Federal Reserve Bank of Dallas

# Lags, Leave-Outs and Fixed Effects

Alexander Chudik, Cameron M. Ellis and Johannes G. Jaspersen

**Working Paper 2536** **September 2025**

Research Department

https://doi.org/10.24149/wp2536

# Lags, Leave-Outs and Fixed Effects*

Alexander Chudik†, Cameron M. Ellis‡ and Johannes G. Jaspersen§

September 19, 2025

## Abstract

To avoid endogeneity, financial economists often construct regressors and/or instruments using values from other observations, with lagged and leave-out variables being common examples. We examine the use of such variables in common settings with fixed effects and show that it can induce bias and distort inference. We illustrate the severity of this problem via simulations and with patent examiner data. Even when scrambling the patent examiners, thus removing any instrument validity, the bias leads to a first-stage F-statistic over 1,000. General and case-specific solutions are provided.

**JEL Codes:** C13, C36, D22, K00

**Keywords:** Lagged Regressors, Leave-Out Instruments, Fixed Effects, Weak Exogeneity Bias, Patents

---

* We thank Scott Cunningham, Ralf Elsas, Keith Ericson, Shan Ge, Paul Goldsmith-Pinkham, Aurobindo Ghosh, Charles Hadlock, Michal Kolesar, Peter Hull, Eli Liebman, David Matsa, Seongjoo Min, Richard Peter, Marc Ragin, Ryan Riordan, Stephen Shore, Suyong Song, Justin Sydnor, and seminar participants at The University of Iowa, LMU Munich, The European Group of Risk and Insurance Economists, The German Finance Association, The Western Finance Association, and The Financial Management Association for providing valuable comments. The views in this paper are solely the authors' and do not reflect the views of the Federal Reserve Bank of Dallas or the Federal Reserve System.

† Alexander Chudik, Federal Reserve Bank of Dallas, alexander.chudik@dal.frb.org.

‡ Cameron M. Ellis, Dept. of Finance, Tippie College of Business, University of Iowa, cameron-ellis@uiowa.edu.

§ Johannes G. Jaspersen, LMU Munich School of Management, LMU Munich, jaspersen@lmu.de.

# 1 Introduction

Consider a financial economist who wants to estimate the causal effect of $x$ on $y$, but worries that $x$ is endogenous. To overcome this endogeneity, researchers will look for another variable $v$ that correlates with $x$, but doesn't directly impact $y$. A common strategy is to use different values of $x$ that share a data dimension, such as the same firm or industry-year, that more plausibly satisfy the exclusion restriction. The economist will use this value (or an average if there are multiple) to "construct" a proxy or instrument for $x$. Perhaps the most common version of this constructed variable strategy is to exploit temporal order and use lagged values instead of the contemporaneous value. More recently, researchers have developed "leave-out" constructions, which average other observations along a relevant data dimension to construct an instrument for $x$. This leave-out strategy originated with the "Hausman" instrument, which averages a firm's prices in other markets in the same year to instrument for prices in the focal market (e.g., Hausman, 1996) but has found many applications in finance and economics, with the most common being the "Judge Fixed Effect" instrument strategy (e.g., Aizer and Doyle, 2015; Dobbie et al., 2018).

Finance researchers often combine these constructed regressors with fixed effects. Most corporate finance papers include both $\mathit{Firm}$ and $\mathit{Year}$ fixed effects, because it is impossible to collect data on all firm (or time) -invariant characteristics relevant to the outcome. With Hausman instruments, it is important to include (at least) $\mathit{Firm}$ effects to ensure that characteristics such as general firm quality do not bleed into the instrument (Nevo, 2001). Judge designs also require fixed effects, since judges are typically only randomly assigned within a court, county or, in the case of patents, art units (Chyn et al., 2024; Farre-Mensa et al., 2020).

We explore a potential pitfall of this common combination of estimation strategies: when constructed variables are combined with fixed effects that "overlap"—where the set of observations used to construct a fixed effect also includes observations that the constructed regressor is meant to avoid—they become mechanically correlated with the exact value they are trying to avoid, leading to estimation bias. This bias even appears if the constructed regressor is not the hypothesis variable but merely included as a control. To show the prevalence of this estimation strategy we examined all papers in issues 75, 76, and 77 (2020 - 2022) of the *Journal of Finance* and found 33 (out of 211) papers that likely contain some bias from the combination of a constructed regressor and an overlapping fixed effect. Table 1 shows the break-down by field, with over a quarter of empirical, non-asset pricing papers—and over 15% of all papers—likely having some bias .[1]

[1] We focused on papers that made causal claims, which excluded most asset pricing papers, and only included papers that had the potential for a non-trivial bias based on the level of overlap. Further details and a complete list of identified papers is given in Online Appendix E.

Table 1: *Journal of Finance* Papers with Constructed Regressors and Overlapping Fixed Effects

| Type | Total | Possible Bias | Percent |
|---|---|---|---|
| Empirical (Non Asset Pricing) | 101 | 28 | 27.72% |
| Empirical (Asset Pricing) | 45 | 3 | 6.66% |
| Structural | 14 | 2 | 14.29% |
| Theoretical | 42 | - | - |
| Other | 9 | - | - |
| — | | | |
| Total | 211 | 33 | 15.64% |

*Note:* Table provides an overview over the categorization of all 211 papers surveyed in the structured literature review. The "other" category is comprised of replications/ corrigenda, methods papers, and similar types of research. Further details on the procedure and the full list of identified papers is given in Online Appendix E.

To clarify the intuition in a simple example, consider a firm-year panel with many firms, indexed by $f = 1, 2, 3, ..., F$, but only four years, indexed by $t = 1, 2, 3, 4$. $x_{ft}$ is known to be endogenous to $y_{ft}$, so the researchers use the lagged value of $x$ instead and estimates

$$y_{ft} = \beta x_{f,t-1} + \eta_f + \varepsilon_{ft}, \tag{1}$$

for $t = 2, 3, 4$, where $\eta_f$ is a firm fixed effect. This is equivalent to estimating $\tilde{y}_{f,t} = \beta\tilde{x}_{f,t-1} + \tilde{\varepsilon}_{f,t}$, with tildes used to denote fixed effects de-meaned variables, namely $\tilde{y}_{f,t} = y_{f,t} - \frac{1}{3}\sum_{\tau=2}^{4} y_{f,\tau}$ and similarly $\tilde{x}_{f,t-1} = x_{f,t-1} - \frac{1}{3}\sum_{\tau=2}^{4} x_{f,\tau-1}$. Take $t = 3$ as an example. Here, the fixed effects de-meaned regressor becomes $\tilde{x}_{f,2} = \frac{2}{3}x_{2,f} - \frac{1}{3}x_{1,f} - \frac{1}{3}x_{3,f}$. Thus, the de-meaned lagged variable, $\tilde{x}_{f,t-1}$, is mechanically correlated with the contemporaneous value $x_{f,t}$ and the researchers have reintroduced the same endogeneity that they were trying to avoid in the first place. And it has to be this way. Firm-level fixed effects make all variation relative to other values of the same firm. Because some of these other values are known to be endogenous, the fixed effects estimator of $\beta$ will be biased and inconsistent as $F \to \infty$.[2] The extent of this bias will be determined by several factors, including the panel dimensions, the number of fixed effects, and degree of overlap.[3]

After showing generally that the bias exists when the observations in the fixed effects overlap with those used for constructing the regressor, and does not exist when they don't, we discuss possible solutions. We first show that interacting the fixed effects' desired dimension with a sep-

[2] In other words, condition $E(x_{f,t-1}\varepsilon_{f,t}) = 0$ is not the sufficient for consistency of the Fixed Effects estimator.

