An Efficient Fixed-Effects Estimator for Corporate Finance

Daniela Osterrieder¶ § Darius Palia¶ § Ge Wu∗

April 10, 2020

Abstract

The use of panel data in corporate finance is ubiquitous to estimate the impact of managers’ and/or shareholders’ choices on firm value. We evaluate the properties of four existing and widely used estimators (pooled OLS, random-effects, first-difference, and fixed-effects), and find them to be lacking consistency, efficiency, or both. Consequently, we introduce the new consistent efficient fixed-effects (EFE) estimator. When examining the relationship between CEO performance-pay sensitivity and firm value, we find the EFE estimator to be most appropriate. All estimators are presented as GMM estimators, allowing us to straightforwardly design and conduct tests for model misspecification, including endogeneity.

‡ Rutgers Business School; ¶ CREATES (Aarhus University); § Columbia Law School; ∗ University of Richmond

We thank participants of the 2018 Triple Crown Conference at Fordham, the Econometrics Workshop at North Carolina State, the Finance Seminar at Rutgers, the 4th International Workshop in Financial Econometrics in Maceio (Brazil), the 2019 Applied Time Series Econometrics Workshop at the Federal Reserve Bank of St. Louis, and especially Bill Greene for valuable comments. All errors remain our responsibility. Corresponding Author: Darius Palia; dpalia@business.rutgers.edu
I Introduction

Many studies in corporate finance use panel data (typically across firms and for a number of years) to examine the impact of various corporate choices made by a firm’s managers and/or shareholders on firm value. These corporate decisions include choosing an optimal capital structure (for example, Fama and French, 1998, Pástor and Veronesi, 2003), board characteristics (for example, Fich and Shivdasani, 2006, Coles, Daniel and Naveen, 2008), shareholder rights (for example, Gompers, Ishii and Metrick, 2003, Bebchuk, Cohen and Ferrell, 2008), insider ownership (for example, Morck, Shleifer and Vishny, 1988, Himmelberg, Hubbard and Palia, 1999), and CEO pay (for example, Palia, 2001, Kim and Lu, 2011). We focus on the relationship between CEO pay and firm value in our investigation of the different panel estimation techniques, but in doing so provide a roadmap for testing other corporate finance hypotheses.

To the best of our knowledge, there is no comprehensive study that jointly examines which assumptions about the regression model are the most appropriate when one uses panel data, and which panel estimators should be used to efficiently estimate the parameters of interest. These regressions are plagued by unobservable, endogenous firm-effects. Endogeneity in empirical corporate-finance studies and the related issues of causality and identification have been described in detail by Roberts and Whited (2013) and Kahn and Whited (2016, 2018). Additionally, these regressions also suffer from changing conditional residual variance across firms, residual serial correlation, and non-constant time-varying residual variances.

In this paper, we make the following contributions. (1) We provide a new panel estimator, which we call the efficient fixed-effects estimator (EFE). It is an efficient estimator that is robust to misspecifications of the endogenous firm fixed-effect. This estimator has not yet been used in the corporate finance literature. We show theoretically that EFE is asymptotically consistent and efficient. We also demonstrate through simulations that the estimator has desirable finite sample properties. Specifically, we find only very small biases
in the parameter estimates that correspond to the covariates that change over time and across firms, and a smaller uncertainty than competing estimators.

(2) When we estimate empirically the relationship between firm value and CEO pay using EFE, we find that a one standard deviation increase in the CEO’s pay-performance sensitivity propels a distressed firm (defined as a company whose firm value lies at the fifth percentile of the empirical distribution) only to the 13th – 14th percentile of the firm-value distribution. This is half the impact found when using the standard fixed-effect estimator (FE) instead, which would promise a large leap from the fifth to the 23rd – 24th percentile in response to a one standard deviation increase in the sensitivity of the CEO towards performance pay.

Figure 1: (Asymptotic) Distribution of Parameter Estimator - The figure plots the asymptotic normal distribution of the EFE and FE estimators considered in this paper, for the same choice of parameters as in Table 1. The population parameter of interest is $\alpha$. $\hat{\alpha}_*$ and $\hat{\alpha}_{**}$ are hypothetical point estimates.

Given these large differences in the point estimates for the estimated effect, it is instructive to review the concept of efficient estimation. Often, efficiency is merely associated with smaller standard errors (at least, asymptotically) and hence with improved statistical tests, e.g. for the significance of an effect. However, as shown in Figure 1, efficiency is first and
foremost about estimating an effect.

Let the unknown population effect of interest be denoted by $\alpha$. The likelihood of finding an estimate in an empirical sample that is very far away from the truth, such as e.g. $\hat{\alpha}_{**}$, is much higher for inefficient estimators than for efficient ones. This means that inefficient estimators have inflated uncertainty\(^1\). Conversely, efficient estimation substantially increases the probability to find a sample estimate, e.g. $\hat{\alpha}_*$, that is very close to the true $\alpha$.

(3) We provide detailed practical guidelines for researchers to evaluate the adequacy of the competing estimators to uncover the dominant approach. Implementing these in our empirical data study, we find that the EFE estimator that is based on mean differences outperforms the EFE estimator for data in first-differences, making the former estimator the preferred estimator in our sample. Furthermore, both EFE estimators are preferable over the standard FE and the first-difference (FD) estimator. Lastly, we find that standard FE, FD, and random-effects (RE) estimators are preferable over pooled OLS (POLS), with the FD and the FE estimator outperforming the RE estimator. This last finding is largely consistent with the previous empirical corporate-finance literature (see e.g. Palia, 2001).

(4) We derive the EFE estimator in a single unified Generalized Method of Moments (GMM) estimation framework. This has the advantage that we can derive a test statistic for overidentification. Accordingly, we propose using the usual $J$-statistic to complete the empirical analysis of a typical corporate finance question. We show that the $J$-test is capable to detect misspecification, i.e. it can be used to check whether the assumptions underlying our econometric regression framework hold. In doing so, we can easily test for endogeneity without structural modeling and without natural experiments designed to uncover valid instrumental variables. In finite samples the $J$-test is found to have good size properties. More importantly, it has very strong power to detect violations of the underlying modeling assumptions, which should alleviate some of the concerns regarding too small rejection rates,

\(^1\) True uncertainty of the estimator, or its standard deviation, would be captured by the standard error of the parameter estimate asymptotically when the number of firms approaches infinity.
described in Roberts and Whited (2013) and Erickson and Whited (2012).

Casting the estimators in the GMM framework helps us understand that if one relies on any of the above mentioned standard panel-data estimators as well as the new EFE, then one is implicitly using past, current, and future values of the covariates as instruments. Relying on the $J$-test, we find no reason to reject this assumption in our data. This finding stands in contrast to e.g. Shue and Townsend (2017), who among others argue that pay-performance sensitivity is endogenous. Thus, we also replicate the analysis of Shue and Townsend (2017), relying on their “external” instruments derived from a natural experiment. We find that while valid, such instruments are not relevant in our empirical analysis. In particular, the estimated effect of interest does not change when adding these instruments, and the instruments are insignificant in explaining pay-performance sensitivity in a first-stage regression. It is further not clear that the sample selection that this estimation approach requires is representative.

(5) In a different empirical analysis, the $J$-test may, of course, also reject the modeling assumptions; for instance, if the empirical corporate-finance data truly has a dynamic panel-data structure. A dynamic model, in which e.g. the dependent variable depends on its own historical values, requires different estimators (for an overview, see e.g. Arellano and Bond, 1991 for linear models, or Bazdresch, Kahn and Whited, 2018 for estimators of nonlinear models). In an extension, we thus propose two additional new GMM estimators that are consistent for the parameters of such a dynamic model: The dynamic fixed-effects estimator (DFE) and the dynamic efficient fixed-effects estimator (DEFE). These two estimators are straightforward extensions of the FE and EFE estimators, respectively. In particular, the DFE and DEFE rely on same set of instruments as all other estimators discussed here, but we add lagged dependent variable terms to the set of regressors. We show that the DFE and the DEFE produce small finite-sample biases. The estimation uncertainty is low, especially for the DEFE, which is asymptotically the efficient estimator, and the $J$-test has good finite sample properties.
Within the GMM estimation framework, we also examine four standard panel estimators that have been utilized in the current literature. Specifically, we first examine the POLS estimator, which has been implemented in the studies of e.g. Edmans, Gabaix and Landier (2009) to estimate effect of firm size on CEO wealth-performance sensitivity, of Frydman and Saks (2010) to find the effect of size on executive compensation, of Chen, Hong and Scheinkman (2015) to detect the impact of size and risk (and, in an extension, past executive compensation) on executive compensation, and of Peng, Roell and Tang (2016) to quantify the effect of lagged size on CEO pay-performance sensitivity. Second, we analyze the RE estimator that has been used e.g. by Anderson and Reeb (2003). Third, we study the FD estimator that has been used, inter alia, by Coughlan and Schmidt (1985) and Hall and Liebman (1998) to examine how stock prices (and sales) affect executive compensation, by Jensen and Murphy (1990) to estimate the impact of shareholder wealth on CEO salary and bonus, and by Frydman and Saks (2010) and Chen et al. (2015) in their alternative estimation results. Lastly, we focus on the standard FE estimator. For instance, Murphy (1985) relies on the FE to analyze how firm performance impacts executive compensation, whereas Palia (2001) uses it to ask the reverse question. Graham, Li and Qiu (2012) also consider the FE to examine determinants of executive compensation. We show that the first two estimators (POLS and RE) do not estimate the true effect of corporate choices on firm value; only the last two (FD and FE) estimators are unbiased (as the number of firms becomes large). We also find that the first-difference estimator and the fixed-effects estimator that are generally used in the extant literature are not efficient.

Our paper is closely related to Petersen (2009). Like Petersen (2009), we consider (conditional) heteroskedasticity across firms and time, and serial correlation in the residuals. Petersen (2009) demonstrates how to adjust standard errors in the presence of such nuisances. The suggested “clustering” of standard errors ensures that statistical inference (e.g. t-tests for significance) is correct, yet it does not improve the efficiency of estimation. Stated differently, the actually estimated numbers may still be very far away from the true relationship.
In contrast, in this paper we show that if we use these realistic assumptions to modify panel estimators, we can minimize the uncertainty in the estimators themselves, thus increasing their chance of accuracy.

Our work also shares insights with Gormley and Matsa (2014). Both papers are explicitly concerned with estimation consistency in the presence of unobserved fixed-effects or unobserved heterogeneity. Gormley and Matsa (2014) conclude that FE estimation unbiasedly uncovers the true effect of interest, whereas POLS, “adjusted-Y” (AdjY) estimation that removes the average from the dependent variable before POLS estimation, and “average effects” (AvgE) estimation that includes the average of the dependent variable as an additional explanatory variable in POLS estimation, all lead to inconsistent estimates. We add to these results that RE estimation is inconsistent in this context, as well. We also take the analysis one step further, by (i) relaxing several assumptions in order to achieve a realistic econometric framework, and (ii) suggesting the new EFE estimator (and the new dynamic DEFE estimator) that estimates the true relationship not only consistently but efficiently.

Lastly, besides addressing the endogeneity of the firm fixed-effects, we also focus on the issue of exogeneity of covariates and the idiosyncratic shocks in the model. In this, we emphasize an important aspect of the extant literature that has recently received renewed attention through the work of Grieser and Hadlock (2019) (see also the many references therein). In particular, Grieser and Hadlock (2019) argue that strict exogeneity is required for consistency of FE estimators, and they provide methods to test for it. We complement their work in two ways: One, we show that within the unified GMM estimation framework strict exogeneity can be replaced by the weaker assumption of orthogonality. Two, the above mentioned $J$-test that we propose can be viewed as a test for this assumption. More precisely, the test’s failure to reject can be interpreted as a confirmation (with the usual statistical error) of the required orthogonality condition. That is, the covariates are not correlated with the contemporaneous, the future, or the past idiosyncratic shocks.

Our analysis proceeds as follows. Section II presents the unified econometric regression
framework and corresponding maintained assumptions. In Section III, we discuss the consistency and efficiency of various panel data estimators, and present simulation evidence for their finite-sample performance. Section IV describes our data and variables, and Section V reports our empirical tests and results. In Section VI we propose two additional new estimators for the extended dynamic panel-date model, and Section VII concludes.

II Common Framework

In this section we describe the general framework in which we nest the various estimation techniques. Let firm value be denoted by \( q_{it} \), where \( q \) stands for a firm value measure such as Tobin’s Q, \( i = 1, 2, \ldots, N \), for the relevant firms, and \( t = 1, 2, \ldots, T \), for the relevant year. In corporate finance, we generally have many more cross-sectional than time-series observations (i.e., \( N >> T \)). Then the data generating process (DGP) is:

\[
q_{it} = \psi + ppse_{it}\alpha + x_{it}'\beta + c_i + \eta_{it} \tag{1}
\]

\[
\begin{bmatrix}
q_{i1} \\
q_{i2} \\
\vdots \\
q_{iT}
\end{bmatrix} =
\begin{bmatrix}
1 \\
1 \\
\vdots \\
1
\end{bmatrix}
ppse_{i1} \\
ppse_{i2} \\
\vdots \\
ppse_{iT}
\begin{bmatrix}
x_{i1}' \\
x_{i2}' \\
\vdots \\
x_{iT}'
\end{bmatrix}
\begin{bmatrix}
\psi \\
\alpha \\
\beta
\end{bmatrix} +
\begin{bmatrix}
c_i \\
c_i \\
\vdots \\
c_i
\end{bmatrix} +
\begin{bmatrix}
\eta_{i1} \\
\eta_{i2} \\
\vdots \\
\eta_{iT}
\end{bmatrix} \tag{2}
\]

\[
q_{i} = [1 \ F_i]
\begin{bmatrix}
\psi \\
\gamma
\end{bmatrix} + \epsilon_{it} = X_i\delta + \epsilon_i, \tag{3}
\]

where \( ppse_{it} \) is the CEO pay-performance sensitivity, i.e. the main variable of interest, \( x_{it} \) is the random vector of other explanatory variables or covariates that vary across years and firms\(^2\), and \( \iota \) is a column of ones. The error term of the model, \( \epsilon_{it} \), is composed of an unobserved firm fixed-effect, \( c_i \), and an idiosyncratic error term, \( \eta_{it} \). Assumption 1 formalizes

\(^2\) Strictly speaking, the vector \( x_{it} \) may also contain variables that only vary across firms, in which case we would have an observed firm fixed-effect. Furthermore, some of \( x_{it} \)'s components may be year dummies, say \( a_t \), that vary only over time. In that case, the model would also have an observed fixed time effect (Petersen, 2009). In the later case we require in our theoretical derivations that \( a_t \) be deterministic and its elements cannot be included in the vector of instruments \( z_i \) (see Assumption 1), however.
the DGP.

**Assumption 1 (DGP).** The assumed data generating process (DGP) in equations (1)-(3) is an error components model. Let \( x_{it} \) have size \( ([L_x - 1] \times 1) \). Then \( X_i \) is a matrix of size \( (T \times [L_x + 1]) \), and \( F_i \) has size \( (T \times L_x) \).

Define a vector of instruments as

\[
z_i \equiv \left[ \text{ppse}_{i1} \ x_{i1}' \ \text{ppse}_{i2} \ x_{i2}' \ \ldots \ \text{ppse}_{iT} \ x_{iT}' \ 1 \right]'.
\]

(4)

\( z_i \) has size \( ([TL_x + 1] \times 1) \).

We assume throughout that \( T \) is fixed, whereas \( N \) is large.

To analyze the behavior of panel estimators, we make three additional mathematical assumptions about the model. Note that these assumptions are very nonrestrictive.

**Assumption 2 (Stationarity and Ergodicity).** Let \( w_i \) be comprised of the elements of \((q_{i1}, q_{i2}, \ldots, q_{iT}, \text{ppse}_{i1}, \text{ppse}_{i2}, \ldots, \text{ppse}_{iT}, x_{i1}, x_{i2}, \ldots, x_{iT})\). The process \( \{w_i\} \) is jointly stationary and ergodic.

Assumption 2 simply states that the observations for firm \( i \) can not be perfectly correlated with firm \( j \), but nonstationary behavior across time is permitted. That means that the data can even have unit roots. As explained below, one of our covariates, \( x_{it} \), is a variable that measures a firm’s holdings in property, plants and equipment. This variable will barely change over time, or only very slowly, leaving it strongly serially correlated.

**Assumption 3 (Orthogonality).** Our orthogonality conditions that the estimators exploit for parameter identification are

\[
E(\text{ppse}_{it} \eta_{is}) = 0 \forall s, t = 1, 2, \ldots, T. \tag{5}
\]

\[
E(x_{it} \eta_{is}) = 0 \forall s, t = 1, 2, \ldots, T. \tag{6}
\]

\[
E(\eta_{it}) = 0 \forall t = 1, 2, \ldots, T. \tag{7}
\]
The important take-away from Assumption 3 is that, while $\eta_{it}$ has to be uncorrelated with the model’s covariates, $x_{it}$ and $ppse_{it}$, we do not require the unobserved firm fixed-effect, $c_i$, to be exogenous. This is in contrast to Petersen (2009). Here, the relation between the unobserved fixed firm-effect, $c_i$, and the covariates may be of arbitrary form and scope.

The following simple example demonstrates why allowing for dependence between $c_i$ and the covariates is of importance. Many determinants of the value of the firm do not change over time (or perhaps only very slowly). Examples include firm culture, firm industry, board quality, location, ethical standards, efficiency of operations, HR strategies, etc. While some of these are quantifiable and observable, many others like firm culture and ethical standards are likely to be unobservable and latent. Thus, they are included in $c_i$. Firm culture, wherein talented people cooperate to share information to effectively work in teams to raise firm value, is likely to correlate with the incentive scheme given to the CEO of a company, $ppse_{it}$. The fact that we do not restrict the relationship between these two effects allows us to replicate a realistic scenario, allowing for endogeneity of $c_i$.

Our model allows for endogenous unobserved heterogeneity at the firm level, $i$, which coincides with the level of cross-sectional observations in the panel. In contrast, Gormley and Matsa (2014) are mostly concerned with unobserved fixed-effects at a superordinate level or group, such as e.g. at the industry level, or the state level. The argument is that there may be heterogeneity across groups, yet not necessarily within groups. Note that such an effect is allowed for in our model setup and is subsumed in $c_i$. In this case, $c_i = c_j$ for firms $i, j$ in the same group, whereas $c_i \neq c_j$ if firm $i$ is in a different group than firm $j$.

The only restriction that our model imposes in this context is that a firm may not change its group membership over time. Given the relatively short time span of typical corporate finance panels, it can be considered rather unlikely that a firm changes the industry in which it operates, or moves to a different state, within total time $T$. 

(8)
The orthogonality assumptions rule out correlations between the explanatory variables and contemporaneous, future, or past idiosyncratic shocks $\eta_{it}$. The framework proposed here does consequently not lend itself to dynamic panel-data modeling. Nevertheless, this assumption is weaker than the strict exogeneity emphasized by Grieser and Hadlock (2019) and thus potentially more broadly applicable\(^3\). Henceforth, when we speak about exogeneity, we refer to the “weak” orthogonality form.

