjustment coefficient rises to 0.98.
The figure also shows that firms with strong positive stock return shocks readjusted, although their readjustment speed was slower than that of firms with strong negative shocks, again confirming the Leary and Roberts (2005) evidence.

A Other Mechanical Biases

There may also be other violations of the SDPP that could be caused by the assumption of additivity of shocks:

Homogeneity of Errors: Consider the effect of a small ϵ shock to equity on the leverage ratio. (Leverage ratios are symmetric, so all arguments that follow also work with shocks to the amount of debt.)

$$\mathsf{L}(\epsilon) \quad = \quad \frac{D}{D + (1+\epsilon)\cdot E} \quad \Longrightarrow \quad \frac{\partial \mathsf{L}}{\partial \epsilon} = \frac{D \cdot E}{(D + E + \epsilon \cdot E)^2} \quad .$$

Normalizing the total value of the firm to 1, the denominator is 1. Thus, a shock to equity has a marginal influence of 0 if the firm is financed fully with equity, a marginal influence of 0.25 if the firm is financed equally with both debt and equity, and a marginal influence of 0.1875 if the firm is financed with 75% debt and 25% equity. Therefore, when firms of different initial leverage ratios experience same-sized shocks (changes to equity), their debt ratio changes will be of different magnitudes. This violates the orthogonality (and homogeneity) of the error assumption. The variance of the error is predictably correlated with initial leverage. These biases are worse when firms are more heterogeneous in their debt ratios

Curvature: There is also a directional second derivative effect. The average over two equal-sized shocks to equity, one positive and the other negative, is

$$0.5 \cdot \frac{D \cdot E}{(D + E + \epsilon \cdot E)^2} \quad + \quad 0.5 \cdot \frac{D \cdot E}{(D + E - \epsilon \cdot E)^2}$$

For example, if a firm starts out with \$25 in debt and \$75 in equity, then a +\$25 shock to equity reduces the leverage ratio to \$25/\$125=20%, while a –\$25 shock to equity increases its leverage ratio to \$25/\$75=33%. The average ratio of 26.67% is above the initial ratio of 25%. It is easy to show that this bias is positive when the debt ratio is below 50% and negative when the debt ratio is above 50%. Thus, this bias works in favor of concluding that there is mean reversion when there is none.

Our process estimations ignore the non-linearity and curvature violations. We are only tackling the limited domain issue in the paper.

B Biases in OLS and FE Estimations

This appendix will not go into the published paper.

The proofs in this appendix relate to the SDPP

$$\mathsf{L}_{i,t} \quad = \quad \rho \cdot \mathsf{L}_{i,t-1} + (1-\rho)\cdot \mathsf{T}_i + \tilde{\epsilon}_{i,t}, \text{ for } i = 1,2,\ldots,N \;\&\; t = 1,2,\ldots,T;$$

where $\tilde{\epsilon}_{i,t} \sim \mathcal{N}(0, \sigma_\epsilon^2)$, $\mathsf{T}_i \sim \mathcal{N}(\mu, \sigma_\mathsf{T}^2)$, $\mathsf{L}_{i,0} = \mathsf{T}_i + \tilde{\epsilon}_{i,0}$, and $\rho < 1$.

B.1 OLS Estimator Bias

A simple OLS estimator is biased, because of an omitted variables problem. Lagged leverage ($\mathsf{L}_{i,t-1}$) is not orthogonal to the error term ($(1-\rho)\cdot \mathsf{T}_i + \tilde{\epsilon}_{i,t}$). Instead, the lagged dependent variable is positively correlated with the unobserved and not included firm-specific target. It therefore overestimates the true correlation coefficient ρ. We now calculate the magnitude of this bias. We can derive this bias.

