Estimation in the Presence of Unobserved Group-Level Heterogeneity*

Todd A. Gormley† and David A. Matsa‡

January 31, 2012

Abstract

Controlling for unobserved heterogeneity, such as industry-specific shocks, is a fundamental challenge in empirical research, as failing to do so can introduce omitted variables biases and preclude causal inference. This paper discusses limitations of two approaches commonly used to control for unobserved group-level heterogeneity in finance—demeaning the dependent variable with respect to the group (e.g., “industry-adjusting”) and adding the group’s mean as a control. We show that these popular techniques typically provide inconsistent coefficients and can lead researchers to incorrect inference. In contrast, the fixed effects estimator is consistent and should be used instead. We also explain how to estimate the fixed effects model when traditional methods are computationally infeasible.

JEL Codes: G3, G2, C01, C13
Keywords: unobserved heterogeneity, group fixed effects, industry-adjust, bias

---

* We thank Michael Anderson, Joshua Angrist, Michael Roberts, Michael R. Wagner, and seminar participants at Wharton for helpful comments. Matthew Denes, Christine Dobridge, Jingling Guan, and Kanis Saengchote provided helpful research assistance. Gormley thanks the Rodney L. White Center for Financial Research – Brandywine Global Investment Management Research Fellowship and the Cynthia and Bennett Golub Endowed Faculty Scholar Award for financial support.

† The Wharton School, University of Pennsylvania, 3620 Locust Walk, Suite 2400, Philadelphia, PA, 19104. Phone: (215) 746-0496. Fax: (215) 898-6200. E-mail: tgormley@wharton.upenn.edu

‡ Kellogg School of Management, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208. Phone: (847) 491-8337. Fax: (847) 491-5719. E-mail: dmatsa@kellogg.northwestern.edu
1. Introduction

Controlling for unobserved heterogeneity is a fundamental challenge in empirical finance research because most corporate policies—including financing and investment—depend on factors that are unobservable to the econometrician. If these factors are correlated with the variables of interest, then without proper treatment, omitted variables bias infects the estimated parameters and precludes causal inference. In many settings, important sources of unobserved heterogeneity are common within groups of observations. For example, unobserved factors—like investment opportunities—are often common across firms in an industry; as investment opportunities affect many corporate decisions, failing to control for them can cause serious identification challenges.¹

While the existing literature uses various estimation strategies to control for unobserved group-level heterogeneity, there is little understanding of how these approaches differ and under what circumstances each provides consistent estimates. Our paper examines this question and shows that some commonly used approaches typically lead to inconsistent estimates and can distort inferences.

We focus on two popular estimation strategies. The first, which we refer to as “adjusted means” (AM), is to demean the dependent variable with respect to the group before estimating the model with ordinary least squares (OLS). A common example is when researchers “industry-adjust” their dependent variable so as to remove common industry factors in a firm-level analysis. A second approach, which we refer to as “average effects” (AE), is to use the group’s mean as a control in the OLS specification. A common implementation of AE uses observations’ state-year mean to control for time-varying differences in local economic environments.

Both AM and AE have been used for at least two decades in empirical finance research, and continue to be used widely. Papers published in top finance journals—including the Journal of Finance, the Journal of Financial Economics, and the Review of Financial Studies—have used both approaches

¹ Potential unobserved factors abound: unobserved differences in local economic environments, management quality, and the cost of capital, to name a few, can pose equally problematic identification problems.
since at least the late 1980s. These estimation strategies continue to be used today; in an analysis of papers published in these three journals in 2008–2010, we found over 30 papers that employed at least one of the two techniques. The techniques are used to study a variety of finance topics, including capital structure, corporate boards, corporate governance, executive compensation, corporate control, banking, and return forecasting. Papers using these econometric techniques have also been published in the *American Economic Review*, the *Journal of Political Economy*, and the *Quarterly Journal of Economics*.

Our paper shows that, despite their popularity, the AM and AE estimators rarely provide consistent estimates of models with unobserved group-level heterogeneity. In the presence of such heterogeneity, the AM estimator suffers from an omitted variable bias if there is any within-group correlation across observations—either among or across independent variables in the model. Such correlations are likely in the presence of an unobserved group-level effect. The AE estimator is inconsistent for the same reason and also suffers from a measurement error bias because the group’s sample mean measures the true unobserved heterogeneity with error.

This shortcoming of the AM and AE estimators stands in stark contrast to the “fixed effect” (FE) estimator, which is another approach available to control for unobserved group-level heterogeneity. The FE estimator, which instead adds group indicator variables to the OLS estimation, is consistent in the presence of unobserved group-level heterogeneity.

The differences between the estimators are important because the AM and AE estimators can lead researchers to make incorrect inferences. We show that AM and AE estimates can be more biased than OLS and even yield estimates with the opposite sign of the true coefficient. AM and AE can also be inconsistent even in circumstances where the original OLS estimates would be consistent. When

---

2 The exact origin of the two estimators in finance is unclear, but one potential source is the event-studies literature where stock returns are regressed on the average market return to construct market-adjusted returns. This approach is qualitatively similar in that it uses a group mean—the average market return for a given period—to account for a potential group effect. In this regard, event studies’ analysis of market-adjusted returns is similar to AM, and estimation of the market model is similar to AE. An important difference in the event studies literature, however, is that the underlying statistical model assumes that each return is a function of the average market return whereas the statistical model underlying unobserved group-level heterogeneity does not.

3 Our analysis compares the consistency of estimators as the number of observations, \( N \), goes to infinity. Any references in this paper to “bias” does not refer to the finite-sample properties of the estimator, but rather the difference between the probability limit of the estimate and the true parameter as \( N \) goes to infinity.
estimating a few textbook finance models using each of the different techniques to control for unobserved heterogeneity, we find large differences between the AM, AE, and FE estimates and confirm that AM and AE can exhibit larger biases than OLS and yield coefficients of the opposite sign as FE. These differences confirm the presence of unobserved group-level heterogeneity in these settings and of correlations within these commonly used data structures that cause the AM and AE estimators to be inconsistent in practice.

Based on these findings, we argue that AM and AE should not be used to control for unobserved group-level heterogeneity. The same is also true of related techniques used in the literature to remove unobserved heterogeneity. In general, estimations that involve transforming the dependent variable, but not the independent variables, will yield inconsistent estimates. Similar to AM, more complicated transformations of the dependent variable, such as subtracting the mean or median of a comparable set of firms, will yield inconsistent estimates by failing to account for correlations across independent variables within these groups. Regression analyses of conglomerates’ diversification “discount” are subject to a similar critique. Even a simple comparison of industry-adjusted means before and after events—as is common in analyses of corporate control transactions, stock issues, and other sets of 0/1 events—does not reveal the true causal effect of the events because the comparison fails to adjust the implicit independent variable, the event indicator, to control for the share of firms in the industry that are affected.\textsuperscript{4} In contrast, FE estimators are consistent because they are equivalent to transforming both the dependent and independent variables so as to remove the unobserved heterogeneity. Of the approaches analyzed here, only FE provides consistent estimates in the presence of unobserved group-level heterogeneity.

The FE estimator, however, also has limitations. While the FE estimator controls for unobserved factors that vary across groups, it is unable to control for unobserved within-group heterogeneities. FE estimation also cannot identify the effect of independent variables that do not vary within groups and is subject to severe attenuation bias in the presence of measurement error. We discuss these limitations and provide guidance on when FE estimation is appropriate. We also describe how researchers can use

\textsuperscript{4} Put differently, by demeaning the data using an industry mean that includes both affected and unaffected firms, the simple comparison incorrectly removes part of the events’ effect from the estimate.
instrumental variable (IV) techniques within the FE framework to address concerns about attenuation bias and to identify the effects of independent variables that do not vary within groups.

We also address how researchers can estimate FE models in the presence of multiple types of unobserved heterogeneity. As the size and detail of datasets has increased, researchers are increasingly interested in controlling for multiple sources of unobserved heterogeneity, which can pose computational problems for fixed effects estimation. For example, in analyses of executive compensation, there is concern about unobserved heterogeneity both across managers and across firms (Graham, Li, Qui, forthcoming; Coles and Li, 2011a). Controlling for both firm- and manager-level fixed effects may also be important in other contexts (Coles and Li 2011b). Likewise, researchers who use firm-level data are increasingly concerned about time-varying heterogeneity across industries, such as industry-level shocks to demand (Matsa 2010). The presence of such heterogeneity may warrant the addition of industry-by-time fixed effects to a specification that already includes firm and time fixed effects. And in identification strategies that make use of regional changes in regulation over time, researchers may wish to include region-by-time fixed effects in addition to firm and time fixed effects to ensure the coefficients are estimated only using variation across observations within the same region and time period.

We discuss techniques that provide consistent estimates for models with multiple, high-dimensional group effects while avoiding the computational constraints of a standard FE estimator. One approach is to interact all values of the multiple group effects to create a large set of fixed effects in one dimension that can be removed by transforming the data. A second approach, which helps avoid potential attenuation biases and allows the researcher to estimate a larger set of parameters, is to maintain the multidimensional structure but to make estimation feasible by reducing the amount of information that needs to be stored in memory. This can be accomplished by using the properties of sparse matrices and/or by employing iterative algorithms. We discuss the relative advantages of each approach and how these techniques can be implemented easily in the widely used statistical package Stata.

---

5 When direct FE estimation requires a large number of indicator variables, the computer memory required can exceed the resources available to most researchers. In these cases, overcoming these difficulties has motivated some researchers to use AM or AE rather than FE.
Overall, our paper provides practical guidance on empirical estimation in the presence of unobserved group-level heterogeneity—a pervasive identification challenge in empirical finance research. A small but impactful set of recent papers have addressed other challenges researchers face. For example, Bertrand, Duflo, and Mullainathan (2004) and Petersen (2009) recommend methods to account for correlation across residuals in computing standard errors; Almeida, Campello, and Galvao (2010) and Erickson and Whited (forthcoming) compare methods used to account for measurement error in investment regressions; and Fee, Hadlock, and Pierce (2011) evaluate the use of F-tests on indicator variables in managerial style regressions.

The remainder of the paper is organized as follows. Section 2 describes the underlying identification concern of estimating a model with unobserved group-level heterogeneity and the FE estimator. Section 3 discusses two alternative estimation techniques—AM and AE—and when these techniques will lead to inconsistent estimates. Section 4 shows that differences between the estimation techniques can be important in practice. Section 5 discusses how one can estimate high-dimensional fixed effects models in practice, and Section 6 discusses limitations of the FE estimator and provides guidance on when its use is appropriate. Section 7 concludes.

2. Identification in the presence of unobserved group-level heterogeneity

We consider the case when an independent variable of interest, $X$, has a causal effect on a dependent variable, $y$, which also has unobserved group-level heterogeneity that is correlated with $X$. Assume the data exhibits the following structure:

\[ y_{i,j} = \alpha + \beta X_{i,j} + f_j + \varepsilon_{i,j} \]
\[ \text{var}(\varepsilon) = \sigma^2_{\varepsilon}, \mu_{\varepsilon} = 0 \]
\[ \text{var}(X) = \sigma^2_X \]
\[ \text{var}(f) = \sigma^2_f \]
\[ \text{corr}(f_j, \varepsilon_{i,j}) = 0 \]
\[ \text{corr}(X_{i,j}, \varepsilon_{i,j}) = \text{corr}(X_{i,j}, \varepsilon_{-i,j}) = 0 \]
\[ \text{corr}(X_{i,j}, f_j) = \rho_{Xf} \neq 0, \]

where the unit of analysis is $i$, these units are organized into larger groups, $j$, and there is some constant
unobserved group effect, \( f_j \), for each group \( j \). There are a total of \( N \) observations for the unit of analysis \( i \) with \( N_j \) observations in each group \( j \). In finance, \( i \) commonly refers to a firm, and \( j \) refers to the firm’s industry or year. In some cases, the group-level omitted factor is time-varying (such as industry shocks that vary over time). This latter example can be captured in the above model by adding an additional subscript \( t \) to each variable. In other cases, researchers worry about two types of unobserved heterogeneity, such as firm and year group effects in panel data. This can be captured in the above model by adding a second type of unobserved heterogeneity, \( \delta_i \).

Because there exists nonzero correlation, \( \rho_{Xf} \), between the unobserved group term and the independent variable of interest, failing to account for the unobserved heterogeneity causes an omitted variable problem. We assume the unobserved group term, \( f \), is the only omitted variable; that is, \( f \) and the independent variable of interest, \( X \), do not co-vary with the residual, \( \varepsilon \). For ease of exposition, we further assume that both the group term and independent variable are mean zero (i.e. \( \mu_X = \mu_f = 0 \)).

It is well known that using ordinary least squares (OLS) to estimate \( \beta \)—the causal effect of \( X \) on \( Y \)—when the data exhibit the structure specified in Equation (1) will yield an inconsistent estimate. OLS estimates the following specification:

\[
y_{i,j} = \alpha^{OLS} + \beta^{OLS} X_{i,j} + u^{OLS}_{i,j}
\]

And, the OLS estimate is

\[
\hat{\beta}^{OLS} = \beta + \rho_{Xf} \left( \frac{\sigma_f}{\sigma_X} \right)
\]

The OLS estimate of \( \beta \) is inconsistent when the correlation between the group term, \( f \), and the independent variable of interest, \( X \), is non-zero because of an omitted variable problem. By failing to control for the group term, \( \beta^{OLS} \) will reflect the causal effects of both \( X \) and \( f \) on the dependent variable \( Y \).

---

6 This assumption only simplifies the analysis and has no effect on the estimate of \( \beta \) under the different estimation techniques we analyze in Section 3. Assuming non-zero means affects the estimate of the constant, \( \alpha \).

7 Throughout the paper, we use the standard, large-sample approach to determine an estimate’s consistency by taking the number of observations, \( N \), to infinity while holding group size, \( N_j \), constant.
A fixed effects (FE) estimator, however, produces consistent estimates. FE estimation can be implemented by inserting an indicator variable for each group directly into the OLS equation. This estimation, which is also referred to as least squares dummy variable (LSDV) estimation, is consistent because it correctly controls for the unobserved group-level heterogeneity, \( f_j \). The FE estimate is consistent even if \( \rho_{Xf} = 0 \) and the original OLS estimate is consistent.

Equivalently, the FE estimator can be implemented by transforming the data to remove the unobserved group-level heterogeneity. This is often implemented by demeaning all of the variables—both the dependent and independent variables—with respect to the group and then estimating OLS on the transformed data. Specifically, fixed effects (FE) estimates

\[
y_{i,j} - \bar{y}_j = \alpha^FE + \beta^FE (X_{i,j} - \bar{X}_j) + u^FE_{i,j}
\]

where

\[
\bar{y}_j = \frac{1}{N_j \text{group}} \sum_{kj} (\alpha + \beta X_{k,i} + f_j + \varepsilon_{k,j})
\]

\[
\bar{X}_j = \frac{1}{N_j \text{group}} \sum_{kj} X_{k,j}.
\]

With a bit of algebra, the dependent variable in the FE specification can be written as:

\[
y_{i,j} - \bar{y}_j = \beta \left( X_{i,j} - \bar{X}_j \right) + \varepsilon_{i,j} - \frac{1}{N_j \text{group}} \sum_j \varepsilon_{k,j}
\]

Because the \( X \)'s covariance with \( \varepsilon \) is zero, as assumed in Equation (1), their covariance with the last term in Equation (6) is also zero and the FE estimate of \( \beta \) will be consistent.\(^8\)

This within-group transformation is particularly useful when the group effect is of high-dimension. Creating and storing a matrix of indicator variables, as is required for LSDV estimation, may be computationally infeasible when the dimension of the group effect is large.\(^9\) The within transformation

---

\(^8\) While the estimates are consistent, an adjustment to the standard errors is necessary to account for the degrees of freedom appropriately. Statistical software programs that estimate FE specifications make this adjustment automatically. See Wooldridge (2001), pp. 265–74, for more details about this adjustment.