**Assumption 4 (Martingale Difference Sequence).** Define a vector sequence

$$g_i \equiv \begin{bmatrix} \eta_{i1}z_i' & \eta_{i2}z_i' & \ldots & \eta_{iT}z_i' \end{bmatrix}' = \eta_i \otimes z_i.$$  \hspace{1cm} (9)

The sequence \(\{g_i\}\) is a joint martingale difference sequence.

Assumption 4 restricts the dependence across firms. The idiosyncratic error term, $\eta_{it}$, may be correlated over time, but not across firms. In fact, we permit the realistic situation wherein the composite regression error may have serial correlation. Additionally, we allow for a non-constant variance of residuals, as there is no reason to expect that the variance of firm value does not change over time. In fact, the riskiness of a firm would vary depending on the conditions in the macroeconomy.

Lastly, the potential presence of conditional heteroskedasticity across firms is not ruled out by any of the assumptions made. Let us examine the appropriateness of this allowance in the context of the relationship between CEO incentive pay and firm uncertainty. The prior literature has found a statistically significant relationship between CEO incentives and firm uncertainty. For example, Lambert and Larcker (1987) and Aggarwal and Samwick (1999) have found a negative relationship between incentives and firm-value uncertainty. On the other hand, Core and Guay (1999) and Oyer and Schaefer (2005) have found a positive relationship. In our setup, the relationship between the variability of firm value and the level of incentives given can be of either sign and magnitude.

\(^3\) By the Law of Iterated Expectations strict exogeneity as in the Grieser and Hadlock (2019), $E(\eta_{it}|X_{is}, pps\epsilon_{is}, c_i) = 0$, implies the orthogonalities (5)-(7), but the reverse is not necessarily true.
III  Panel Estimators

In this section, we start by discussing several well-known estimators for panel data in corporate finance.\footnote{We focus on POLS, RE, FD, and FE. While the latter are very popular in the literature, other estimators have also been used. We do not discuss the Fama-MacBeth method (Fama and MacBeth, 1973), as it “was designed to deal with time effects in a panel data set, not firm effects” (Petersen, 2009), and is hence not applicable in our setup. We also ignore the AdjY and AvgE estimators, given the unfavorable findings in Gormley and Matsa (2014).} We do this to set a benchmark, and to point out the weaknesses in existing estimation methods. Following that, we introduce a new estimator, the efficient fixed-effect estimator (EFE).

In our analysis of the various estimators, we lay a strong focus on the issue of estimation efficiency. An estimator is efficient if it consistently estimates the true parameters with the smallest possible degree of uncertainty. The new EFE estimator satisfies this property.

All estimators are explained in the Generalized Method of Moments framework (GMM) for comparability. Adapting all popular (and new) estimators to the GMM estimation structure, affords a straightforward possibility to detect model misspecification and to verify the endogeneity of regressors with respect to the idiosyncratic errors. Importantly, however, it is not necessary for the researcher to find valid instrumental variables, or to specify a structural model.

3.1  Pooled OLS (POLS) and Random-Effects (RE)

In Appendix A, we analyze the pooled OLS and the random-effects estimator assuming that Assumptions 1-4 are satisfied. Theorem 1 shows that both estimators can be written as GMM estimators with particular weighting matrices. We find that both estimators are inconsistent. That is, it is unlikely that the estimators converge to the true parameters $\delta = [\psi \alpha \beta]'$. This finding is not surprising, as consistency of these two estimators would require exogeneity (or, more precisely, orthogonality) of the fixed firm-effect, $c_i$. Since the estimators are inconsistent, they cannot be efficient. In addition, the estimators are not likely to have the smallest possible (asymptotic) standard deviations, i.e. estimation uncertainty.
is not necessarily small.

3.2 First-Differences (FD) and Fixed-Effects (FE)

The first-difference estimator and the fixed-effects estimator have been used often for panel data in the corporate finance literature. In Appendices B and C, we show that the respective FD and FE estimator can be viewed as a GMM estimator of a transformed model (12). More precisely, denote by \( I \) the identity matrix and define a matrix of size \((T \times T)\):

\[
M = I_T - \iota (\iota' \iota)^{-1} \iota'.
\]  

Further let \( C \) be a full rank matrix of size \((T \times [T - 1])\) such that \( C' \iota = 0 \). The matrix \( C_1 \), which is equal to \( M \) excluding its last column is one choice for \( C \) that has appeared in the fixed-effects literature. Alternatively, the matrix \( C_2 \) has been used:

\[
C_2 = \begin{bmatrix}
-1 & 0 & 0 & \ldots & 0 & 0 \\
1 & -1 & 0 & \ldots & 0 & 0 \\
0 & 1 & -1 & \ldots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \ldots & 1 & -1 \\
0 & 0 & 0 & \ldots & 0 & 1 \\
\end{bmatrix}. 
\]  

Either option for \( C \) leads to a transformation of equations (1)-(3) resulting in

\[
C'q_i = C'X_i \delta + C' \varepsilon_i = C'F_i \gamma + C' \eta_i. 
\]  

If we let \( C = C_1 \), then the data in the regression is demeaned (i.e. time-series averages are removed). Conversely, if \( C = C_2 \) then the data is in first-differences.

The advantage of the above transformation of the data is that the endogeneity in the error term has been removed. In particular, pre-multiplication with the matrix \( C \) eliminates
the unobserved fixed firm-effect. The remaining error term is exogenous. It is thus not surprising that the FD and the FE estimator is consistent for $\gamma = [\alpha \ \beta']'$.

On the downside, we now have a system of only $(T - 1)$ equations in estimating (12). In particular, we are loosing information about the time-series average of the variables (if $C = C_1$), or the level of the time series (if $C = C_2$). Additionally, we will no longer be able to identify the overall intercept, $\psi$.

Theorem 2 shows that the FD estimator is a GMM estimator with particular weighting matrix of the transformed equation (12), but only when $C = C_2$. Under the Assumptions 1-4, the estimator is consistent. Yet, it is not efficient unless the error of the transformed regression is conditionally homoskedastic (across firms), uncorrelated over time, and has a time-constant variance. However, as argued above neither assumption is very realistic, which is why the standard errors of the FD estimator need to be “adjusted”, i.e. clustered (see Petersen, 2009).

In Theorem 2, we also present a $J$-test statistic that can be used to verify the outlined assumptions. Under the null hypothesis, Assumptions 1-4 are jointly satisfied. Therefore, a failure to reject can, for instance, be interpreted as statistical evidence for the exogeneity of covariates and idiosyncratic error.

More popular than the first-difference estimator, especially in more recent empirical corporate-finance work, is the fixed-effects estimator. In fact, Grieser and Hadlock (2019) call it the “most common estimator choice in” the likely presence of a firm fixed-effect. In Appendix C we analyze the FE estimator within the GMM framework in detail. We show that the FE estimator is more general than the FD estimator, since it can be defined for an arbitrary choice of the matrix $C$. Thus, for the FE estimator, matrices $C_1$ and $C_2$ above are just two examples, but the results presented here hold for any full rank matrix $C$ whose columns sum to zero.

Theorem 3 shows that while consistent, the FE estimator is not efficient, under the Assumptions 1-4. If the idiosyncratic shocks were conditionally homoskedastic (across firms),
uncorrelated over time, and had a time-constant variance, then the FE estimator would be efficient. As argued before, these are strong assumptions that are not likely to hold in empirical corporate-finance studies, which is why standard errors need to be adjusted for correct inference. We further derive a $J$-statistic in Theorem 3, which under the null hypothesis of no misspecification has the usual $\chi^2$-distribution.

### 3.3 Efficient Fixed-Effects (EFE)

Clustering standard errors, as suggested in Petersen (2009), can be viewed of an *ex-post* adjustment for the presence of conditional heteroskedasticity (across firms), serial error correlation, and time-varying variance. Thus, these data characteristics are treated as nuisances that need to be corrected. An alternative approach is to view these properties of the data as information that can improve the efficiency of the estimation of the parameters.

In Appendix D we follow this idea and develop an efficient fixed-effects estimator under the small set of Assumptions 1-4. We chose to extend the FE estimator, instead of the FD estimation, since the former is more general and more popular. We present the new EFE estimator as a GMM estimator of a transformed model (12).

The EFE estimator for $\gamma$, derived in Theorem 4, is given by

$$
\hat{\gamma} = \left\{ \left( \frac{1}{N} \sum_{i=1}^{N} (F_i' C \otimes z_i') \right) \left( \frac{1}{N} \sum_{i=1}^{N} (C' \hat{\eta}_i \hat{\eta}_i' C \otimes z_i z_i') \right)^{-1} \left( \frac{1}{N} \sum_{i=1}^{N} (C' F_i \otimes z_i) \right)^{-1} \right\}^{-1} \left( \frac{1}{N} \sum_{i=1}^{N} (C' \hat{\eta}_i \hat{\eta}_i' C \otimes z_i z_i') \right)^{-1} \left( \frac{1}{N} \sum_{i=1}^{N} (C' q_i \otimes z_i) \right),
$$

where $C' \hat{\eta}_i$ are the regression residuals of (12). Theorem 4 shows that the EFE estimator is consistent, under the Assumptions 1-4.

Clustered standard errors that are robust against conditional heteroskedasticity across firms, serial correlation, and heteroskedasticity across years are the diagonal elements of the
\[ \sqrt{\text{Var}(\hat{\gamma})} = \left( \frac{1}{N} \sum_{i=1}^{N} (C'F_i \otimes z_i)' \left( \frac{1}{N} \sum_{i=1}^{N} (C'\hat{\eta}_i\hat{\eta}_i'C \otimes z_i'z_i') \right)^{-1} \frac{1}{N} \sum_{i=1}^{N} (C'F_i \otimes z_i) \right)^{-1}, \] (14)

multiplied by \(1/\sqrt{N}\). Note that the estimator and the standard errors will be different, depending on the choice of matrix \(C\). For any valid choice, the estimator is efficient in the sense that the uncertainty in the estimator for \(\alpha\) and \(\beta\) in model (12) is minimized.

Lastly, Theorem 4 also present a \(J\)-test statistic that can be used to verify the outlined assumptions.

### 3.4 Simulations of Balanced Panels

To investigate the finite sample behavior of all estimators, we carry out a Monte Carlo simulation study.

We simulate the error components model in equations (1)-(3). More precisely, we assume that the true population parameters \(\psi\), \(\alpha\), and \(\beta\) are random draws from a \(\chi^2\)-distribution. In every one of the 50,000 simulations, we randomly generate the covariates in \(X_i\) by drawing from a logistic distribution. Next, we generate the endogenous unobserved firm fixed-effect as a function of the time-series averages of \(X_i\) and a random error term. Thus, \(c_i\) is correlated with the explanatory variables (more precisely with the average of the variables \(ppse_{it}\) and \(x_{it}\)).

In the next step of the simulations, we generate the exogenous idiosyncratic error \(\eta_{it}\) as an autoregressive process of order one, with autoregressive coefficient of 0.6. In addition, \(\eta_{it}\) has a GARCH structure (the constant is set to 0.5, the \(a\)-coefficient is 0.2, and \(b = 0.55\)) across \(i\) and the variance also varies randomly across \(t\). Note that \(\eta_{it}\) is thus heteroskedastic over time and firms, and has autoregressive time-series dynamics, i.e. exhibits serial correlation, but is independent across \(i\). Finally, we generate the dependent variable, \(q_{it}\), as in equation (1). We then use the simulated series to compute the various estimators described in the
The results for \( N = 1,000 \) firms are in Table 1. As expected, the overall bias in \( \hat{\alpha} \) and \( \hat{\beta} \) from the consistent estimators (FD, FE, and EFE) is small, whereas the inconsistent estimators, POLS and RE, produce large biases in the parameter estimates. Interestingly, the bias in POLS and RE is always positive, suggesting that both inconsistent estimators will always produce estimates that are too large. While POLS heavily overestimates the magnitude of the true parameter, the bias from RE is more moderate; yet still it is approximately between 2 and 175 times larger than e.g. the bias from the EFE \((C = C_1)\) estimator. For all but one element of the coefficient vector composed of \( \alpha \) and \( \beta \), the efficient estimators (EFE) produce a substantially smaller bias in the estimates than their inefficient counterpart (FD and FE), whereas for one element the picture is reversed. There is no clear dominance in terms of bias reduction when comparing the EFE in first-differences \((C = C_2)\) with the demeaned EFE \((C = C_1)\) estimator. Similarly, there is no consistent dominance between the FD and the FE estimator, when analyzing the parameter bias. For the constant term, \( \psi \), only POLS and RE produce an estimator. Table 1 shows that it is strongly overestimated.
the inconsistent RE estimator has a standard deviation that competes with the one of the EFE.

[ Insert Figure 3 ]

As for the bias, we also plot the convergence properties of the pooled estimators’ standard deviations. The results are in figure 3. As $N$ increases, all estimators have a decreasing dispersion. Yet, the standard deviation of POLS remains significantly larger than any other estimator. Inefficient FD is less volatile than FE for small $N$, but this discrepancy becomes smaller as $N$ increases, and both of these inefficient estimators have substantially larger uncertainty than the remaining estimators. The standard deviations of RE and EFE are the lowest and seem to converge to one another in magnitude as $N \to \infty$.

Finally, Table 1 shows the behavior of the $J$-test, which is the standard GMM test for overidentification or misspecification. At a nominal size of 5%, the actual size of the $J$-test for the FD, the FE, and the EFE estimation techniques is good, albeit a little too small. The sizes (i.e. the probability for a Type I statistical error) vary from 3.6% for the FD and the FE estimation to 4.2% for the EFE estimators, and thus exhibit a small degree of undersizing. However, this is only a small sample problem, as figure 4 demonstrates. We plot the actual size of the $J$-test as $N$ increases. As can be seen, the size of the test converges to the nominal level of 5%.

[ Insert Figure 4 ]

For the POLS and the RE estimator the results in Table 1 summarize the power of the $J$–test (i.e. one minus the probability for a Type II statistical error). As can be seen, the assumptions of the model are strongly rejected, which confirms that these estimators require the unobserved firm fixed-effect to be exogenous. Simulations in figure 4 further confirm that the power of the test is very strong, even for a very small number of firms.
3.5 Simulations of Unbalanced Panels

The theory for the estimators developed above and in Appendices A, B, C, and D assumes that panels are balanced. However, in many studies in corporate finance, the data set is unbalanced. In what follows, we present simulation evidence for the efficient and inefficient fixed-effects estimators for unbalanced panels\(^5\). We omit the POLS and the RE estimator since they are inconsistent in our setup. We further no longer focus on the FD estimator, since it is less commonly used and restricts the C-matrix to equal \(C_2\).

Adjustments for unbalanced panels are minimal. In particular, we modify the raw data by replacing missing observations by time-series averages of the remaining observations. Theorem 5 in Appendix E shows that the resulting fixed-effects estimator is the correct estimator for unbalanced panels\(^6\).

The simulation setup is comparable to the one for balanced panels, with one exception. To replicate an unbalanced panel with \(T = 5\) years, we set one time period as missing for 50\% of firms, two time periods missing for 20\% of firms, and three time periods missing for 10\% of firms.

[ Insert Table 2 ]

The results are in Table 2. As in the balanced-panel simulations, the overall bias in the consistent estimators (EFE and FE) is also small in the unbalanced-panel scenario. Again, there is no clear dominance of one of the estimators (FE, demeaned EFE (\(C = C_1\)), and EFE for first-differences (\(C = C_2\))) over the others in terms of bias in the parameter estimates for \(\alpha\) and \(\beta\). As can further be seen in table 2, EFE is clearly more efficient than FE for parameters for \(\alpha\) and \(\beta\). That is, the standard deviation over repeated samples of the EFE estimators is visibly smaller than the one of the FE estimator.

\(^5\) Matlab codes for all fixed effects estimators can be downloaded here: https://sites.google.com/site/danielaosterrieder/Research. \(^6\) Note that the same is not necessarily true for the POLS, the RE, and the FD estimators. The development of these three estimators as GMM estimators for unbalanced panels is beyond the scope of this paper, and left for future research. In the empirical analysis below, POLS, RE, and FD are computed as standard pooled estimators.
Lastly, Table 2 summarizes the $J$-test. At a nominal size of 5%, the actual size of the $J$-test is somewhat too small for all estimation techniques. Its size varies from 3.29% for the FE estimation to 3.45% for the EFE estimations. As argued above, this is a small sample problem that vanishes with increasing $N$. To safeguard against being overly optimistic about the results for small $N$, we suggest the usage of a conservative significance level of 10%.

### 3.6 Simulated Power of the $J$-Test

One of the advantages of integrating commonly used panel-data estimators into the GMM estimation framework, is that we can define a $J$-test for misspecification. More precisely, the $J$-tests developed in the appendices, assume that the Assumptions 1-4 hold jointly under the null hypothesis. The alternative hypothesis manifests itself if either one of the assumptions fails, such as e.g. the exogeneity of idiosyncratic shocks, or the uncorrelatedness of $\eta_i$ across firms. Detecting a failure of the assumptions is thus of paramount importance, which implies that the test needs to have sufficient power. Focusing on the FE and the EFE estimators, the following simulations show that the corresponding $J$-tests achieve very high rejection rates of incorrect null hypotheses in finite samples, thus alleviating the concerns about statistical power that have been voiced in the literature (see e.g. Roberts and Whited, 2013, or Erickson and Whited, 2012).

As mentioned previously, the presented estimators are not robust to dynamic models. To see that a dynamic-panel specification would violate our assumptions, assume the following alternative DGP, where past realizations of $q_{it}$ enter into the model:

$$q_{it} = \psi + ppse_{it} + x'_{it} \beta + \sum_{p=1}^{P} q_{i(t-p)} \phi_{p} + c_i + \eta_{it}$$
If the researcher continues to estimate this model with the same GMM estimators as above, using $X_i$ as regressor matrix and instruments $z_i$ from Equation (4), then all estimators including FE and EFE will be inconsistent. To see this, note that the error term of a regression of $q_i$ on $X_i$ can be written as $\tilde{\epsilon}_i = u\hat{c}_i + \tilde{\eta}_i$, or more specifically\footnote{We require that $(1 - \sum_{p=1}^P \phi_p z^p) = 0$, where $z$ is a complex number, has to satisfy $|z| > 1$, i.e. all roots of the system are outside of the unit circle. Then, $(1 - \sum_{p=1}^P \phi_p L^p)^{-1} = \sum_{j=0}^{\infty} \theta_j L^j$ and $\theta_0 = 1$.}

$$q_i = \begin{bmatrix} q_{i(P+1)}^T \\ q_{i(P+2)}^T \\ \vdots \\ q_{i(P+T)}^T \end{bmatrix} = \begin{bmatrix} 1 & ppse_i(P+1) & x_i'(P+1) & q_i(P) & q_i(P-1) & \cdots & q_1 \\ 1 & ppse_i(P+2) & x_i'(P+2) & q_i(P+1) & q_i(P) & \cdots & q_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & ppse_i(P+T) & x_i'(P+T) & q_i(P+T-1) & q_i(P+T-2) & \cdots & q_T \end{bmatrix} \begin{bmatrix} \psi \\ \alpha \\ \beta \\ \phi_1 \\ \phi_2 \\ \vdots \\ \phi_P \end{bmatrix} + \begin{bmatrix} \epsilon_i \\ c_i \\ \eta_i(P+1) \\ \eta_i(P+2) \\ \eta_i(P+T) \end{bmatrix}$$

$$q_i = \begin{bmatrix} \iota & F_i & Q_i^{(-)} \end{bmatrix} \begin{bmatrix} \psi \\ \gamma \\ \phi \end{bmatrix} + \epsilon_i = \begin{bmatrix} X_i & Q_i^{(-)} \end{bmatrix} \begin{bmatrix} \delta \\ \phi \end{bmatrix} + \epsilon_i \tag{15}$$

$$\tilde{\epsilon}_{it} = \frac{c_i}{1 - \sum_{p=1}^P \phi_p} + \left( \sum_{j=1}^{\infty} ppse_{i(t-j)} \alpha \theta_j + \sum_{j=1}^{\infty} x_{i(t-j)}' \beta \theta_j + \sum_{j=0}^{\infty} \eta_{i(t-j)} \theta_j \right) = \tilde{c}_i + \tilde{\eta}_{it}, \tag{16}$$

where the parameters $\theta_j$ are functions of the autoregressive parameters $\phi_p$. It is easy to see that the new error term, $\tilde{\eta}_i$, does not satisfy the orthogonality conditions (5)-(7). That is $E(ppse_{it}\tilde{\eta}_{is})$, $E(x_{it}\tilde{\eta}_{is})$, and $E(\tilde{\eta}_{it})$ are not likely to be zero even if Assumption 3 continues to hold.