[3] Importantly, it is not necessarily determined by the number of observations used for constructing the regressor or fixed effect. In the extreme case of a Hausman IV with Firm-Year fixed effects all of the non-focal observations fully cancel out, regardless of how many markets there are, and the estimator is numerically equivalent to instrumenting the focal observation with itself. This one-to-one correlation is noted by Gormley and Matsa (2014).

arate fixed effect that is orthogonal to how the regressor is constructed removes the bias. For example, in the Hausman IV, the researcher could include $Market \times Firm$ fixed effects instead of simply $Firm$. We also discuss other potential methods for removing the overlap of fixed effects and constructed regressors that leads to the bias as well as detail case-specific solutions for the most popular construction strategies (lags and leave-outs) that do not seem to have made their way to finance research.[4]

We then examine, through extensive Monte Carlo simulations, the size of the bias, the consequences for inference, and the cost of solutions. Finally, we provide an empirical example with commonly used patent examiner data. We find that the inclusion of "overlapping fixed effects" biases the coefficient by over 20% relative to the bias of OLS. Even when scrambling the patent examiners, thus removing any identifying variation, the bias induces a first-stage F-statistic of over 1,000.

Our primary contribution is to further financial economists' understanding of the potential bias in this extremely common empirical design. In our structured literature review, we find that over 20% of empirical, non-asset pricing, papers used constructed regressors either as instruments or hypothesis variables and combined them with overlapping fixed effects. A further 12% used them as controls, which we show can also introduce bias. While several of these papers feature data dimensions that could make the bias relatively small, essentially none of them consider the bias's existence or attempt any solutions. We thus join the literature aimed at addressing the challenges empirical finance researchers face. This includes Petersen's (2008) work on clustering standard errors, Gormley and Matsa's (2014) work on fixed effects vs. de-meaning vs. controlling for averages, Jiang's (2017) work on weak instruments, Grieser and Hadlock's (2019) work on weak exogeneity in panel settings, and Baker et al.'s (2022) work on difference-in-differences approaches in staggered adoption settings.

Our work is timely because of the rapidly rising popularity of leave-out constructions in causal inference. In addition to Hausman instruments and Judge fixed effects, "Peer effects" are constructed in a similar manner through averaging other observations along a dimension such as social network or school, but are often used directly instead of as an instrument (e.g., Fruehwirth et al., 2019; Lavy and Megalokonomou, 2024). "Supply restriction" instruments also have a similar construction but rely on a different kind of exogenous shock, are also increasingly common (e.g., Ma et al., 2022; Gabaix and Koijen, 2024).[5] Though largely originating in other fields, leave-out

[4]In the case of leave-out instruments, researchers can use the IJIVE or UJIVE estimators of Ackerberg and Devereux (2009) and Kolesár (2013). For lagged variables, researchers can use the Double-Filter IV of Hayakawa et al. (2019).

[5]For supply restriction instruments, the correlation relies on aggregate supply constraints such that a positive shock to other observations in the same year will lead to a negative shock to the focal observation. These variables can also

instruments have been increasingly exploited by financial economists across a diverse range of topics including: innovation (Farre-Mensa et al., 2020), mortgage servicing (Aiello, 2022), bank supervision (Eisenbach et al., 2022), broker financial crimes (Honigsberg and Jacob, 2021), and manager promotion (Benson et al., 2019).

We also make a novel contribution to the methodological literature. Case-specific versions of the bias we explore have been identified by other work (e.g., Nickell, 1981; Angrist et al., 1999; Ackerberg and Devereux, 2009; Kolesár, 2013; Chudik et al., 2018). While that work proves that the bias exists for a general set of control variables, with the bias increasing in the number of controls, our focus on fixed effects (the most common "many controls" setting) allows us to create a general theory that encompasses all cases. We use this general theory to show that, in contrast to the results of existing work, it not just the number of fixed effects that determines the magnitude of the bias, but also the degree of overlap. Indeed, there are cases where multiplying two fixed effects together can fully remove the bias, even though the number of controls increases dramatically.

# 2 Theory

## 2.1 General Result

We consider fixed effects (FE) estimation of a linear regression model featuring a constructed regressor. This estimation can either be the direct test of a hypothesis, or the first stage of a two-stage least squares design, in which case the constructed regressor acts as the instrument. Both types of analyses are common in the literature. We assume the following general model to derive our main results:

$$y_i = \beta z_i + \boldsymbol{x}_i'\boldsymbol{\delta} + \eta_{C(i)} + \varepsilon_i, \tag{2}$$

$$z_i = \frac{1}{|G(i)|} \sum_{j \in G(i)} v_j, \tag{3}$$

for $i = 1, 2, ..., n$. Here, index $i$ uniquely identifies one of a total of $n$ observations. $z_i$ is the constructed regressor which is based on the underlying, endogenous variable $v$. The constructed regressor for observation $i$ is calculated using a set of observations that we call construction "group", denoted by $G(i)$. The constructed regressor is calculated as the mean of variable $v$ in $G(i)$. Note that by the nature of constructed regressors, $i \notin G(i)$. We denote the cardinality of any set $A$ with

be used for direct inference, as is the case in Paravisini et al. (2023). We also note that Gabaix and Koijen (2024) are very careful in properly detailing how their instrument should be constructed in the presence of fixed effects.

the commonly used operator $|A|$. Note that our set-up allows for the case of $v_i = y_i$ such that our results subsume the Judge Fixed Effect instrument for case outcomes, among other applications. $\boldsymbol{x}_i = (x_{i1}, x_{i2}, ..., x_{ik})'$ is a $k \times 1$ vector of further regressors and $\boldsymbol{\delta}$ is the associated coefficient vector. The estimation features fixed effects. For observation $i$, the set of all observations used for the calculation of the fixed effect is called the "cell", denoted by $C(i) \in \mathcal{C}$. $\mathcal{C}$ is a partition of the data such that each observation belongs to exactly one cell. In Equation (2), the fixed effect for cell $C(i)$ is denoted by $\eta_{C(i)}$. While not typically used in this fashion, this notation subsumes several common panel structures, including Firm-Year panels, Firm-Market-Year panels, and repeated cross-sections, while allowing for a possibility of multiplicative fixed effects (e.g. Industry $\times$ Year Fixed Effects).[6] Important special cases of the model (2)-(3) are discussed in the subsections below. Using this notation is helpful for our purposes, because it makes transparent which fixed effect cell is used to de-mean observations in the estimation.