\(^9\) In Section 5, we describe why these indicator variable estimators can be computationally infeasible, and we discuss estimation techniques that can be used to circumvent these computational problems.
reduces the number of parameters by eliminating the need to construct indicator variables. Because of this, statistical programs typically use this transformation to estimate models with fixed effects.\footnote{In panel datasets where the unit of analysis, $i$, is time (i.e., the “group” is a set of observations over time), first differencing can also be used to remove the unobserved group-level heterogeneity. First differencing only varies from this within transformation in its assumption about idiosyncratic errors. See Wooldridge (2001), pg. 284 for details on the relative efficiency of the two transformations.}

3. Alternative estimation techniques

Two other approaches are also used to control for unobserved group-level heterogeneity: “adjusted-means” (AM), which demeans the dependent variable with respect to the group before estimating the model with OLS, and “average effects” (AE), which uses the group’s mean of the dependent variable as a control in an OLS specification.

While both the AM and AE approaches are used widely in empirical finance research, the motivation for their use, rather than using FE, is often unclear. When there is more than one source of unobserved group heterogeneity, AM or AE are sometimes used to overcome computational difficulties in the FE estimator. The within transformation of the FE estimator only removes one group effect and using indicator variables to remove additional group effects may be infeasible.\footnote{In Section 5, we will discuss how specifications with two or more group effects can be estimated using more complicated transformations of the data, memory saving procedures, and iterative procedures. These alternative approaches yield consistent estimates of $\beta$, whereas, as shown shortly, the AM and AE estimators are typically inconsistent.} Because of this, the AM and AE estimates are sometimes combined with a FE estimator to control for two group effects: AM or AE is used to ‘adjust’ or ‘control’ for one group, and FE is used to remove the second group. In many other cases, however, the motivation for using AM or AE is less clear. Of the published papers we found using either AM or AE, applying a FE estimator appears feasible in almost all of the papers.\footnote{Of the more than thirty papers published in the Journal of Finance, Journal of Financial Economics, and Review of Financial Studies between 2008 and 2010 that use either AM or AE estimation, there were only three papers where computational problems might arise. In all but two of the other papers, there was only one fixed effect, which is easily handled using the within transformation, and in the other two cases, the amount of required memory to estimate the FE model would not pose a computational problem.}

In this section, we describe the AM and AE estimators in a simple model with one group effect, as in Equation (1). We show how these approaches differ from FE, describe the specific circumstances in which they yield consistent estimates, and discuss why they typically lead to inconsistent estimates of the
coefficient of interest, $\beta$, in the presence of unobserved group-level heterogeneity. The intuition for the bias extends to models with more complicated data structures.

### 3.1. Adjusted-means estimation

The AM estimator attempts to remove the influence of the group term on the dependent variable of interest by demeaning the dependent variable within each group. This is different than the fixed effects transformation, which deems both the dependent and independent variables. Adjusted-means estimation is applied, for example, at the industry level in firm-level panel datasets by subtracting out the industry-mean from the dependent variable of interest. When this adjustment is applied at the industry or industry-year level, researches typically refer to the dependent variable as being “industry-adjusted.”

While there are a variety of approaches used in practice, a common implementation of AM is to demean the dependent variable using the sample group’s mean, excluding the observation at hand.\(^{13}\)

More specifically, the researcher calculates the group mean, $\bar{y}_{-i,j}$, as

$$\bar{y}_{-i,j} = \frac{1}{N_j - 1} \sum_{k \in \text{group}_j} (\alpha + \beta X_{k,i,j} + f_j + \varepsilon_{k,j})$$

and estimates the following model using OLS:

$$y_{i,j} - \bar{y}_{-i,j} = \alpha + \beta X_{i,j} + u_{i,j}$$

Unlike FE, the adjusted-means estimation does not generally provide a consistent estimate of $\beta$ in the presence of unobserved group-level heterogeneity because the estimation suffers from an omitted variable problem. To see this, it is helpful to re-express the sample group mean as

$$\bar{y}_{-i,j} = \alpha + f_j + \bar{r}_{i,j},$$

where

$$\bar{r}_{i,j} = \frac{1}{N_j - 1} \sum_{k \in \text{group}_j} \left( \beta X_{k,j} + \varepsilon_{k,j} \right)$$

\(^{13}\)In some cases, the authors do not exclude the observation at hand, and in other cases, the median is used. Both of these alternative approaches yield similarly inconsistent estimates, and proof of this in the case when the mean of all observations is used is provided in the Appendix A6.
In the presence of unobserved group-level heterogeneity as in Equation (1), the dependent variable in AM estimation can be written as:

$$y_{i,j} - \bar{Y}_{i} = \beta X_{i,j} - \bar{X}_{i,j} + e_{i,j}$$  \hspace{1cm} (11)

Comparing Equations (9) and (11), we see that the adjusted-mean estimation fails to control for \( \bar{X}_{i,j} \), leading to a biased estimate for \( \beta \) if \( \bar{X}_{i,j} \) is correlated with the independent variable, \( X \). In other words, the covariance between \( X_{i,j} \) and the estimation error, \( u_{i,j}^{AM} \), is not zero when the correlation between \( \bar{X}_{i,j} \) and \( X_{i,j} \) is non-zero. But because \( \bar{X}_{i,j} \) includes the summation of the other \( X \) observations in the group, it will be correlated with \( X_{i,j} \) if the \( X \)’s are correlated within groups. Such a correlation is likely when the group effect, \( f_j \), is correlated with \( X_{i,j} \).

Letting \( \rho_{X_{i,j},X_{i,j}} \) represent the correlation between \( X \)’s, we can derive the sign and magnitude of the potential bias in the AE estimate of Equation (1).

Proposition 1: In the presence of unobserved group-level heterogeneity as in Equation (1), the adjusted means estimator yields an inconsistent estimate for \( \beta \). Specifically,

$$\hat{\beta}^{AM} = \beta \left( 1 - \rho_{X_{i,j},X_{i,j}} \right).$$

The proof of Proposition 1 is provided in the Appendix.

The direction and magnitude of the bias depends on the correlation between \( X \)’s within groups. \( \hat{\beta}^{AM} \) captures the effects of both \( X \) and \( \bar{X}_{i,j} \) on \( Y \). If the \( X \)’s are positively correlated within groups, such that \( \rho_{X_{i,j},X_{i,j}} \in (0,1) \), then the adjusted-mean estimate for \( \beta \) exhibits an attenuation bias, because \( \bar{X}_{i,j} \) has a negative effect on \( Y \) (see Equation (11)). And vice versa, if the \( X \)’s are negatively correlated within groups, the adjusted-mean estimate of \( \beta \) is biased away from zero. In practice, the bias on \( \beta \) is typically attenuating because it is unusual for the \( X \)’s to be negatively correlated within groups when there is also a group component, \( f \), that has non-zero correlation with \( X \) (as in Equation (1)).
The bias in the adjusted-mean estimation is present even with very large groups and even when standard OLS estimates are consistent. Because the AM estimator suffers from an omitted variable bias, increasing group size does not lessen the identification problem—the estimation’s error term will always contain the summation of the other within-group $X$’s, which are correlated with the independent variable $X_{i,j}$ when $\rho_{X_{i,j}X_{i-1}} \neq 0$. This is also why AM is inconsistent even when the correlation between $X$ and $f$ is zero and the OLS estimate is consistent. In this case, AM introduces a new omitted variable problem in its attempt to control for a non-existent omitted variable problem in the original OLS specification.

Another way to understand why AM provides inconsistent estimates is to compare it to a regression of $Y$ onto two independent variables, $X$ and $Z$. As is well known, a researcher interested in the effect of $X$ on $Y$ controlling for $Z$ can also identify this effect by regressing the residuals from a regression of $Y$ on $Z$ onto the residuals from a regression of $X$ on $Z$. Partialing out the effect of $Z$ from both $X$ and $Y$ before regressing $Y$ on $X$ is equivalent to regressing $Y$ on $X$ controlling for $Z$ (see Greene 2000, pp. 231-233 for more detail). The within transformation is simply the result of partialing out the effect of $Z$, which is a collection of indicator variables in the FE estimation, from both the independent and dependent variables. The AM estimation, however, is equivalent to only partialing out the effect of $Z$ from the dependent variable $Y$, which is not equivalent to regressing $Y$ on $X$ controlling for $Z$. By failing to transform the independent variable, $X$, the AM estimation fails to control for the variation in $X$ that is correlated with the unobserved group-level heterogeneity.\textsuperscript{14}

Using the same logic, other ways of demeaning the dependent variable in AM estimation, such as subtracting off a group median or subtracting off a value-weighted group-level mean, will also result in inconsistent estimates. Subtracting off the mean or median of a matched control sample (or a variable constructed based on a set of matched controls as is customary, for example, in studies of conglomerates’

\textsuperscript{14} Yet another way to understand why adjusted-mean estimation is inconsistent is to compare its approach to that of fixed effects estimation. Comparing the FE estimation in Equation (4) and the AM estimation in Equation (8) with the true underlying structure of the demeaned dependent variable in Equation (6), we see that the FE estimator correctly controls for the demeaned independent variable, $X_{i,j} - \bar{X}_j$. AM, however, only controls for $X_{i,j}$, which introduces a bias whenever the group mean is correlated with $X_{i,j}$, because this causes a non-zero correlation between $X_{i,j}$ and $u_{i,j}$. 

11
diversification “discount”) presents a similar issue. Demeaning the dependent variable but not accounting for the demeaned component of the independent variable creates an omitted variable bias whenever this demeaned component is correlated with a group member’s own independent variable.

Even if the true underlying data structure did not exhibit unobserved group-level heterogeneity and exactly matched the adjusted-mean specification, the AM estimator would still be biased. Suppose the underlying data exhibited the following structure:

\[ y_{ij} = \alpha + \beta X_{ij} + \bar{Y}_{-i,j} + \epsilon_{ij} \]  

(12)

This data structure contains a peer effect such that each observation within a group influences the other observations in the group. As shown in Manski (1993) and Leary and Roberts (2010), however, OLS estimation of the peer effects model does not reveal causal effects because of a reflection problem. Because the dependent variable has a causal effect on the dependent variable of other group members, using the group mean as an independent variable in OLS introduces an endogeneity problem.\(^{15}\)

**MULTIPLE INDEPENDENT VARIABLES**—The bias of the AM estimator becomes considerably more complex when there is more than one independent variable of interest. Suppose the true model is given by the following:

\[ y_{ij} = \alpha + \beta X_{ij} + \gamma Z_{ij} + f_j + \epsilon_{ij} \]

\[ \text{var}(\epsilon) = \sigma^2_{\epsilon}, \mu_\epsilon = 0 \]

\[ \text{var}(X) = \sigma^2_{X} \]

\[ \text{var}(f) = \sigma^2_{f} \]

\[ \text{var}(Z) = \sigma^2_{Z} \]

\[ \text{co} \text{rr}(f_j, \epsilon_{i,j}) = 0 \]

\[ \text{co} \text{rr}(X_{i,j}, \epsilon_{i,j}) = \text{co} \text{rr}(X_{i,j}, \epsilon_{-i,j}) = 0 \]

\[ \text{co} \text{rr}(Z_{i,j}, \epsilon_{i,j}) = \text{co} \text{rr}(Z_{i,j}, \epsilon_{-i,j}) = 0 \]

\[ \text{co} \text{rr}(X_{i,j}, f_j) = \rho_{Xf} \]

\[ \text{co} \text{rr}(X_{i,j}, Z_{ij}) = \rho_{XZ} \]

\[ \text{co} \text{rr}(Z_{i,j}, f_j) = \rho_{ZF}. \]

(13)

There is still just one type of unobserved group-level heterogeneity, \( f_j \), but there are two independent variables of interest, \( X \) and \( Z \). As before, assume that the independent variables, \( X_{ij} \) and \( Z_{ij} \), and the

\(^{15}\)For the same reason, both OLS and FE would also yield inconsistent estimates of the peer effects model.
unobserved group-level effect, \( f_j \), are uncorrelated with the errors. The two independent variables, however, are correlated with each other and with the unobserved group-level heterogeneity, \( f_j \). For ease of exposition, again assume without loss of generality that both the group term and independent variables are mean zero (i.e., \( \mu_X = \mu_Z = \mu_f = 0 \)).

**Proposition 2:** In the presence of unobserved group-level heterogeneity and two independent variables as in Equation (13), the adjusted means estimator yields inconsistent estimates for both \( \beta \) and \( \gamma \). Specifically,

\[
\hat{\beta}^{AM} = \beta + \gamma \left( \frac{\rho_{zX} \rho_{z_iX_{ij}} - \rho_{z_iX_{ij}}}{1 - \rho_{zX}^2} \right) + \beta \left( \frac{\rho_{zX} \rho_{z_iX_{ij}} - \rho_{z_iX_{ij}}}{1 - \rho_{zX}^2} \right)
\]

\[
\hat{\gamma}^{AM} = \gamma + \gamma \left( \frac{\rho_{zX} \rho_{z_iX_{ij}} - \rho_{z_iX_{ij}}}{1 - \rho_{zX}^2} \right) + \beta \left( \frac{\rho_{zX} \rho_{z_iX_{ij}} - \rho_{z_iX_{ij}}}{1 - \rho_{zX}^2} \right),
\]

where \( \rho_{z_iX_{ij}} \) and \( \rho_{z_iX_{ij}} \) represent the within-group correlations among the independent variables and \( \rho_{z_iX_{ij}} \) and \( \rho_{z_iX_{ij}} \) represent the within-group correlations across these variables.

The proof of Proposition 2 is provided in the Appendix.

The direction and magnitude of the bias in the AM estimator is ambiguous when there are two or more independent variables. The sign of \( \hat{\beta}^{AM} \) may not even match the sign of the true \( \beta \). The expression for \( \hat{\beta}^{AM} \) is more complicated in Proposition 2 than in Proposition 1, because there are now two omitted variables, the group means \( \bar{X}_{i_{-j}} \) and \( \bar{Z}_{i_{-j}} \), and each omitted variable may be correlated with each independent variable, \( X_{i_{-j}} \) and \( Z_{i_{-j}} \). FE, however, is still consistent when there are multiple independent variables because the estimator directly controls for the unobserved heterogeneity.

**DIFFERENCE-IN-DIFFERENCES**—Using AM in the context of a difference-in-differences estimator also yields inconsistent estimates. A properly specified difference-in-differences estimator compares the differences in means between treated and untreated observations across pre- and post-treatment periods. In many applications in finance, however—such as studies of mergers and acquisitions and leveraged buyouts—researchers instead compare the means of an industry-adjusted dependent
variable for treated firms across the two periods. This AM-type comparison, however, does not reveal the true causal effect of the event being analyzed because AM removes a weighted-average of treated and untreated firms in the industry, whereas the difference-in-differences estimator removes the mean of just the untreated firms.

Proposition 3: An AM-type comparison of pre- versus post-treatment period means for treated firms does not reveal the causal effect of the treatment. Specifically,

$$\hat{\beta}_{AM} = \theta \beta,$$

where $\theta$ is the average fraction of untreated firms in a treated firm’s group.

The proof of Proposition 3 is provided in the Appendix.

The AM-type comparison suffers an attenuation bias because it incorrectly removes part of the treatment’s effect on the dependent variable when it demeans the data using a weighted average of treated and untreated firms. The source of this bias is the same as the more general AM bias described above. In a regression context, the AM approach to difference-in-differences suffers an omitted variable problem in that it fails to control for the share of firms in the industry that are treated.