The fully pooled OLS estimator for rho is

$$\hat{\rho}_{\text{OLS}} \quad = \quad \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} \mathsf{L}_{i,t}\cdot \mathsf{L}_{i,t-1}}{\sum_{i=1}^{N}\sum_{t=1}^{T} \mathsf{L}_{i,t-1}^2} \quad = \quad \rho + \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}[(1-\rho)\cdot \mathsf{T}_i + \tilde{\epsilon}_{i,t}]\cdot \mathsf{L}_{i,t-1}}{\sum_{i=1}^{N}\sum_{t=1}^{T} \mathsf{L}_{i,t-1}^2} \qquad (1)$$

The numerator in this equation is

$$\begin{aligned}
\text{Numerator} &\equiv \lim_{N\to\infty} \frac{1}{N\cdot T}\sum_{i=1}^{N}\sum_{t=1}^{T}[(1-\rho)\cdot \mathsf{T}_i + \tilde{\epsilon}_{i,t}]\cdot \mathsf{L}_{i,t-1} \\
&= \lim_{N\to\infty} \frac{1}{N\cdot T}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\epsilon}_{i,t}\cdot \mathsf{L}_{i,t-1} \;+\; \lim_{N\to\infty} \frac{1}{N\cdot T}\sum_{i=1}^{N}(1-\rho)\cdot \mathsf{T}_{i,t}\cdot \sum_{t=1}^{T}\mathsf{L}_{i,t-1}
\end{aligned}$$

The first part is zero. The second part requires expansion of $\mathsf{L}_{i,t-1}$. Using $1+\rho+\cdots+\rho^{T-1} = (1-\rho^T)/(1-\rho)$, leverage $\mathsf{L}_{i,t}$ can be rewritten as

$$\mathsf{L}_{i,t} = \tilde{\epsilon}_{i,t} + \rho\cdot\tilde{\epsilon}_{i,t-1} + \cdots + \rho^{t-1}\cdot\tilde{\epsilon}_{i,1} + \rho^{t}\cdot\tilde{\epsilon}_{i,o} + \mathsf{T}_i \tag{2}$$

Sum equivalent terms in this equation over t to obtain

$$\sum_{t=1}^{T}\mathsf{L}_{i,t-1} = T\cdot\mathsf{T}_i + \left(\frac{1-\rho^T}{1-\rho}\right)\cdot\tilde{\epsilon}_{i,0} + \left(\frac{1-\rho^{T-1}}{1-\rho}\right)\cdot\tilde{\epsilon}_{i,1} + \cdots + \left(\frac{1-\rho^2}{1-\rho}\right)\cdot\tilde{\epsilon}_{i,T-2} + \tilde{\epsilon}_{i,T-1}\,. \tag{3}$$

Thus, as N goes to infinity,

$$\begin{aligned}
\text{Numerator} &= \lim_{N\to\infty}\frac{1}{N\cdot T}\sum_{i=1}^{N}(1-\rho)\cdot\mathsf{T}_{i,t}\cdot\sum_{t=1}^{T}\mathsf{L}_{i,t-1} \\
&= \lim_{N\to\infty}\frac{1}{N\cdot T}\sum_{i=1}^{N}(1-\rho)\cdot\mathsf{T}_{i,t}\cdot\Bigg[T\cdot\mathsf{T}_i + \left(\frac{1-\rho^T}{1-\rho}\right)\cdot\tilde{\epsilon}_{i,0} + \left(\frac{1-\rho^{T-1}}{1-\rho}\right)\cdot\tilde{\epsilon}_{i,1} + \cdots + \left(\frac{1-\rho^2}{1-\rho}\right)\cdot\tilde{\epsilon}_{i,T-2} + \tilde{\epsilon}_{i} \\
&= (1-\rho)\cdot\sigma^2_{\mathsf{T}}
\end{aligned}$$

The denominator in (1) can be similarly derived. Squared leverage can be expanded into

$$\mathsf{L}^2_{i,t} = \tilde{\epsilon}^2_{i,t} + \rho^2\cdot\tilde{\epsilon}^2_{i,t-1} + \cdots + \rho^{2\cdot t-2}\cdot\tilde{\epsilon}^2_{i,1} + \rho^{2\cdot t}\cdot\tilde{\epsilon}^2_{i,o} + \mathsf{T}^2_i + \text{cross terms}\,,$$

The cross-terms have probability limits of zero as $N\to\infty$. Thus, the sum over T such squares is

$$\sum_{t=1}^{T}\mathsf{L}^2_{i,t-1} = T\cdot\mathsf{T}^2_i + \left(\frac{1-\rho^{2\cdot T}}{1-\rho^2}\right)\cdot\tilde{\epsilon}^2_{i,0} + \left(\frac{1-\rho^{(2\cdot T-2)}}{1-\rho^2}\right)\cdot\tilde{\epsilon}^2_{i,1} + \cdots + \left(\frac{1-\rho^4}{1-\rho^2}\right)\cdot\tilde{\epsilon}^2_{i,T-2} + \tilde{\epsilon}^2_{i,T-1} + \text{cross terms}\,. \tag{4}$$