For ease of notation, we use lower-case bold fonts for vectors and upper-case bold fonts for matrices. All vectors are column vectors. We may rewrite Equation (2) in matrix notation, such that $\boldsymbol{y} = \beta\boldsymbol{z} + \boldsymbol{X}\boldsymbol{\delta} + \boldsymbol{\eta} + \boldsymbol{\varepsilon}$.[7] Let the average of any variable $x_i$ over elements of set $C(i)$ be $\bar{x}_{C(i)} = |C(i)|^{-1} \sum_{j \in C(i)} x_j$. Then we can define FE-de-meaned versions of any variable $x_i$ as $\tilde{x}_i = x_i - \bar{x}_{C(i)}$. Lastly, define $c_{ij} = I\left[i \in G(j)\right]$ with $I(.)$ being an indicator function.

We consider a simple illustrative set-up for the endogeneity of $v$, given by Assumption 1 below, and two general technical requirements regarding existence of the estimator and regularity on idiosyncratic errors and regressors given by Assumptions 4 and 5 in Appendix A.1.

**Assumption 1** (Endogenous $v$)**.** *For all $i, j = 1, 2, ..., n$, we have*

$$E(v_i \varepsilon_j) = \begin{cases} 0 & \text{for } i \neq j \\ \sigma_{v\varepsilon i} & \text{for } i = j \end{cases} \tag{4}$$

*where $\sigma_{v\varepsilon i} \neq 0$.*

Our endogeneity assumption is not just for ease of exposition, but more importantly our aim is to show that the bias we explore here can appear even if conditions for estimation are otherwise ideal. We thus explore the case where $v_i$ is only correlated with the error term of observation $i$ and not with any other error term. Note that this correlation does not have to be homogeneous across

[6] Additive fixed effects, such as $Firm + Year$ two-way fixed effects, would require the addition of extra $\eta$ terms in (2). Inclusion of additional fixed effects terms in (2) would not result in any new useful insights. In Online Appendix A we consider a single extension to one of our applications that further emphasizes this point. For clarity of exposition, we thus consider a single $\eta_{C(i)}$ term in the derivations below.

[7] $\boldsymbol{y} = (y_1, y_2, ..., y_n)'$, $\boldsymbol{z} = (z_1, z_2, ..., z_n)'$, $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n)'$, $\boldsymbol{\eta} = \left(\eta_{C(1)}, \eta_{C(2)}, ..., \eta_{C(n)}\right)'$, and $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)'$.

different observations, but can take different values for different $i$. This assumption corresponds to the best possible situation one could have while using constructed regressors in form of lags or leave-out variables. If $\sigma_{v\varepsilon i} = 0$ for all observations, then constructed regressors would be unnecessary in the first place. The very idea of the constructed regressor is to take advantage of $E(v_i \varepsilon_j) = 0$ for $i \neq j$.

Let $\hat{\beta}$ be the FE estimator of $\beta$ in model (2),

$$\hat{\beta} = \left(\tilde{\boldsymbol{z}}' \boldsymbol{M}_{\tilde{X}} \tilde{\boldsymbol{z}}\right) \tilde{\boldsymbol{z}}' \boldsymbol{M}_{\tilde{X}} \tilde{\boldsymbol{y}}, \tag{5}$$

where $\boldsymbol{M}_{\tilde{X}} = \boldsymbol{I}_n - \tilde{\boldsymbol{X}} \left(\tilde{\boldsymbol{X}}' \tilde{\boldsymbol{X}}\right)^{-1} \tilde{\boldsymbol{X}}'$.[8] We begin by stating the general result for the consistency of the FE estimator when the sets $C(i)$ and $G(i)$ overlap in an arbitrary way. The result allows for cases in which each set has observations that are not in the other but also for cases in which one is a subset of the other.

**Proposition 1.** *Let $y_i, z_i$ be given by (2)-(3), and suppose Assumptions 1, 4 and 5 hold. Consider the FE estimator $\hat{\beta}$ given by (5). Then, as $n \to \infty$,*

$$\hat{\beta} \to_p \beta - Q_{zx}^{-1} \Delta_\beta, \tag{6}$$

*where*

$$\Delta_\beta = \lim_{n\to\infty} \Delta_{\beta,n} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \frac{\sum_{j \in C(i)} c_{ji} \sigma_{v\varepsilon j}}{|G(i)|\,|C(i)|},$$

*$c_{ji} = I\left[j \in G(i)\right]$, $I(.)$ is an indicator function, and $Q_{zx} = plim_{n\to\infty} n^{-1} \tilde{\boldsymbol{z}}' \boldsymbol{M}_{\tilde{X}} \tilde{\boldsymbol{z}}$ is given by Assumption 5 in Appendix A.1.*

All proofs are provided in Appendix A.2.

This proposition shows that the overlap between the sets $C(i)$ and $G(i)$, given by $\sum_{j \in C(i)} c_{ji} = |G(i) \cap C(i)|$, can lead to inconsistency of $\hat{\beta}$.[9] For illustrative purposes, assume $\sigma_{v\varepsilon i} = \sigma_{v\varepsilon}$ for all $i$. Then $\Delta_{\beta,n}$ reduces to

$$\Delta_{\beta,n} = \sigma_{v\varepsilon} \frac{1}{n} \sum_{i=1}^{n} \frac{\sum_{j \in C(i)} c_{ji}}{|G(i)|\,|C(i)|} = \sigma_{v\varepsilon} \frac{1}{n} \sum_{i=1}^{n} \theta_i = \sigma_{v\varepsilon} \bar{\theta},$$

[8] Similarly to $\boldsymbol{y}$, $\boldsymbol{z}$, and $\boldsymbol{X}$, we define $\tilde{\boldsymbol{y}} = (\tilde{y}_1, \tilde{y}_2, ..., \tilde{y}_n)'$, $\tilde{\boldsymbol{z}} = (\tilde{z}_1, \tilde{z}_2, ..., \tilde{z}_n)'$, and $\tilde{\boldsymbol{X}} = (\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2, ..., \tilde{\boldsymbol{x}}_n)'$.

[9] Note that if Assumption 1 is relaxed to allow for $E(v_i \varepsilon_j) \neq 0$ for some $i \neq j$, the expression for the bias will be different and likely more complex.

where

$$\bar{\theta} = n^{-1} \sum_{i=1}^{n} \theta_i, \text{ and } \theta_i = \frac{|G(i) \cap C(i)|}{|G(i)|\,|C(i)|}, 0 \leq \theta_i \leq 1. \tag{7}$$

Clearly, $\bar{\theta}$ depends on $n$ and it can tend to zero or a positive nonzero value ($\leq 1$), as $n \to \infty$, depending on given empirical setting. For instance, as long as $|C(i)|$ and $|G(i)|$ do not change with $n$ then the bias of $\hat{\beta}$ will not dimish in $n$ and $\hat{\beta}$ will be inconsistent so long as $|G(i) \cap C(i)| \neq 0$. A specific example of this case is when the constructed regressor is calculated from lagged past observations of the same firm and there are $Firm$ level fixed effects, then adding new firms changes nothing about the size of the construction group or the fixed effect cell. Only adding more years, and thus increasing $|C(i)|$, will decrease the bias. This example is discussed further in Subsection 2.2 below.