3.2. Average effects estimation

Average effects estimation is related to the adjusted-means approach. Subtracting the group’s sample mean, $\bar{y}_{i,j}$, from the dependent variable is equivalent to including the sample mean as a control variable in the regression while constraining its coefficient to equal 1. AE instead allows this coefficient to vary and estimates it in the regression. In this approach, the group’s sample mean, $\bar{y}_{i,j}$, is used as a proxy for the group fixed effect, $f_j$. It is again common to exclude the current observation when calculating the group sample mean for each observation. Specifically, when there is one independent variable of interest, the average effects approach estimates the following regression:

$$y_{i,j} = \alpha^{AE} + \beta^{AE} X_{i,j} + \lambda^{AE} \bar{y}_{i,j} + u_{i,j}^{AE}$$  \hspace{1cm} (14)

As shown in Appendix A6, not excluding the observation at hand also yields inconsistent estimates.
In the presence of unobserved group heterogeneity as in Equation (1), average effects estimation is inconsistent. The underlying problem is that AE suffers from a measurement error bias. As seen in Equation (9), \( f_j = \overline{y}_{i,j} - \alpha - \overline{r}_{i,j} \), meaning the groups’ sample means measure the group effects with error. Plugging this expression for \( f_j \) into the true model specified in Equation (1), we obtain

\[
y_{i,j} = \beta X_{i,j} + \overline{y}_{i,j} - \overline{r}_{i,j} + \epsilon_{i,j}
\]

(15)

Such measurement error does not only bias the coefficient on the mismeasured variable; it also biases coefficients for other variables in the estimated equation that are correlated with the mismeasured variable. In particular, the measurement error will lead to a bias for \( \hat{\beta}^{AE} \) because the mismeasured variable, \( f_j \), is correlated with the independent variable, \( X_{i,j} \).

Proposition 4: In the presence of unobserved group-level heterogeneity as in Equation (1), the average effects estimator yields an inconsistent estimate for \( \beta \). Specifically,

\[
\hat{\beta}^{AE} = \beta + \frac{\rho_{Xf} \left( \frac{\sigma_f}{\sigma_X} \right) \left( \beta^2 (1 - \rho_{X_{i,j}X_{i,j}}) + \frac{\sigma^2}{\sigma_X^2} \right) + \beta \left( \frac{\sigma^2_f}{\sigma_X^2} \right) (\rho_{Xf}^2 - \rho_{X_{i,j}X_{i,j}}) }{(1 - \rho_{X_{i,j}X_{i,j}}) \left( \frac{\sigma_f}{\sigma_X} \right)^2 - (\rho_{Xf}^2 - \rho_{X_{i,j}X_{i,j}}) \left( \beta^2 (1 - \rho_{X_{i,j}X_{i,j}}) + \frac{\sigma^2}{\sigma_X^2} \right) + \frac{\beta^2 \left( 1 - \rho_{X_{i,j}X_{i,j}} \right) + \frac{\sigma^2}{\sigma_X^2} }{N_j - 1}}.
\]

The proof of Proposition 4 is provided in the Appendix.

As before, another way to understand why AE provides inconsistent estimates is to compare it to a regression of \( Y \) onto two independent variables, \( X \) and \( Z \), where \( Z \) in this case happens to be a set of indicator variables. The AE approach is equivalent to regressing \( Y \) on \( X \) and the fitted values from a regression of \( Y \) on \( Z \), which is not the same as regressing \( Y \) on \( X \) and \( Z \). The AE estimation thus fails to control for the variation in \( X \) that is correlated with \( Z \), the unobserved heterogeneity.

The sign and magnitude of the bias for \( \hat{\beta}^{AE} \) is more complicated than for \( \hat{\beta}^{AM} \) because there are two distinct channels through which the bias arises. This is because part of the measurement error, \( \overline{r}_{i,j} \),...
is not just correlated with the mismeasured variable, \( \bar{y}_{-i,j} \), but is also correlated with \( X_{i,j} \) when the within-group correlation among the \( X \)'s is non-zero. Therefore, in addition to measurement error bias, there is also an omitted variable bias similar to that of AM.\(^{17}\)

Because of these two channels for bias, the AE estimate can be inconsistent even when OLS is consistent and when group sizes are very large. Even when \( \rho_{ij} = 0 \) (and OLS is consistent), AE will still be inconsistent if \( \rho_{x_{i,j},x_{-i,j}} \neq 0 \) because of the omitted variable. Likewise, large group size, \( N_j \), does not eliminate the bias. Similar to AM, the omitted variable bias is unrelated to group size, because \( \text{cov}(X_{i,j}, \bar{y}_{-i,j}) \) is not a function of \( N_j \). Large group size, \( N_j \), does decrease the separate measurement error bias, but it does not disappear completely.\(^{18}\)

3.3. Type 1 and 2 errors in OLS, AM, AE, and FE

**TYPE 1 ERROR**—The potential for incorrect inferences differs across the four approaches: OLS, AM, AE, and FE. In the presence of unobserved group-level heterogeneity, FE provides a consistent estimate of \( \beta \), whereas OLS, AM, and AE yield inconsistent estimates. When \( \beta = 0 \), however, both the AM and AE estimation techniques provide less biased coefficients than OLS.

Proposition 5: Assume \( \beta = 0 \) in Equation (1). The fixed effects and adjusted-means estimators for \( \beta \) are consistent. The average-effects estimator for \( \beta \) is inconsistent when \( N_j < \infty \), but is closer than the ordinary least squares estimator to the true \( \beta \).

The proof of Proposition 5 is provided in the Appendix.

AM avoids Type 1 errors entirely because AM’s bias is multiplicative.\(^{19}\) AE, however, can incorrectly reject the null hypothesis because the bias in AE is also additive, similar to the bias in OLS.

---

\(^{17}\) And as with AM, even if the true underlying data exhibit a peer effect, the AE estimator will not reveal causal effects because of the reflection problem.

\(^{18}\) The variance of the measurement error is given by \( \beta^2 \sigma_{x_{i,j}}^2 + \{\beta^2 (\sigma_{x_{i,j}} + \sigma_{x_{-i,j}}^2) + \sigma_{y_{i,j}}^2\}/(N_j - 1) \).

\(^{19}\) This is not true, however, when there is more than one independent variable. As shown in Proposition 2, the bias in the AM estimate is more complicated when there are two independent variables. With two independent variables as in Equation (13), AM only avoids Type 1 errors if both \( \beta = 0 \) and \( \gamma = 0 \).
When $\beta = 0$, the bias for AE, however, is less than that for OLS, and unlike OLS, the bias approaches zero when the number of observations per group, $N_j$, increases.

**TYPE 2 ERROR**—While both AM and AE are less biased than OLS when $\beta = 0$, this is not necessarily the case when $\beta \neq 0$. In some cases, AE and AM will provide a less biased estimate of $\beta$ than OLS, and in other cases, they will be *more* biased and possibly even have the wrong sign. Some examples of this are provided in Table I. Under certain parameters, OLS will actually be less biased then both AM and AE. But as shown in Table I, there are also cases where OLS is less biased than AE, but more biased then AM, and other cases where OLS is less biased than AM, but more biased then AE. There are also cases where the AE estimate actually has the opposite (and incorrect) sign.\(^{20}\)

The correlation between the independent variable of interest, $X$, and the unobserved group-level heterogeneity, $f$, has a large effect on the relative performance of each estimator. Figure 1 graphs OLS, AM, and AE estimates of Equation (1) as functions of various parameter values when $\beta = 1$. Each figure panel shows the effect of varying a specific parameter in the data structure while holding the rest constant. When not being varied, the default parameters values are: $\rho_{xf} = 0.25$; $\rho_{x,x} = 0.5$; $N_j = 10$; $\sigma_x / \sigma_x = 1$; and $\sigma_f / \sigma_x = 1$. Panel A plots the impact of the correlation between the independent variable, $X$, and the unobserved group-level heterogeneity, $f$, on each estimator. AM is less biased than the OLS only when the absolute magnitude of the correlation between $X$ and $f$ is large. This is because the AM bias is unaffected by $\rho_{xf}$, whereas the magnitude of the OLS bias increases linearly in the absolute value of $\rho_{xf}$. The AE bias, in contrast, is nonlinear in $\rho_{xf}$. Under these parameters, AE is extremely biased and can even reverse the sign of the coefficient for low values of $\rho_{xf}$, whereas for high values of $\rho_{xf}$, the AE is less biased then both OLS and AM.

---

\(^{20}\) In the case of just one independent variable, as in Table I, the AM estimator cannot incorrectly flip the sign of $\beta$. But as seen in Proposition 2 (and our later applications in Section 4), this is no longer true when there is more than one independent variable. In this case, both the AM and AE estimators can return an inconsistent estimate with the incorrect sign while the OLS estimate (also inconsistent) has the correct sign.
The correlation between observations within a group, \( \rho_{x_{ij}x_{ij}} \), is also important in terms of the relative performance of the estimators, but only for AM and AE. As shown in Panel B, the OLS estimate is unaffected by \( \rho_{x_{ij}x_{ij}} \), whereas the sign and magnitude of the bias for both AE and AM depend on \( \rho_{x_{ij}x_{ij}} \). As shown earlier, both AM and AE fail to control for the group mean of the independent variable, which is negatively correlated with the demeaned dependent variable and correlated with \( X_i \) when \( \rho_{x_{ij}x_{ij}} \neq 0 \). Because of this omitted variable problem, both AM and AE are positively biased when \( \rho_{x_{ij}x_{ij}} \) is sufficiently low (or negative in the case of AM), and negatively biased when \( \rho_{x_{ij}x_{ij}} \) is high.

An increase in the number of observations per group, \( N_j \), does not necessarily improve the performance of the various estimators. As shown in Panel C, the OLS and AM bias are unaffected by the number of observations per group since both estimators suffer from an omitted variable bias that is independent of \( N_j \). Bias in the AE estimate, however, is affected by \( N_j \), but under the parameters displayed in Figure 1, an increase in observations per group actually increases the bias. While increasing \( N_j \) reduces the measurement error noise, the AE estimator still suffers from an omitted variable bias similar to that of OLS and AM. On net, the reduction in noise leads the AE estimator to asymptote to a biased estimate. The same occurs when the variance of the error, \( \sigma^2 \), approaches zero. As shown in Panel D, the AE estimate asymptotes to the same biased coefficient as the noise from measurement error declines. For high values of \( \sigma^2 \), the AE estimate asymptotes to a different biased coefficient, which under these parameters, has a bias of the opposite sign as when \( \sigma^2 \) is low.

The relative variation of the unobserved group-level heterogeneity to that of the independent variable, \( \sigma_f / \sigma_x \), also has different implications for the various estimators. As shown in Panel E, the AM estimate is unaffected, whereas the magnitude of the bias is increasing with \( \sigma_f / \sigma_x \) for both OLS and AE—albeit in opposite directions.

In sum, whether AM or AE provides an improvement over OLS when \( \beta > 0 \) depends on the exact parameter values, but in all cases FE is consistent and preferable to the other three estimators.
4. Comparing approaches in common finance applications

In this section, we examine whether the differences between the various approaches are actually important in practice. We estimate standard empirical finance models using each of the various estimators—OLS, FE, AM, and AE—and compare the resulting estimates.

We start by estimating a standard capital structure OLS regression:

\[(D / A)_{i,t} = \alpha + \beta \mathbf{X}_{i,t} + \epsilon_{i,t}\]  

(16)

where \((D/A)_{i,t}\) is book leverage for firm \(i\) in year \(t\), and \(\mathbf{X}_{i,t}\) is a vector of variables thought to affect leverage. We use data for 1950–2010 from Compustat, and we include five independent variables: fixed assets/total assets, Ln(sales), return on assets, modified Altman-Z score, and market-to-book ratio. All variables are winsorized at the one percent level, and the standard errors are adjusted for clustering at the firm level. To account for possible unobserved firm-specific factors in the residual—\(\epsilon_{i,t}\)—that covary with \(\mathbf{X}_{i,t}\), we then estimate the model using the three different techniques analyzed here: FE, AM, and AE. The estimates are reported in Table II.

The various estimation techniques lead to very different estimates for \(\beta\). The OLS estimates, reported in Column (i) of Table II, differ considerably from the FE estimates in Column (ii) and the AM and AE estimates also differ [Columns (iii)-(iv)]. In some cases, the AM and AE estimates match the OLS estimate but differ from the FE estimate. For example, OLS, AM, and AE all estimate a coefficient of 0.011 for Ln(sales), while the FE estimate is more than 50% larger. In other cases, AM and AE produce estimates that are much further from the FE estimate than OLS. For the coefficient on the proportion of fixed assets, the OLS estimate is 9% larger than the FE estimate of 0.248, while the AM and AE estimates are 73% and 58% lower, respectively; and for z-score, OLS matches the FE estimate of -0.017, while AM and AE are off by about 40%. There is even a case where AM and AE yield an estimate that has the opposite sign as OLS and FE. As reported in Table II, the OLS estimate for return on assets is -0.015. Relying on the AM and AE estimates of 0.051 and 0.039, respectively, one might conclude that unobserved firm-level heterogeneity imposes a large downward bias on the OLS coefficient. But the FE estimate of -0.028—almost twice the OLS estimate—suggests the bias actually works the other way.
We next estimate the following model for executive compensation:

\[ \ln(\text{Total Compensation})_{i,j,t} = \alpha + \beta' X_{i,j,t} + \delta_t + \epsilon_{i,j,t} \]  \tag{17} 

where \( \ln(\text{Total Compensation})_{i,j,t} \) is the natural log of total compensation for manager \( i \), at firm \( j \), in year \( t \), \( X \) is a vector of variables thought to affect compensation, and \( \delta_t \) is a year fixed effect. Using data for 1992-2010 from Execucomp, Compustat, and CRSP, we estimate the model using each of the different estimators—OLS, FE, AM, and AE—to account for unobserved heterogeneity across managers. We include ten commonly included independent variables in our regressions: \( \ln(\text{total assets}) \), market-to-book ratio, contemporaneous and lagged stock returns, contemporaneous and lagged return on assets, volatility of daily log stock returns, and separate indicators for being a CEO, chairman, or female. The estimates are reported in Table III.

Similar to the earlier capital structure regressions, the various techniques lead to very different estimates of \( \beta \). For the coefficient on lagged return on assets, AM and AE estimates are statistically insignificant and an order of magnitude smaller than the statistically significant OLS and FE estimates. AM and AE estimates are as also considerably smaller than both the OLS and FE estimates for \( \ln(\text{Total Assets}) \), market-to-book ratio, stock returns, return on assets, and the CEO indicator. In many cases, the difference between AM/AE and FE is considerably larger than the difference between OLS and FE.

Lastly, we estimate the following model for firm value, as measured using Tobin’s Q:

\[ Q_{i,j,t} = \alpha + \beta' X_{i,j,t} + \delta_t + \epsilon_{i,j,t} \]  \tag{18} 

where \( Q_{i,j,t} \) is Tobin’s Q for firm \( i \), in 4-digit SIC industry \( j \), in year \( t \), \( X \) is a vector of variables thought to affect firm value, and \( \delta_t \) is a year fixed effect. Using Compustat data from 1962-2000, we estimate the model using each of the different estimators—OLS, FE, AM, and AE—to account for unobserved heterogeneity across industry-year combinations. Using AM to account for such heterogeneity is often referred to as analyzing “industry-adjusted” data. We include four commonly included independent variables in our regressions: an indicator to capture being incorporated in Delaware, \( \ln(\text{sales}) \), R&D expenses / assets, and return on assets. All variables are winsorized at the one percent level, and the
standard errors are adjusted for clustering at the firm level. The estimates are reported in Table IV.