The series can be expanded into

$$\begin{aligned}
\left(\frac{1-\rho^2}{1-\rho^2}\right) + \left(\frac{1-\rho^4}{1-\rho^2}\right) + \cdots + \left(\frac{1-\rho^{2\cdot T}}{1-\rho^2}\right) &= \frac{T-(\rho^2+\rho^4+\cdots+\rho^{2\cdot T})}{1-\rho^2} \\
= \frac{T-\rho^2\left(\frac{1}{1-\rho^2}-\frac{\rho^{2\cdot T}}{1-\rho^2}\right)}{1-\rho^2} &= \frac{T-T\cdot\rho^2-\rho^2+\rho^{2\cdot T+2}}{(1-\rho^2)^2},
\end{aligned}$$

Substitute to find that the probability limit of the denominator becomes

$$\lim_{N\to\infty}\frac{1}{N\cdot T}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathsf{L}^2_{i,t-1} = \sigma^2_{\mathsf{T}} + \left[\frac{T-T\cdot\rho^2-\rho^2+\rho^{2\cdot T+2}}{T\cdot(1-\rho^2)^2}\right]\cdot\sigma^2_{\epsilon}\,. \tag{5}$$

And the OLS estimator is

$$\begin{aligned}
\lim_{N\to\infty}\hat{\rho}_{\text{OLS}} = \frac{\text{Numerator}}{\text{Denominator}} &= \rho + \frac{(1-\rho)\cdot\sigma^2_{\mathsf{T}}}{\sigma^2_{\mathsf{T}} + \left[\frac{T-T\cdot\rho^2-\rho^2+\rho^{2\cdot T+2}}{T\cdot(1-\rho^2)^2}\right]\cdot\sigma^2_{\epsilon}} \\
&= \rho + \frac{(1-\rho)}{1+\left[\frac{T-T\cdot\rho^2-\rho^2+\rho^{2\cdot T+2}}{T\cdot(1-\rho^2)^2}\right]\cdot\left(\frac{\sigma^2_{\epsilon}}{\sigma^2_{\mathsf{T}}}\right)}
\end{aligned} \tag{6}$$

as the number of firms tends to infinity.

When $\rho = 0$, $\lim_{N\to\infty}\ \hat{\rho} = \frac{\sigma^2_{\mathsf{T}}}{\sigma^2_{\mathsf{T}}+\sigma^2_{\epsilon}}$. The intuition is that when the true model is $\mathsf{L}_{i,t} = \mathsf{T}_i + \tilde{\epsilon}_{i,t}$, the lagged dependent variable $\mathsf{L}_{i,t-1} = \mathsf{T}_i + \tilde{\epsilon}_{i,t-1}$ becomes a noisy proxy for the omitted target T_i.

When $\rho = 1$, the probability limit of ρ is 1, because the numerator of the bias is zero and the denominator is finite. The intuition is that we no longer have an omitted target error term. $\mathsf{L}_{i,t} = \mathsf{L}_{i,t-1} + \tilde{\epsilon}_{i,t}$, and the lagged dependent variable is orthogonal to the current error term.

B.2 Fixed Effects Estimator Bias

The fixed effect estimator is obtained by estimating the model in deviations from firm-specific means.[31] Thus,

$$\begin{aligned}\hat{\rho}_{\text{FE}} &= \frac{\frac{1}{N\cdot T}\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathsf{L}_{i,t}-\overline{\mathsf{L}}_i)\cdot(\mathsf{L}_{i,t-1}-\overline{\mathsf{L}}_{i,-1})}{\frac{1}{N\cdot T}\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathsf{L}_{i,t-1}-\overline{\mathsf{L}}_{i,-1})^2} \\ &= \rho + \frac{\frac{1}{N\cdot T}\sum_{i=1}^{N}\sum_{t=1}^{T}(\tilde{\epsilon}_{i,t}-\overline{\epsilon}_i)\cdot(\mathsf{L}_{i,t-1}-\overline{\mathsf{L}}_{i,-1})}{\frac{1}{N\cdot T}\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathsf{L}_{i,t-1}-\overline{\mathsf{L}}_{i,-1})^2}\end{aligned} \tag{7}$$

where $\overline{\mathsf{L}}_i = \frac{1}{T}\sum_{t=1}^{T}\mathsf{L}_{i,t}$ and $\overline{\mathsf{L}}_{i,-1} = \frac{1}{T}\sum_{t=1}^{T}\mathsf{L}_{i,t-1}$.