In general, $\bar{\theta}$ can be of order $n^{-\alpha}$, for some value of the exponent $\alpha$ in the range $0 \leq \alpha \leq 1$, denoted as $\bar{\theta} \approx n^{-\alpha}$. Then $\hat{\beta}$ is consistent if $\alpha > 0$. However, even if $\hat{\beta}$ is consistent, this does not mean inference would be valid. Under the usual regularity assumptions for $\sqrt{n}$ convergence rate of $\hat{\beta}$, we would need $\alpha > 1/2$ for the asymptotic distribution to be correctly centered at zero. This is a much stronger requirement than simply $\alpha > 0$. In the Monte Carlo section we illustrate that inference could be wrong even if the bias of $\hat{\beta}$ is asymptotically negligible (namely when $\hat{\beta} \to_p \beta$).

When thinking of Equation (2) as the first stage in an instrumental variable estimation, then an important implication of Proposition 1 is that the first stage F-Statistic can be spuriously large even when $\beta = 0$. If the data sample is increased along the $C(i)$ dimension, holding $|C(i)|$ and $|G(i)|$ constant, then the bias does not diminish in $n$ while the standard error of the estimated coefficient will approach zero, leading the first-stage F-Statistic to approach infinity. A common case where the data increase but $|C(i)|$ and $|G(i)|$ are constant is in the Judge Fixed Effects case, where $C(i)$ is constructed at the Court-Year Level. $G(i)$ is constructed Judge-Year level. Adding more Court-Years while keeping the number of cases per Judge-Year and Judges per Court-Year constant will not change the bias, but it will lower the standard error.[10] Leave-out construction examples are further discussed in Subsection 2.3 below.

Lastly it needs to be mentioned that while we show the bias and inference problems in a model with a simple constructed regressor, such variables are often transformed in applied empirical finance. Transformations such as a change in units for normalization or the application of logarithms will, however, not change the underlying problem. While the exact result from Proposition 1 might change (e.g., in the case of a nonlinear transformation of the constructed regressor), the underlying mechanism will still lead to a biased estimator and problematic inference.

[10] A similar argument is made by Kolesár (2013).

### 2.2 Lagged Construction Example

One way in which regressors can be constructed is the use of lags. We consider a very simple panel data model that uses a single lag, this time indexing observations as $(f,t)$ to indicate both the cross section unit $f$ and the time period $t$. One can think of the data structure as a panel of $F$ firms (indexed by $f = 1, 2, ..., F$) observed for $T$ years (indexed by $t = 1, 2, ..., T$), but the data can also be any other type of panel. For illustrative purposes, we focus on the case without additional control variables. The equation to be estimated is

$$y_{ft} = \beta v_{f,t-1} + \eta_f + \varepsilon_{ft}, \tag{8}$$

which is a special case of (2)-(3) with the datapoint $i$ given by the pair $(f,t)$, $\boldsymbol{x}_i = \boldsymbol{x}_{f,t} = 0$ (no additional regressors), $z_{ft} = v_{f,t-1}$, and $G(f,t) = \{(f, t-1)\}$.

The constructed regressor is thus given by $z_{f,t} = v_{f,t-1}$, so in terms of Equation (3), we are taking the average over a single observation (formally: $|G(f,t)| = 1$). This is the most common way of including lags in the empirical literature, though there are some applications in which the average is taken over multiple lags (see, e.g., Gao et al., 2023) and the result extends to such estimations.

We focus on a balanced panel, which is again for illustrative purposes. Assuming $v_{f,0}$ is not observed, then $T-1$ periods are available for estimation ($t = 2, 3, ..., T$), and we have $C(f,t) = \{(f,2), (f,3), ..., (f,T)\}$. Hence $|C(f,t)| = T-1$. The following corollary shows that the FE estimator is not consistent when $T$ is fixed, and the bias is of order $T^{-1}$.

**Corollary 1.** Consider the special case of model (2)-(3) given by (8), and suppose conditions of Proposition 1 hold, with $E(v_{ft}\varepsilon_{ft}) = \sigma_{v\varepsilon,f}$ for all $t$. Let $\hat{\beta}$ be the FE estimator of $\beta$ in (8), using a balanced sample on $T$ time periods and $F$ firms. Then, as $F \to \infty$ and $T$ is fixed,

$$\hat{\beta} \to_p \beta - Q_{v,T}^{-1}\Delta_{\beta,T}, \tag{9}$$

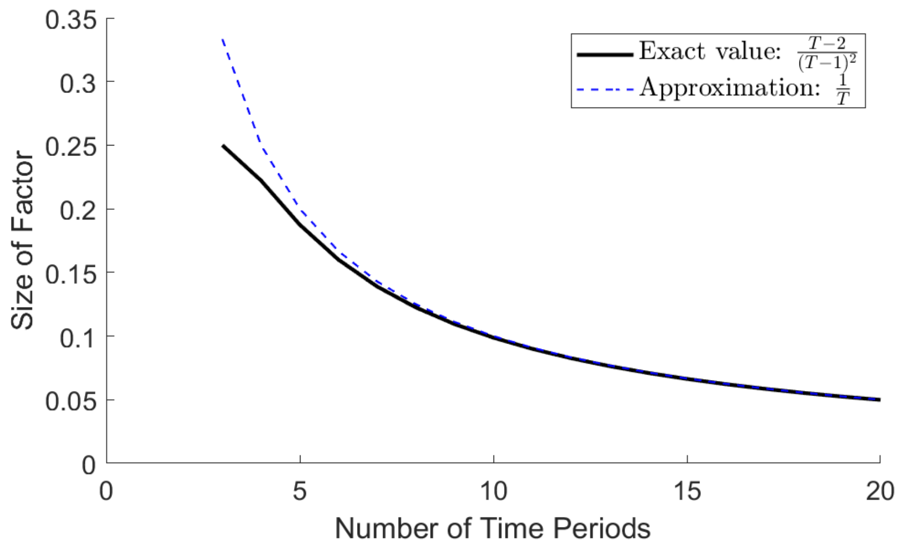
where

$$\Delta_{\beta,T} = \frac{T-2}{(T-1)^2}\bar{\sigma}_{v\varepsilon},$$

$\bar{\sigma}_{v\varepsilon} = \lim_{F\to\infty} F^{-1}\sum_{f=1}^{F}\sigma_{v\varepsilon,f}$ and $Q_{v,T} = plim_{F\to\infty}F^{-1}(T-1)^{-1}\sum_{f=1}^{F}\sum_{t=2}^{T}\tilde{v}_{f,t-1}^2$.

We can thus see that the bias described in Section 2.1 leads to a very tangible result in the commonly used regression specification where lagged regressors are combined with unit-level fixed effects. The overlap between the fixed effect cell (all observations for a unit $f$) and the construction group (lagged value of $v$ for unit $f$) is given by $(T-2)(T-1)^{-2}$. This is approximately equal to $1/T$ and only slightly smaller for very small values of $T$.[11] As can be seen in Figure 1, the bias can be quite sizeable and is particularly strong for panels with short $T$.