As in the other examples, the various techniques lead to significantly different estimates of $\beta$. In particular, it is apparent that analyzing industry-adjusted data rather than using industry-year FE can distort inference. For the coefficient on Delaware incorporation, the AM and AE estimates are statistically insignificant and considerably smaller than the statistically significant OLS and FE estimates. A researcher relying on AM or AE would conclude that incorporation in Delaware is uncorrelated with firm value after accounting for industry trends, when the opposite is true. The AM and AE estimates for Ln(sales) and R&D expenses also differ considerably from the FE estimates.

Overall, the large differences between AM, AE, and FE suggest that researchers’ choice of which approach to use is important in practice. The differences between OLS and FE in the specifications we study here suggest that unobserved heterogeneity across groups in the OLS equation is correlated with the independent variables. As shown in Proposition 4, such correlations will cause AE estimates to be inconsistent. Based on Propositions 1 and 2, the large differences between AM and FE also imply there is considerable correlation among independent variables within groups (e.g., firm size is serially correlated). Such correlations will cause both the AM and AE estimates to be inconsistent in the presence of unobserved heterogeneity and can lead to severe biases and incorrect inferences.

5. High-dimensional fixed effects in finance

As highlighted in Section 2, controlling for group effects is necessary whenever the researcher believes there may be some unobserved heterogeneity across groups that may be correlated with both the dependent variable and independent variables. Estimating the model using OLS fails to account for these group effects and leads to inconsistent estimates. The AM and AE estimation techniques are also typically not consistent, and in practice, these approaches can lead to substantially different results than FE, which provides consistent estimates in the presence of unobserved group-level heterogeneity. Given this, it is clear that FE is preferred to AM and AE when FE is computationally feasible.

As researchers work with large datasets and face multiple sources of unobserved heterogeneity,
however, they face computational hurdles when trying to estimate FE models. Computational constraints can affect the estimation of FE models with two or more high-dimension group effects (i.e., each set of groups includes many distinct occurrences) because of the need to include many indicator variables in the estimation. The large number of parameters to be estimated in these models can lead to computer memory requirements that exceed the available computer resources.\footnote{Following Cornellissen (2008), assume that 8 bytes of memory is required to store each element of the design matrix $X$, which is used to calculate the OLS estimates, $\hat{\beta}= (X'X)^{-1}X'y$. With N observations and K independent variables, the total memory required to store the design matrix in an FE estimation with G indicator variables would be $N \times (G+K) \times 8$. This number can be quite large. For example, suppose that $N = 300,000$ (which is roughly the number of firm-year observations in Compustat from 1970-2008), $G = 30,000$ (which is roughly the number of firms in Compustat from 1970-2008), and $K = 50$ (which might include both year dummies and other variables of interest). The computer would require 72.12 gigabytes of memory to create the design matrix, which exceeds the amount computer memory available to most researchers.}

There are an increasing number of examples where finance researchers argue the need to control for multiple, high-dimensional fixed effects. For example, in the analysis of executive compensation, there may be concern about unobserved heterogeneity across managers (such as skill, risk aversion, personality) and unobserved heterogeneity across firms (such as firm culture and organization capital) that might also correlate with the variables of interest (Graham, Li, Qui, forthcoming, Coles and Li, 2011a). The inclusion of manager fixed effects, in addition to firm fixed effects, can be used to remove this unobserved heterogeneity and allow the researcher to remove potential omitted variable biases introduced by such unobserved heterogeneity at the manager- or firm-level. The inclusion of firm- and manager-level fixed effects may also be important in other contexts (Coles and Li 2011b). Likewise, researchers that use firm-level data are increasingly concerned about time-varying heterogeneity across industries, such as industry-level shocks to demand (Matsa 2010). The presence of such heterogeneity may warrant the addition of industry-by-time fixed effects to a specification that already includes firm fixed effects. And in identification strategies that make use of regional changes in regulation over time, there is a concern that these changes in regulation may coincide with other time-varying heterogeneity across observations in these regions (Cetorelli and Strahan 2006). The inclusion of region-by-time fixed effects in addition to unit of analysis fixed effects can ensure the researcher is only estimating coefficients using variation across observations within a given region-time.
In this section, we discuss alternative estimation techniques that can be used to arrive at consistent estimates even when the high-dimensionality of the group effects makes standard FE estimation techniques computationally infeasible. We also discuss some limitations of the fixed effects estimator and provide guidelines on how researchers can assess whether controlling for multiple, high-dimensional fixed effects is warranted.

5.1. Estimating a model with multiple high-dimensional fixed effects

When there is just one type of unobserved group effect, the fixed effects estimation is always computationally feasible if the original OLS estimation is feasible. This is because the data can be transformed by demeaning with respect to the group component and estimating OLS on the transformed data. In a specification with K independent variables of interest and G groups, the within transformation reduces the number of parameters to estimate from G+K to the original K parameters of OLS.

When there are two unobserved group effects of dimension G1 and G2, however, the transformation needed to reduce the number of parameters generally does not exist. If the data is a balanced panel, meaning there is consistent set of observations for each subgroup, the data can be transformed by demeaning the dependent variable and each independent variable with respect to each group sequentially and then estimating the regression using OLS (see Greene 2000, pp. 564-565 for more detail). This transformation reduces the number of parameters from G1+G2+K to K. But if the panel is unbalanced, which is far more common in practice, such a transformation does not generally exist.\(^{22}\)

In practice, unbalanced models with two unobserved group effects, such as firm and year group effects, are estimated using a partial transformation. The researcher inserts indicator variables for the smaller group directly in the specification and performs a within transformation with respect to the group with higher dimension. This combination of indicator variables and a within transformation yields consistent estimates for the K parameters of interest, and eases computational difficulties by reducing the

\(^{22}\) Wansbeek and Kapteyn (1989) proposed a transformation for cases where the data are unbalanced but patterned, such as with individual and time group effects, but this transformation is not typically used in practice because the number of time periods is usually not large enough to cause computational hurdles. See Baltagi (1995), pp.159–160, for further discussion.
number of estimated parameters from $G_1 + G_2 + K$ to $G_2 + K$, where $G_2 < G_1$. A common application of this is to insert year indicators in a panel estimation and demean with respect to the firm.

This partial transformation is only computational feasible, however, if one of the two groups is of low dimension such that creating a design matrix with $G_2 + K$ parameters is feasible. There are many examples for which this is not the case. For instance, suppose the researcher is working with a firm-level panel dataset and worries that there are both unobserved firm effects and time-varying industry shocks. In this case, both group effects are of high-dimension, particularly when the level of industrial classification is the three- or four-digit SIC, and estimation is infeasible.\(^{23}\)

As shown earlier, both average effects and adjusted-means provide an enticing (but incomplete) solution to this problem. With average effects, the researcher adds a control for each of the two high-dimensional group effects to the estimation. This reduces the number of parameters to be estimated from $G_1 + G_2 + K$ to just $K + 2$. And with adjusted-means, the researcher first demeans the dependent variable of interest with respect to one of the group-effects, and then does a within transformation to eliminate the second group effect, reducing the number of parameters to be estimated to $K$. As shown above, however, both the average effects and adjusted means approaches yield inconsistent estimates.

Fortunately, there exist computational techniques that provide consistent estimates for models with more than one high-dimensional group effect but avoid the need to store large matrices in memory. One approach is to interact the multiple types of unobserved heterogeneity into a one-dimensional set of fixed effects, which is accounted for using a within transformation. A second approach is to reduce the size of the matrix that needs to be stored in memory, and a third is to employ iterative algorithms. Each of these approaches has benefits and limitations.

\(^{23}\) For example, suppose a researcher wanted to estimate a model with both firm and 4-digit SIC*year fixed effects in Compustat using data from 1970-2008. The required memory would equal $N \times (G_2 + K) \times 8$, where $N$ is the number of observations. Since Compustat covers 450 4-digit SIC industries from 1970-2008, the FE estimation that uses a within transformation to remove the firm fixed effects would still require roughly $G_2 = 450 \times 38 = 17,100$ industry-year dummies. If the number of observations, $N$, is 300,000 (which roughly corresponds to the number of observations in Compustat) and the number of independent variables of interest, $K$, is equal to 5, then the required memory to create the design matrix is 38.23 gigabytes, where one gigabyte = $1,024^3$ bytes.
5.1.1. **Interacted fixed effects**

The first procedure for reducing memory requirements is to interact the multiple fixed effects into a one-dimensional set of fixed effects, which is accounted for using a within transformation. For example, in a model that is thought to include unobserved heterogeneity at the firm and industry-year level, the two types of fixed effects (firm and industry-year fixed effects) could be replaced with one firm-industry-year fixed effect and removed using a within transformation at the level of the firm-year. This within transformation removes both firm and industry-year unobserved heterogeneity and avoids the computational problems of trying to estimate a model with separate fixed effects for firms and industry-years since only the K original OLS parameters need to be estimated.

The interacted fixed effects approach, however, has potentially serious limitations. The first limitation is that interacted fixed effects remove more heterogeneity than necessary and may, as a result, severely limit the types of parameters that can be estimated. In the above example with firm-year fixed effects, a researcher would only be able to identify the effect of variables that vary within firm-years, whereas the original specification provides estimates for variables that vary within firms and within industry-years. Given most finance datasets only contain one observation per firm-year, using interacted fixed effects in this case is infeasible because there is no within variation left after including firm-year fixed effects! And even when some variation remains after transforming the data using interacted fixed effect, the estimates may suffer from serious attention biases because of measurement error.

A second limitation of interacted fixed effects estimation is that it does not allow the researcher to recover the uninteracted fixed effects. When the researcher seeks to analyze the distribution, correlation, and importance of the fixed effects for specific groups (such as manager/worker fixed effects, as in Abowd, Kramarz, and Margolis 1999; Abowd, Creecy, and Kramarz 2002; Coles and Li 2011a; Graham, 2016),

---

24 In the context of worker-firm matched data sets, this approach is referred to as the spell method (e.g., Andrews, Schank, and Upward 2006).
25 One example where interacted fixed effects are feasible is firm-manager matched data. An interacted manager-firm fixed effect can be used to remove unobserved heterogeneity both within managers and within firms. Such estimations are possible when managers work at a firm for more than one year. See Coles and Li (2011a, 2011b) and Graham, Li, and Qui (forthcoming) for examples of this approach.
26 The potential for severe attention biases with FE is discussed in Section 6.1.
Li, and Qui 2011), other estimation techniques that allow the researcher to recover the estimates on the separate fixed effects are required.

5.1.2. Memory saving procedures

Other approaches maintain the unobserved heterogeneity’s multidimensional structure but modify the algorithm used to estimate the original FE model. One possibility is to use memory-saving procedures. Memory requirements can be reduced by recognizing that the design matrix in estimations with many fixed effects is a sparse matrix. Sparse matrices—matrices with many zeros—can be compressed into smaller matrices that require less memory. This is accomplished by only storing the non-zero values of the matrix since the zero values will not contribute to later calculations. By compressing the matrix of indicator variables, the required memory can be significantly reduced (Abowd, Creecy, and Kramarz 2002). Cornelissen (2008) provides a detailed description of how this can be implemented, and provides a user-written program, FELSDVREG, that uses this estimation technique in Stata for a model with two high-dimensional fixed effects. The program uses indicator variables for the fixed effect of lower dimension and then applies the within transformation to remove to other fixed effect but makes use of the fact that the indicator variables represent a sparse matrix to save memory.27

Using this technique reduces the needed memory in the model with two high-dimensional fixed effects of dimension G1 and G2, where G2 < G1, from N×(G2+K)×8 to (G2+K)2×8 bytes. If N = 300,000, G2=17,100 industry-year dummies, and K = 5, then the required memory to create the design matrix is reduced from 38.23 gigabytes to 2.17 gigabytes. Unlike interacted fixed effects, using such memory-saving procedures allow the researcher to recover estimates for the original fixed effect using techniques described in Abowd, Kramarz, and Margolis (1999) and Abowd, Creecy, and Kramarz (2002).

While these procedures significantly reduce memory requirements, they still require a significant amount of computer memory in many cases. The memory required often still exceeds the resources available to most researchers. The next approach eliminates memory requirements entirely.

27 This approach was recently applied to firm-manager panel data sets by Coles and Li (2011a, 2011b) and Graham, Li, and Qui (forthcoming).
5.1.3. Solving FE using an iterative algorithm

A third approach avoids creating and saving to memory a large design matrix by estimating the FE model using an iterative algorithm. FE estimation of the linear model \( Y = X\beta + \varepsilon \), where \( X \) is the \( N \times (K+G_1+G_2) \) matrix of \( N \) observations, \( K \) independent variables, and \( G_1+G_2 \) indicators, and \( \beta \) is the \( (K+G_1+G_2) \times 1 \) vector of parameters to be estimated \( \beta_1, \beta_2, \ldots, \beta_K, f_1, \ldots, f_{I_1}, \delta_1, \ldots, \delta_{I_2} \) is given by the set of parameters, \( \hat{\beta}^{FE} \), that minimize the squared residuals, given by \( \hat{e}^{FE} = Y - X\hat{\beta}^{FE} \). It can be shown that \( \hat{\beta}^{FE} = (X'X)^{-1}X'Y \). An alternative way to solve for \( \hat{\beta}^{FE} \), however, is using a Gauss-Seidel algorithm. Specifically, an initial guess \( \hat{\beta}^{(0)} \) is made (i.e. \( \beta_1^{(0)}, \beta_2^{(0)}, \ldots, \beta_K^{(0)}, f_1^{(0)}, \ldots, f_{I_1}^{(0)}, \delta_1^{(0)}, \ldots, \delta_{I_2}^{(0)} \)). Using this guess, solve for the \( \beta_1^{(1)} \) that minimizes the squared residuals using \( \beta_2^{(0)}, \ldots, \beta_K^{(0)}, f_1^{(0)}, \ldots, f_{I_1}^{(0)}, \delta_1^{(0)}, \ldots, \delta_{I_2}^{(0)} \). Then update the initial guess and solve for the \( \beta_2^{(1)} \) that minimizes the squared residuals using the updated guess for \( \beta_1^{(1)} \). Repeat this zigzag algorithm until it converges.\(^{28}\)

A key benefit of this iterative algorithm is that it does not need to construct and store into memory the potentially large design matrix \( X \) in order to obtain the consistent OLS estimates. This method thus avoids memory limitations that bind even after implementing memory saving procedures. Guimarães and Portugal (2010) apply such a method to show how one can estimate the OLS coefficients and standard errors in a model with two high-dimensional fixed effects, and the user-written program REG2HDFE implements this algorithm in Stata. When computer memory is not a binding constraint, however, a limitation of the iterative algorithm is that it can take longer to compute when a large number of iterations are required.

5.1.4. Comparing estimation techniques

These alternative estimation techniques for models with more than one high-dimension group effect are computationally feasible and relatively quick. To illustrate this, we attempted to estimate a

\(^{28}\) See Smyth (1996) and Guimarães and Portugal (2010) for more details on how this algorithm is performed, particularly in the presence of two high-dimension fixed effects.
capital structure regression using the standard FE approach and the alternative techniques. Specifically, we estimated the following regression:

\[
\left(\frac{D}{A}\right)_{ijt} = \alpha + \beta X_{ijt} + f_i + \delta_{jt} + u_{ijt}
\]  

(19)

where \(X_{ijt}\) is a vector of five time-varying controls (i.e. \(K=5\)) for firm \(i\), in 4-digit SIC industry \(j\), and year \(t\). The five independent variables we use are fixed assets/total assets, ln(sales), return on assets, modified Altman-Z score, and market-to-book ratio. The regression includes two high-dimension group fixed effects: firm fixed effects, \(f_i\), and industry-by-year fixed effects, \(\delta_{jt}\). We use a sample of Compustat that covers 1970-2008, which contains \(N = 318,808\) firm-year observations, \(G_1 = 28,365\) unique firms, 450 4-digit SIC industries, and \(G_2 = 16,769\) unique industry-years.