To clarify, estimating model (15) by FE or EFE, as defined in Section III, generates an omitted variables problem, where the estimation inconsistency arises because the omitted variables, $Q_i^{(-)}$, are correlated with the instruments, $z_i$.

Fortunately, such a violation of the assumptions can easily be detected by relying on the $J$-statistic. To demonstrate this, we carry out the same set as simulations for balanced
panels as described in Section 3.4, with the exception that \( q_{it} \) is also driven by its lagged term \( q_{it(t-1)} \), scaled by a parameter \( \phi \in (0, 1) \) that is randomly drawn from a uniform distribution. We find large biases in the parameters estimates of all estimators, and the uncertainty in the estimators is big. The EFE is further no longer guaranteed to have a smaller variances than the standard FE estimator. However, the power of the \( J \)-test of all estimators is 100\%, which implies that we can be confident to detect such a violation of our assumptions. The exact same picture arises when we run the simulations for unbalanced panels, instead.

A common concern is that adding “useless” instruments, i.e. instruments that do not aid the identification of the model parameters, to the vector \( z_i \) will negatively impact the GMM estimation. To see that our estimators are rather robust to such a scenario, we again rely on the same balanced-panel (unbalanced-panel) simulation design as in Section 3.4 (3.5), but we add \( L_u = 2 \) valid instruments, \( u_{it} \), that are uncorrelated with \( X_i \). In particular, we rely on the alternative instrument vector

\[
\tilde{z}_i = \begin{bmatrix}
ppse_{i1} & x'_{i1} & u'_{i1} & ppse_{i2} & x'_{i2} & u'_{i2} & \cdots & ppse_{iT} & x'_{iT} & u'_{iT} & 1
\end{bmatrix}',
\]

(17)

where \( \tilde{z}_i \) has size \( ([T(L_x + L_u) + 1] \times 1) \). We find that the small-sample biases in the parameter estimates remain numerically about the same as in Table 1 (2), yet the efficiency gain of the EFE relative to the FE estimator is even more pronounced. In particular, the standard deviation of the bias is approximately 0.010 (0.012) for the FE and only 0.003 (0.007) for the EFE. Finally, at a nominal test size of 5\% the actual sizes of the \( J \)-test are 2.51\% (2.56\%) for FE, 3.18\% (2.65\%) for EFE with \( C = C_1 \), and 3.14\% (2.64\%) for EFE with \( C = C_2 \), when adjusting the degrees of freedom of the \( \chi^2 \)-distribution accordingly\(^8\).

More importantly, the usual worry about “useless” instruments is that they negatively affect the statistical power of identification test (Roberts and Whited, 2013). Combining the two simulation designs described in this section, we investigate the power of the \( J \)-test with “useless” instruments against a dynamic model. We find that the \( J \)-tests corresponding to

\(^8\) Note that in this case, we have \([T-1]/[T(L_x + L_u) + 1] \) moment conditions to identify \( L_x \) parameters.
all estimators have a perfect rejection rate of 100% (100%). Thus, even in this challenging scenario the misspecification tests proposed here are a powerful tool to detect violations of the assumptions.

IV Data Description

4.1 Data Description

To construct our sample, we obtain CEO compensation data from 2005 to 2018 from ExecuComp\textsuperscript{9}. Over the same period, each firm’s stock price at fiscal year-end is obtained from CRSP. All other firm-specific variables (e.g. Tobin’s Q, research and development expenses, advertising expenses, total debt, total assets and property, plant and equipment) are from Compustat. There are slightly more than eight hundred unique firms in the sample for the 14 years 2005-2018, and almost one and a half thousand unique CEOs.

4.2 Variables Used

4.2.1 Firm value

We follow the empirical corporate finance literature and proxy for firm value by using Tobin’s Q. As in Smith Jr and Watts (1992) and Shin and Stulz (1998), we calculate Tobin’s Q as the ratio of the market value of equity minus the book value of equity plus the book value of assets to the book value of assets.

4.2.2 CEO pay-performance sensitivity

Following Core and Guay (2002) and Coles et al. (2006), we calculate the CEO incentive measure \( ppse \) as the dollar change in CEO wealth for a one percentage point change in firm stock price.

\textsuperscript{9} Hayes, Lemmon and Qiu (2012) find that the regulatory changes in stock option expensing mandated by FAS 123R in 2005 resulted in a substantial decrease in options usage. Accordingly, we begin our sample in 2005.
4.2.3 Control variables

Based on the prior literature, we create a number of firm characteristic variables that might affect firm value. Following Morck et al. (1988), and McConnell and Servaes (1990), we first include research and development expenses (xrd) and advertising expenses (adv) as proxies of future growth opportunities. We define xrd and adv as the ratio of research and development expenses to total assets, and the ratio of advertising expenses to total assets, respectively. Note that Compustat has a fair amount of missing values for these two intangible assets. We also control for the debt equity ratio (debt) which has been shown to be beneficial for firm value. We define debt as the book value of total debt to book value of assets. Our measure of firm size (lnassets) is defined as the natural logarithm of total assets. Finally, we include the variable prop, defined as the ratio of property, plant and equipment to book value of assets. Detailed definitions of all variables constructed in the paper are given in Table 3.

[ Insert Table 3 ]

4.3 Summary Statistics

Table 4 presents the summary statistics of all variables used in our empirical analysis. We winsorize Tobin’s Q and two CEO pay-performance measures at the 1st and 99th percentiles. Our sample of firms has an average Tobin’s Q of 2.262 and a median value of 1.852. The CEO’s pay-performance sensitivity variable, ppse, has an average value of 654.642 and a median value of 199.328. The sample firms have an average of 7.370 of lnassets, a mean ratio of research and development to assets of 0.051, and an average ratio of advertising to assets of 0.031. We find the average debt-to-equity ratio in our sample is 0.214, with a median value of 0.175. On average firms have 20.162% of total assets in property, plants and equipment.

[ Insert Table 4 ]
V Empirical Results

5.1 Variance Decomposition

In Table 5, we examine the within-firm (time-series) and between-firm (cross-sectional) variation for all our variables. For Tobin’s Q \( q_{it} \), the proportion of total variation that is given by within-firm is 36.922%, and the between-firm variation is 63.078%. For the CEO pay-performance sensitivity measure \( ppse \), we find a similar one-to-two ratio: 33.448% and 66.552% are due to the time-series and the cross-sectional dimension, respectively. Since a non-negligible part of the variation in the variable of interest comes from the time-series dimension, we expect the effect of \( ppse_{it} \) on \( q_{it} \) to be estimated well by techniques that explicitly exploit within-group variation over time, such as the fixed-effects estimator. However, for firm characteristic variables such as research and development expenses \( xrd \), advertising expenses \( adv \), total assets \( lnassets \) and property, plant and equipment \( prop \), most of the variation come from cross-sectional variation.

[ Insert Table 5 ]

5.2 Estimation Results

We now use the different estimators presented to estimate the effect of CEO pay-performance sensitivity on firm value. We start the discussion with the “standard” estimators that have been used in the literature, to set a benchmark. Next, we analyze the results from the new efficient fixed-effects estimator (EFE).

We start by looking at the most naïve estimator, the pooled OLS (POLS). The results are given in Table 6. We find that the coefficient on \( ppse \) is positive and statistically significant at the one-percent level, suggesting that higher CEO incentives result in higher firm value. This effect is also economically important: a one standard deviation increase in \( ppse \) would result in a 0.273 standard deviation increase in firm value. This corresponds, for instance,
to a jump from the median to the 64th percentile of the distribution of Tobin’s Q. For a distressed firm at the fifth percentile of the firm-value distribution, the same increase would imply a large leap to the 24th percentile.

[ Insert Table 6 ]

We also find that the intangible firm-specific variables \( xrd \), \( adv \), \( debt \), and \( prop \) are positively related to firm value, although only the former is statistically different from zero. Furthermore, a reduction in firm size seems beneficial with respect to firm value.

Next, we run regressions of firm value on the CEO pay-performance sensitivity measure using the standard random-effect estimator and report the results in Table 6. We find that \( ppse \) is positively related to firm value. The effect is strongly statistically significant and it is economically important. A one standard deviation increase in \( ppse \) would result in a 0.276 standard deviation increase in Tobin’s Q. This numbers correspond to a leap from the median to the 64th percentile, or from the fifth to the 24th percentile, in the firm value distribution. Thus, the estimated impact of the variable of interest on firm value based on POLS and RE is almost identical.

Among the firm-specific variables, we also find that \( xrd \) is significantly positively related to firm value, whereas \( adv \) produces a positive effect that is statistically insignificant. Conversely, \( lnassets \), \( debt \), and \( prop \) are negatively related to firm value, yet only the former is significantly negative.

Table 6 also summarizes the estimation results from the first-difference estimator (FD). The main variable of interest, \( ppse \), is positively and statistically significantly related to firm value. The magnitude of the estimated impact is, once again, very similar to the two previous estimators, albeit marginally larger. If \( ppse \) increased by one standard deviation, it would result in a 0.284 standard deviation increase in Tobin’s Q, or a leap from the fifth to the 24th percentile for distressed firms. Interestingly, FD estimation further suggests that all firm-specific variables are statistically significant; \( xrd \), \( adv \), \( lnassets \), and \( prop \) seem to
have a positive effect, and \textit{debt} has a negative impact on firm value.

In Table 7, we report our estimation results for the effect of CEO pay-performance sensitivity on firm value using the standard fixed-effects (FE) estimator. We still find the statistically strong and positive association between the CEO pay-performance sensitivity measure, \textit{ppse}, and firm value. The economic impact is again almost identical to the three previously discussed estimates. A one standard deviation increase in \textit{ppse} would result in a 0.269 standard deviation increase in firm value; these numbers represent an increase from the median to the 63\textsuperscript{rd} percentile of the Tobin’s Q distribution. Correspondingly, it would boost a distressed firm from the fifth to the 23\textsuperscript{rd} percentile. The effect of \textit{xrd}, which before was found to be positive, is no longer statistically relevant. The same is true for the impact of \textit{adv} and \textit{debt}. The negative relation between \textit{lnassets} and firm value found with POLS and RE estimation still holds, and \textit{prop} also negatively impacts Tobin’s Q, albeit only marginally significantly so.

[ Insert Table 7 ]

To summarize our findings so far, all of the benchmark estimators that are commonly used in the literature would find a statistically and economically important positive relation between the sensitivity of the CEO towards performance pay and firm value. The estimated impacts are also very similar in magnitude. Optimizing the CEO’s compensation structure in the sense that they are given a larger monetary incentive for performance could seemingly substantially increase firm performance. This is particularly true for firms that are in relative distress.

Conversely, if we estimate the effect of CEO pay-performance sensitivity on firm value using the efficient fixed-effects estimator (EFE), we can paint a very different picture for the firms in our sample. The results are in Table 8. While still statistically significant, the estimated impact of \textit{ppse} on Tobin’s Q is substantially smaller. Efficient estimation thus suggests that improving CEO’s pay-performance contracts further does not have a large
impact of firm value. When relying on the demeaned EFE estimator \((C = C_1)\), we find that a one standard deviation increase in \(ppse\) results in a 0.154 standard deviation increase in Tobin’s Q. This number is only approximately half as large as the estimated impact with the benchmark estimators. The estimated impact becomes even smaller when we estimate the EFE with the data in first-differences \((C = C_2)\). We find that a one standard deviation increase in \(ppse\) result in a 0.137 standard deviation increase in firm value, which is only 49%-50% of the magnitude estimated by the benchmark estimators. A distressed firm at the fifth percentile of the Tobin’s Q distribution will only leap to the 13\(^{th}\) – 14\(^{th}\) percentile, in response to such a increase in \(ppse\).

[ Insert Table 8 ]

Interestingly, the two EFE estimators disagree of the impact of \(xrd\) on Tobin’s Q. In addition, \(lnassets\) and \(prop\) seem to negatively affect Tobin’s Q. The impacts of \(adv\) and \(debt\) are positive and significant at a one-percent level. Note that all effects stay strongly significant, even if we increase the standard errors by adjusting them for the degrees of freedom.

5.3 The Dominant Estimator

Given the different estimation outcomes reported in the previous section, we now evaluate the adequacy of the respective estimation techniques for our data set. Again, we start the analysis with the most naïve estimators, the POLS and the RE, and end with the new efficient (EFE) estimator. The goal is to find the dominant estimator for our data set.

As we demonstrate in Appendix A, the two estimators POLS and RE are the same if the 
\((T \times T)\)-variance-covariance matrix \(\hat{\Sigma}\) of the regression residuals \(\hat{\varepsilon}_{it}\) is a diagonal matrix with identical diagonal elements. This is true if the residuals have no temporal correlation and a time-constant variance. The Lagrangian Multiplier (LM) test of Breusch and Pagan (1980), and refined by Baltagi and Li (1990) for unbalanced panels, is designed to investigate
this hypothesis. More precisely, the LM test evaluates the null hypothesis that the variance of $c_i$ is zero. In the literature it is common to rely on this test, and hence we follow this approach, even though it relies on some assumptions that may be too restrictive given our setup\textsuperscript{10}. We report the outcome of the test in Table 6. The LM-statistic is equal to 8,833.280. Thus, based on $\chi^2(1)$-inference the null hypothesis is rejected at the one-percent significance level. The results suggest that the random-effects is a more efficient, and hence preferable, estimator than the POLS (if $c_i$ is exogenous).

In a second step, we compare the adequacy of the FE and the FD estimators to the two initial estimators, the POLS and the RE. The $F$-test reported in Table 7 evaluates the null hypothesis that the unobserved fixed firm-effects are constant, i.e. $c_i = c$. If this were true then the endogeneity of the system would vanish. In this situation, POLS would be preferable over FE since in the former approach no information is discarded. We strongly reject this hypothesis, however, with an $F$-statistic of 135.850. Since there is no evidence that $c_i$ is constant, POLS is also not likely to dominate FD.

We then conduct a Hausman specification test (Hausman, 1978) to compare the standard fixed-effects (FE) and the random-effects (RE) estimator. Under the null hypothesis that the unobservable fixed firm effects, $c_i$, are exogenous, both the RE and the FE estimators are consistent, yet RE is more efficient (this follows directly from Appendices A and C). The corresponding statistics are reported in Table 7 and the null hypothesis is rejected at the one-percent significance level.\textsuperscript{11} Note that this finding can further be interpreted as implying a preference of FD over RE, since only the former is consistent in the presence of endogenous $c_i$ (see Appendix B).

Now, we conduct a $J$-test for the assumptions behind the fixed-effects estimator, which

\textsuperscript{10} Recall that the error term of the composite model is $\epsilon_{it} = c_i + \eta_{it}$. The LM test is derived under the assumption of error normality, and - more importantly - of no serial correlation or time-varying variance in $\eta_{it}$. Thus, in the view of the LM test, the only cause of non-spherical composite error variances could be the variance of $c_i$. The null hypothesis $\text{Var}(c_i) = 0$ guarantees that $\epsilon_{it}$ is serially uncorrelated and has time-constant variance, in which case the POLS is the preferred estimator (and the efficient one if, additionally, $c_i$ were exogenous).

\textsuperscript{11} Note that the Hausman test relies on the assumption of conditional homoskedasticity (across $i$).
are the same as the underlying assumptions of the first-difference estimator. The value of the $J$-statistic in Table 7 is 718.369. We fail to reject the overidentification assumptions even at conservative significance levels, with a $p$-value of 0.510. This indicates that the required exogeneity of the covariates with respect to current, future, and past idiosyncratic shocks is satisfied by the data.

The results so far suggest that the FE and the FD estimators are the most appropriate one among the above-mentioned four standard panel estimation methods. The evidence of the hypothesis tests confirms our assumption that the unobserved firm fixed-effect seems to be present as well as endogenous. This leaves the POLS and RE estimators inconsistent. Idiosyncratic shocks seems to be exogenous, however, which suggests consistency of the FD, the FE, and the EFE estimator.

Next, we contrast the FD and the FE estimators, with the EFE estimators. As demonstrated in Appendices C and D, the EFE estimator is more efficient than the FE estimator if (a) the idiosyncratic-shock variance across firms is (conditionally) heteroskedastic, (b) if the latter is the same or constant over time, and (c) if there is no time-series correlation in the idiosyncratic errors $\eta_t$.\footnote{Note that the LM test that compares the POLS and the RE test above, can be viewed as a test of implications (b) and (c). Yet, its inference is based on several unrealistic assumptions.} Similarly, EFE is more efficient than the FD estimator, if assumptions (a) to (c) hold for transformed shocks, $\Delta \eta_t$, as shown in Appendix B.

We evaluate implication (b) by extracting the FE residuals $C_1'\hat{\eta}_t$ ($C_2'\hat{\eta}_t$) for every year separately, and then testing for equal variances across the series\footnote{From the derivations in Appendix C it follows that the variances of $C_1'\eta_t$ should equal $k(T-1)/T \forall t = 1, 2, \ldots, T-1$, and the variances of $C_2'\eta_t$ should equal $2k \forall t = 1, 2, \ldots, T-1$, where $k$ is some constant, for efficiency of the FE estimator.}. The Bartlett test for equality of variances has 12 degrees of freedom (d.o.f), the value of the test statistic is 271.036 (294.029), and its $p$-value is 0.000 (0.000). Thus, we strongly reject the equality of variances. Similarly, the Levene test with 12,10504 d.o.f. has a test statistic of 3.438 (5.738) with corresponding $p$-value of 0.000 (0.000). Lastly, the Brown-Forsythe test has 12,10504 d.o.f., a value of 3.438 (5.738) and $p$-value of 0.000 (0.000). Hence, variances are
not constant over time. For the FD residuals, \( C_2' \hat{\eta}_i \), the value of the three respective variance test statistics is 356.873, 5.200, and 5.408. We reject the time invariance of variances strongly with corresponding \( p \)-values of zero.