Figure 1: Extent of bias in the FE estimation of an illustrative panel regression with a single lag




*Note:* The graph shows the size of the factor $(T-2)(T-1)^{-2}$ from the result in Corollary 1 for a given number of time periods in the estimation. The blue dashed line shows the approximation $1/T$. The graph starts at $T = 3$. When $T = 2$, only one period is used in the estimation, thus not allowing for Firm-level fixed effects.

Corollary 1 is not qualitatively new. It is reminiscent of Nickell bias (Nickell, 1981) and it is a special case of the weak exogeneity bias derived in Chudik et al. (2018). This shows how our general result in Proposition 1 generalizes several known results from the literature. The intuition is the same as for the general result. Including the unit-level fixed effects is equivalent to demeaning the lagged regressor. Because about $1/T$ of this mean is the focal observation's value of the regressor, this share of the covariance between $v_{f,t}$ and $\varepsilon_{f,t}$ biases the estimate of the coefficient $\beta$.

Under the usual regularity conditions, the FE estimator of $\beta$ in Equation (8) will converge at the rate of $\sqrt{FT}$. As long as both dimensions $F, T \to \infty$, $\hat{\beta}$ is consistent. However, the asymptotic distribution of $\hat{\beta}$ will be correctly centered at zero only if $\sqrt{FT}Q_{v,T}^{-1}\Delta_{\beta,T} \to 0$. Since $\Delta_{\beta,T}$ in Corollary 1 is of order $1/T$, this condition is met only when $F/T \to 0$ as $F, T \to \infty$. Hence, the

[11] It is not exactly $1/T$ because the $v_{f,T}$ does not appear in any constructed regressors and $v_{f,1}$ never appears in a fixed effect cell so there is no overlap for these observations.

inference will be valid only if $F$ is sufficiently small relative to $T$. This rules out many applications in finance where $F$ (e.g., the number of firms) is typically quite large in comparison with $T$ (e.g., the number of years).

### 2.3 Jackknife IVs: The Leave-Out Construction Example

One of the other main applications of constructed regressors are leave-out means. These underlay the common "judge fixed effects" estimation strategy that is popular in both finance and economics (e.g., Dobbie et al., 2018; Farre-Mensa et al., 2020). Angrist et al. (1999) show that they are equivalent to Jackknife IV estimators (or JIVEs) which is why they are also sometimes referred to as such. The difference to the general result of Proposition 1 is that the construction groups are now set in a less flexible structure. Specifically, the estimation uses a leave-out setting if the following assumption is fulfilled

**Assumption 2** (Leave-out setting). *For each $i$ it holds that $G(i) = J(i) \setminus i$ with $J(i) \in \mathcal{J}$ and $\mathcal{J}$ being a partition of the data.*

We call $J(i)$ the jackknife set. In the leave-out setting the construction group of observation $i$ is comprised of all observations in $J(i)$ except $i$ itself. Thus, within a jackknife set, the construction groups of two observations differ only regarding a single observation: their own.

The idea is best exemplified by a concrete data application. For this, we borrow the setting of our empirical application described in Section 5. Here, examiners grant or refuse patent applications. Examiners differ in their strictness and the allocation of patent applications to examiners is random within the so-called art units. For causal inference, researchers estimate the average strictness of an application's examiner leaving out only the focal application. The applications handled by one examiner thus forms the jackknife set.[12]

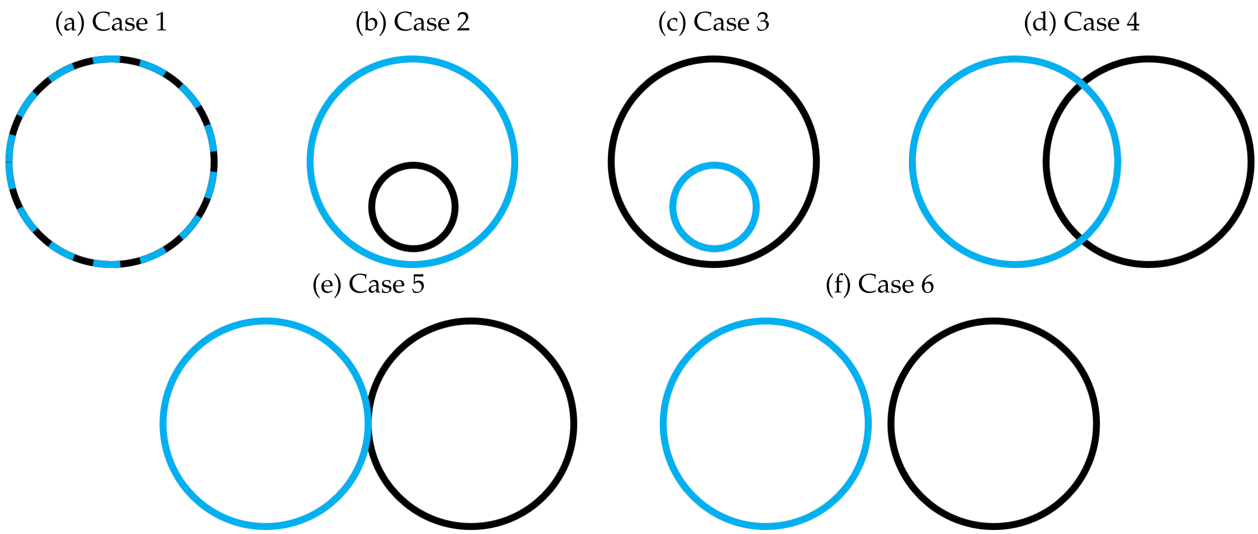
Because the leave-out setting is a special case of the general model, Proposition 1 carries over. The extent of the bias is now determined by the overlap of the jackknife set with the fixed effect cell. In the above example, if the estimation was combined with $Art\ Unit$ fixed effects, then the bias is determined by the share of applications an examiner handles within their own art unit.

This special case does, however, allow for interesting further results. Specifically, if the data is balanced in the sense that each jackknife set is of equal cardinality, then we can derive a number of

[12] A counterexample can make clear how the leave-out setting differs from the general case: When corporate finance researchers are interested in peer effects, they might define a company's peers as the competitors for the company's main product. Because company A being a competitor for company B's main product does not have to mean that said product is also company A's main product, this definition would violate Assumption 2.

illustrative analytical results. In particular, we can derive equivalent estimators that show how the fixed effects allow for the focal observation to bleed over into the constructed regressor. The small sample results cover six possible cases in which the data can be arranged. The cases are depicted in Figure 2 and cover different arrangements of how $C(i)$ and $J(i)$ can overlap. If the relationship between these sets is the same for all observations, the six shown cases are exhaustive.