Estimating this model using a combination of a within transformation and indicator variables is not computationally feasible in a standard econometrics program, Stata. Using the within-transformation to remove the firm fixed effects still leaves the researcher with 16,769 industry-year dummies. If the memory required to store each element of the design matrix is 8 bytes, the total memory required to store the design matrix of the transformed data would be \(N \times (G_2 + K) \times 8\). In our setup, this would require 39.84 gigabytes of memory. Even if we reduce the industry to the three-digit SIC level before constructing the industry-year fixed effects, we are still left with 10,527 industry-year dummies, and the required memory is 26.86 gigabytes, which still exceeds the memory available to most researchers (including the authors).

The memory-saving and iteration techniques we discuss above avoid such limitations. (Using interacted fixed effects in this case is not possible because there does not exist any within firm-year variation.) Using the FELSDVREG algorithm to estimate the FE model reduces the required memory to \((G_2 + K)^2 \times 8 = 2.09\) gigabytes. While computational times will obviously vary based on computing resources, the FELSDVREG estimation was able to successfully generate the OLS estimates with standard errors clustered at the firm level on the author’s computer in about 7 hours and 50 minutes. The REG2HDFE command was considerably faster. Using the iterative approach, the REG2HDFE command successfully returned estimates with clustered standard errors in less than five minutes.
6. When is fixed effects estimation appropriate?

In this section, we discuss limitations of the FE estimator and under what criteria using FE is indicated. We also discuss estimation strategies to use when FE is inappropriate.

6.1. Limitations of fixed effects estimation

While fixed effects estimation is effective in controlling for unobserved group-level heterogeneity, it has limitations. First, FE is unable to control for unobserved heterogeneity within groups. For example, industry-level fixed effects do not control for unobserved geographic heterogeneity across local markets within an industry, such as differences in investment opportunities. Likewise, firm-level fixed effects do not control for unobserved firm heterogeneity that varies over time, such as the accumulation of organization capital. Second, FE cannot directly identify the effect of independent variables that do not vary within groups. For example, a researcher interested in understanding the effect of a gender on managers’ total compensation would typically not be able to do so when manager-level fixed effects are included in the estimation because such time-invariant characteristics are perfectly collinear with the manager fixed effect.\(^{29}\)

FE estimation is also subject to potential attenuation biases because of measurement error. While FE estimators successfully remove unobserved heterogeneity that would otherwise bias estimates, they can also remove a substantial amount of the meaningful variation in an independent variable of interest (all the more so when interacted fixed effects are used). If some of the within-group variation is due to measurement error, then the share of variation being exploited that is noise can rise sharply in FE estimation. This can cause severe attenuation bias and lead the researcher to infer that the independent variable of interest does not have an effect on the dependent variable, when the opposite may be true.

For example, a researcher interested in estimating the effect of a regulatory change on banks’ allocation of credit to low income households could obtain data from a credit reporting bureau and control

\(^{29}\) Under some circumstances, however, the effect of variables that do not vary within groups can be recovered from the FE estimation (Hausman and Taylor 1981). See Section 6.2 for more detail.
for unobserved shocks to credit demand by including zip code-by-quarter fixed effects. But if there is measurement error in the credit reporting data (e.g., some loan originations and payoffs are not recorded in a timely manner), then the importance of this error could be magnified by the FE strategy. The FE estimation could indicate that the regulatory change had little or no effect on low income borrowing, when the true effect was large but mostly absorbed by the fixed effects, resulting in severe attenuation bias. This magnification of a measurement error bias is particularly acute when much of the variation in the independent variable of interest occurs across, rather than within, groups.30

6.2. So what should I do?

AM and AE should not be used to control for unobserved group-level heterogeneity, because they introduce additional biases relative to FE. But given the potential limitations of FE (which are also limitations of AM and AE), when should researchers estimate a model with fixed effects and what are the alternatives? There are four criteria that indicate estimating a model with fixed effects:

1. There likely exists unobserved group-level heterogeneity.
2. The heterogeneity is potentially correlated with an independent variable of interest.
3. There exists within-group variation in the independent variable of interest.
4. The independent variable of interest is well measured.

Conditions (1) and (2) represent existence criteria for unobserved group-level heterogeneity that could cause an omitted variable bias. If the factor is observable, then the researcher can control for potential omitted variable directly using OLS.31 If the heterogeneity is unobservable but uncorrelated with the independent variable of interest, then OLS without the group indicator variables is also consistent. Conditions (1) and (2) provide the motivation for using a FE estimator, which, unlike the AM and AE estimators, can successfully control for the unobserved group-level heterogeneity.

31 If the researcher is unsure about the functional form of the relationship between this group-level variable and the dependent variable, however, using fixed effects to absorb this variation can still be useful because it allows for a more general set of correlations.
Conditions (3) and (4) refer to the ability to obtain correct inferences from FE estimation. If there is no within-group variation in the independent variable of interest, it is not possible to disentangle the group component from that of the independent variable in FE estimation. Instead, an alternative estimation strategy, such as instrumental variables (IV), is required. And even if there is some variation, the researcher needs to proceed with caution when the FE estimation suggests the independent variable has a small or zero effect on the dependent variable. If the within-group variation is partially driven by measurement error, there may be a severe attenuation bias.

In some cases, violations of conditions (3) and (4) can be addressed using IV strategies within the FE estimation framework. As shown in Hausman and Taylor (1981), the coefficients on variables with no within-group variation can be recovered using a two-step procedure when some of the covariates that vary within groups are uncorrelated with the unobserved group-level heterogeneity. In the first step, the FE estimation is used to estimate the coefficients for variables that vary within groups. In the second step, group-average residuals from the first step are regressed on the covariates that do not vary within groups using as instruments covariates that vary within groups and are not correlated with the unobserved group-level heterogeneity. This estimation can be implemented in Stata using the XTHTAYLOR command.

When a researcher suspects measurement error bias may be present, standard techniques to address measurement error can be applied. The typical solution is to identify an instrument for within-group variation in the independent variable and implement standard IV methods. The instrument will typically be a variable outside the data structure that meets the standard relevance and exclusion criteria. In panel data, another option is to recover the true $\beta$ from the biased coefficients obtained from OLS estimation on different transformations of the data, such as first differences or within transformation. If the measurement errors are serially uncorrelated, the different data transformations affect the estimate in predictable ways that can be used to analytically solve for the bias and recover the true $\beta$ (see Griliches and Hausman 1986).\footnote{Even if the measurement errors are serially correlated, it may be possible to identify the true parameters if the researcher is willing to make additional assumptions. See Griliches and Hausman (1986) for more details.}
7. Conclusion

As empirical researchers, it is well understood that we must overcome the identification challenge posed by unobserved heterogeneity if we hope to infer causal statements from data we analyze. It is less clear, however, how researchers can best account for such heterogeneity. In practice, there are numerous methodologies that are used widely to account for unobserved heterogeneity that is common across groups of observations. One approach is to control for correlations with the mean of the group, and a second approach is to remove the unobserved heterogeneity by demeaning the dependent variable within groups. A third approach—fixed-effects estimation—is to include indicator variables for each group as additional controls, or equivalently, demean all of the model variables within groups (not just the dependent variable). This paper provides explores how these various approaches differ and under what circumstances they will provide consistent estimates of the parameters of interest.

We find that only the fixed effects approach yields consistent estimates in the presence of unobserved group-level heterogeneity while the other widely used approaches typically yield inconsistent estimates. Demeaning the dependent variable within groups yields inconsistent estimates because it suffers from an omitted variable problem by failing to account for the within-group mean of the independent variables. Controlling directly for the within-group mean of the dependent variable suffers from a similar omitted variable bias, and also exhibits measurement error bias because the within-group mean measures the unobserved group-level heterogeneity with error.

The difference between the various approaches is important in practice. Besides providing inconsistent estimates, the alternative approaches can lead to severe biases and incorrect inferences. They have the potential to generate estimates whose magnitude and sign do not match the true parameter. Estimating textbook finance models using each of the approaches, we confirm that these biases can be severe in practice. Compared to the FE estimates, the alternative approaches often result in very different estimates, and occasionally, return statistically significant estimates of the opposing sign.

While we show fixed effect estimation to be the best way to account for unobserved
heterogeneity, the estimation strategy has limitations. It cannot control for unobserved heterogeneity that occurs within groups; nor is it able to identify the causal effect of independent variables that do not vary within groups. The estimates are also subject to severe attenuation biases when the within-group variation is measured with error. Given these limitations, we provide guidance on when fixed effects are warranted and how one can potentially overcome concerns about measurement error in the estimation.

We also address how researchers can overcome computational difficulties when estimating fixed effects models that include multiple sources of unobserved heterogeneity across many groups. This type of estimation is increasingly common as researchers work with large datasets and attempt to account for more sources of unobserved group-level heterogeneity in their analyses. We describe new techniques that can be used to estimate the FE model when standard approaches are computationally infeasible. These new techniques are likely to be of increasing importance and of practical use to researchers.

References


Figure 1 -- Comparative Statics on Bias of OLS, AM, and AE when $\beta = 1$

This figure presents analytical solutions for the OLS, AM, and AE estimates of the $\beta$ in Equation (1),

$$y_{ij} = \alpha + \beta X_{ij} + f_j + \epsilon_{ij},$$

as a function of underlying parameters ($\rho_{Xf}$, $\rho_{X(i)X(-i)}$, $\sigma_f/\sigma_X$, and $\sigma_e/\sigma_X$) and the number of observations in a group ($N_j$) when $\beta = 1$. Each figure panel shows the effect of varying a specific parameter in the data structure while holding the rest constant. The vertical axis in each graph displays the estimated $\beta$. When not varying along the horizontal axis, the default parameter values are: $\sigma_f/\sigma_X=1; \sigma_e/\sigma_X=1; \rho_{X(i)X(-i)}=0.5; \rho_{Xf}=0.5; N_j=10$.

**Panel A. Correlation between $X$ and $f$**

**Panel B. Correlation between $X_i$ and $X_{-i}$**
Figure 1 continued

Panel C. Number of observations per group

Panel D. Relative variation in $\varepsilon$ and $X$

Panel E. Relative variation in $f$ and $X$
Table I
Analytical Examples of Bias with OLS, AM, and AE

This table reports analytical solutions for the OLS, AM, and AE estimators of Equation (1),
\[ y_{ij} = \alpha + \beta x_{ij} + f_j + \epsilon_{ij}, \]
under different assumptions about the underlying data structure. In all cases, the true coefficient (\( \beta \)) equals 1 and the number of observations per group (Nj) equals 10, and \( \sigma_{\epsilon}/\sigma_X \) equals 1. The assumed \( \text{corr}(X_{ij}, X \text{-} i,j), \text{corr}(X_{ij}, f_j), \) and \( \sigma_f/\sigma_X \) are given in columns (i)-(iii), and the OLS, AM, and AE estimates are given in columns (vii)-(ix).

<table>
<thead>
<tr>
<th>Underlying Data Structure</th>
<th>Estimates of ( \beta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \rho_{X_{ij},X_{i,j}} )</td>
<td>( \rho_{X_f} )</td>
</tr>
<tr>
<td>(i)</td>
<td>(ii)</td>
</tr>
<tr>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>-0.05</td>
<td>0</td>
</tr>
</tbody>
</table>

**OLS is less biased than AM, AE**

| 0.5 | -0.7 | 1 | 1 | 10 | 1 | 0.30 | 0.50 | -3.88 |
| 0.25 | -0.5 | 1 | 1 | 10 | 1 | 0.50 | 0.75 | -0.56 |
| 0.1 | -0.4 | 1 | 1 | 10 | 1 | 0.60 | 0.90 | 0.18 |

**AM least biased, AE most biased**

| 0.5 | 0.35 | 1 | 1 | 10 | 1 | 1.35 | 0.50 | 0.81 |
| 0.75 | 0.5 | 1 | 1 | 10 | 1 | 1.50 | 0.25 | 0.68 |
| 0.25 | 0.45 | 1 | 1 | 10 | 1 | 1.22 | 0.75 | 1.05 |

**AE least biased, AM most biased**

| 0.25 | 0.5 | 1 | 1 | 10 | 1 | 1.50 | 0.75 | 1.05 |
| 0.5 | 0.6 | 1 | 1 | 10 | 1 | 1.60 | 0.50 | 0.98 |
| -0.05 | 0.2 | 1 | 1 | 10 | 1 | 1.20 | 1.05 | 1.09 |
Table II
OLS, FE, AM, and AE Estimates in a Capital Structure Regression with Unobserved Group-Level Heterogeneity Across Firms

This table reports coefficients from firm-panel regressions of book leverage on fixed assets / total assets, Ln(sales), return on assets, modified Altman Z-score, and market-to-book ratio using different methodologies to account for unobserved group-level heterogeneity across firms. The data are from Compustat from 1950-2010 and exclude financial and regulated industries. All variables were winsorized at the one percent tails. Column (i) reports the OLS estimates; Column (ii) reports the FE estimates; Column (iii) reports the AM estimates; and Column (iv) reports the AE estimates. Standard errors, clustered at the firm level, are reported in parentheses. *** significant at 1% level.

<table>
<thead>
<tr>
<th>Dependent Variable = Book Leverage</th>
<th>OLS (i)</th>
<th>FE (ii)</th>
<th>AM (iii)</th>
<th>AE (iv)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fixed Assets/ Total Assets</strong></td>
<td>0.270***</td>
<td>0.248***</td>
<td>0.066***</td>
<td>0.103***</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.014)</td>
<td>(0.004)</td>
<td>(0.004)</td>
</tr>
<tr>
<td><strong>Ln(sales)</strong></td>
<td>0.011***</td>
<td>0.017***</td>
<td>0.011***</td>
<td>0.011***</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td><strong>Return on Assets</strong></td>
<td>-0.015***</td>
<td>-0.028***</td>
<td>0.051***</td>
<td>0.039***</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.004)</td>
<td>(0.004)</td>
</tr>
<tr>
<td><strong>Z-score</strong></td>
<td>-0.017***</td>
<td>-0.017***</td>
<td>-0.010***</td>
<td>-0.011***</td>
</tr>
<tr>
<td></td>
<td>0.000</td>
<td>(0.001)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td><strong>Market-to-book Ratio</strong></td>
<td>-0.006***</td>
<td>-0.003***</td>
<td>-0.004***</td>
<td>-0.004***</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td><strong>Observations</strong></td>
<td>166,974</td>
<td>166,974</td>
<td>166,974</td>
<td>166,974</td>
</tr>
<tr>
<td><strong>R-squared</strong></td>
<td>0.29</td>
<td>0.66</td>
<td>0.14</td>
<td>0.56</td>
</tr>
</tbody>
</table>
Table III
OLS, FE, AM, and AE Estimates in a Compensation Regression with Unobserved Group-Level Heterogeneity Across Managers

This table reports coefficients from manager-level panel regressions of Ln(Total Compensation) on year fixed effects and control variables using different methodologies to account for unobserved group-level heterogeneity across managers. The data are from Execucomp, Compustat, and CRSP from 1992-2010. Column (i) reports the OLS estimates; Column (ii) reports the FE estimates; Column (iii) reports the AM estimates; and Column (iv) reports the AE estimates. The included control variables are: Log(Total Assets), lagged market-to-book ratio, contemporary and lagged stock returns, which is calculated using a 12 month holding period, contemporary and lagged return on assets, which is calculated using income before extraordinary items / total assets, volatility of annualized daily log stock returns, an indicator for whether the CEO is chairman of the board, an indicator for whether the manager is the CEO, and an indicator for being female. Standard errors, clustered at the firm level, are reported in parentheses. *** significant at 1% level; ** significant at 5% level; * significant at 10% level.