Next, we evaluate assumption (c) that there is no residual correlation over time. More precisely, for the efficiency of the FE estimator, it should hold that \( \eta_i \) is uncorrelated across time. Our FE estimation residuals are estimates for \( C' \eta_i \), however. From the derivations in Appendix C it can be deduced that for the efficiency of the FE estimator the autocorrelations of \( C_1' \eta_i \) would have to equal \(-1/(T-1) < 0\) at all lags, and the autocorrelations of \( C_2' \eta_i \) would have to equal \(-1/2\) at lag 1, and 0 at all higher lags. To check assumption (c), we analyze the \([T-1] \times [T-1]\) correlation matrix of the extracted FE residuals \( C_1' \hat{\eta}_i \) (\( C_2' \hat{\eta}_i \)). The smallest estimated correlation over all permutations is -0.475 (-0.438) and the largest is 0.592 (0.122). Out of 78 autocorrelation estimates from \( C_1' \hat{\eta}_i \), a simple \( t \)-test for the null hypothesis that the correlation is zero against the alternative hypothesis that it is negative rejects only in 45 of the cases at a one percent significance level, in 47 of the cases at a five percent level, and in 48 of the permutations at a ten percent level. Hence, roughly 40% of the autocorrelations are not statistically negative, which violates assumption (c) for the FE estimation with \( C = C_1 \). To test assumption (c) for the FE errors \( C_2' \eta_i \), we consider the autocorrelation estimates for lags 2 and higher. Out of the 66 such autocorrelation estimates, a simple \( t \)-test for the null hypothesis that the correlation is zero fails to rejects only in 43 of the cases at the standard 5% significance level. Thus, approximately 35% of the autocorrelation estimates violate assumption (c) for the FE estimation with \( C = C_2 \). If instead we analyze FD residuals, \( C_2' \hat{\eta}_i \), assumption (c) would require the absence of any serial correlation at all lags. However, we find a statistically significant non-zero autocorrelation estimate in about 32% of all permutation, suggesting that also FD estimation does not fulfill this efficiency requirement.

As a last step of the analysis, we note that above FE regression may also be viewed as an OLS regression of \( \mathbf{Mq}_i \) on \( \mathbf{MF}_i \). Similarly, FD regression is an OLS regression for
transformed dependent variable $C_2'q_i$ and independent variable $C_2'F_i$. The residuals of such pooled regressions are standard uncorrelated OLS errors that we subject to White’s test for heteroskedasticity (White, 1980). We obtain a $\chi^2(21)$-test statistic of 854.645 with a corresponding $p$-value of 0.000 for the FE estimation. For the FD regression the corresponding test statistic equals 1828.522, and the $p$-value is zero. We strongly reject the null hypothesis of conditional homoskedasticity, which confirms that condition (a) is also satisfied by the data.

We conclude that the EFE estimator is the preferable estimator among the estimators analyzed. As residuals are neither serially uncorrelated, nor have a time-constant variance, nor are conditionally homoskedastic, the EFE is efficient. The $J$–test applied to the EFE estimation results in Table 8 supports the assumptions of the model and fails to reject the overidentification restrictions, even at a conservative significance level of 10%.

Finally, we compare the two versions of the EFE estimators with one another. The results of the EFE estimations differ, depending on whether we rely on the EFE in first-differences with $C = C_1$, or on the demeaned version with $C = C_2$. In order to determine which estimator is preferable, we rely on the $R^2$ measures reported in Table 8. These values should be interpreted with care, since in the GMM context the total sum of squares is not necessarily equal to the sum of the explained and the residual sum of squares. The reported values all are pseudo $R^2$ values, obtained as the squared correlation between the observed untransformed Tobin’s Q measure, $q_i$, and the corresponding fitted values obtained from the estimates. We observe that the estimates obtained with $C = C_1$ correlate substantially more with the dependent variable, than predicted values from $C = C_2$. For this particular data set, we therefore conclude that demeaned EFE estimation is preferable and dominates all other estimation techniques considered.
5.4 Robustness: CEO Fixed-Effect

In Section II, we argued that the unobserved fixed-effect, $c_i$, absorbs time-invariant heterogeneity across firms, but alternatively also heterogeneity across superordinate groups such as industries. What happens, however, if there is a fixed-effect across CEOs? CEOs may be viewed as a subordinate level relative to firms; one firm may have more than one CEO over time. In its current form, our model does not permit such an effect, and the estimators are hence not robust to its potential presence.

For robustness, we therefore repeat our estimations at the CEO level, focusing only on the main estimators of interest: The FE and the EFE. More precisely, let $i = 1, 2, \ldots, N$ denote the $i$th CEO. Now $c_i$ denotes the CEO fixed-effect. Note that $c_i$ could alternatively capture the firm fixed-effect (or a fixed-effect at an even higher aggregation level), as long as the same CEO does not head two or more different firms within the sample period $T$.

[ Insert Table 9 ]

Table 9 shows that the results remain largely unaltered. We find that a one standard deviation increase in $ppse$ results in a predicted 0.351 standard deviation increase in firm value, when we estimate by standard fixed-effects. The corresponding economic magnitude of the estimated impact with the demeaned EFE estimator ($C = C_1$) is 0.222. Similarly to our previous results, the EFE effect is substantially smaller, amounting only to roughly 60% of the corresponding FE impact. With the EFE estimator using $C = C_2$, the estimated standard-deviation change in Tobin’s $Q$ triggered by a one-standard deviation increase in $ppse$ is also small, equal to 0.237. The estimated impacts are overall economically marginally larger than the analogous firm-level estimates, but as in the previously obtained results, we find that the efficient estimators produce a significantly smaller effect of CEO pay on Tobin’s $Q$ relative to FE.

As before, we extract the residuals $C'_1 \hat{\eta}_i$ ($C'_2 \hat{\eta}_i$) from the FE estimation results to test the conditions under which the EFE estimator is more efficient than the FE estimator. Using the
Bartlett test, the Levene test, and the Brown-Forsythe test for equality of variances across time, we find respective test statistics of 380.924 (415.293), 5.060 (5.149), and 5.060 (5.149) respectively, and all corresponding p-values are equal to zero, i.e. we strongly reject. Next, we test whether there is residual correlation over time. We do not find a statistically negative autocorrelation at the 1%, 5%, and 10% significance level, for 32%, 26%, and 26% of the respective autocorrelations of the FE residuals $C_1 \hat{\eta}_i$. Further, we reject the hypothesis that the residual $C'_2 \hat{\eta}_i$ at lags $j \geq 2$ is not autocorrelated at the 1%, 5%, and 10% significance level in a large number of cases, more precisely for roughly 36%, 52%, and 53% of the possible annual correlations, respectively. Lastly, we carry out a White test for heteroskedasticity and find a test statistic of 2,409.032. With corresponding p-values of zero, we confirm the presence of conditional heteroskedasticity. Hence, the EFE dominates the FE estimator.

Finally, similarly to our results for the firm-level estimates, we find evidence for the necessary exogeneity of the idiosyncratic errors $\eta_i$. The $J$-tests summarized in Table 9 all fail to reject the overidentification restrictions, with p-values ranging from 0.24 to 0.35. The pseudo $R^2$ values in Table 9 suggest that the demeaned EFE estimation is to be preferred over EFE estimation with data in first-differences. All in all, the EFE with $C = C_1$ is thus the preferred estimator to use for our data, when modeling at CEO level, as well as at firm level.

5.5 Robustness: External Instrument

In recent years, there has been a growing concern in the empirical corporate finance literature that causal effects cannot be identified without relying on exogenous shocks or institutional discontinuities to define instrumental variables (IVs). While we agree that endogeneity is a valid concern, we do not believe that this is a “one size fits all” solution. In many cases, it may still be possible to estimate the ceteris paribus effect of a variable on the dependent variable in a “standard” panel regression, when using the correct techniques and carefully implementing the approach.
One prominent example of the literature that uses institutional compensation plan characteristics to define IVs is Shue and Townsend (2017), who argue that CEO pay performance sensitivity is endogenous in explaining firms’ volatility, leverage, and investment. As demonstrated in the previous sections, when estimating the effect of \( ppse \) on \( q \) by POLS, RE, FD, FE, and the new EFE estimator, one is implicitly using the current values of \( ppse \) as well as its leads and lags as IVs. Thus, the suggested endogeneity of Shue and Townsend (2017) would invalidate any of the estimation techniques presented. In the following, we show that our estimation results remain valid.

First and foremost, all \( J \)-tests presented in the previous sections fail to reject the model specification. That is, there is not enough sample evidence to doubt the joint null hypothesis, which includes the restriction that the idiosyncratic shocks are orthogonal to all past, current, and future values of \( ppse \).

Secondly, we repeat Shue and Townsend’s (2017) analysis with our data. The cross-sectional observation unit, \( i \), is the CEO. We implement the IV strategy-I and strategy-II of Shue and Townsend (2017), which rely on two types of multiyear CEO compensation plans: fixed-number (FN) plans and fixed-value (FV) plans. In particular, under strategy-I, the authors suggest to instrument the supposedly endogenous pay-performance variable with a variable that predicts the first year of a new FV-compensation contract period, \( IV^{(1)}_{it} \). As only CEOs on FV plans are considered under strategy-I, the sample size is significantly reduced to 915 CEO-year observations (over the same \( T = 14 \) years 2005-2018). Strategy-II looks only at participants of FN or FV plans (with some additional restrictions), leading to a small sample of 508 CEO-year observations. Three IVs are considered in strategy-II: A dummy variable identifying FN-plan members, the industry return over the previous year, and the product of the previous two. Let these three IVs be summarized in the vector \( IV^{(2)}_{it} \).

For the strategy-I sample (strategy-II sample), we start by re-estimating our model by “standard” FE. The FE estimate for the effect of \( ppse \) on \( q \), \( \hat{\alpha} \), is equal to \( 2.343 \times 10^{-4} \) (\( 2.209 \times 10^{-4} \)), which is quite a bit smaller than the entire sample FE estimate, but marginally
larger than whole sample EFE estimate. The effect is significantly positive with a corresponding p-value of $4.527 \times 10^{-4}$ ($1.439 \times 10^{-4}$). The $J$–statistic from the FE estimation equals 256.988 (165.3458). With a 5% critical value of 1,177.236, we again have no reason to doubt our modeling assumptions, including the orthogonality of past, current, and future $ppse$. We also estimate our benchmark model by EFE with $C = C_1$ for the reduced sample. The estimate for $\alpha$ is $1.746 \times 10^{-4}$ ($1.821 \times 10^{-4}$), the corresponding standard error is $3.907 \times 10^{-7}$ ($1.242 \times 10^{-6}$), and the $J$-statistic is 257.912 (164.756). Similarly, if we instead estimate by EFE with $C = C_2$ we find the estimated impact of $ppse$ on $q$ to equal $1.829 \times 10^{-4}$ ($1.783 \times 10^{-4}$), the corresponding standard errors are $4.511 \times 10^{-7}$ ($1.040 \times 10^{-6}$), and $J = 257.981$ ($J = 164.404$). Thus, just like in the full sample, EFE estimation produces a substantially smaller, albeit still significantly positive, estimated effect than FE estimation. If we compare the $J$-statistics form the EFE estimation to the critical value above, we still have no reason to believe that anything (besides the fixed effect $c_i$) is endogenous in our setup. Nevertheless, in the next step we add the contemporaneous Shue and Townsend (2017) IVs to our set of instruments and re-estimate efficiently. The addition of $IV_{it}^{(1)}$ ($IV_{it}^{(2)}$) does not reveal any new or different information; the modified EFE with $C = C_1$ estimate for $\alpha$ is $1.744 \times 10^{-4}$ ($1.858 \times 10^{-4}$), and hence almost identical to the number above. Similarly, the modified EFE with $C = C_2$ estimate is $\hat{\alpha} = 1.804 \times 10^{-4}$ ($\hat{\alpha} = 1.774 \times 10^{-4}$), and all estimates are strongly statistically significant. The $J$-statistic for $C = C_1$ is 258.000 (164.459), and for $C = C_1$ it is 256.951 (165.070). The corresponding 5% critical value is 1,190.690 (1,217.592), and hence we again fail to reject, which suggests that the IVs are valid.

We draw four conclusions from our replication of the Shue and Townsend (2017) IV estimation approach. First, all estimates for the effect of $ppse$ on $q$, $\hat{\alpha}$, are substantially smaller when estimated for the Shue and Townsend (2017) subsamples than when taking the entire sample into account. Second, the pattern that EFE estimation produces noticeably smaller point estimates for $\alpha$ than FE estimation is robust also in the subsamples. Third,
including the Shue and Townsend (2017) IVs does not seem to alter our conclusions, or provide new insights. Fourth, the Shue and Townsend (2017) IVs are valid instruments.

Since both IV strategies of Shue and Townsend (2017) only consider CEOs that are on certain specific multiyear compensations plans, we discard a large amount of data. Given that the estimates for $\alpha$ in these subsamples differ considerably from the ones in Table 9, we now investigate the representativeness of the subsamples by comparing full-sample and subsample summary statistics. The average Tobin’s Q in the full sample is 2.280 and in the subsamples for IV strategy-I and strategy-II it is 2.340 and 2.340, respectively. The relative sample dispersion of $q$, measured by the coefficient of variation (CV=average/standard deviation), equals 59.863% for the full sample; 53.543% and 53.019% for the two subsamples. Whereas the distributional properties of $q$ thus seem to remain relatively stable throughout the different samples, the opposite is true for our main explanatory variable of interest. The average of $ppse$ in the full sample is 656.843. Based on the full-sample statistics, the 99% confidence interval for the mean of $ppse$ is [612.200, 701.487]. The averages of $ppse$ in the strategy-I and strategy-II samples, which equal 1,068.441 and 1,171.920, are not anywhere close to the interval. Similarly, the dispersion in $ppse$ differs widely, with a CV of 223.342% for the full sample, but only 175.768% and 183.932% for the subsamples of strategy-I and strategy-II. We conclude that the subsamples do not seem representative for $ppse$.

Above, we find that the Shue and Townsend (2017) IVs are valid, but as a final step of this analysis we want to examine the relevance of $IV_{it}^{(1)}$ and $IV_{it}^{(2)}$ as instruments for $ppse$. A decomposition of variances leads to some initial doubts as to the “commonalities” between regressor and instruments. In the strategy-I and strategy-II subsamples, the within-CEO variation is cut down to 19.2-20.0%, whereas the between-CEO variation makes up for 80.0-80.8% of the total subsample variation in $ppse$. Conversely, the IVs seem to have most of their variation in the time-series dimension. 90.5% of the variation in $IV_{it}^{(1)}$ is within-CEO variation; for the interaction instrument of strategy-II the corresponding percentage is 67.5%. For a more formal test of IV relevance, we return to using the entire data sample.
We construct two dummy variables, $D_{it}^{(1)}$ and $D_{it}^{(2)}$, that equal one if an observations would have been included in the strategy-I and strategy-II subsample, respectively. We replicate the “first-stage” regression of Shue and Townsend (2017), by regressing $ppse_{it}$ on its own future and past values, on current, future, and past realizations of the other explanatory variables, on the four Shue and Townsend (2017) IVs, $IV_{it}^{(1)}$ and $IV_{it}^{(2)}$ (missing values are replaced by zeros), and lastly also on $D_{it}^{(1)}$ and $D_{it}^{(2)}$. We estimate this regression by FE and FD, including a constant term, year fixed-effects, and CEO fixed-effects. We then conduct a Wald test for the hypothesis that $IV_{it}^{(1)}$, $IV_{it}^{(2)}$, $D_{it}^{(1)}$, and $D_{it}^{(2)}$ are jointly insignificant. The Wald statistic resulting from FE estimation is 9.115, and it is equal to 1.505 from the FD estimation. With a critical value of 10.645, we fail to reject the irrelevance of the instruments (and dummies) even at a conservative level of 10%. Conversely, a test of the null hypothesis that all remaining right-hand-side variables are insignificant, results in a Wald statistic of 346.676 for FE estimation and 548.082 for FD estimation. The corresponding critical value at a 1% significance level is 122.94; thus we strongly reject.

In summary, our estimation methodology works optimally in the data. In contrast, using compensation plan characteristics to aid the identification of the effect of pay-performance sensitivity on firm value, produces valid but irrelevant (or weak) instruments. In addition, this IV approach requires a very specific sample selection that is not necessarily representative of the statistical population, and the resulting small sample size is not optimal for statistical inference.

VI Extension

Up until this point, we ruled out dynamic models. In this section, we propose a straightforward extension of the fixed-effects and the efficient fixed-effects estimators that is robust to dynamic models of the form (15).

In section 3.6, we show that FE and EFE are inconsistent if the researcher omits the lags
of the dependent variable, $Q_i(-)$, from the set of regressors. However, if the researcher realizes that the lags of the dependent variable contribute importantly to the dependent variable, she can easily include them in the estimation\(^{14}\). The transformed model

$$C'q_i = C'X_i\delta + C'Q_i(-)\phi + C'\varepsilon_i = C'F_i\gamma + C'Q_i(-)\phi + C'\eta_i,$$

(18)
can then be estimated by GMM, using $z_i$ from (4) as instruments. The resulting estimators, the dynamic fixed effects estimator (DFE) and the dynamic efficient fixed effects estimator (DEFE) rely on the same weighting matrices as defined in Theorems 3 and 4 in Appendices C and D, respectively. Both estimators are consistent for the parameter vector $[\gamma, \phi]'$, as long as the maintained assumptions hold. If the data is conditionally heteroskedastic across $i$, with serially correlated errors and non-constant residual variance, then only DEFE is efficient, for model (18) and instruments $z_i$.

It is important to note that the elements of the matrix of lagged values, $Q_i(-)$, may not be included in the set of instruments. The latter are endogenous and thus not valid instruments, because $E(q_i(t-p)\eta_{is})$

$$= E\left(\frac{\psi}{1 - \sum_{p=1}^{P} \phi_p} + \sum_{j=0}^{\infty} ppse_i(t-p-j)\alpha\theta_j + \sum_{j=0}^{\infty} x'i(t-p-j)\beta\theta_j + \tilde{c}_i + \sum_{j=0}^{\infty} \eta_i(t-p-j)\theta_j \right) \eta_{is}$$

$$= E\left(\tilde{c}_i + \sum_{j=p}^{\infty} \eta_i(t-j)\theta_{j-p} \right) \eta_{is} \neq 0 \ \forall \ s, t = 1, 2, \ldots, T \ \text{and} \ \forall \ p = 1, 2, \ldots, P,$$

(19)

where the second row in equation (19) assumes that Assumption 3 continues to hold. Even considering transformed lagged values, $C'Q_i(-)$, as part of the instrument vector, will still lead to a violation of the orthogonality assumption.

In defining the dynamic model in equation (15), we specified $T$ as the effective number of time series observations, that is $T = (\text{total # of time-series observations} - P)$. The DFE and the DEFE estimate $L_x + P$ parameters using $(T - 1)(TL_z + 1)$ moment conditions.