Figure 2: Venn diagrams of possible cases for $J(i)$ (Jackknife set) and $C(i)$ (fixed effect cell)




*Note:* The diagrams show the schematic possibilities for the sets $J(i)$ (in blue) and $C(i)$ (in black). The tangency of the two sets in Case 5 indicates a single observation shared between both sets. The jackknife set $J(i)$ is used to calculate the constructed regressor in a leave-out mean, while the estimator is demeaned over the fixed effect cell $C(i)$. For all cases we only consider non-trivial structures such that $|J(i)| \geq 2$ and $|C(i)| \geq 2$ and in Case 4 that $|J(i)| \geq 3$ and $|C(i)| \geq 3$. The six cases can be described as $J(i) = C(i)$ (Case 1), $C(i) \subset J(i)$ (Case 2), $J(i) \subset C(i)$ (Case 3), $|J(i) \cap C(i)| \geq 2 \wedge J(i) \setminus C(i) \neq \varnothing \wedge C(i) \setminus J(i) \neq \varnothing$ (Case 4), $|J(i) \cap C(i)| = 1$ (Case 5), and $J(i) \cap C(i) = \varnothing$ (Case 6).

The cases are sorted by the severity of the problem. In Cases 1 and 2, de-meaning due to fixed effects leads to an instrument solely consisting of the focal value, $v_i$.[13] In Cases 3 and 4, the focal value is still contained in the equivalent instrument, but to a lesser degree. Here, Case 4 has a less severe bias than Case 3. Cases 5 and 6 show no bias. Case 5 allows combining a leave-out instrument with fixed effects, but requires additional dimensions in the panel. Case 6, which does not occur naturally, structures the instrument in a way such that the overlap is removed.

We now state these results formally. For notational convenience, we split $J(i)$ into two subsets. That which has shared elements with $C(i)$, denoted as $\hat{J}(i)$, and that which does not, which we denote $\check{J}(i)$. Formally, $\hat{J}(i) = J(i) \cap C(i)$ and $\check{J}(i) = J(i) \setminus \hat{J}(i)$. Because observation $i$ always needs

[13]This was already noted by Gormley and Matsa (2014) and is shown formally here.

to be in $C(i)$, it is clear that $i \in \hat{J}(i)$ and $i \notin \check{J}(i)$. The small sample properties of the first stage of the JIVE estimator with fixed effects can be derived under a balanced data structure. Letting $k_1$ and $k_2$ denote a generic constants that do not depend on the data structure, this is summarized in the next assumption.

**Assumption 3** (Balanced leave-out data)**.** *For each observation $i$ it holds that $|J(i)| = k_1$ and $|\hat{J}(i)| = k_2$. We exclude trivial cases by assuming that $k_1 \geq 2$.*

We can then state the results.

**Proposition 2.** *Under Assumptions 1 to 4, estimating Equation* (2) *with the FE estimator is equivalent to*

1. *estimating Equation* (2) *with the FE estimator and using*

$$z_i = -\frac{1}{|J(i)|} v_i$$

   *if $J(i) = C(i)$ (Case 1) or $C(i) \subset J(i)$ (Case 2).*

2. *estimating Equation* (2) *with the FE estimator and using*

$$z_i = -\frac{1}{|J(i)|} \left( v_i + \sum_{j \in C(i) \setminus J(i)} v_j \right)$$

   *if $J(i) \subset C(i)$ (Case 3).*

3. *estimating Equation* (2) *with the FE estimator and using*

$$z_i = -\frac{1}{|J(i)|} \left( v_i + \sum_{j \in C(i) \setminus J(i)} v_j - \sum_{j \in \check{J}(i)} v_j \right)$$

   *if $|J(i) \cap C(i)| \geq 2$, $J(i) \setminus C(i) \neq \varnothing$, and $C(i) \setminus J(i) \neq \varnothing$ (Case 4).*

*Further, the estimation is unbiased if $|J(i) \cap C(i)| = 1$ (Case 5) or $J(i) \cap C(i) = \varnothing$ (Case 6).*

There are several takeaways from this proposition. Item 1 shows that the leave-out instrument with fixed effects on the same or smaller level as the jackknife set is not an instrument at all. Instead, using it is equivalent to a multiplicative transformation on the focal value. Intuitively, the mean of the leave-out instrument contains all observations in $J(i)$ equally often. The leave-out instrument itself also includes all of these observations except the focal one. By subtracting one from the other, what is left is the focal observation and its mean across the set $J(i)$. If $v_i = y_i \; \forall \; i$, then the first stage is perfectly collinear.[14]

The intuition from Cases 1 and 2 carries over to Cases 3 and 4 in items 2 and 3 of Proposition 2, but the problem is not as severe as it is with more granular fixed effects. Because the de-meaning is now on a broader scale – there is less shared information between construction group and fixed effect cell – the instrument does not fully reduce to information about the focal observation.

Case 3 is particularly likely to appear. The patent examiner application introduced above is one example for this. Another one is a set of Hausman instruments for firms operating in multiple markets where the estimation features $Firm$ or $Year$ fixed effects, or both.[15] The commonplace appearance of Case 3 highlights the importance of our result. Even though the bias can disappear if the data becomes larger, this is only the case if the instrument itself is informative. The problem is, however, that if the instrument is not informative, the other element of the weighted sum will simply add noise orthogonal to the value that is to be instrumented. In this case, the focal market component will be the main determinant of the instrument's coefficient, leading to a bias. Increasing the sample along the relevant dimension will only increase the standard error. Thus, adding more data potentially only makes the fit of the first stage worse. A researcher unaware of our result might remove (or re-weight, as recommended by Coussens and Spiess, 2021) portions of their data to seemingly achieve a better fit in the first stage, while in actuality, all explanatory variation could be coming from the bias introduced by the focal market component.

Case 4 (and thus item 3 of Proposition 2) appears when there are observations in the Jackknife set which are not contained in the fixed effect cell. This makes the equivalent instrument include a part which is not biased by the de-meaning process, improving the performance of the estimator. In the example of the patent examiners Case 4 would appear if individual examiners handled patents in multiple art units.

The last part of the proposition shows a way towards potential solutions for the bias. If there is only a single observation shared by the Jackknife set and the fixed effect cell (Case 5), then that observation has to be the focal one. Since that implies that there are no shared observations between $J(i)$ and $C(i)$, there will be no bias in the estimation as can be seen in Equation (6). The intuition for this is straightforward. If $i$ is the only observation shared between the two sets, then it never enters the mean of the other observations within the same fixed effect cell. Stated differently, none

[14]It should be noted that most statistical programs will not treat this estimation as perfectly collinear. Rather, it will return a highly significant first stage and proceed with the estimation. For $v_i = x_i$ and $|J_i| = k_1 \; \forall \; i$, an easy way to see whether the instrument reduces to the focal value is to check whether the coefficient on the instrument in the first stage is equal to $-\frac{1}{k_1-1}$.

[15]We provide a technical treatment of the two-way fixed effects case in Online Appendix A.

of the observations in $C(i)$ use observation $i$ in the calculation of their leave-out instrument. Observation $i$ is only used in the leave-out instrument for observations that are both not observation $i$ and in $J(i)$ at the same time. By the definition of the case, these observations are not part of $C(i)$. Similarly, no bias appears in Case 6 when there are no shared observations between $J(i)$ and $C(i)$. An interesting aspect of both Case 5 and 6 is that these solutions to the biased estimation are not tied to the number of fixed effect cells, i.e., they are not associated with fewer control variables. This is a new result in comparison to the extant literature on JIVE estimators, which generally considers the bias to be increasing in the number of control variables (Ackerberg and Devereux, 2009; Kolesár, 2013).