<table>
<thead>
<tr>
<th>Dependent Variable = Ln(Total Compensation)</th>
<th>OLS (i)</th>
<th>FE (ii)</th>
<th>AM (iii)</th>
<th>AE (iv)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ln(Total Assets)</td>
<td>0.341***</td>
<td>0.240***</td>
<td>0.021***</td>
<td>0.066***</td>
</tr>
<tr>
<td>(0.007)</td>
<td>(0.014)</td>
<td>(0.002)</td>
<td>(0.005)</td>
<td></td>
</tr>
<tr>
<td>Market-to-Book Ratio [t-1]</td>
<td>0.093***</td>
<td>0.032***</td>
<td>0.009***</td>
<td>0.021***</td>
</tr>
<tr>
<td>(0.011)</td>
<td>(0.008)</td>
<td>(0.002)</td>
<td>(0.003)</td>
<td></td>
</tr>
<tr>
<td>Stock Return [t]</td>
<td>0.120***</td>
<td>0.091***</td>
<td>0.039***</td>
<td>0.050***</td>
</tr>
<tr>
<td>(0.015)</td>
<td>(0.014)</td>
<td>(0.008)</td>
<td>(0.009)</td>
<td></td>
</tr>
<tr>
<td>Stock Return [t-1]</td>
<td>0.041***</td>
<td>0.076***</td>
<td>0.053***</td>
<td>0.051***</td>
</tr>
<tr>
<td>(0.011)</td>
<td>(0.010)</td>
<td>(0.007)</td>
<td>(0.009)</td>
<td></td>
</tr>
<tr>
<td>Return on Assets [t]</td>
<td>0.287***</td>
<td>0.268***</td>
<td>0.092***</td>
<td>0.120***</td>
</tr>
<tr>
<td>(0.062)</td>
<td>(0.083)</td>
<td>(0.032)</td>
<td>(0.033)</td>
<td></td>
</tr>
<tr>
<td>Return on Assets [t-1]</td>
<td>0.135**</td>
<td>0.193***</td>
<td>0.004</td>
<td>0.023</td>
</tr>
<tr>
<td>(0.068)</td>
<td>(0.067)</td>
<td>(0.034)</td>
<td>(0.036)</td>
<td></td>
</tr>
<tr>
<td>Volatility of Daily Ln(Returns)</td>
<td>0.132***</td>
<td>0.002</td>
<td>0.002</td>
<td>0.021</td>
</tr>
<tr>
<td>(0.030)</td>
<td>(0.023)</td>
<td>(0.012)</td>
<td>(0.013)</td>
<td></td>
</tr>
<tr>
<td>CEO = Chairman Indicator</td>
<td>0.225***</td>
<td>0.028*</td>
<td>0.045***</td>
<td>0.071***</td>
</tr>
<tr>
<td>(0.017)</td>
<td>(0.016)</td>
<td>(0.012)</td>
<td>(0.012)</td>
<td></td>
</tr>
<tr>
<td>CEO Indicator</td>
<td>0.723***</td>
<td>0.431***</td>
<td>0.141***</td>
<td>0.224***</td>
</tr>
<tr>
<td>(0.017)</td>
<td>(0.017)</td>
<td>(0.012)</td>
<td>(0.015)</td>
<td></td>
</tr>
<tr>
<td>Female Indicator</td>
<td>-0.115***</td>
<td>-0.023***</td>
<td>-0.036***</td>
<td></td>
</tr>
<tr>
<td>(0.026)</td>
<td>(0.005)</td>
<td>(0.006)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>96,719</td>
<td>96,719</td>
<td>96,719</td>
<td>96,719</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.47</td>
<td>0.82</td>
<td>0.08</td>
<td>0.76</td>
</tr>
</tbody>
</table>
Table IV

OLS, FE, AM, and AE Estimates in a Determinants of Tobin's Q Regression with Unobserved Group-Level Heterogeneity Across Industry-Years

This table reports coefficients from firm-level regressions of Tobin's Q on an indicator for being incorporated in Delaware, Ln(sales), R&D expenses / assets, return on assets, and year fixed effects using different methodologies to account for unobserved group-level heterogeneity across industry-years at the 4-digit SIC industry level. The data are from Compustat from 1962-2000 and exclude financial and regulated industries. All variables were winsorized at the one percent tails. Column (i) reports the OLS estimates; Column (ii) reports the FE estimates; Column (iii) reports the AM estimates; and Column (iv) reports the AE estimates. Standard errors, clustered at the firm level, are reported in parentheses. *** significant at 1% level; ** significant at 5% level.

<table>
<thead>
<tr>
<th>Dependent Variable = Tobin’s Q</th>
<th>OLS</th>
<th>FE</th>
<th>AM</th>
<th>AE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(i)</td>
<td>(ii)</td>
<td>(iii)</td>
<td>(iv)</td>
</tr>
<tr>
<td>Delaware Incorporation</td>
<td>0.100***</td>
<td>0.086**</td>
<td>0.019</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td>(0.036)</td>
<td>(0.039)</td>
<td>(0.032)</td>
<td>(0.032)</td>
</tr>
<tr>
<td>Ln(sales)</td>
<td>-0.125***</td>
<td>-0.131***</td>
<td>-0.054***</td>
<td>-0.072***</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.011)</td>
<td>(0.008)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>R&amp;D Expenses / Assets</td>
<td>6.724***</td>
<td>5.541***</td>
<td>3.022***</td>
<td>3.968***</td>
</tr>
<tr>
<td></td>
<td>(0.260)</td>
<td>(0.318)</td>
<td>(0.242)</td>
<td>(0.256)</td>
</tr>
<tr>
<td>Return on Assets</td>
<td>-0.559***</td>
<td>-0.436***</td>
<td>-0.526***</td>
<td>-0.535***</td>
</tr>
<tr>
<td></td>
<td>(0.108)</td>
<td>(0.117)</td>
<td>(0.095)</td>
<td>(0.097)</td>
</tr>
<tr>
<td>Observations</td>
<td>55,792</td>
<td>55,792</td>
<td>55,792</td>
<td>55,792</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.22</td>
<td>0.37</td>
<td>0.08</td>
<td>0.34</td>
</tr>
</tbody>
</table>

40
Appendix

A1 – Proof of Proposition 1

Given $\mu_X = \mu_f = 0$ and Equation (1), when the number of observations, $N$, goes to infinity and the number of observations per group, $N_j$, remains constant, the AM estimates $\hat{\alpha}^{AM}$ and $\hat{\beta}^{AM}$ are given by

$$\lim_{N \to \infty} b = \lim_{N \to \infty} \left( \frac{1}{N} X'X \right)^{-1} \frac{1}{N} X'y,$$

where

$$b = \begin{bmatrix} \hat{\alpha}^{AM} \\ \hat{\beta}^{AM} \end{bmatrix}, \quad X' = \begin{bmatrix} 1 & \cdots & 1 \\ X_{i,j} & \cdots & X_{N,j} \end{bmatrix}, \quad y = \begin{bmatrix} y_{i,j} - \bar{y}_{i,j} \\ \vdots \\ y_{N,j} - \bar{y}_{N,j} \end{bmatrix} = \begin{bmatrix} \beta X_{i,j} - \bar{r}_{i,j} + \epsilon_{i,j} \\ \vdots \\ \beta X_{N,j} - \bar{r}_{N,j} + \epsilon_{N,j} \end{bmatrix},$$

and

$$\bar{r}_{i,j} = \alpha + f_j + \bar{r}_{i,j}, \quad \bar{r}_{i,j} = \frac{1}{N_j - 1} \sum_{k \neq i} (\beta X_{k,j} + \epsilon_{k,j}).$$

It can then be shown that

$$\lim_{N \to \infty} \left( \frac{1}{N} X'X \right)^{-1} = \lim_{N \to \infty} \left( \frac{1}{N} \begin{bmatrix} 1 & \cdots & 1 \\ X_{i,j} & \cdots & X_{N,j} \end{bmatrix} \right)^{-1} = \lim_{N \to \infty} \left( \frac{1}{N} \begin{bmatrix} 1 & \cdots & 1 \\ X_{i,j} & \cdots & X_{N,j} \end{bmatrix} \right)^{-1}$$

$$= \lim_{N \to \infty} \left( \frac{1}{N} \sum_{i=1}^{n} (X_{i,j})^2 \right)^{-1} = \left( \frac{1}{0} \frac{1}{\sigma_x^2} \right)^{-1} = \left( \frac{1}{0} \frac{1}{\sigma_x^2} \right)^{-1}$$

And since the covariance between $X_{i,j}$ and $\bar{r}_{i,j}$, $\sigma_{Xr}$, is given by
\[
\text{cov}(X_{i,j}, \bar{X}_{i,j}) = \text{cov}
\left(X_{i,j} - \frac{1}{N_j - 1} \sum_{k \in \text{group}_j} \left( \beta X_{k,j} + \epsilon_{k,j} \right) \right)
\]
\[
= \text{cov}
\left(X_{i,j} - \frac{\beta}{N_j - 1} \sum_{k \in \text{group}_j} X_{k,j} \right) + \text{cov}
\left(X_{i,j} - \frac{1}{N_j - 1} \sum_{k \in \text{group}_j} \epsilon_{k,j} \right),
\]
\[
= \beta \sigma_{X_{i,j}X_{i,j}}
\]

it can be shown that

\[
\text{plim} \left\{ \frac{1}{N} X'y \right\} = \left[ \begin{array}{c}
\beta \sigma_{X_{i,j}} - \sigma_{X_{i,j}} \\
\sum_{i} \beta X_{i,j} - \bar{X}_{i,j} + \epsilon_{i,j} \\
\sum X_{i,j} (\beta X_{i,j} - \bar{X}_{i,j} + \epsilon_{i,j}) \\
\end{array} \right] = \left[ \begin{array}{c}
0 \\
0 \\
0
\end{array} \right]
\]
\[
\beta \left[ \sigma^2_X - \sigma_{X_iX_{i,j}} \right]
\]
\[
\beta \left[ 1 - \rho_{X_{i,j}X_{i,j}} \right].
\]

Therefore,

\[
b = \begin{bmatrix} \hat{\alpha}^{AM} \\ \hat{\beta}^{AM} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{\sigma_X^2} \end{bmatrix} \begin{bmatrix} 0 \\ \beta \left[ \sigma^2_X - \sigma_{X_iX_{i,j}} \right] \\ \beta \left[ 1 - \rho_{X_{i,j}X_{i,j}} \right] \end{bmatrix}.
\]

QED

**A2 – Proof of Proposition 2**

Given \( \mu_X = \mu_Z = \mu_f = 0 \) and Equation (13), when \( N \) goes to infinity and the number of observations per group, \( N_j \), remains constant, the AM estimates \( \hat{\alpha}^{AM}, \hat{\beta}^{AM}, \) and \( \hat{\rho}^{AM} \) are given by

\[
\text{plim}_{N \to \infty} b = \text{plim}_{N \to \infty} \left( \frac{1}{N} X'X \right)^{\frac{1}{2}} \frac{1}{N} X'y,
\]

where
\[ \mathbf{b} = \begin{bmatrix} \hat{\alpha}^{AM} \\ \hat{\beta}^{AM} \\ \hat{\phi}^{AM} \end{bmatrix} \]

\[ \mathbf{X}' = \begin{pmatrix} 1 & \ldots & 1 \\ X_{i,j} & \ldots & X_{N,j} \\ Z_{i,j} & \ldots & Z_{N,j} \end{pmatrix} \]

\[ \mathbf{y} = \begin{pmatrix} y_{i,j} - \bar{y}_{i,j} \\ \vdots \\ y_{N,j} - \bar{y}_{N,j} \end{pmatrix} = \begin{pmatrix} \beta X_{i,j} - \bar{r}_{i,j} + \epsilon_{i,j} \\ \vdots \\ \beta X_{N,j} - \bar{r}_{N,j} + \epsilon_{N,j} \end{pmatrix} \]

and

\[ \bar{y}_{i,j} = \alpha + f_j + r_{i,j} \]

\[ \bar{r}_{i,j} = \frac{1}{N_j - 1} \sum_{k_{group}} \left( \beta X_{k,j} + \gamma Z_{k,j} + \epsilon_{k,j} \right). \] (A10)

It can then be shown that

\[
\begin{aligned}
\operatorname{plim}_{N \to \infty} \left( \frac{1}{N} \mathbf{X}' \mathbf{X} \right)^{-1} &= \operatorname{plim}_{N \to \infty} \left( \frac{1}{N} \begin{pmatrix} 1 & \ldots & 1 \\ X_{i,j} & \ldots & X_{N,j} \\ Z_{i,j} & \ldots & Z_{N,j} \end{pmatrix} \right)^{-1} \\
&= \operatorname{plim}_{N \to \infty} \begin{pmatrix}
1 & \frac{\sum X_{i,j}}{N} & \frac{\sum Z_{i,j}}{N} \\
\frac{\sum X_{i,j}}{N} & \frac{\sum (X_{i,j})^2}{N} & \frac{\sum X_{i,j} Z_{i,j}}{N} \\
\frac{\sum Z_{i,j}}{N} & \frac{\sum X_{i,j} Z_{i,j}}{N} & \frac{\sum (Z_{i,j})^2}{N}
\end{pmatrix}^{-1} \\
&= \begin{pmatrix}
1 & 0 & 0 \\
0 & \sigma_x^2 & \sigma_{xz} \\
0 & \sigma_{xz} & \sigma_z^2
\end{pmatrix}^{-1} \begin{pmatrix}
1 & 0 & 0 \\
0 & 1 + \frac{\sigma_{xz}^2}{\sigma_x^2} & \frac{\sigma_{xz}}{\sigma_x^2} \\
0 & -\frac{\sigma_{xz}}{\sigma_x^2} & \frac{\sigma_z^2}{\sigma_{xz}^2} - \frac{\sigma_x^2}{\sigma_{xz}^2} \\
\end{pmatrix}^{-1} \begin{pmatrix}
0 & \sigma_x^2 & \sigma_{xz} \\
0 & \frac{\sigma_x^2}{\sigma_{xz}^2} - \sigma_{xz}^2 & -\frac{\sigma_x^2}{\sigma_{xz}^2} \\
\end{pmatrix}^{-1}
\end{aligned} \] (A11)

where \( \sigma_{xz} \) is the covariance between \( X_{i,j} \) and \( Z_{i,j} \).