When N goes to infinity, the numerator of the second term for the bias in (7) becomes

$$\begin{aligned}\text{Numerator} &= \lim_{N\to\infty}\frac{1}{N\cdot T}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathsf{L}_{i,t-1}-\overline{\mathsf{L}}_{i,-1}\right)\left(\tilde{\epsilon}_{i,t}-\overline{\tilde{\epsilon}}_i\right) \\ &= \lim_{N\to\infty} -\frac{1}{N}\sum_{i=1}^{N}\overline{\mathsf{L}}_{i,-1}\cdot\overline{\tilde{\epsilon}}_i = -\left[\frac{T-1-T\cdot\rho+\rho^T}{T^2\cdot(1-\rho)^2}\right]\cdot\sigma_\epsilon^2 .\end{aligned}$$

When N goes to infinity, the denominator of (7) is

$$\begin{aligned}\text{Denominator} &= \lim_{N\to\infty}\frac{1}{N\cdot T}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathsf{L}_{i,t-1}-\overline{\mathsf{L}}_{i,-1}\right)^2 \\ &= \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}\mathsf{L}_{i,t-1}^2-\overline{\mathsf{L}}_{i,-1}^2\right) \\ &= \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{T}\sum_{t=1}^{T}\mathsf{L}_{i,t-1}^2-\frac{1}{T^2}\left(\sum_{t=1}^{T}\mathsf{L}_{i,t-1}\right)^2\right].\end{aligned}$$

We already derived the expressions for $\sum L_{i,t}$ in (3) and for $\sum L_{i,t-1}^2$ in (4). Collect the terms, and obtain the probability limit of

$$\left[\frac{T-T\cdot\rho^2-\rho^2+\rho^{2\cdot T+2}}{T\cdot(1-\rho^2)^2}\right]\cdot\sigma_\epsilon^2-\frac{1}{T\cdot(1-\rho)^2}\sigma_\epsilon^2+\frac{2\rho\cdot(1-\rho^T)}{T^2\cdot(1-\rho)\cdot(1-\rho)^2}\cdot\sigma_\epsilon^2-\frac{\rho^2\cdot(1-\rho^{2\cdot T})}{T^2\cdot(1-\rho^2)\cdot(1-\rho)^2}\cdot\sigma_\epsilon^2 .$$