\(^{14}\) Lag selection can be done, e.g. by one of the popular information criteria, AIC or BIC.
A necessary condition for overidentification, and thus the use of the \( J \)-statistic, is that 
\[ (T - 1)(TL_x + 1) > L_x + P. \]

In Table 10 we report simulation results for the DFE and the DEFE estimator. The simulation design is the same as for the dynamic models in section 3.6. The bias in the parameter estimates for the DFE and the DEFE is small and comparable to the FE and the EFE estimators. Similarly, the uncertainty in the DFE and the DEFE is small, albeit slightly higher in magnitude than the standard deviation of the sampling error of the FE and the EFE, respectively. This is to be expected, as the DFE and DEFE are estimating more parameters than FE and EFE, but they rely on the same number of identification conditions. Importantly, the standard deviation of DEFE estimators is always visibly smaller than the corresponding value from DFE estimation, suggesting that the latter is inefficient. Lastly, the size of the \( J \)-test is similar to previous findings, exhibiting a small degree of undersizing for samples of size \( N = 1,000 \).

Figures 2, 3, and 4 plot the finite-sample behavior of the relative bias, the estimation uncertainty, and the size of the \( J \)-test, respectively. The bias in the pooled estimate for \([\alpha, \beta, \phi]^\prime\) shrinks very quickly to zero as \( N \) increases, especially for the DEFE estimation. Whereas the variability of the DFE pooled estimator is relatively big, the DEFE estimation is very efficient. Lastly, the size of the \( J \)-test based on DEFE estimation approaches the nominal size of 5\% very quickly, whereas \( N \) needs to grow quite large to correct the undersizing in the \( J \)-test that corresponds to DFE estimation.

As a final note, it is necessary to point out that the suggested DFE and DEFE estimators are not the same as the Arellano-Bond estimator (Arellano and Bond, 1991), or the Blundell-Bond estimator (Blundell and Bond, 1998). The estimators differ both, in terms of the instruments used as well as the required assumptions. Whereas the Arellano-Bond estimator

\[ E(\eta_i [ \tilde{p} \tilde{p}_{it} \tilde{X}_i \tilde{q}_{it(t-1)} \ldots \tilde{q}_{it(t-P)} ]), \]

where \( \tilde{\cdot} \) denotes a variable that has been pre-multiplied with the matrix \( C^\prime \), must be of rank \( L_x + P \) for some \( t \) (see Hayashi, 2000).

---

\[ \text{A sufficient condition for identification is that the matrix } E(\eta_i \left[ \begin{array}{c} \tilde{p} \tilde{p}_{it} \tilde{X}_i \tilde{q}_{it(t-1)} \ldots \tilde{q}_{it(t-P)} \end{array} \right]), \]

where \( \tilde{\cdot} \) denotes a variable that has been pre-multiplied with the matrix \( C^\prime \), must be of rank \( L_x + P \) for some \( t \) (see Hayashi, 2000).
relies on the lags of $q_i$ and $X_i$ as instruments, and the Blundell-Bond estimators adds lagged differences of the variables to the set of instruments, the DFE and the DEFE use only the instruments $z_i$, which are comprised of all elements of $X_i$. Furthermore, the Arellano-Bond and the Blundell-Bond estimator are derived under the assumption that the errors $\eta_i$ are serially uncorrelated with a constant variance. In contrast, when using the DFE and the DEFE estimators, these restrictions may be violated, which is in line with the properties of our empirical data set.

VII Conclusions

The use of panel data wherein the researcher examines data across many firms and few years is common practice in corporate finance. This paper evaluates the properties of several regression estimators when one uses panel data. We begin by presenting a unified econometric regression model in which we address the appropriateness of various assumptions that are implicitly underlying the various estimation techniques. Specifically, we analyze four existing regression estimators (POLS, RE, FD, and FE), and introduce an additional estimator (EFE). Theoretically and in simulations we find that the EFE estimator is consistent and has the lowest uncertainty, the FD and FE estimators are consistent but have high uncertainty, and the POLS and RE estimators are inconsistent. In an extension, we also suggest two estimators for dynamic corporate-finance panels: the DFE and the DEFE.

When we use empirical panel data to examine the relationship between CEO pay and firm value, we find the EFE estimator to be the most appropriate, followed by the FD or FE estimator. The data also suggests that the assumptions for consistency of the POLS and RE estimators do not hold. For a one standard deviation increase in the CEO’s pay-performance variable, we find that the preferred EFE estimator shows a very small economic impact of 0.15 of CEO pay on firm value, whereas the FD and FE estimators produce an effect that is about twice as large. In summary, we suggest the use of the EFE estimator when one
examines the relationship between CEO pay and firm value.

The above results can be used by future papers in the following ways. Theory and simulations suggest that EFE estimator is consistent with the lowest uncertainty. Given typical features of corporate finance data, one may often safely assume that there is an endogenous firm fixed-effect, (conditional) heteroskedasticity across time and firms, and that the innovations are serially correlated. Thus, one should use the EFE estimator in applications\textsuperscript{16}. Following that, the developed $J$-test should be carried out. If the $p$-values are large, one can be confident that the EFE is the best choice of estimator. If not, the researcher should consider alternatives such as, for instance, the dynamic panel-data estimators that we develop in the extension.

To conclude, we point out some potential avenues for future research. Whereas this paper develops the FE, EFE, DFE, and DEFE estimators explicitly as GMM estimators in balanced panels as well as unbalanced panels, the POLS, RE, and FD are presented as GMM estimators for balanced panels, only. We leave it for future research to find the appropriate GMM representation of the latter three estimators in an unbalanced-panel model. Furthermore, the inclusion of year fixed-effects in the GMM estimation framework for unbalanced panels is not straightforward and needs more work.

Appendices

Appendix A  Pooled OLS and RE as GMM Estimators

\textbf{Theorem 1.} Let Assumptions 1-3 be satisfied. The OLS and the RE estimators for $\delta = [\psi, \alpha, \beta]'$ are given by

$$\hat{\delta}(\hat{W}_{OLS}) = \left( \sum_{i=1}^{N} X_i'X_i \right)^{-1} \sum_{i=1}^{N} X_i'q_i \quad (A1)$$

\textsuperscript{16} Even if the above mentioned ‘nuisances’ are not present in the data, the EFE will still be a consistent estimator, albeit not necessarily efficient.
\[
\hat{\delta}(W_{RE}) = \left( \sum_{i=1}^{N} X_i' \hat{\Sigma}^{-1} X_i \right)^{-1} \sum_{i=1}^{N} X_i' \hat{\Sigma}^{-1} q_i,
\]

(A2)

where \( \hat{\Sigma} \) is a matrix of size \((T \times T)\). Its \((t, s)\)-element is given by \( \frac{1}{N} \sum_{i=1}^{N} \hat{e}_{it} \hat{e}_{is} \), where \( \hat{e}_{it} \) are the regression residuals. Also introduce a scalar constant, \( k \), that may for instance be equal to \( \hat{\delta}^{-2} \equiv \left( \frac{1}{T} \frac{1}{N} \sum_{i=1}^{T} \sum_{i=1}^{N} \hat{e}_{it}^2 \right)^{-1} \).

The estimators are GMM estimators with particular choices of the weighting matrices \( \hat{W}_{OLS} \) and \( \hat{W}_{RE} \), given by

\[
\hat{W}_{OLS} = kI_T \otimes \left( \frac{1}{N} \sum_{i=1}^{N} z_i z'_i \right)^{-1}
\]

and

\[
\hat{W}_{RE} = \hat{\Sigma}^{-1} \otimes \left( \frac{1}{N} \sum_{i=1}^{N} z_i z'_i \right)^{-1}.
\]

(A3)

The estimators are not consistent for \( \delta \).

In addition, let Assumption 4 be satisfied. Define the matrix of size \((T[TL_x + 1] \times T[TL_x + 1])\)

\[
\Omega_{OLS} = \Omega_{RE} \equiv E \left( [\varepsilon_i \otimes z_i] [\varepsilon_i \otimes z_i]' \right) = E (\hat{\varepsilon}_i \hat{\varepsilon}_i' \otimes z_i z'_i),
\]

(A4)

which is the asymptotic variance of \( \varepsilon_i \otimes z_i \). Assume that \( \Omega_{OLS} \) is nonsingular.

The estimators \( \hat{\delta}(\hat{W}_{OLS}) \) and \( \hat{\delta}(\hat{W}_{RE}) \) are not likely to have the smallest possible uncertainty, since the sufficient condition that \( \hat{W}_{OLS} \rightarrow P \Omega_{OLS}^{-1} \) (where \( \rightarrow P \) denotes convergence in probability as \( N \rightarrow \infty \)) cannot be proven. The same is true for \( \hat{W}_{RE} \rightarrow P \Omega_{RE}^{-1} \).

Now, let

\[
S^{(OLS)}_{xz} = S^{(RE)}_{xz} \equiv \frac{1}{N} \sum_{i=1}^{N} (X_i \otimes z_i) \quad \text{and} \quad S^{(OLS)}_{zq} = S^{(RE)}_{zq} \equiv \frac{1}{N} \sum_{i=1}^{N} (q_i \otimes z_i).
\]

(A5)

The J-statistics to test for overidentification are computed as

\[
J = N \left( S^{(OLS)}_{zq} - S^{(OLS)}_{xz} \hat{\delta}(\hat{W}_{OLS}) \right)' P^{-1} \left( S^{(OLS)}_{zq} - S^{(OLS)}_{xz} \hat{\delta}(\hat{W}_{OLS}) \right)
\]

(A6)

\[
J = N \left( S^{(RE)}_{zq} - S^{(RE)}_{xz} \hat{\delta}(\hat{W}_{RE}) \right)' P^{-1} \left( S^{(RE)}_{zq} - S^{(RE)}_{xz} \hat{\delta}(\hat{W}_{RE}) \right).
\]

(A7)

Note that \( P = UVU' \), with \( V = \left( \frac{1}{N} \sum_{i=1}^{N} \hat{e}_i \hat{e}_i' \otimes z_i z'_i \right) \) and \( U = \left( I - S_{xz}(S_{xz} W S_{xz})^{-1} S_{xz} W \right) \).

Under the null hypothesis of correct specification, \( J \) is \( \chi^2 \)-distributed with \( rank(P) = T(TL_x + 1) - (L_x + 1) \) degrees of freedom.

**Proof:** For conciseness, we suppress the superscripts \( (\cdot)^{OLS} \) and \( (\cdot)^{RE} \). Define \( \Sigma_{xz} \equiv E(X_i \otimes
\(z_i\), \(\sigma_{zq} \equiv E(q_i \otimes z_i)\). The standard GMM estimator is given by

\[
\hat{\delta}(\hat{W}) = \left(S'_{ex} \hat{W} S_{ex}\right)^{-1} \left(S'_{ex} \hat{W} s_{zq}\right),
\]

which for OLS and RE is equal to:

\[
= \left(\sum_{i=1}^{N} (X'_i \otimes z'_i) \left[ D \otimes \left(\sum_{i=1}^{N} z'_i z'_i\right)^{-1}\right] \sum_{i=1}^{N} (X_i \otimes z_i) \right) \left(\sum_{i=1}^{N} (X'_i \otimes z'_i) \left[ D \otimes \left(\sum_{i=1}^{N} z'_i z'_i\right)^{-1}\right] \sum_{i=1}^{N} (X_i \otimes z_i)\right)^{-1}
\]

\[
\times \left(\sum_{i=1}^{N} \left[ E'_i z'_i E'_2 z'_i \ldots E'_T z'_i\right] \left[ D \otimes \left(\sum_{i=1}^{N} z'_i z'_i\right)^{-1}\right] \sum_{i=1}^{N} (q_i \otimes z_i)\right)
\]

\[
= \left[ \sum_{i=1}^{N} \sum_{i=1}^{T} d'_{i1} \sum_{i=1}^{T} d_{i2} E'_i \ldots \sum_{i=1}^{T} d_{iT} E'_i \right] \left[ \sum_{i=1}^{N} \sum_{i=1}^{T} d'_{i1} \sum_{i=1}^{T} d_{i2} E'_i \ldots \sum_{i=1}^{T} d_{iT} E'_i \right]^{-1}
\]

\[
\times \left(\sum_{i=1}^{N} \sum_{i=1}^{T} d'_{i1} E'_i \sum_{i=1}^{T} d_{i2} E'_i \ldots \sum_{i=1}^{T} d_{iT} E'_i \right) \left[ \sum_{i=1}^{N} \sum_{i=1}^{T} d'_{i1} \sum_{i=1}^{T} d_{i2} E'_i \ldots \sum_{i=1}^{T} d_{iT} E'_i \right]^{-1}
\]

\[
= \left(\sum_{i=1}^{N} \sum_{i=1}^{T} \sum_{i=1}^{T} d_{ts} E'_i z'_i E'_s \right)^{-1} \left(\sum_{i=1}^{N} \sum_{i=1}^{T} d_{ts} E'_i z'_i q_{is}\right)
\]

\[
= \left(\sum_{i=1}^{N} \sum_{i=1}^{T} \sum_{i=1}^{T} d_{ts} \begin{bmatrix} 1 & ppse_{it} \\ x_{it} \end{bmatrix} \right)^{-1} \left(\sum_{i=1}^{N} \sum_{i=1}^{T} \sum_{i=1}^{T} d_{ts} \begin{bmatrix} 1 & ppse_{it} \\ x_{it} \end{bmatrix} q_{is}\right)
\]

43
\[
\left( \sum_{i=1}^{N} X_i' D X_i \right)^{-1} \sum_{i=1}^{N} X_i' D q_i, \quad (A9)
\]

where \( D \) denotes either \( k I_T \) or \( \Sigma^{-1} \) and \( d_{ab} \) is the \((a,b)\)-element of \( D \). The last line follows from Hayashi, 2000 (page 292). This proves that the OLS and the RE estimators are GMM estimators with weighting matrices \( \hat{W} \) given in (A3). Note that above we have used

\[
E_l = \begin{bmatrix}
0 & 0 & 1 & 1 \\
0_{(T-1)L_x \times 1} & 0_{L_x \times L_x} & 0_{(T-1)L_x \times L_x} & 0_{1 \times L_x} \\
0_{L_x \times 1} & 0_{L_x \times L_x} & 0_{L_x \times L_x} & 0_{1 \times L_x}
\end{bmatrix}.
\quad (A10)
\]

The sampling error is equal to

\[
\hat{\delta}(\hat{W}) - \delta = \left( S'_{ex} \hat{W} S_{xx} \right)^{-1} S'_{ex} \hat{W} \left[ \frac{1}{N} \sum_{i=1}^{N} (c_i \ell \otimes z_i) + \frac{1}{N} \sum_{i=1}^{N} (\eta_i \otimes z_i) \right].
\quad (A11)
\]

Even if the terms to the left of the square brackets converged in probability to their finite populations moments, the sampling error would still not converge to zero. By the Ergodic Theorem, the second term in the brackets converges to \( E(\eta_i \otimes z_i) = 0 \) almost surely. The convergence properties of the first term are unknown, however. Furthermore, even if the term converges, the population moments \( E(c_i \ell \otimes z_i) \) is not likely to be zero.

Next, we analyze the uncertainty in the OLS and the RE estimator. It is well known that the efficient GMM estimator is the one where \( \hat{W} \overset{P}{\rightarrow} \Omega^{-1} \). It is easy to see that without additional assumptions, \( \hat{W} \) in (A3) is not likely to converge to the inverse of \( \Omega \) in (A4). However, if we make the additional assumptions of conditional homoskedasticity and serial uncorrelatedness of \( \varepsilon_{it} \), with a time-constant variance \( \sigma^2 \) that can be consistently estimated by \( 1/k \), then \( \Omega_{OLS} \) becomes \( \sigma^2 I_T \otimes E(z_i z'_i) \), which is the probability limit of \( \hat{W}_{OLS}^{-1} \). For the RE estimator, we would need to make the additional assumption of conditional homoskedasticity of \( \varepsilon_{it} \). Then \( \Omega_{RE} \) becomes \( E(\varepsilon_i \varepsilon'_i) \otimes E(z_i z'_i) = \Sigma \otimes E(z_i z'_i) \). Yet, this may still not be the probability limit of \( \hat{W}_{RE}^{-1} \), since \( \Sigma \) may not converge to \( \Sigma \), because of the inconsistency of the RE estimator.

Finally, the proof for the functional form and convergence properties of the \( J \)—statistics (A6) and (A7) follows directly from Cochrane, 2005 (page 204).

\[\text{\footnote{The above is a sufficient condition for efficiency. A necessary and sufficient condition is that there exists a matrix } B \text{ such that } \Sigma_{xx} = B \Sigma_{xx} \Omega^{-1}, \text{ where } W \text{ is the probability limit of } \hat{W}, \text{ according to Newey and McFadden, 1994 (page 2165).}}\]
Appendix B  Standard FD as a GMM Estimator

Theorem 2. Let Assumptions 1-3 be satisfied. The FD estimator for $\gamma = [\alpha, \beta]'$ is given by

$$\hat{\gamma}(\hat{W}_{FD}) = \left( \sum_{i=1}^{N} F'_i C_2 C'_2 F_i \right)^{-1} \sum_{i=1}^{N} F'_i C_2 C'_2 q_i$$  \hspace{1cm} (B1)

The estimator is a GMM estimator with a particular choice of the weighting matrix $\hat{W}_{FD}$, given by

$$\hat{W}_{FD} = k I_{T-1} \otimes \left( \frac{1}{N} \sum_{i=1}^{N} z_i z'_i \right)^{-1}.$$  \hspace{1cm} (B2)

The estimator is consistent for $\gamma$ if the matrix $\Sigma_{z^2}^{FD} \equiv \mathbb{E}(C_2' F_i \otimes z_i) = (C_2' \otimes I) \mathbb{E}(F_i \otimes z_i)$ of size $([T - 1][TL_x + 1] \times L_x)$ has full column rank $L_x$. Furthermore, $\mathbb{E}(z_i z'_i)^{-1}$ must exist.

In addition to the assumptions above, let Assumption 4 be satisfied and define the matrix of size $([T - 1][TL_x + 1] \times [T - 1][TL_x + 1])$

$$\Omega_{FD} \equiv \mathbb{E} \left( [C_2' \eta_i \otimes z_i][C_2' \eta_i \otimes z_i]' \right) = \mathbb{E} \left( C_2' \eta_i, C_2 \otimes z_i \right), \hspace{1cm} (B3)$$

which is the asymptotic variances of $C_2' \eta_i \otimes z_i$. Then, as $N \to \infty$

$$\sqrt{N}(\hat{\gamma}(\hat{W}_{FD}) - \gamma) \overset{D}{\to} N(0, \mathbb{E}(F'_i C_2 C'_2 F_i)^{-1} \mathbb{E}(F'_i C_2 C'_2 \eta_i, C_2 C'_2 F_i) \mathbb{E}(F'_i C_2 C'_2 F_i)^{-1})$$  \hspace{1cm} (B4)

(where $\overset{D}{\to}$ denotes convergence in distribution as $N \to \infty$).