And since the covariance between \( X_{i,j} \) and \( \bar{r}_{i,j} \), \( \sigma_{xy} \), is given by

A-3
\[
\text{cov}(X_{i,j}, \tilde{r}_{i,j}) = \text{cov}\left( \frac{1}{N_j} \sum_{k = 1}^{N_j} (\beta X_{i,k} + \gamma Z_{i,k} + \epsilon_{i,k}) \right)
\]
\[
= \beta \sigma_{x_i x_{i-j}} + \gamma \sigma_{x_i z_{i-j}}
\] (A12)

where \( \sigma_{x_i z_{i-j}} \) is the covariance between \( X_{i,j} \) and \( Z_{i-j,j} \), and the covariance between \( Z_{i,j} \) and \( \tilde{r}_{i,j} \), \( \sigma_{z_i} \), is given by

\[
\text{cov}(Z_{i,j}, \tilde{r}_{i,j}) = \text{cov}\left( \frac{1}{N_j} \sum_{k = 1}^{N_j} (\beta X_{i,k} + \gamma Z_{i,k} + \epsilon_{i,k}) \right)
\]
\[
= \beta \sigma_{z_i x_{i-j}} + \gamma \sigma_{z_i z_{i-j}}
\] (A13)

where \( \sigma_{z_i z_{i-j}} \) is the covariance between \( Z_{i,j} \) and \( Z_{i-j,j} \) and \( \sigma_{z_i x_{i-j}} \) is the covariance between \( Z_{i,j} \) and \( X_{i-j,j} \), it can be shown that

\[
\text{plim}_{N \to \infty} \frac{1}{N} X'y = \begin{bmatrix}
1 & \cdots & 1 \\
X_{1,j} & \cdots & X_{N,j} \\
Z_{1,j} & \cdots & Z_{N,j}
\end{bmatrix} \begin{bmatrix}
\beta X_{1,j} + \gamma Z_{1,j} - \tilde{r}_{i,j} + \epsilon_{i,j} \\
\vdots \\
\beta X_{N,j} + \gamma Z_{N,j} - \tilde{r}_{i,j} + \epsilon_{i,j}
\end{bmatrix} N
\]
\[
= \text{plim}_{N \to \infty} \frac{1}{N} \sum_j \left( \beta X_{i,j} + \gamma Z_{i,j} - \tilde{r}_{i,j} + \epsilon_{i,j} \right)
\]
\[
= \begin{bmatrix}
\beta \sigma_{x_i}^2 + \gamma \sigma_{x_i z_i} - \sigma_{x_i} \\
\beta \sigma_{x_i z_i} + \gamma \sigma_{z_i}^2 - \sigma_{z_i}
\end{bmatrix}
\] (A14)
Therefore,

\[
\mathbf{b} = \begin{bmatrix} \hat{\alpha}_{AM} \\ \hat{\beta}_{AM} \\ \hat{\gamma}_{AM} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{\sigma_{ij}^2}{\sigma_{ij}^2 - \sigma_{XZ}} & -\frac{\sigma_{XZ}}{\sigma_{ij}^2 - \sigma_{XZ}} \\ 0 & -\frac{\sigma_{XZ}}{\sigma_{ij}^2 - \sigma_{XZ}} & \frac{\sigma_{XZ}^2}{\sigma_{ij}^2 - \sigma_{XZ}} \end{bmatrix} \begin{bmatrix} \beta [\sigma_{XX}^2 - \sigma_{XX,ij}^2] + \gamma [\sigma_{XZ} - \sigma_{XZ,ij}] \\ \gamma [\sigma_{XZ}^2 - \sigma_{XZ,ij}^2] + \beta [\sigma_{XZ} - \sigma_{XZ,ij}] \end{bmatrix}
\]

(A15)

\[
= \beta + \beta \left( \frac{\rho_{XZ} \rho_{z_{ij}z_{ij}} - \rho_{X_i z_{ij}}}{1 - \rho_{XZ}^2} \right) + \gamma \left( \frac{\rho_{XZ} \rho_{z_{ij}z_{ij}} - \rho_{X_i z_{ij}}}{1 - \rho_{XZ}^2} \right)
\]

\[
\gamma + \gamma \left( \frac{\rho_{XZ} \rho_{z_{ij}z_{ij}} - \rho_{X_i z_{ij}}}{1 - \rho_{XZ}^2} \right) + \beta \left( \frac{\rho_{XZ} \rho_{z_{ij}z_{ij}} - \rho_{X_i z_{ij}}}{1 - \rho_{XZ}^2} \right)
\]

where \( \rho_{z_{ij}z_{ij}} \) is the correlation between \( Z_{i,j} \) and \( Z_{i,j} \), \( \rho_{X_i z_{ij}} \) is the correlation between \( X_{i,j} \) and \( Z_{i,j} \), and \( \rho_{X_i X_i z_{ij}} \) is the correlation between \( X_{i,j} \) and \( Z_{i,j} \).

QED

A3 – Proof of Proposition 3

Suppose a researcher is interested in determining the effect of a treatment, \( T \), on a variable \( y \), when the true underlying structure of the data is given by

\[
y_{ij} = \beta_0 + \beta_1 P_t + \beta_2 T_y + \beta_3 (P_t \times T_y) + \varepsilon_{ij},
\]

(A16)

where \( y_{ij} \) is the outcome for unit \( i \), in group \( j \), and period \( t \); \( P_t \) is an indicator equal to 1 if treatment has occurred by period \( t \); and \( T_y \) is an indicator equal to 1 if unit \( i \) is treated.

The difference-in-differences estimator, which is a direct estimation of Equation (A16), compares the mean of \( y \) for the untreated and treated units in the pre- and post-treatment periods. This estimator provides a consistent estimate of \( \beta_3 \), which is the causal effect of the treatment on the outcome \( y \).

The AM-style approach of estimating \( \beta_3 \) is to compare the pre- and post-treatment group-adjusted means for the treated units. The AM estimator is

\[
y_{ij} - \bar{y}_j = \alpha^{AM} + \beta^{AM} P_t + \varepsilon_{ij}^{AM}
\]

(A17)

Using Equation (A16), we can see that the group mean is given by

A-5
\[ \bar{y}_{j,t} = \beta_0 + \beta_1 T_{j,t} + \beta_2 (P_t \times T_{j,t}) + \bar{e}_{j,t}, \quad (A18) \]

and the group-adjusted mean is given by
\[ y_{jt} - \bar{y}_{jt} = \beta_1 (T_{jt} - \bar{T}_j) + \beta_2 (T_{jt} - \bar{T}_j) P_t + (e_{jt} - \bar{e}_j). \quad (A19) \]

Thus, the group mean for treated firms, where \( T_{jt} = 1 \), is equal to
\[ y_{jt} - \bar{y}_{jt} = \beta_2 (1 - \bar{T}_j) + \beta_3 (1 - \bar{T}_j) P_t + (e_{jt} - \bar{e}_j). \quad (A20) \]

Comparing the AM estimator in Equation (A17) and the true underlying data structure in Equation (A20) reveals that the AM estimator is not consistent. Specifically, \( \hat{\beta}^{AM} = \beta_3 (1 - \bar{T}) \), where \( \bar{T} \) is the average share of untreated firms in a treated firm’s group. Intuitively, the AM estimator exhibits an attenuation bias because it incorrectly demeans the data using an average of treated and untreated observations, which removes some of the treatment effect on \( y \). The difference-in-differences estimator only removes the mean of untreated observations.

Equivalently, one can describe the bias in the AM estimator as an omitted variable bias. Equation (A20) can be written as
\[ y_{jt} - \bar{y}_{jt} = \beta_2 (1 - \bar{T}_j) + \beta_3 (1 - \bar{T}_j) P_t + (e_{jt} - \bar{e}_j). \quad (A21) \]

Comparing (A21) to Equation (A17), the AM estimator fails to control for \( P_t \bar{T}_j \), which affects the group-adjusted mean, \( y_{jt} - \bar{y}_{jt} \), and is correlated with the independent variable of interest, \( P_t \). Biased estimates for the causal effect, \( \beta_3 \), result. QED

**A4 – Proof of Proposition 4**

Given \( \mu_X = \mu_T = 0 \) and Equation (1), when \( N \) goes to infinity and the number of observations per group, \( N_j \), remains constant, the AE estimates \( \hat{\alpha}^{AE} \), \( \hat{\beta}^{AE} \), and \( \hat{\lambda}^{AE} \) are given by
\[ plim_{N \to \infty} b = plim_{N \to \infty} \left( \frac{1}{N} X'X \right)^{1/2} \frac{1}{N} X'y, \quad (A22) \]

where
\[
\mathbf{b} = \begin{bmatrix}
\hat{\alpha} \\
\hat{\beta} \\
\hat{\sigma}^2
\end{bmatrix},
\]

\[
\mathbf{X}' = \begin{bmatrix}
1 & \cdots & 1 \\
X_{1j} & \cdots & X_{Nj}
\end{bmatrix},
\]

\[
y = \begin{bmatrix}
y_{1j} \\
\vdots \\
y_{Nj}
\end{bmatrix}
\]

and

\[
\bar{y}_{-i,j} = \alpha + f_j + \bar{r}_{i,j}
\]

\[
\bar{r}_{-i,j} = \frac{1}{N_j - 1} \sum_{k \in \text{group}} \left( \beta X_{kj} + \epsilon_{kj} \right).
\]

It can then be shown that

\[
\text{plim}_{N \to \infty} \frac{1}{N} \mathbf{X}'\mathbf{X} = \text{plim}_{N \to \infty} \begin{bmatrix}
1 & 0 & \alpha \\
0 & \sigma_X^2 & \sigma_{yx} + \sigma_{yx} \\
\alpha & \sigma_{yx} + \sigma_{yx} & \alpha^2 + \sigma_f^2 + 2\sigma_{\bar{r}} + \sigma_{\bar{r}}^2
\end{bmatrix}^{-1}
\]

where \( \sigma_{yx} \) is the covariance between \( X_{ij} \) and \( \bar{r}_{ij} \), \( \sigma_{\bar{r}} \) is the covariance between \( \alpha \) and \( \bar{r}_{ij} \), and \( \sigma_{\bar{r}}^2 \) is the variance of \( \bar{r}_{ij} \).

And, it can be shown that

\[
\text{plim}_{N \to \infty} \frac{1}{N} \mathbf{X}'y = \begin{bmatrix}
1 & \cdots & 1 \\
X_{1j} & \cdots & X_{Nj}
\end{bmatrix} \begin{bmatrix}
\alpha + \beta X_{1j} + f_j + \epsilon_{1j} \\
\vdots \\
\alpha + \beta X_{Nj} + f_j + \epsilon_{Nj}
\end{bmatrix} \begin{bmatrix}
\alpha + \beta X_{1j} + f_j + \epsilon_{1j} \\
\vdots \\
\alpha + \beta X_{Nj} + f_j + \epsilon_{Nj}
\end{bmatrix}
\]

\[A-7\]
\[
\begin{aligned}
\hat{b}^{\text{AE}} &= \begin{bmatrix}
\hat{a}^{\text{AE}} \\
\hat{\beta}^{\text{AE}} \\
\hat{\lambda}^{\text{AE}}
\end{bmatrix} = \begin{bmatrix}
1 & 1 & \ldots & 1 \\
0 & 0 & \ldots & 0 \\
\alpha & \sigma_{\text{XY}} + \sigma_{\text{XR}} & \ldots & \alpha + \sigma_{\text{XY}} + \sigma_{\text{XR}} + \sigma_{\text{PR}}
\end{bmatrix}^{-1}
\begin{bmatrix}
\alpha \\
\beta \sigma_{\text{XY}} + \sigma_{\text{XY}} \\
\alpha^2 + \beta \sigma_{\text{XY}} + \sigma_{\text{PR}} + \sigma_{\text{PR}}
\end{bmatrix}
\end{aligned}
\] (A26)

Therefore,
\[
\begin{aligned}
b &= \begin{bmatrix}
\alpha - \alpha \\
\sigma_{\text{XY}} \left( \sigma_{\text{PR}}^2 + \sigma_{\text{PR}}^2 - \frac{\left( \sigma_{\text{XY}} + \sigma_{\text{XR}} \right)^2}{\sigma_{\text{X}}^2} \right) - \left( \sigma_{\text{XY}} + \sigma_{\text{XR}} \right) \left[ \sigma_{\text{PR}}^2 - \frac{\left( \sigma_{\text{XY}} + \sigma_{\text{XR}} \right)^2}{\sigma_{\text{X}}^2} \right] \\
\sigma_{\text{XY}} \left( \sigma_{\text{PR}}^2 + \sigma_{\text{PR}}^2 - \frac{\left( \sigma_{\text{XY}} + \sigma_{\text{XR}} \right)^2}{\sigma_{\text{X}}^2} \right) \left( \sigma_{\text{PR}}^2 + \sigma_{\text{PR}}^2 - \frac{\left( \sigma_{\text{XY}} + \sigma_{\text{XR}} \right)^2}{\sigma_{\text{X}}^2} \right) \\
\sigma_{\text{PR}} + \sigma_{\text{PR}} - \frac{\left( \sigma_{\text{XY}} + \sigma_{\text{XR}} \right)^2}{\sigma_{\text{X}}^2} \\
\sigma_{\text{PR}}^2 + 2 \sigma_{\text{PR}} + \sigma_{\text{PR}}^2 - \frac{\left( \sigma_{\text{XY}} + \sigma_{\text{XR}} \right)^2}{\sigma_{\text{X}}^2}
\end{bmatrix}
\end{aligned}
\] (A27)

This expression can be simplified by recognizing that,
\[
\sigma_{\text{PR}} = \text{cov}\left\{ f_j, \frac{1}{N-j-1} \sum_{k \in \text{group } j} \left( \beta X_{k,j} + \epsilon_{k,j} \right) \right\}
\]
\[
= \text{cov}\left\{ f_j, \frac{\beta}{N-j-1} \sum_{k \in \text{group } j} X_{k,j} \right\} + \text{cov}\left\{ f_j, \frac{1}{N-j-1} \sum_{k \in \text{group } j} \epsilon_{k,j} \right\}
\]

\[
= \beta \sigma_{\text{XY}}
\] (A28)
\[ \sigma^2 = \text{var}\left( \frac{1}{N_j - 1} \sum_{k \in \text{group } j} (\beta X_{ij} + \varepsilon_{ij}) \right) \]

\[ = \text{var}\left( \frac{\beta}{N_j - 1} \sum_{k \in \text{group } j} X_{ij} + \left( \frac{1}{N_j - 1} \sum_{k \in \text{group } j} \varepsilon_{ij} \right) \right) \]

\[ = \left( \frac{\beta}{N_j - 1} \right)^2 \text{var}\left( \sum_{k \in \text{group } j} X_{ij} \right) + \left( \frac{1}{N_j - 1} \right)^2 \text{var}\left( \sum_{k \in \text{group } j} \varepsilon_{ij} \right) \]

\[ = \left( \frac{\beta}{N_j - 1} \right)^2 \left[ \sum_{k \in \text{group } j} \sigma^2_X + \sum_{\substack{n \in \text{group } j \atop m \in \text{group } j}} \text{cov}(X_{ij}, X_{mj}) \right] + \left( \frac{1}{N_j - 1} \right)^2 \left( \sum_{k \in \text{group } j} \sigma^2_\varepsilon \right) \]

\[ = \left( \frac{\beta}{N_j - 1} \right)^2 \left[ (N_j - 1)\sigma^2_X + (N_j - 1)(N_j - 2) \text{cov}(X_{ij}, X_{ij}) \right] + \frac{\sigma^2_\varepsilon}{N_j - 1} \]

\[ = \frac{\beta^2 \sigma^2_X}{N_j - 1} + \frac{\beta^2 (N_j - 2) \sigma^2_{X_{ij},X_{ij}}}{N_j - 1} + \sigma^2_\varepsilon \]

\[ = \frac{\beta^2 \sigma^2_X}{N_j - 1} + \frac{\left( \frac{N_j - 2}{N_j - 1} \right) \left( \beta^2 \sigma^2_{X_{ij},X_{ij}} \right)}{N_j - 1} + \frac{\sigma^2_\varepsilon}{N_j - 1} \]

\[ = \beta^2 \sigma^2_X + \frac{\beta^2 \sigma^2_X}{N_j - 1} - \beta^2 \sigma^2_{X_{ij},X_{ij}} + \sigma^2_\varepsilon \]

\[ = \frac{\beta^2 \sigma^2_X}{N_j - 1} + \frac{\beta^2 \left( \sigma^2_X - \sigma^2_{X_{ij},X_{ij}} \right) + \sigma^2_\varepsilon}{N_j - 1} \]

With these results and some algebra, it can be shown that
A5 – Proof of Proposition 5

Given Equation (1) and Propositions 1 and 4, the estimates of $\beta$ when $\beta = 0$ are given by

$$
\begin{align*}
\begin{bmatrix}
\hat{\beta}_{AE} \\
\hat{\beta}_{AM} \\
\hat{\beta}_{OLS}
\end{bmatrix}
= \frac{
\alpha
}{\left( 1 - \rho_{x_{i},x_{-i}} \right)}
\begin{bmatrix}
\frac{\sigma_{f}}{\sigma_{x}}