# 3 Solutions

Proposition 1 and the discussion of Cases 5 and 6 in Section 2.3 provide an obvious path to a solution for the bias. Any estimation with a data structure such that $G(i)$ and $C(i)$ do not overlap will be unbiased. There are several generic ways in which this solution can be implemented and we introduce them with the help of examples in Subsection 3.1 below. However, these solutions are not always possible. In addition, specific situations may have specific solutions, which we discuss in Subsection 3.2.

## 3.1 Avoiding overlap between the sets $G(i)$ and $C(i)$

If the data is sufficiently rich, one way to ensure that there is no overlap between construction group $G(i)$ and fixed effect cell $C(i)$ is to interact the existing fixed effect with another fixed effect that is orthogonal to the construction group. Orthogonal in this context means that for each construction group, each observation is in a different cell of the new fixed effect. Consider the Hausman instruments case as an example. Let the regressor be constructed from observations of the same firm in the same year but in different markets and let the estimation feature $Firm$ fixed effects. If, instead of only including the $Firm$ fixed effects, we interact them with fixed effects for the market dimension (so $Firm \times Market$ fixed effects), then the $G(i)$ and $C(i)$ will not overlap for any observation $i$.[16] Thus, the estimation will be unbiased as long as the other assumptions necessary for leave-out instruments are met.[17] This is a good example of how adding *more* fixed effects as control variables can actually reduce (or, in this case, eliminate) the bias.

[16]We illustrate this solution in Section 4. Note that using $Firm + Market$ fixed effects will not reduce the bias.

[17]See, among others, Angrist (2014), Betz et al. (2018), and Borusyak and Hull (2023) for the discussion of the necessary assumptions.

If the data are not rich enough to introduce a fixed effect orthogonal to the construction group, one can attempt to separate the construction group and fixed effect cell altogether. In the leave-out setting, this would be Case 6. However, this data structure does not occur naturally, and must be induced by the econometrician through alternate construction of the instrument or fixed effects. We detail two potential ways to do this.

The easiest way to separate construction group and fixed effects cell can be used when there exists "outside-sample" data that is irrelevant to the direct research question, but helpful in estimating the identifying shock. In this case, the researcher can define the construction group on this outside-sample, which is data that is not used for the later estimation of Equation (2). This reduces the overlap of $G(i)$ and $C(i)$ to zero by constructions. A concrete example of this is given by Sampat and Williams (2019) who study the effect of patent protection on follow-on innovation in the human genome. They compare follow-on innovation across successful and unsuccessful patent applications, instrumenting for the success of applications using the strictness of the assigned patent examiner for patents that do not relate to the human genome. Because there is no overlap between the instrument construction and the fixed effects, there is no bias. This solution is a type of two-sample, two-stage least squares estimation (Angrist and Krueger, 1992; Inoue and Solon, 2010). Another example of such a procedure is given in Section 5.

When there is no outside-sample instrument available and the identifying variation is a subset of the necessary fixed effect, a split-sample approach might be possible. The basic idea is to separate the data into a regressor construction sample and an estimation sample. Data used for the construction of the regressor is then simply not used for the estimation of the equation and there can be no overlap between $G(i)$ and $C(i)$. As an example, consider the lagged regressor of Section 2.2. If the constructed regressor is the simple one-period lag, and it is known that there is no feedback from the error term $\varepsilon_{ft}$ to $v_{ft}$ beyond the contemporaneous value then the econometrician can limit the estimation to every second period. This way, the fixed effects de-meaning will not introduce correlation between the de-meaned error term and de-meaned regressor. However, such a strategy is no longer satisfactory if there is a feedback from the error term $\varepsilon_{f,t}$ to future values of $v_{ft}$ (to be discussed in the next subsection below). In this case, the bias will not be fully eliminated.

### 3.2 Case-Specific Solutions

Specific models have idiosyncratic solutions available to them. One important example is a Firm-Year panel where future values of regressors are not independent from the regression error term, such as model (8). An easy-to-implement solution to mitigate the bias of the FE estimator of $\boldsymbol{\beta}$ in Firm-Year panel data model is the split-panel jackknife approach by Dhaene and Jochmans (2015) and Chudik et al. (2018). This approach involves FE estimation of the full panel and its sub-samples, most commonly two half-panels. The half-panel jackknife FE estimator is given by

$$\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} - \widehat{bias} = 2\hat{\boldsymbol{\beta}} - \left(\hat{\boldsymbol{\beta}}_a + \hat{\boldsymbol{\beta}}_b\right)/2, \tag{10}$$

where $\hat{\boldsymbol{\beta}}$ is the FE estimator using the full sample, $\hat{\boldsymbol{\beta}}_a$ is the FE estimator using only the first half of the time periods, and $\hat{\boldsymbol{\beta}}_b$ is the FE estimator using the second half of the time periods. Since the bias that we discussed in Section 2.2, is of order $1/T$, then estimates using only one half of available time periods will feature a bias of order $1/(T/2)$, namely the bias of $\hat{\boldsymbol{\beta}}_a$ and $\hat{\boldsymbol{\beta}}_b$ is approximately twice larger than the bias of $\hat{\boldsymbol{\beta}}$. This allows for $\widehat{bias} = \left(\hat{\boldsymbol{\beta}}_a + \hat{\boldsymbol{\beta}}_b\right)/2 - \hat{\boldsymbol{\beta}}$ to be a good estimate of the FE estimator bias, up to the first order of approximation. The jackknife FE estimator $\tilde{\boldsymbol{\beta}}$ is still not guaranteed to be unbiased, but its bias is significantly reduced to order $T^{-2}$.

While straightforward to implement, there are few disadvantages to this jackknife bias correction. First, it is implicitly assumed that the correlation between the errors and future regressors does not meaningfully change over time. The second disadvantage is that while the bias is reduced by an order of magnitude, it is not fully eliminated. In settings where $T$ is quite small, $\tilde{\beta}$ will still be inconsistent as $F \to \infty$. Nevertheless, $\tilde{\beta}$ could still be helpful as a first pass in checking the potential size of the bias.

There are also readily available solutions in the literature to achieve consistent estimation when $T$ is quite small, and $F$ is large. Our suggested approach in Firm-Year panel data models is the double-filtered IV (DFIV) estimator proposed by Hayakawa et al. (2019).[18] The main idea of this approach is to avoid the bias from overlapping fixed effects by using deviations of regressors from their past means as instruments, and to transform the model by taking deviations from future means. These transformations fully remove Firm fixed effects, and the resulting estimator will be consistent even as $T$ is fixed and $F \to \infty$, which is the prevalent setting in finance and economic applications. We provide details on the DFIV estimator in Appendix A.3, and we illustrate benefits and drawbacks of DFIV as a solution to the bias problem in our Monte Carlo simulations below.