Add the finite fourth moments assumption, i.e. that $\mathbb{E}(\tilde{f}_{it}^{(a)}, \tilde{z}_{ti}^{(c)}, \tilde{z}_{ti}^{(d)})$ exists and is finite for all $t, s = 1, 2, \ldots, T-1$, $a, b = 1, 2, \ldots, L_x$, and $c, d = 1, 2, \ldots, TL_x + 1$. Note that $\tilde{f}_{it}^{(a)}$ is the $(t, a)$-element of $C_2 F_i$ and $\tilde{z}_{ti}^{(c)}$ the $c$th element of vector $z_i$. We can consistently estimate the asymptotic variance of the estimator by

$$\widehat{\text{Var}}(\hat{\gamma}(\hat{W}_{FD})) = \left( \frac{1}{N} \sum_{i=1}^{N} F'_i C_2 C'_2 F_i \right)^{-1} \left( \frac{1}{N} \sum_{i=1}^{N} F'_i C_2 \hat{\eta}_i, \hat{\eta}_i', C_2 C'^2 F_i \right) \left( \frac{1}{N} \sum_{i=1}^{N} F'_i C_2 C'_2 F_i \right)^{-1}.$$  \hspace{1cm} (B5)

The estimator $\hat{\gamma}(\hat{W}_{FD})$ is not likely to be efficient, since the sufficient condition that $\hat{W}_{FD} \overset{D}{\to} \Omega_{FD}^{-1}$ cannot be proven.

Now, let

$$S_{z^2}^{(FD)} \equiv \frac{1}{N} \sum_{i=1}^{N} (C_2' F_i \otimes z_i) \quad \text{and} \quad S_{zq}^{(FD)} \equiv \frac{1}{N} \sum_{i=1}^{N} (C_2' q_i \otimes z_i). \hspace{1cm} (B6)$$
The $J$-statistic to test for overidentification is computed as

$$J = N \left( s_{zz}^{(FD)} - s_{zz}^{(ED)} \gamma(\hat{W}_{FD}) \right)' P^{-1} \left( s_{zz}^{(FD)} - s_{zz}^{(ED)} \gamma(\hat{W}_{FD}) \right).$$

(B7)

Note that $P = UVU'$, with $U = \left( I - s_{zx}^{(FD)} (S_{zx}^{(FD)})^{-1} s_{zx}^{(FD)} \right) \hat{W}_{FD}$ and $V = (\frac{1}{N} \sum_{i=1}^{N} C_i' \hat{\eta}_i \hat{\eta}_i C_i \otimes z_i' z_i)$. Under the null hypothesis of correct specification, $J$ is $\chi^2$-distributed with rank($P$) = $(T - 1)(T L_x + 1) - L_x$ degrees of freedom.

**Proof:** For conciseness, we suppress the superscript $(\cdot)^{FD}$. The standard GMM for $\gamma$ given in (A10) is equal to:

$$= \left( \sum_{i=1}^{N} (C_i' F_i \otimes z_i) \right)' \left[ k I \otimes \left( \sum_{i=1}^{N} z_i z_i' \right)^{-1} \right] \sum_{i=1}^{N} (C_i' F_i \otimes z_i) \quad -1$$

$$\sum_{i=1}^{N} \left( C_i' F_i \otimes z_i \right)' \left[ k I \otimes \left( \sum_{i=1}^{N} z_i z_i' \right)^{-1} \right] \sum_{i=1}^{N} (C_i' q_i \otimes z_i) \quad -1$$

$$= \sum_{i=1}^{N} \left( F_i' \otimes z_i' \right) \left[ k C_2 C_i' \otimes \left( \sum_{i=1}^{N} z_i z_i' \right)^{-1} \right] \sum_{i=1}^{N} (F_i \otimes z_i) \quad -1$$

$$\sum_{i=1}^{N} \left( F_i' \otimes z_i' \right) \left[ k C_2 C_i' \otimes \left( \sum_{i=1}^{N} z_i z_i' \right)^{-1} \right] \sum_{i=1}^{N} (q_i \otimes z_i) \quad -1$$

$$= \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{t=1}^{T} d_{ts} E_i z_i E_s \quad -1 \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{t=1}^{T} d_{ts} E_i z_i q_{is} = \sum_{i=1}^{N} F_i' C_2 C_i X_i \quad -1 \sum_{i=1}^{N} F_i' C_2 C_i q_i,$$

where $d_{ab}$ is the scalar $(a, b)$-element of $k C_2 C_i'$. This proves that the FD estimator is a GMM estimator with weighting matrix $\hat{W}$ given in (B2). Note that above we have used matrix $\hat{E}_t$, which is the same as matrix $E_t$ defined in (A10) in Appendix A, excluding the first column.

It is straightforward to see that the FD-GMM estimator satisfies all necessary assumptions outlined in Hayashi, 2000 (Chapters 3 and 4) for consistency and asymptotic normality. The asymptotic variance of a GMM estimator is generally given by

$$\text{Avar}(\hat{\gamma}(\hat{W})) = (\Sigma_{zz}') W \Sigma_{zz} W \Omega W \Sigma_{zz} (\Sigma_{zz}')^{-1} W \Omega W \Sigma_{zz}^{-1},$$

which for the FD estimator is equal to:

$$= \{ E(C_i' F_i \otimes z_i)' [\sigma^{-2} I \otimes E(z_i z_i')^{-1}] E(C_i' F_i \otimes z_i) \}^{-1} \Sigma_{zz} W \Omega W \Sigma_{zz}$$

46
\[ \times (C_2^*F_i \otimes z_i)' \left[ \sigma^{-2}I \otimes E(z_i'z_i)^{-1} \right] E(C_2^*F_i \otimes z_i) \right]^{-1} \]

\[ = (\sigma^{-2}E(F_i'C_2^*C_2F_i))^{-1} \Sigma_{xx} W W \Sigma_{xx} \left( E(\sigma^{-2}F_i'C_2^*C_2F_i) \right)^{-1} \]

\[ = (\sigma^{-2}E(F_i'C_2^*C_2F_i))^{-1} E(F_i' \otimes z_i')(C_2 \otimes I)(\sigma^{-2}I \otimes E(z_i'z_i)^{-1})E(C_2^*\eta_i'\eta_C^*C_2 \otimes z_i'z_i) \]

\[ \times (\sigma^{-2}I \otimes E(z_i'z_i)^{-1})(C_2' \otimes I)E(F_i \otimes z_i)(\sigma^{-2}E(F_i'C_2^*C_2F_i))^{-1} \]

\[ = (E(F_i'C_2^*C_2F_i))^{-1} E \left\{ E(F_i' \otimes z_i') (C_2'C_2^*\eta_i'\eta_i' \otimes E(z_i'z_i)^{-1}z_i'z_i[E(z_i'z_i)^{-1}] E(F_i \otimes z_i) \right\} \left( E(F_i'C_2^*C_2F_i) \right)^{-1} \]

\[ = (E(F_i'C_2^*C_2F_i))^{-1} E \left\{ \sum_{s=1}^{T} \sum_{t=1}^{T} d_{ts}^{(i)} \hat{E}_i z_i z_i' \hat{E}_s \right\} \left( E(F_i'C_2^*C_2F_i) \right)^{-1} \]

\[ = (E(F_i'C_2^*C_2F_i))^{-1} E \left\{ \sum_{s=1}^{T} \sum_{t=1}^{T} d_{ts}^{(i)} \left[ \begin{array}{c} \text{ppse}_{it} \\ x_{it} \end{array} \right] \right\} \left( E(F_i'C_2^*C_2F_i) \right)^{-1}, \quad (B10) \]

where \( d_{ab}^{(i)} \) is the scalar \((a, b)\)-element of \( C_2'C_2^*\eta_i'\eta_i' \). Hence it holds that

\[ \text{Avar}(\hat{\gamma}(\hat{W}_{FD})) = E(F_i'C_2^*C_2F_i)^{-1} E(F_i'C_2^*C_2^*\eta_i'\eta_i' \otimes C_2C_2F_i) E(F_i'C_2^*C_2F_i)^{-1}, \quad (B11) \]

which follows from Hayashi, 2000 (page 292). The convergence of the estimator in (B5) to (B11) follows directly from the Ergodic Theorem.

Next, we show the analyze the uncertainty in the FD estimator. It is easy to see that without additional assumptions, \( \hat{W} \) in (B2) is not likely to converge to the inverse of \( \Omega \) in (B3). However, if we make the additional assumptions of conditional homoskedasticity and serial uncorrelatedness in the residual changes \( \Delta \eta_{it} \), with a time-constant variance \( \sigma^2 \), then \( \Omega_{FD} \) becomes \( \sigma^2I \otimes E(z_i'z_i) \), which is the probability limit of \( \hat{W}_{FD}^{-1} \).

Finally, the proof for the functional form and convergence properties of the \( J \)-statistic (B7) follows directly from Cochrane, 2005 (page 204).

**Appendix C  Standard FE as a GMM Estimator**

**Theorem 3.** Let Assumptions 1-3 be satisfied. The FE estimator for \( \gamma = [\alpha, \beta]' \) is given by

\[ \hat{\gamma}(\hat{W}_{FE}) = \left( \sum_{i=1}^{N} F_i'MF_i \right)^{-1} \sum_{i=1}^{N} F_i'Mq_i. \quad (C1) \]

The estimator is a GMM estimator with a particular choice of the weighting matrix \( \hat{W}_{FE} \), given by

\[ \hat{W}_{FE} = k(C'C)^{-1} \otimes \left( \frac{1}{N} \sum_{i=1}^{N} z_i'z_i \right)^{-1}. \quad (C2) \]
The estimator is consistent for $\gamma$ if the matrix $\Sigma_{zz}^{FF} \equiv E(C'F_i \otimes z_i) = (C' \otimes I)E(F_i \otimes z_i)$ of size $([T-1][TL_x+1] \times L_x)$ has full column rank $L_x$. Furthermore, $E(z_i z_i')^{-1}$ must exist.

In addition to the assumptions above, let Assumption 4 be satisfied and define the matrix of size $([T-1][TL_x+1] \times [T-1][TL_x+1])$

$$\Omega_{FE} \equiv E \left( [C'z_i] [C'z_i]' \right) = E (C'z_i C \otimes z_i z_i') ,$$

which is the asymptotic variances of $C'z_i \otimes z_i$. Then, as $N \to \infty$

$$\sqrt{N}(\hat{\gamma}(W_{FE}) - \gamma) \overset{D}{\to} N(0, E(F_i'MF_i)^{-1}E(F_i'M\hat{z}_i/MF_i)E(F_i'MF_i)^{-1})$$

(where $\overset{D}{\to}$ denotes convergence in distribution as $N \to \infty$).

Add the finite fourth moments assumption, i.e. that $E(\hat{z}_i^{(a)} \hat{z}_i^{(b)} \hat{z}_i^{(c)} \hat{z}_i^{(d)})$ exists and is finite for all $t, s = 1, 2, \ldots, T-1, a, b = 1, 2, \ldots, L_x$, and $c, d = 1, 2, \ldots, TL_x + 1$. Note that $f_{it}^{(a)}$ is the $(t, a)$-element of $C'F_i$ and $z_i^{(c)}$ the $c$th element of vector $z_i$. We can consistently estimate the asymptotic variance of the estimator by

$$\hat{\text{Avar}}(\hat{\gamma}(W_{FE})) = \left( \frac{1}{N} \sum_{i=1}^{N} F_i'MF_i \right)^{-1} \left( \frac{1}{N} \sum_{i=1}^{N} F_i'M\hat{z}_i/MF_i \right) \left( \frac{1}{N} \sum_{i=1}^{N} F_i'MF_i \right)^{-1} .$$

The estimator $\hat{\gamma}(W_{FE})$ is not likely to be efficient, since the sufficient condition that $W_{FE} \overset{P}{\to} \Omega_{FE}^{-1}$ cannot be proven.

Now, let

$$S_{z^x}^{(FE)} \equiv \frac{1}{N} \sum_{i=1}^{N} (C'F_i \otimes z_i) \quad \text{and} \quad S_{z^q}^{(FE)} \equiv \frac{1}{N} \sum_{i=1}^{N} (C'q_i \otimes z_i).$$

The $J$-statistics to test for overidentification are computed as

$$J = N \left( S_{z^q}^{(FE)} - S_{z^x}^{(FE)} \gamma(W_{FE}) \right)' P^{-1} \left( S_{z^q}^{(FE)} - S_{z^x}^{(FE)} \gamma(W_{FE}) \right) .$$

Note that $P = UVU'$, with $U = \left( I - S_{z^x}^{(FE)} (S_{z^x}^{(FE)})' \hat{W}_{FE} S_{z^x}^{(FE)} \right)^{-1} S_{z^x}^{(FE)}' \hat{W}_{FE}$ and $V = \left( \frac{1}{N} \sum_{i=1}^{N} C'z_i z_i' C \otimes z_i z_i' \right)$.

Under the null hypothesis of correct specification, $J$ is $\chi^2$-distributed with rank($P$) = $(T - 1)(TL_x + 1) - L_x$ degrees of freedom.

**Proof:** For conciseness, we suppress the superscript $(\cdot)^{FE}$. The standard GMM for $\gamma$ given
in (A8) is equal to:

\[
\begin{align*}
&= \left( \sum_{i=1}^{N} (C'F_{i} \otimes z_{i})' \right) \left[ k (C'C)^{-1} \otimes \left( \sum_{i=1}^{N} z_{i}z_{i}' \right)^{-1} \right] \sum_{i=1}^{N} (C'F_{i} \otimes z_{i})^{-1} \\
&\times \sum_{i=1}^{N} (C'F_{i} \otimes z_{i})' \left[ k (C'C)^{-1} \otimes \left( \sum_{i=1}^{N} z_{i}z_{i}' \right)^{-1} \right] \sum_{i=1}^{N} (C'q_{i} \otimes z_{i}) \\
&= \left( \sum_{i=1}^{N} (F_{i}' \otimes z_{i}') \left[ M \otimes \left( \sum_{i=1}^{N} z_{i}z_{i}' \right)^{-1} \right] \sum_{i=1}^{N} (F_{i}' \otimes z_{i}) \right)^{-1} \sum_{i=1}^{N} (F_{i}' \otimes z_{i}') \left[ M \otimes \left( \sum_{i=1}^{N} z_{i}z_{i}' \right)^{-1} \right] \sum_{i=1}^{N} (q_{i} \otimes z_{i}) \\
&= \left( \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{t=1}^{T} m_{t,s} \tilde{E}_{i}z_{i}z_{i}' \tilde{E}_{s} \right)^{-1} \sum_{s=1}^{T} \sum_{t=1}^{T} m_{t,s} \tilde{E}_{i}z_{i}q_{is} = \left( \sum_{i=1}^{N} F_{i}'MX_{i} \right)^{-1} \sum_{i=1}^{N} F_{i}'Mq_{i},
\end{align*}
\]

where \( m_{ab} \) is the scalar \((a,b)\)-element of \( M \). We have used the property that \( M = C(C'C)^{-1}C' \) to prove that the FE estimator is a GMM estimator with weighting matrix \( \tilde{W} \) given in (C2). Note that above we have used matrix \( \tilde{E}_{i} \), which is the same as matrix \( E_{i} \) defined in (A10) in A, excluding the first column.

It is straightforward to see that the FE-GMM estimator satisfies all necessary assumptions outlined in Hayashi, 2000 (Chapters 3 and 4) for consistency and asymptotic normality. The asymptotic variance of a GMM estimator is generally given by

\[
\text{Avar}(\hat{\gamma}(\tilde{W})) = (\Sigma_{xz}W\Sigma_{xz})^{-1} \Sigma_{xz}W\Omega W\Sigma_{xz} (\Sigma_{xz}W\Sigma_{xz})^{-1},
\]

which for the FE estimator is equal to:

\[
\begin{align*}
&= \{ E(C'F_{i} \otimes z_{i})' [\sigma^{-2}(C'C)^{-1} \otimes E(z_{i}z_{i}')^{-1}] E(C'F_{i} \otimes z_{i}) \}^{-1} \Sigma_{xz}W\Omega W\Sigma_{xz} \\
&\times \{ E(C'F_{i} \otimes z_{i})' [\sigma^{-2}(C'C)^{-1} \otimes E(z_{i}z_{i}')^{-1}] E(C'F_{i} \otimes z_{i}) \}^{-1} \\
&= (\sigma^{-2}E(F_{i}'MF_{i}))^{-1} \Sigma_{xz}W\Omega W\Sigma_{xz} (\sigma^{-2}E(F_{i}'MF_{i}))^{-1} \\
&= (\sigma^{-2}E(F_{i}'MF_{i}))^{-1} E(F_{i}'z_{i})(C \otimes I)(\sigma^{-2}(C'C)^{-1} \otimes E(z_{i}z_{i}')^{-1}) E(C'\eta_{i}\eta_{i}'C \otimes z_{i}z_{i}') \\
&\times (\sigma^{-2}(C'C)^{-1} \otimes E(z_{i}z_{i}')^{-1})(C' \otimes I)E(F_{i} \otimes z_{i}) (\sigma^{-2}E(F_{i}'MF_{i}))^{-1} \\
&= (E(F_{i}'MF_{i}))^{-1} E \left\{ E(F_{i}'z_{i})(M\eta_{i}\eta_{i}'M \otimes E(z_{i}z_{i}')^{-1}z_{i}z_{i}'E(z_{i}z_{i}')^{-1}) E(F_{i} \otimes z_{i}) \right\} (E(F_{i}'MF_{i}))^{-1} \\
&= (E(F_{i}'MF_{i}))^{-1} \left\{ \sum_{s=1}^{T} \sum_{t=1}^{T} m_{t,s} \tilde{E}_{i}z_{i}z_{i}' \tilde{E}_{s} \right\} (E(F_{i}'MF_{i}))^{-1} \\
&= (E(F_{i}'MF_{i}))^{-1} \left\{ \sum_{s=1}^{T} \sum_{t=1}^{T} m_{t,s} \left[ ppse_{it} x_{it} \right] \right\} (E(F_{i}'MF_{i}))^{-1},
\end{align*}
\]
where \(d_{ab}^{(i)}\) is the scalar \((a, b)\)-element of \(M\eta_i\eta'_iM\). Hence it holds that

\[
\text{Avar}(\hat{\gamma}(\hat{W}_{FE})) = \mathbb{E}(F_i'MF_i)^{-1}\mathbb{E}(F_i'M\eta_i\eta'_iMF_i)\mathbb{E}(F_i'MF_i)^{-1}, \tag{C11}
\]

which follows from Hayashi, 2000 (page 292). The convergence of the estimator in (C5) to (C11) follows directly from the Ergodic Theorem.

Next, we show the analyze the uncertainty in the FE estimator. It is easy to see that without additional assumptions, \(\hat{W}\) in (C2) is not likely to converge to the inverse of \(\Omega\) in (C3). However, if we make the additional assumptions of conditional homoskedasticity and serial uncorrelatedness of \(\eta_{it}\), with a time-constant variance \(\sigma^2\), then \(\Omega_{FE}\) becomes

\[
\sigma^2 C'C \otimes \mathbb{E}(z_i'z_i'),
\]

which is the probability limit of \(\hat{W}_{FE}^{-1}\).

Finally, the proof for the functional form and convergence properties of the \(J\)-statistic (C7) follows directly from Cochrane, 2005 (page 204).