\left( \rho_{y} \left( 1 - \rho_{x_{i},x_{-i}} \right) \right)
+ \left( \frac{\sigma_{f}}{\sigma_{x}} \right) \left( 1 - \rho_{x_{i}x_{-i}} \right)
\end{bmatrix}
\end{align*}
$$

(A30)

Therefore, the AM estimate is unbiased, and the bias of the AE estimate approaches zero as $N_j \to \infty$.

\[QED\]
It can also be shown that the AE estimate of $\beta$ is less biased than the OLS estimate of $\beta$. When $\rho_{xy} < 0$, 

$\hat{\beta}^{OLS} < \hat{\beta}^{AE} < 0$ since

$$\hat{\beta}^{AE} = \rho_{xy} \left( \frac{\sigma_f}{\sigma_x} \right) \left( \frac{\sigma_y^2}{\sigma_y^2 + \rho_{xy}^2 \sigma_x^2} \right) \left( \frac{N_j}{N_j - 1} \right) < 0$$  \hspace{1cm} (A32)

and

$$\hat{\beta}^{AE} > \hat{\beta}^{OLS}$$

$$\rho_{xy} \left( \frac{\sigma_f}{\sigma_x} \right) \left( \frac{\sigma_y^2}{\sigma_y^2 + \rho_{xy}^2 \sigma_x^2} \right) \left( \frac{N_j}{N_j - 1} \right) > \rho_{xy} \left( \frac{\sigma_f}{\sigma_x} \right)$$

$$\left(1 - \rho_{xy}^2 \right) \left( \frac{\sigma_y^2}{\sigma_y^2 + \rho_{xy}^2 \sigma_x^2} \right) \left( \frac{N_j}{N_j - 1} \right) < 1$$

$$\left(1 - \rho_{xy}^2 \right) \left( \frac{\sigma_y^2}{\sigma_y^2 + \rho_{xy}^2 \sigma_x^2} \right) \left( \frac{N_j}{N_j - 1} \right) < 1$$

$$0 < \left(1 - \rho_{xy}^2 \right) \left( \frac{\sigma_y^2}{\sigma_y^2 + \rho_{xy}^2 \sigma_x^2} \right).$$ \hspace{1cm} (A33)

When $\rho_{xy} > 0$, $0 < \hat{\beta}^{AE} < \hat{\beta}^{OLS}$ since

$$\hat{\beta}^{AE} = \rho_{xy} \left( \frac{\sigma_f}{\sigma_x} \right) \left( \frac{\sigma_y^2}{\sigma_y^2 + \rho_{xy}^2 \sigma_x^2} \right) \left( \frac{N_j}{N_j - 1} \right) > 0$$  \hspace{1cm} (A34)
and

\[
\rho_{xy} \left( \frac{\sigma_y}{\sigma_x} \right) \left( \frac{\sigma_y^2}{\sigma_x^2} \frac{N_j - 1}{N_j} \right) < \rho_{xy} \left( \frac{\sigma_y}{\sigma_x} \right)
\]

\[
(1 - \rho_{xy}^2) \left( \frac{\sigma_y^2}{\sigma_x^2} \right) + \left( \frac{\sigma_y^2}{\sigma_x^2} \frac{N_j - 1}{N_j} \right) < 1
\]

\[
(1 - \rho_{xy}^2) \left( \frac{\sigma_y^2}{\sigma_x^2} \right) + \left( \frac{\sigma_y^2}{\sigma_x^2} \frac{N_j - 1}{N_j} \right)
\]

0 < (1 - \rho_{xy}^2) \left( \frac{\sigma_y^2}{\sigma_x^2} \right).
\]

(A35)

QED

A6 – Proof that the AM and AE estimators are still biased when the observation at hand is not excluded when calculating group averages

Given \( \mu_X = \mu_f = 0 \) and Equation (1), the AM estimates when the observation at hand is not excluded, \( \hat{\alpha}_{AM} \) and \( \hat{\beta}_{AM} \), are given by the same calculation as in Appendix A1, except that

\[
y = \begin{bmatrix} y_{i,j} - \bar{y}_j \\ \vdots \\ y_{N,j} - \bar{y}_j \end{bmatrix} = \begin{bmatrix} \beta X_{i,j} - \bar{r}_j + \epsilon_{i,j} \\ \vdots \\ \beta X_{N,j} - \bar{r}_j + \epsilon_{N,j} \end{bmatrix},
\]

(A36)

where

\[
\bar{y}_j = \alpha + f_j + \bar{r}_j
\]

\[
\bar{r}_j = \frac{1}{N_j} \sum_{k \in \text{group } j} \left( \beta X_{k,j} + \epsilon_{k,j} \right).
\]

(A37)
The $\text{plim} \left( \frac{1}{N} \mathbf{X} \mathbf{X}^{-1} \right)$ remains the same as in Appendix A1, and because the covariance between $X_{i,j}$ and $\overline{r}_j$, $\sigma_{Xr}$, is given by

$$\text{cov}(X_{i,j}, \overline{r}_j) = \text{cov} \left( X_{i,j}, \frac{1}{N_j} \sum_{k \in \text{group}_j} (\beta X_{k,j} + \xi_{k,j}) \right) = \text{cov} \left( X_{i,j}, \frac{\beta}{N_j} \sum_{k \in \text{group}_j} X_{k,j} \right) + \text{cov} \left( X_{i,j}, \frac{1}{N_j} \sum_{k \in \text{group}_j} \xi_{k,j} \right) = \text{cov} \left( X_{i,j}, \frac{\beta}{N_j} \sum_{k \in \text{group}_j} X_{k,j} \right) + \frac{\beta}{N_j} \sigma_{x}^{2} = \beta \left( \frac{N_j - 1}{N_j} \right) \sigma_{x_{i,j}x_{-i,j}} + \frac{\beta}{N_j} \sigma_{x}^{2},$$

(A38)

it can be shown that

$$\text{plim} \left( \frac{1}{N} \mathbf{X}' \mathbf{y} \right) = \begin{bmatrix}
1 & \cdots & 1 \\
X_{1,j} & \cdots & X_{N,j} \\
\vdots & & \vdots \\
\beta X_{i,j} - \overline{r}_j + \xi_{i,j} & \cdots & \beta X_{N,j} - \overline{r}_j + \xi_{N,j}
\end{bmatrix} = \begin{bmatrix}
0 \\
0 \\
0
\end{bmatrix} = \begin{bmatrix}
\beta \sigma_{x}^{2} - \sigma_{x} \overline{r}_j \\
\beta \sigma_{x}^{2} - \sigma_{x} \overline{r}_j \\
\beta \sigma_{x}^{2} - \sigma_{x} \overline{r}_j
\end{bmatrix} = \begin{bmatrix}
\beta \left( \frac{N_j - 1}{N_j} \right) [\sigma_{x}^{2} - \sigma_{x_{i,j}x_{-i,j}}] \\
\beta \left( \frac{N_j - 1}{N_j} \right) [\sigma_{x}^{2} - \sigma_{x_{i,j}x_{-i,j}}] \\
\beta \left( \frac{N_j - 1}{N_j} \right) [\sigma_{x}^{2} - \sigma_{x_{i,j}x_{-i,j}}]
\end{bmatrix}.$$

(A39)

Therefore,

$$\mathbf{b} = \begin{bmatrix}
\hat{\alpha}^{AM} \\
\hat{\beta}^{AM}
\end{bmatrix} = \begin{bmatrix}
1 & 0 \\
0 & 1/\sigma_{x}^{2}
\end{bmatrix} \begin{bmatrix}
0 \\
\beta \left( \frac{N_j - 1}{N_j} \right) [\sigma_{x}^{2} - \sigma_{x_{i,j}x_{-i,j}}]
\end{bmatrix} = \begin{bmatrix}
0 \\
\beta \left( \frac{N_j - 1}{N_j} \right) [\sigma_{x}^{2} - \sigma_{x_{i,j}x_{-i,j}}]
\end{bmatrix} = \begin{bmatrix}
0 \\
\beta \left( \frac{N_j - 1}{N_j} \right) [1 - \rho_{x_{i,j}x_{-i,j}}]
\end{bmatrix}.$$

(A40)

Comparing (A40) to (A7), it is apparent that including the observation at hand just multiplies the AM estimate by a factor of $(N_j - 1)/N_j$. 
Given $\mu_X = \mu_f = 0$ and Equation (1), the AE estimates $\tilde{\alpha}^{AE}$, $\tilde{\beta}^{AE}$, and $\tilde{\lambda}^{AE}$ are given by the same calculation as in Appendix A4, except that $X'$ is now given by

$$X' = \begin{pmatrix} 1 & \ldots & 1 \\ X_{i,j} & \ldots & X_{N,j} \\ \bar{y}_j & \ldots & \bar{y}_j \end{pmatrix}. \quad (A41)$$

It can then be shown that

$$\text{plim}_{N \to \infty} \left( \frac{1}{N} X'X \right)^{-1} = \text{plim}_{N \to \infty} \left[ \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \left( \begin{array}{cc} 1 & \ldots & 1 \\ X_{i,j} & \ldots & X_{N,j} \\ \alpha + f_j + \bar{r}_j & \ldots & \alpha + f_j + \bar{r}_j \end{array} \right) \right]^{-1} \left( \begin{array}{c} 1 \\ \vdots \\ 1 \end{array} \right) \left( \begin{array}{c} X_{i,j} \\ \vdots \\ X_{N,j} \\ \alpha + f_j + \bar{r}_j \end{array} \right), \quad (A42)$$

where $\sigma_{\beta_f}$ is the covariance between $f_j$ and $\bar{r}_j$, and $\sigma_{\beta_f}^2$ is the variance of $\bar{r}_j$.

And, it can be shown that

$$\text{plim}_{N \to \infty} \frac{1}{N} X'y = \text{plim}_{N \to \infty} \left[ \begin{pmatrix} 1 \\ X_{i,j} \\ \vdots \\ X_{N,j} \\ \alpha + f_j + \bar{r}_j & \ldots & \alpha + f_j + \bar{r}_j \end{pmatrix} \left( \begin{array}{c} \alpha + f_j + \bar{r}_j + \epsilon_{i,j} \\ \vdots \\ \alpha + f_j + \bar{r}_j + \epsilon_{N,j} \\ \alpha + f_j + \bar{r}_j + \epsilon_{i,j} \end{array} \right) \right] \left( \begin{array}{c} \alpha + f_j + \bar{r}_j + \epsilon_{N,j} \\ \alpha + f_j + \bar{r}_j + \epsilon_{i,j} \end{array} \right) \left( \begin{array}{c} 1 \\ X_{i,j} \\ \vdots \\ X_{N,j} \\ \alpha + f_j + \bar{r}_j \end{array} \right), \quad (A43)$$

$$= \text{plim}_{N \to \infty} \left( \begin{array}{c} \sum_j (\alpha + f_j + \bar{r}_j) \\ \sum_j (\alpha + f_j + \bar{r}_j) \cdot (\alpha + f_j + \bar{r}_j) \end{array} \right) \frac{N}{\sum_j (\alpha + f_j + \bar{r}_j) (\alpha + f_j + \bar{r}_j) \cdot (\alpha + f_j + \bar{r}_j)}.$$
Therefore,

\[
\mathbf{b} = \begin{bmatrix}
\hat{\alpha}^M \\
\hat{\beta}^M \\
\hat{\lambda}^M
\end{bmatrix} = \begin{bmatrix}
1 & 0 \\
0 & \sigma_x^2 \\
\alpha & \sigma_{xy} + \sigma_{y^2}
\end{bmatrix}^{-1} \begin{bmatrix}
\alpha \\
\beta \sigma_x^2 + \sigma_{y^2}
\end{bmatrix}
\]

\[
= \beta + \frac{\alpha - \alpha}{\sigma_{y^2}^2 + 2\sigma_{y^2} + \sigma_{y^2}^2 - \sigma_{x^2}^2 \left( \frac{\sigma_{x^2}^2 + \sigma_{y^2}^2}{\sigma_{x^2}^2} \right)}
\]

\[
= \beta + \frac{\sigma_{y^2}^2 \left( \frac{\sigma_{y^2}^2 + \sigma_{y^2}^2}{\sigma_{x^2}^2} \right) - \left( \frac{\sigma_{x^2}^2 + \sigma_{y^2}^2}{\sigma_{x^2}^2} \right)^2}{\sigma_{y^2}^2 + 2\sigma_{y^2} + \sigma_{y^2}^2 - \sigma_{x^2}^2 \left( \frac{\sigma_{x^2}^2 + \sigma_{y^2}^2}{\sigma_{x^2}^2} \right)}
\]

Comparing (A44) to (A27), we see that, by including the observation at hand, \( \sigma_y^2 \) replaces \( \sigma_x^2 \), \( \sigma_{xy} \) replaces \( \sigma_{yx} \), and \( \sigma_{y^2} \) replaces \( \sigma_{x^2} \) in the expression for the AE estimate. But as shown in Equations (A5) and (A38),

\[
\sigma_{y^2} = \left( \frac{N_j}{N_j - 1} \right) \sigma_{y^2} + \left( \frac{\beta}{N_j} \right) \sigma_x^2,
\]

and it can be shown that \( \sigma_{y^2} = \sigma_{x^2} \). Moreover,
\[
\sigma^2_e = \text{var}\left(1 \frac{1}{N_j} \sum_{k \in \text{group } j} \left( \beta X_{ij} + \epsilon_{ij} \right) \right)
= \text{var}\left( \frac{\beta}{N_j} \sum_{k \in \text{group } j} X_{ij} + \frac{1}{N_j} \sum_{k \in \text{group } j} \epsilon_{ij} \right)
= \left(\frac{\beta}{N_j}\right)^2 \text{var}\left( \sum_{k \in \text{group } j} X_{ij} \right) + \left(\frac{1}{N_j}\right)^2 \text{var}\left( \sum_{k \in \text{group } j} \epsilon_{ij} \right)
= \left(\frac{\beta}{N_j}\right)^2 \left[ \sum_{k \in \text{group } j} \sum_{m \in \text{group } j} \text{cov}(X_{ij}, X_{mj}) \right] + \left(\frac{1}{N_j}\right)^2 \sum_j \sigma^2_j
= \left(\frac{\beta}{N_j}\right)^2 \left[ N_j \sigma^2_x + N_j (N_j - 1) \text{cov}(X_{ij}, X_{mj}) \right] + \left(\frac{1}{N_j}\right)^2 N_j \sigma^2_e
= \frac{\beta^2 \sigma^2_x}{N_j} + \frac{\beta^2 (N_j - 1) \sigma_{X_{ij}X_{mj}}}{N_j} + \sigma^2_e
= \frac{\beta^2 \sigma^2_x}{N_j} + \frac{\beta^2 (N_j - 1)}{N_j} \left( \frac{\beta^2 \sigma_{X_{ij}X_{ij}}}{N_j} \right) + \sigma^2_e
= \beta^2 \sigma^2_{X_{ij}X_{ij}} + \frac{\beta^2 \sigma^2_x - \beta^2 \sigma_{X_{ij}X_{ij}} + \sigma^2_e}{N_j}
= \beta^2 \sigma^2_{X_{ij}X_{ij}} + \frac{\beta^2 \left( \sigma^2_x - \sigma_{X_{ij}X_{ij}} \right) + \sigma^2_e}{N_j}
\]

which is similar to \( \sigma^2_e \), as defined in Equation (A29); the expression only differs in that the denominator in the second term is \( N_j \) rather than by \( N_j - 1 \).

QED