Putting numerator and denominator together and simplifying yields

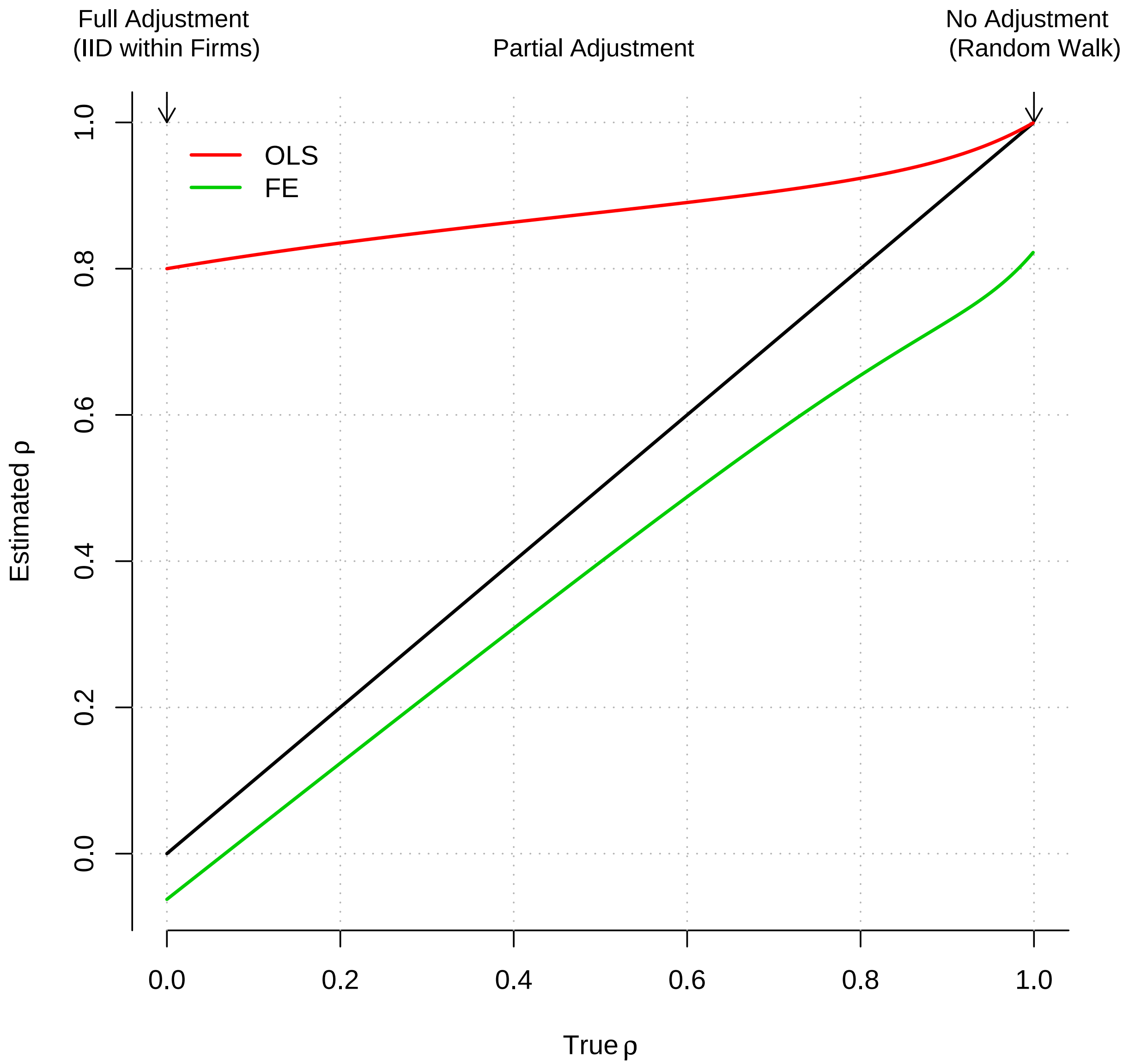
$$\lim_{N\to\infty}\hat{\rho}_{\text{FE}} = \rho - \frac{T-1-T\cdot\rho+\rho^T}{\frac{T^2-T\cdot(T+1)\cdot\rho^2+T\cdot\rho^{2\cdot T+2}}{(1+\rho)^2}+\frac{2\cdot\rho\cdot(1-\rho^T)}{(1-\rho)}-\frac{\rho^2\cdot(1-\rho^{2\cdot T})}{(1-\rho^2)}-T} \tag{8}$$

This does not simplify further. When $\rho = 0$ the bias $\rho_{\text{FE}} - 0$ becomes $-\frac{1}{T}$. The bias does not depend on the variance of the error terms or the variance in the target leverage. The bias comes from having to eliminate the individual fixed effect, which creates a correlation between the demeaned explanatory variable and the demeaned residual.

When $\rho = 1$, the numerator probability limit becomes $\sigma_{\mathsf{T}}^2+\frac{T-1}{2\cdot T}\cdot\sigma_\epsilon^2$, and the denominator probability limit becomes $\sigma_{\mathsf{T}}^2+\frac{T+1}{2}\cdot\sigma_\epsilon^2-\frac{(T+1)\cdot(2\cdot T+1)}{6\cdot T}\cdot\sigma_\epsilon^2$. Therefore, the bias becomes $-\frac{3}{T+1}$.

[31] The results here differs slightly from Nickell (1981) because in our case the fixed effects depends on the AR1 coefficient ρ.

Figure A-1: Bias in the OLS and Fixed Effects Estimations.



Explanation: The graph plots the bias in the OLS and Fixed Effects estimators derived in equation 6 and 8. $t = 16$, $\sigma_\epsilon^2 = 0.125^2$, and $\sigma_\Upsilon^2 = 0.25^2$.