### Appendix D The EFE Estimator

**Theorem 4.** Let Assumptions 1-3 be satisfied. The EFE estimator for \(\gamma = [\beta, \alpha]'\) is given by

\[
\hat{\gamma}(\hat{W}_{EFE}) = \left\{ \left( \frac{1}{N} \sum_{i=1}^{N} (F_i'C \otimes z'_i) \right) \left( \frac{1}{N} \sum_{i=1}^{N} (C'\hat{\eta}_i\hat{\eta}'_iC \otimes z_i z'_i) \right)^{-1} \left( \frac{1}{N} \sum_{i=1}^{N} (C'F_i \otimes z_i) \right) \right\}^{-1} \left( \frac{1}{N} \sum_{i=1}^{N} (C'q_i \otimes z_i) \right), \tag{D1}
\]

where \(C'\hat{\eta}_i = C'(q_i - F_i\hat{\gamma}(\hat{W}_{EFE}))\). The estimator is a GMM estimator with a particular choice of the weighting matrix \(\hat{W}_{EFE}\), given by

\[
\hat{W}_{EFE} = \left( \frac{1}{N} \sum_{i=1}^{N} (C'\hat{\eta}_i\hat{\eta}'_iC \otimes z_i z'_i) \right)^{-1}. \tag{D2}
\]

The estimator is consistent for \(\gamma\) if the matrix \(\Sigma_{xz}^{(EFE)} \equiv \Sigma_{xz}^{(FE)}\) has full column rank \(L_x\). Furthermore, we have to assume that \(\mathbb{E}(f_{it}^{(a)} f_{it}^{(b)} z_{ic}^{(c)} z_{id}^{(d)})\) exists and is finite for all \(t, s = 1, 2, \ldots, T - 1, a, b = 1, 2, \ldots, L_x\), and \(c, d = 1, 2, \ldots, TL_x + 1\). Note that \(f_{it}^{(a)}\) is the \((t, a)\)-element of \(C'F_i\) and \(z_{ic}^{(c)}\) the \(c\)th element of vector \(z_i\). Finally, assume that the matrix \(\Omega_{EFE} \equiv \Omega_{FE}\) is nonsingular.
In addition to the assumptions above, let Assumption 4 be satisfied. Then, as \( N \to \infty \)
\[
\sqrt{N}(\hat{\gamma}(\hat{W}_{EFE}) - \gamma) \overset{D}{\to} N(0, \left( \mathbb{E}(C'F_i \otimes z_i)' \left( \mathbb{E}(C'\eta_i\eta_i'C \otimes z_i z_i')^{-1} \mathbb{E}(C'F_i \otimes z_i) \right)^{-1} \right)). \quad (D3)
\]
We can consistently estimate the asymptotic variance of the estimator by
\[
\bar{\text{Var}}(\hat{\gamma}(\hat{W}_{EFE})) = \left( \frac{1}{N} \sum_{i=1}^{N} (C'F_i \otimes z_i)' \left( \frac{1}{N} \sum_{i=1}^{N} (C'\eta_i\eta_i'C \otimes z_i z_i')^{-1} \frac{1}{N} \sum_{i=1}^{N} (C'F_i \otimes z_i) \right)^{-1} \right).
\quad (D4)
\]
The estimator \( \hat{\gamma}(\hat{W}_{EFE}) \) is efficient, since \( \hat{W}_{EFE} \overset{P}{\to} \Omega_{FE}^{-1} \).

Now, let
\[
S_{zxn}^{(FE)} = \frac{1}{N} \sum_{i=1}^{N} (C'F_i \otimes z_i) \quad \text{and} \quad S_{zqn}^{(FE)} = \frac{1}{N} \sum_{i=1}^{N} (C'q_i \otimes z_i). \quad (D5)
\]
The \( J \)-statistic to test for overidentification is computed as
\[
J = N \left( S_{zy}^{(FE)} - S_{zxn}^{(FE)} \hat{\delta}(\hat{W}_{EFE}) \right)' \hat{W}_{EFE} \left( S_{zy}^{(FE)} - S_{zxn}^{(FE)} \hat{\delta}(\hat{W}_{EFE}) \right). \quad (D6)
\]
Under the null hypothesis of correct specification, \( J \) is \( \chi^2 \)-distributed with rank(\( \hat{W}_{EFE} \)) – \( L_x = (T - 1)(TL_x + 1) - L_x \) degrees of freedom.

Proof: The proof of Theorem 4 follows directly from the proofs in Appendices A, B, C, and from Cochrane, 2005 (page 205).

Appendix E  FE as a GMM Estimator for Unbalanced Panels

Theorem 5. The usual FE estimator for unbalanced panels, where missing observations are zeroed out, is identical to the GMM-FE estimator for balanced panels developed in Appendix C, where the missing observations in the regressand and regressor matrices are replaced by the time-series averages of the available observations.

Proof: According to Hayashi (2000), we can re-write our data generating process (1)-(3) for unbalanced panels as follows
\[
\tilde{q}_i = e_i \psi + \tilde{F}_i \gamma + e_i c_i + \tilde{\eta}_i, \quad (E1)
\]
where \( \tilde{q}_i, \tilde{F}_i, \) and \( \tilde{\eta}_i \) are vectors and matrices where the missing observations take on the value zeros. The vector \( e_i \) is a \( T \times 1 \)-selection vector that zeroes out the missing observations,
i.e.,

\[
\begin{bmatrix}
    e_{i1} \\
    e_{i2} \\
    \vdots \\
    e_{iT}
\end{bmatrix}, \quad e_{it} \equiv \begin{cases} 1 \text{ if observation } i, t \text{ is available} \\ 0 \text{ otherwise} \end{cases}.
\] (E2)

Hayashi (2000) shows that the usual FE estimator for unbalanced panels is given by

\[
\hat{\gamma}_{FE(\text{zeros})} = \left( \sum_{i=1}^{N} \bar{F}_i' M_i \bar{F}_i \right)^{-1} \sum_{i=1}^{N} \bar{F}_i' M_i \bar{q}_i,
\] (E3)

where \( M_i \equiv I - e_i (e_i'e_i)^{-1} e_i' \). Note that the difference to the balanced scenario is that the sandwich matrix \( M_i \) is now dependent on the cross-sectional units \( i \). Thus, the estimator (E3) is not a GMM estimator.

Instead of zeroing out, we can also replace the missing observations by time-series averages, and then define the FE estimator to be identical to the balanced case in Appendix C. To that end, note that \( \psi \bar{q}_i = e_i' \bar{q}_i \) is the time-series sum of the available observations, and \( (e_i'e_i)^{-1} e_i' \bar{q}_i \) is the time-series average of the available observations. Hence, \( \bar{q}_i + (\tau - e_i)(e_i'e_i)^{-1} e_i' \bar{q}_i \) is a vector that replaces the missing zeroed out observations with the time-series average of the available observations. We can re-write the unbalanced model from Equation (E1) further as

\[
\bar{q}_i + (\tau - e_i)(e_i'e_i)^{-1} e_i' \bar{q}_i = e_i \psi + \bar{F}_i \gamma + e_i c_i + \bar{\eta}_i + (\tau - e_i)(e_i'e_i)^{-1} e_i' \bar{q}_i
\]

\[
\hat{\gamma}_{FE(\text{averages})} = \left( \sum_{i=1}^{N} \bar{F}_i' \left( \bar{I} + (\tau - e_i)(e_i'e_i)^{-1} e_i' \right) \bar{M} \left( \bar{I} + (\tau - e_i)(e_i'e_i)^{-1} e_i' \right) \bar{F}_i \right)^{-1} \sum_{i=1}^{N} \bar{F}_i' \left( \bar{I} + (\tau - e_i)(e_i'e_i)^{-1} e_i' \right) \bar{M} \left( \bar{I} + (\tau - e_i)(e_i'e_i)^{-1} e_i' \right) \bar{q}_i
\] (E4)

Now, we apply the balanced FE estimator from Appendix C to the transformed model (E4) above

\[
\hat{\gamma}_{FE(\text{averages})} = \left( \sum_{i=1}^{N} \bar{F}_i' \left( \bar{I} + (\tau - e_i)(e_i'e_i)^{-1} e_i' \right) \bar{M} \left( \bar{I} + (\tau - e_i)(e_i'e_i)^{-1} e_i' \right) \bar{F}_i \right)^{-1} \sum_{i=1}^{N} \bar{F}_i' A_i \bar{F}_i \bar{q}_i.
\] (E5)
Note the following

$$A_i = \left( I + (\ell - e_i) (e_i' e_i)^{-1} e_i' \right)' M \left( I + (\ell - e_i) (e_i' e_i)^{-1} e_i' \right)$$

$$= \left( I - \ell (\ell')^{-1} \ell' + e_i (e_i' e_i)^{-1} e_i' (\ell' e_i) (\ell - e_i) (e_i' e_i)^{-1} e_i' \right)$$

$$= \left( I - \ell (\ell')^{-1} \ell' + e_i (e_i' e_i)^{-1} e_i' (\ell' e_i) (\ell - e_i) (e_i' e_i)^{-1} e_i' \right)$$

$$= \left( I - \ell (\ell')^{-1} \ell' + e_i (e_i' e_i)^{-1} e_i' (\ell' e_i) (\ell - e_i) (e_i' e_i)^{-1} e_i' \right)$$

$$= \left( I - \ell (\ell')^{-1} \ell' + e_i (e_i' e_i)^{-1} e_i' (\ell' e_i) (\ell - e_i) (e_i' e_i)^{-1} e_i' \right)$$

$$= \left( I - \ell (\ell')^{-1} \ell' + e_i (e_i' e_i)^{-1} e_i' (\ell' e_i) (\ell - e_i) (e_i' e_i)^{-1} e_i' \right)$$

Now, note that $\ell' e_i = e_i' e_i$. Then, we can continue re-writing (E6) as

$$A_i = M_i - \ell (\ell')^{-1} \ell' + e_i (e_i' e_i)^{-1} e_i' (\ell' e_i) (\ell - e_i)' - e_i (e_i' e_i)^{-1} e_i' (\ell' e_i) (\ell - e_i)'$$

Hence, the balanced panel estimator in (E5) becomes

$$\hat{\gamma}_{FE(averages)} = \left( \sum_{i=1}^{N} \hat{F}_i' \left[ M_i - (\ell - e_i) (e_i' e_i)^{-1} (\ell - e_i)' \right] \hat{F}_i \right)^{-1} \sum_{i=1}^{N} \hat{F}_i' \left[ M_i - (\ell - e_i) (e_i' e_i)^{-1} (\ell - e_i)' \right] \hat{q}_i.$$

Finally, we note that it holds that $(\ell - e_i)' \hat{F}_i$ and $(\ell - e_i)' \hat{q}_i$ is zero, since the vector in parenthesis selects only those elements of $\hat{F}_i$ and $\hat{q}_i$ that are zero (i.e. the missing zeroed out observations). The estimator is thus equal to the usual FE estimator of unbalanced panels in (E3), that is

$$\hat{\gamma}_{FE(averages)} = \left( \sum_{i=1}^{N} \hat{F}_i' M_i \hat{F}_i \right)^{-1} \sum_{i=1}^{N} \hat{F}_i' M_i \hat{q}_i = \hat{\gamma}_{FE(zeros)}.$$
References


Grieser, W. D. and Hadlock, C. J. (2019). Panel-Data Estimation in Finance: Testable As-


Table 1: Simulation Results for Balanced Panels

This table reports results from 50,000 simulations of the data generating process in equations (1)-(3). $c_i$ is an endogenous function of the time-series averages of the covariates plus an error term. $\eta_{it}$ is heteroskedastic over time and firms, and has AR(1) time-series dynamics, with autoregressive coefficient 0.6, but is independent across $i$. We let $T = 5$, $L_x = 5$, and $N = 1,000$. In ‘PANEL A’, we report the relative bias (i.e. the average of $\frac{\text{[estimate-true parameter]}}{\text{[true parameter]}}$). In ‘PANEL B’, we report the dispersion of the bias (i.e. the standard deviation of $\frac{\text{[estimate-true parameter]}}{\text{[true parameter]}}$). In ‘PANEL C’, we report the size/power of the $J$-test with nominal size 5%.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>POLS</th>
<th>RE</th>
<th>FD</th>
<th>FE</th>
<th>EFE demeaned</th>
<th>EFE first-differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\alpha}$</td>
<td>45,193.217</td>
<td>2.545</td>
<td>-0.225</td>
<td>-1.300</td>
<td>-1.333</td>
<td>-1.686</td>
</tr>
<tr>
<td>$\hat{\beta}_1$</td>
<td>209,158.666</td>
<td>22.706</td>
<td>3.339</td>
<td>4.144</td>
<td>2.651</td>
<td>2.124</td>
</tr>
<tr>
<td>$\hat{\beta}_2$</td>
<td>260,030.425</td>
<td>24.659</td>
<td>2.004</td>
<td>1.132</td>
<td>-0.140</td>
<td>-0.114</td>
</tr>
<tr>
<td>$\hat{\beta}_3$</td>
<td>188,839.250</td>
<td>17.463</td>
<td>-1.047</td>
<td>-1.936</td>
<td>0.741</td>
<td>0.439</td>
</tr>
<tr>
<td>$\hat{\beta}_4$</td>
<td>43,146.166</td>
<td>3.980</td>
<td>0.861</td>
<td>-0.202</td>
<td>-0.125</td>
<td>-0.160</td>
</tr>
<tr>
<td>$\psi$</td>
<td>71.165</td>
<td>88.878</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**PANEL B: Dispersion**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>POLS</th>
<th>RE</th>
<th>FD</th>
<th>FE</th>
<th>EFE demeaned</th>
<th>EFE first-differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\alpha}$</td>
<td>74.690</td>
<td>1.760</td>
<td>2.622</td>
<td>3.231</td>
<td>1.953</td>
<td>1.931</td>
</tr>
<tr>
<td>$\hat{\beta}_1$</td>
<td>80.782</td>
<td>1.755</td>
<td>2.636</td>
<td>3.249</td>
<td>1.953</td>
<td>1.932</td>
</tr>
<tr>
<td>$\hat{\beta}_2$</td>
<td>90.211</td>
<td>1.748</td>
<td>2.614</td>
<td>3.222</td>
<td>1.951</td>
<td>1.931</td>
</tr>
<tr>
<td>$\hat{\beta}_3$</td>
<td>83.334</td>
<td>1.745</td>
<td>2.625</td>
<td>3.243</td>
<td>1.949</td>
<td>1.928</td>
</tr>
<tr>
<td>$\hat{\beta}_4$</td>
<td>75.545</td>
<td>1.750</td>
<td>2.626</td>
<td>3.242</td>
<td>1.953</td>
<td>1.931</td>
</tr>
<tr>
<td>$\psi$</td>
<td>2.120</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**PANEL C: Inference**

| $J$-test | 100% | 100% | 3.642% | 3.576% | 4.294% | 4.188% |

59
Table 2: Simulation Results for Unbalanced Panels

This table reports results from 50,000 simulations of the data generating process in equations (1)-(3). $c_i$ is an endogenous function of the time-series averages of the covariates plus an error term. $\eta_{it}$ is heteroskedastic over time and firms, and has AR(1) time-series dynamics, with autoregressive coefficient 0.6, but is independent across $i$. We let $T = 5$, $L_x = 5$, and $N = 1,000$. We simulate unbalanced panels, where for 50% of firms we assume one time period is missing, for 20% of firms we assume two time periods are missing, and for 10% of firms we assume three time periods are missing. In ‘PANEL A’, we report the relative bias (i.e. the average of $[\text{estimate-true parameter}]$/[true parameter]). In ‘PANEL B’, we report the dispersion of the bias (i.e. the standard deviation of $[\text{estimate-true parameter}]$). In ‘PANEL C’, we report the size of the $J$-test with nominal size 5%.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>FE</th>
<th>EFE demeaned</th>
<th>EFE first-differences</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\times 10^{-6}$</td>
</tr>
<tr>
<td>PANEL A: Bias</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\alpha}$</td>
<td>-0.030</td>
<td>1.884</td>
<td>1.714</td>
</tr>
<tr>
<td>$\hat{\beta}_1$</td>
<td>0.730</td>
<td>0.943</td>
<td>0.577</td>
</tr>
<tr>
<td>$\hat{\beta}_2$</td>
<td>4.787</td>
<td>-0.623</td>
<td>-0.329</td>
</tr>
<tr>
<td>$\hat{\beta}_3$</td>
<td>-2.710</td>
<td>7.830</td>
<td>8.459</td>
</tr>
<tr>
<td>$\hat{\beta}_4$</td>
<td>-2.331</td>
<td>-1.639</td>
<td>-1.463</td>
</tr>
<tr>
<td>PANEL B: Dispersion</td>
<td></td>
<td></td>
<td>$\times 10^{-3}$</td>
</tr>
<tr>
<td>$\hat{\alpha}$</td>
<td>6.462</td>
<td>3.657</td>
<td>3.665</td>
</tr>
<tr>
<td>$\hat{\beta}_1$</td>
<td>6.446</td>
<td>3.670</td>
<td>3.668</td>
</tr>
<tr>
<td>$\hat{\beta}_2$</td>
<td>6.506</td>
<td>3.690</td>
<td>3.691</td>
</tr>
<tr>
<td>$\hat{\beta}_3$</td>
<td>6.471</td>
<td>3.666</td>
<td>3.673</td>
</tr>
<tr>
<td>$\hat{\beta}_4$</td>
<td>6.507</td>
<td>3.704</td>
<td>3.704</td>
</tr>
<tr>
<td>PANEL C: Inference</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$J$-test</td>
<td>3.290%</td>
<td>3.452%</td>
<td>3.450%</td>
</tr>
</tbody>
</table>
Table 3: Definitions of Variables and Sources

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dependent variable:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$q$</td>
<td>Tobin’s Q: (market value of equity - book value of equity + book value of assets) / book value of assets</td>
<td>Compustat</td>
</tr>
<tr>
<td><strong>Independent variable:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$ppse$</td>
<td>The dollar change in CEO wealth for a one percentage point change in firm stock price (for example, Coles et al., 2006)</td>
<td>ExecuComp</td>
</tr>
<tr>
<td><strong>Control variables:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$xrd$</td>
<td>Ratio of R&amp;D expenses to book value of assets</td>
<td>Compustat</td>
</tr>
<tr>
<td>$adv$</td>
<td>Ratio of advertising expenses to book value of assets</td>
<td>Compustat</td>
</tr>
<tr>
<td>$debt$</td>
<td>Ratio of book value of total debt to book value of assets</td>
<td>Compustat</td>
</tr>
<tr>
<td>$lnassets$</td>
<td>The natural logarithm of total assets</td>
<td>Compustat</td>
</tr>
<tr>
<td>$prop$</td>
<td>Ratio of property, plants and equipment to book value of assets</td>
<td>Compustat</td>
</tr>
</tbody>
</table>

Table 4: Descriptive Statistics

This table reports the descriptive statistics for all the variables in our sample. All variables are defined in Table 3. The sample consists of 6,408 firm-year observations for the sample for the 14 years 2005-2018.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Median</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$q$</td>
<td>2.262</td>
<td>1.852</td>
<td>1.366</td>
</tr>
<tr>
<td>$ppse$</td>
<td>654.642</td>
<td>199.328</td>
<td>1489.456</td>
</tr>
<tr>
<td>$xrd$</td>
<td>0.051</td>
<td>0.025</td>
<td>0.072</td>
</tr>
<tr>
<td>$adv$</td>
<td>0.031</td>
<td>0.012</td>
<td>0.065</td>
</tr>
<tr>
<td>$debt$</td>
<td>0.214</td>
<td>0.175</td>
<td>0.231</td>
</tr>
<tr>
<td>$lnassets$</td>
<td>7.370</td>
<td>7.195</td>
<td>1.732</td>
</tr>
<tr>
<td>$prop$</td>
<td>0.202</td>
<td>0.149</td>
<td>0.173</td>
</tr>
</tbody>
</table>
This table reports the within-firm (time-series) variation and between-firm (cross-sectional) variation for the different variables in our sample. In parenthesis, we present the proportion of total variation that is given by within-firm and between-firm variation. All variables are defined in Table 3. The sample consists of 6,408 observations for the sample for the 14 years 2005-2018.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total Variation</th>
<th>Within-Firm Variation (Percentage)</th>
<th>Between-Firm Variation (Percentage)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( q )</td>
<td>15,172.063</td>
<td>5,601.788 (36.922 %)</td>
<td>9,570.275 (63.078 %)</td>
</tr>
<tr>
<td>( pps (\times10^9) )</td>
<td>17.104</td>
<td>5.721 (33.448 %)</td>
<td>11.383 (66.552 %)</td>
</tr>
<tr>
<td>( xrd )</td>
<td>39.905</td>
<td>8.354 (20.935 %)</td>
<td>31.551 (79.065 %)</td>
</tr>
<tr>
<td>( adv )</td>
<td>30.285</td>
<td>6.113 (20.184 %)</td>
<td>24.172 (79.816 %)</td>
</tr>
<tr>
<td>( debt )</td>
<td>429.388</td>
<td>122.923 (28.628 %)</td>
<td>306.465 (71.372 %)</td>
</tr>
<tr>
<td>( lnassets )</td>
<td>24,369.801</td>
<td>1,548.601 (6.355 %)</td>
<td>22,821.200 (93.645 %)</td>
</tr>
<tr>
<td>( prop )</td>
<td>242.949</td>
<td>17.076 (7.029 %)</td>
<td>225.873 (92.971 %)</td>
</tr>
</tbody>
</table>
Table 6: Standard Pooled OLS (POLS), Random-Effects Estimation (RE), and First-Difference (FD) Regressions of Firm Value on CEO Pay

This table reports the POLS, RE, and FD regression results of firm value \( q_{it} \) on the CEO pay-performance sensitivity measure, \( ppse \). The sample consists of \( N = 809 \) unique firms for the 14 years 2005-2018. All dependent and independent variables are defined in Table 3. Year-dummy effects are estimated in a first-stage auxiliary FE regression but their coefficients are not reported. Robust standard errors that incorporate firm-level clustering are reported in parentheses. ***, **, * indicate that the parameter estimate is significantly different from zero at the 1%, 5%, or 10% level, respectively.

<table>
<thead>
<tr>
<th>Variables</th>
<th>POLS</th>
<th>RE</th>
<th>FD</th>
</tr>
</thead>
<tbody>
<tr>
<td>( ppse )</td>
<td>0.0002505 ***</td>
<td>0.0002529 ***</td>
<td>0.0002606 ***</td>
</tr>
<tr>
<td></td>
<td>(0.0000354)</td>
<td>(0.0000323)</td>
<td>(0.0000380)</td>
</tr>
<tr>
<td>( xrd )</td>
<td>4.275 ***</td>
<td>2.146 **</td>
<td>3.314 ***</td>
</tr>
<tr>
<td></td>
<td>(0.987)</td>
<td>(1.029)</td>
<td>(0.988)</td>
</tr>
<tr>
<td>( adv )</td>
<td>1.186</td>
<td>0.399</td>
<td>1.420 **</td>
</tr>
<tr>
<td></td>
<td>(0.733)</td>
<td>(0.320)</td>
<td>(0.568)</td>
</tr>
<tr>
<td>( debt )</td>
<td>0.207</td>
<td>-0.109</td>
<td>-0.622 **</td>
</tr>
<tr>
<td></td>
<td>(0.233)</td>
<td>(0.183)</td>
<td>(0.243)</td>
</tr>
<tr>
<td>( lnassets )</td>
<td>-0.126 ***</td>
<td>-0.294 ***</td>
<td>0.193 ***</td>
</tr>
<tr>
<td></td>
<td>(0.025)</td>
<td>(0.034)</td>
<td>(0.042)</td>
</tr>
<tr>
<td>( prop )</td>
<td>0.084</td>
<td>-0.273</td>
<td>0.898 **</td>
</tr>
<tr>
<td></td>
<td>(0.224)</td>
<td>(0.256)</td>
<td>(0.433)</td>
</tr>
<tr>
<td>constant</td>
<td>3.173 ***</td>
<td>4.617 ***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.203)</td>
<td>(0.281)</td>
<td></td>
</tr>
</tbody>
</table>

Observations 6,408 6,408 5,984
Lagrangian Multiplier (LM) \( \chi^2 \)-statistic 8833.280
\( (p \text{-value}) \) (0.000)
\( R^2 \) 16.1639% 12.520% 5.602%
Year dummies Yes Yes Yes
Table 7: Standard Fixed-Effects Estimation (FE) of Firm Value on CEO Pay

This table reports the FE regression results of firm value \( q_{it} \) on CEO pay-performance sensitivity measure, \( ppse \). The sample consists of \( N = 809 \) unique firms for the 14 years 2015-2018. All dependent and independent variables are defined in Table 3. Year-dummy effects are estimated in a first-stage auxiliary FE regression but their coefficients are not reported. Robust standard errors that incorporate firm-level clustering are reported in parentheses. \*, **, *** indicate that the parameter estimate is significantly different from zero at the 1%, 5%, or 10% level, respectively.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>( ppse )</td>
<td>0.0002464</td>
<td>( 0.000359 )</td>
</tr>
<tr>
<td>( xrd )</td>
<td>1.226</td>
<td>( 1.186 )</td>
</tr>
<tr>
<td>( adv )</td>
<td>-0.002</td>
<td>( 0.519 )</td>
</tr>
<tr>
<td>( debt )</td>
<td>-0.158</td>
<td>( 0.196 )</td>
</tr>
<tr>
<td>( lnassets )</td>
<td>-0.508***</td>
<td>( 0.062 )</td>
</tr>
<tr>
<td>( prop )</td>
<td>-0.885*</td>
<td>( 0.504 )</td>
</tr>
</tbody>
</table>

Observations 6,408

F-test statistic 135.850 (p-value) (0.000)

Hausman test \( \chi^2 \) statistic 142.420 (p-value) (0.000)

J-statistic 718.369 (p-value) (0.510)

R\(^2\) 7.818%

Year dummies Yes
Table 8: Efficient Fixed-Effects Estimation (EFE) of Firm Value on CEO Pay

This table reports the EFE regression results of firm value ($q_{it}$) on CEO pay-performance sensitivity measure, $ppse$. The sample consists of $N = 809$ unique firms for the 14 years 2005-2018. All dependent and independent variables are defined in Table 3. Year-dummy effects are estimated in a first-stage auxiliary FE regression but their coefficients are not reported. Robust standard errors that incorporate firm-level clustering are reported in parentheses. The standard errors are also adjusted for the degrees of freedom of the unbalanced panel. ***, **, * indicate that the parameter estimate is significantly different from zero at the 1%, 5%, or 10% level, respectively.

<table>
<thead>
<tr>
<th>Variables</th>
<th>demeaned</th>
<th>first-differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ppse$</td>
<td>0.0001412 ***</td>
<td>0.0001253 ***</td>
</tr>
<tr>
<td></td>
<td>( 0.0000001 )</td>
<td>( 0.0000002 )</td>
</tr>
<tr>
<td>$xrd$</td>
<td>0.3610 ***</td>
<td>-2.0536 ***</td>
</tr>
<tr>
<td></td>
<td>( 0.0071 )</td>
<td>( 0.0071 )</td>
</tr>
<tr>
<td>$adv$</td>
<td>0.6196 ***</td>
<td>1.0702 ***</td>
</tr>
<tr>
<td></td>
<td>( 0.0023 )</td>
<td>( 0.0019 )</td>
</tr>
<tr>
<td>$debt$</td>
<td>0.1403 ***</td>
<td>0.3265 ***</td>
</tr>
<tr>
<td></td>
<td>( 0.0026 )</td>
<td>( 0.0034 )</td>
</tr>
<tr>
<td>$lnassets$</td>
<td>-0.4648 ***</td>
<td>-0.4615 ***</td>
</tr>
<tr>
<td></td>
<td>( 0.0008 )</td>
<td>( 0.0009 )</td>
</tr>
<tr>
<td>$prop$</td>
<td>-0.9355 ***</td>
<td>-0.5668 ***</td>
</tr>
<tr>
<td></td>
<td>( 0.0088 )</td>
<td>( 0.0094 )</td>
</tr>
</tbody>
</table>

Observations | 6,408 | 6,408 |
$J$-statistic | 719.0068 | 720.0290 |
($p$-value) | ( 0.4405 ) | ( 0.4299 ) |
$R^2$ | 6.4286% | 4.7386% |
Year dummies | Yes | Yes |

65
Table 9: Estimation of Firm Value on CEO Pay at CEO Level

This table reports the regression results of firm value \((q_{it})\) on CEO pay-performance sensitivity measures, where \(i = 1, 2, \ldots, N\) denotes the different CEOs in our sample. The sample consists of \(N = 1,469\) unique CEOs for the 14 years 2005-2018. All variables and estimates are defined analogous to Tables 6, 7, and 8.

<table>
<thead>
<tr>
<th>Variables</th>
<th>FE demeaned</th>
<th>EFE first-differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>(ppse)</td>
<td>0.0003270 ***</td>
<td>0.0002063 ***</td>
</tr>
<tr>
<td></td>
<td>( 0.0000596 )</td>
<td>( 0.0000001 )</td>
</tr>
<tr>
<td>(xrd)</td>
<td>0.0830</td>
<td>-1.0446 ***</td>
</tr>
<tr>
<td></td>
<td>( 1.1925 )</td>
<td>( 0.0006 )</td>
</tr>
<tr>
<td>(adv)</td>
<td>0.2643</td>
<td>0.6686 ***</td>
</tr>
<tr>
<td></td>
<td>( 0.3817 )</td>
<td>( 0.0004 )</td>
</tr>
<tr>
<td>(debt)</td>
<td>-0.2825</td>
<td>0.2752 ***</td>
</tr>
<tr>
<td></td>
<td>( 0.2074 )</td>
<td>( 0.0011 )</td>
</tr>
<tr>
<td>(lnassets)</td>
<td>-0.6601 ***</td>
<td>-0.6367 ***</td>
</tr>
<tr>
<td></td>
<td>( 0.0677 )</td>
<td>( 0.0001 )</td>
</tr>
<tr>
<td>(prop)</td>
<td>-0.6636</td>
<td>-1.4350 ***</td>
</tr>
<tr>
<td></td>
<td>( 0.5177 )</td>
<td>( 0.0014 )</td>
</tr>
</tbody>
</table>

Observations 6,408 6,408 6,408

\(J\)-statistic 930.4408 939.4767 939.1225

\(p\)-value 0.3454 0.2422 0.2375

\(R^2\) 6.8568% 5.4000% 3.5079%

Year dummies Yes Yes Yes
Table 10: Simulation Results for Dynamic Estimators

This table reports results from 50,000 simulations of the dynamic data generating process in equation (15). The number of lags is $p = 1$ with a random AR-coefficient $\phi \in (0, 1)$ drawn from a uniform distribution. $c_i$ is an endogenous function of the time-series averages of the covariates plus an error term. $\eta_{it}$ is heteroskedastic over time and firms, and has AR(1) time-series dynamics, with autoregressive coefficient 0.6, but is independent across $i$. We let the effective time-series sample size be $T = 5$, $L_x = 5$, and $N = 1,000$. We simulate balanced as well as unbalanced panels. For the latter, we generate one missing time period for 50% of firms, for 20% of firms we assume two time periods are missing, and for 10% of firms we assume three time periods are missing. In ‘PANEL A’, we report the relative bias (i.e. the average of [estimate-true parameter]/[true parameter]) for balanced panels; in brackets we report the corresponding unbalanced-panel results. In ‘PANEL B’, we report the dispersion of the bias (i.e. the standard deviation of [estimate-true parameter]) for balanced panels; in brackets we report the corresponding unbalanced-panel results. In ‘PANEL C’, we report the size of the $J$-test with nominal size 5% for balanced panels; in brackets we report the corresponding unbalanced-panel results.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DFE</th>
<th>DEFE demeaned</th>
<th>DEFE first-differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\alpha}$</td>
<td>5.498</td>
<td>-5.598</td>
<td>-5.673</td>
</tr>
<tr>
<td></td>
<td>[13.642]</td>
<td>[4.273]</td>
<td>[4.378]</td>
</tr>
<tr>
<td>$\hat{\beta}_1$</td>
<td>-1.668</td>
<td>-3.519</td>
<td>-4.468</td>
</tr>
<tr>
<td></td>
<td>[14.202]</td>
<td>[3.422]</td>
<td>[3.562]</td>
</tr>
<tr>
<td>$\hat{\beta}_2$</td>
<td>-1.270</td>
<td>-2.704</td>
<td>-3.175</td>
</tr>
<tr>
<td></td>
<td>[12.542]</td>
<td>[2.612]</td>
<td>[2.769]</td>
</tr>
<tr>
<td>$\hat{\beta}_3$</td>
<td>4.575</td>
<td>1.382</td>
<td>0.838</td>
</tr>
<tr>
<td></td>
<td>[9.308]</td>
<td>[1.885]</td>
<td>[2.066]</td>
</tr>
<tr>
<td>$\hat{\beta}_4$</td>
<td>-2.277</td>
<td>0.627</td>
<td>0.139</td>
</tr>
<tr>
<td></td>
<td>[8.760]</td>
<td>[1.872]</td>
<td>[2.152]</td>
</tr>
<tr>
<td>$\hat{\phi}$</td>
<td>1.113</td>
<td>-1.628</td>
<td>-1.717</td>
</tr>
<tr>
<td></td>
<td>[9.139]</td>
<td>[1.139]</td>
<td>[1.069]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DFE</th>
<th>DEFE demeaned</th>
<th>DEFE first-differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\alpha}$</td>
<td>8.182</td>
<td>6.082</td>
<td>6.078</td>
</tr>
<tr>
<td></td>
<td>[25.479]</td>
<td>[8.067]</td>
<td>[8.053]</td>
</tr>
<tr>
<td>$\hat{\beta}_1$</td>
<td>7.893</td>
<td>5.796</td>
<td>5.793</td>
</tr>
<tr>
<td></td>
<td>[25.066]</td>
<td>[7.954]</td>
<td>[7.937]</td>
</tr>
<tr>
<td>$\hat{\beta}_2$</td>
<td>8.334</td>
<td>6.264</td>
<td>6.262</td>
</tr>
<tr>
<td></td>
<td>[28.874]</td>
<td>[8.999]</td>
<td>[8.990]</td>
</tr>
<tr>
<td>$\hat{\beta}_3$</td>
<td>8.226</td>
<td>6.139</td>
<td>6.139</td>
</tr>
<tr>
<td></td>
<td>[25.153]</td>
<td>[8.053]</td>
<td>[8.037]</td>
</tr>
<tr>
<td>$\hat{\beta}_4$</td>
<td>7.964</td>
<td>5.807</td>
<td>5.803</td>
</tr>
<tr>
<td></td>
<td>[24.763]</td>
<td>[7.944]</td>
<td>[7.927]</td>
</tr>
<tr>
<td>$\hat{\phi}$</td>
<td>8.990</td>
<td>7.352</td>
<td>7.332</td>
</tr>
<tr>
<td></td>
<td>[18.147]</td>
<td>[7.006]</td>
<td>[6.974]</td>
</tr>
<tr>
<td></td>
<td>PANEL C: Inference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>-------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( J - )test</td>
<td>3.646%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[ 2.370% ]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.742%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[ 3.030% ]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.734%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[ 3.006% ]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 11: Within-firm (Time-Series) Variation, and Between-firm (Cross-Sectional) Variation

This table reports the within-CEO (time-series) variation and between-CEO (cross-sectional) variation for the different variables in the IV 1 (IV 2) sample. In parenthesis, we present the proportion of total variation that is given by within-CEO and between-CEO variation. All variables are defined in Table 3. The sample consists of 915 observations for the sample for the 14 years 2005-2018.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total Variation</th>
<th>Within-Firm Variation (Percentage)</th>
<th>Between-Firm Variation (Percentage)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$q$</td>
<td>1436.17843775947</td>
<td>303.720385285374 (21.1478168241544 %)</td>
<td>1132.45805247410 (78.8521831758459 %)</td>
</tr>
<tr>
<td>$ppse$ ($\times 10^9$)</td>
<td>3226586946.13783</td>
<td>644966571.114262 (19.9891272691807 %)</td>
<td>2581620375.02356 (80.0108727308191 %)</td>
</tr>
<tr>
<td>$I^{\text{Predicted First Year}}$</td>
<td>183.934426229509</td>
<td>166.458119658120 (90.4986212045036 %)</td>
<td>17.4763065713886 (9.50137879549645 %)</td>
</tr>
</tbody>
</table>
Figure 2: Convergence of estimation error - The figure plots the average relative estimation error for the pooled coefficients $\alpha$ and $\beta$ from 999 simulations of the data generating process described in Table 1. It demonstrates the convergence properties of the POLS, RE, FD, FE, and EFE estimators that are discussed in the main text, as the number of firms $N$ increases from 200 to 5,000 in increments of 100. The figure also plots the average relative estimation error for the pooled coefficients $\alpha$, $\beta$, and $\phi$ from the dynamic data generating process described in Table 10, relying on the DFE and the DEFE estimators. The POLS estimation error is plotted on the right y-axis; all other estimators are on the left y-axis.
Figure 3: Standard deviation of estimation error - The figure plots the standard deviation of the estimation error for the pooled coefficients $\alpha$ and $\beta$ from 999 simulations of the data generating process described in Table 1. It demonstrates the variability of the POLS, RE, FD, FE, and EFE estimators that are discussed in the main text, as the number of firms $N$ increases from 200 to 5,000 in increments of 100. The figure also plots the variability of estimation error of the pooled coefficients $\alpha$, $\beta$, and $\phi$ from the dynamic data generating process described in Table 10, relying on the DFE and the DEFE estimators. The POLS estimation dispersion is plotted on the right y-axis; all other standard deviations are on the left y-axis.
Figure 4: Size/Power of the $J$-Statistic - The figure plots the rejection rates (in %) of the $J$-statistic from 999 simulations of the data generating process described in Table 1. It demonstrates the actual size/power of the $J$-test resulting from the POLS, RE, FD, FE, and EFE estimators that are discussed in the main text, as the number of firms $N$ increases from 200 to 5,000 in increments of 100. The figure also plots the size of the $J$-test resulting from the DFE and the DEFE estimator in the dynamic data generating process described in Table 10. The nominal test size is 5%. The power of the POLS and RE $J$-test is plotted on the right y-axis; the size of the $J$-test from all other estimators are on the left y-axis.