[18] This is an extremely developed literature. For panel autoregressive (AR) models, which is a special case of Firm-Year panels with lagged variables, there have been many solutions considered in the literature, including the IV/GMM methods (Anderson and Hsiao, 1981 and 1982, Arellano and Bond, 1991, Arellano and Bover, 1995, Blundell and Bond, 1998, Ahn and Schmidt, 1995 and 1997, Chudik and Pesaran, 2022), X-differencing method by Han et al. (2014), the first difference least square estimator by Han and Phillips (2010), the likelihood based methods (Hsiao et al., 2002, Lancaster, 2002, Moral-Benito, 2013, Hayakawa and Pesaran, 2015, Dhaene and Jochmans, 2016), or numerous bias correction methods. These methods differ in their assumptions on the initial values, Firm fixed effects, and assumptions for relative panel dimensions, $F/T$. Many of these approaches extend to panel AR models augmented with additional regressors/covariates.

The main downside of the DFIV estimator is a loss of precision compared to the FE estimator, which we document in Section 4. Since there is a cost in terms of efficiency, it is advisable to rely on DFIV only when the FE estimator bias is non-negligible. Thus, the researcher might want to conduct a test for the FE bias, in empirical applications where it is *a priory* unknown if there is any correlation between the error term and future values of the regressors. A straightforward test to consider in this case is a Hausman (1978) specification test, constructed based on the difference between the FE and the DFIV estimators,

$$H_F \;=\; (\hat{\boldsymbol{\beta}}_{\mathrm{DFIV}} - \hat{\boldsymbol{\beta}})' \Big[ \hat{\mathrm{Var}}(\hat{\beta}_{\mathrm{DFIV}}) - \hat{\mathrm{Var}}(\hat{\beta}) \Big]^{-1} (\hat{\boldsymbol{\beta}}_{\mathrm{DFIV}} - \hat{\boldsymbol{\beta}}), \tag{11}$$

assuming $\hat{\mathrm{Var}}(\hat{\beta}_{\mathrm{DFIV}}) - \hat{\mathrm{Var}}(\hat{\beta}) > 0$, where $\hat{\boldsymbol{\beta}}_{\mathrm{DFIV}}$ is DFIV estimator, $\hat{\boldsymbol{\beta}}$ is the FE estimator, and $\hat{\mathrm{Var}}(\hat{\beta}_{\mathrm{DFIV}})$ and $\hat{\mathrm{Var}}(\hat{\beta})$ are, respectively, consistent estimators of their asymptotic variances. Under the null hypothesis of no correlation between the error term and future values of regressors, both DFIV and FE estimators are consistent, but only DFIV is consistent when such correlation is present. Under the null, $H_F$ is asymptotically distributed as $\chi^2(k)$ for a fixed $T$ and as $F \to \infty$, where $k$ is the dimension of $\boldsymbol{\beta}$.

In the case of leave-out settings considered in Proposition 2, we know *a priori* that bias exists. Broad solutions to avoid bias in leave-out settings by avoiding the overlap between the fixed effects cell and the construction group are outlined in Subsection 3.1. If such solutions are not feasible due to data constraints, than three alternative estimation methods have been developed for the leave-out case in the literature: (1) the IJIVE (Ackerberg and Devereux, 2009), (2) the UJIVE (Kolesár, 2013), and (3) the CJIVE Frandsen et al. (2025b).[19] The IJIVE works by projecting out the fixed effects before performing the JIVE. The UJIVE and CJIVE instead directly remove the influence of the focal observation after the JIVE step has occurred, with the CJIVE also removing a researcher-specified "cluster."

# 4 Monte Carlo Evidence

To illustrate the magnitude of the potential bias outlined in Section 2 and its consequences for inference, we employ a series of Monte Carlo simulations in simple settings comparable to our examples above. We consider three sets of experiments in Subsections 4.1 to 4.3 below. The first set of experiments investigates the bias from including both a lagged regressor and fixed effects

[19] A STATA module for the CJIVE is available at Frandsen et al. (2025a). An (in-progress) R package is available at Butts (2025).

in a standard Firm-Year panel, such as (8). The second set of experiments showcases that bias also appears when the lagged variable is only used as a control. The last set of experiments adopts a Firm-Market-Year panel and demonstrates the bias of leave-out instruments under the various options for fixed effects discussed in Section 2.3.[20] In addition to illustrating the theoretical arguments made earlier, we also showcase the small sample performance of possible solutions. We intentionally choose a transparent and simple data generating processes to illustrate our main points as clearly as possible.

### 4.1 Experiments with a lagged regressor

We label the dependent variable as $y_{ft}$ for firm $f = 1, 2, ..., F$, and year $t = 1, 2, ..., T$. In all experiments, we consider a large number of firms, $F = 4,000$ or $20,000$ and a small number of years, $T = 5$ or 10. We simulate data according to[21]

$$\textbf{(DGP1)} \quad y_{ft} = \beta x_{f,t-1} + \varkappa \varepsilon^y_{ft}, \tag{12}$$

$$x_{ft} = \rho_x x_{f,t-1} + \varepsilon^x_{ft}, \quad \text{and} \quad \varepsilon^x_{ft} = \kappa_x \varepsilon^y_{ft} + e^x_{ft}, \tag{13}$$

where $\varepsilon^y_{ft}, e^x_{ft} \sim IIDN\left(0, 1\right)$ and the initial values are $x_{f,-1} = 0$. We assume $x_{f,0}$ is observed, so that $(F, T)$ is the effective sample size used for estimation. We set $\rho_x = 0.8$, and $\varkappa = 4$.[22] Note that here and in all DGPs below, we generate $x$ to be serially correlated, which we find appropriate for applied finance settings. Results are qualitatively comparable when setting $\rho_x = 0$.

We are interested in recovering $\beta$. The value of $\beta$, however, does not actually affect the small sample performance of the individual estimators reported below. The parameters of (12)-(13) are chosen such that the regression would provide a reasonable fit for $\beta = 1$. Nevertheless, we chose $\beta = 0$ for generating our data. This allows us to report a particularly intuitive measure for the accuracy of the inference in addition to other metrics. Specifically, we report the frequency of (falsely) statistically significant $\hat{\beta}$ across replications. Using any other value of $\beta$ leads to numerically identical small sample results for bias. This is in line with Proposition 1, where the bias expression does not depend on $\beta$.

[20] In Online Appendix B, we also consider the bias when instrumenting an endogenous regressor with lags.

[21] DGP1 can be seen as a special case of the general model (2)-(3) by setting $i = (f, t)$, $G\left(f, t\right) = \{(f, t-1)\}$, $v_{ft} = x_{ft}$ and the set defining Firm fixed effects is $C\left(f, t\right) = \{(f, 1), (f, 2), ..., (f, T)\}$. There is an overlap $|G\left(f, t\right) \cap C\left(f, t\right)| = 1$ and, which implies $\bar{\theta}$ given by (7) is of order $1/T$, and we can expect the bias to be of the same order when $\kappa_x \neq 0$, according to Proposition 1.

[22] Value $\rho_x = 0.8$ delivers a moderate serial correlation of the regressor. We set $\varkappa = 4$ so that the fit of the regression would not be unreasonably large if $\beta$ was equal to one.

We vary $\kappa_x$ between $0$, $0.2$ and $1$. This key parameter determines the relationship between $\varepsilon^y$ and $\varepsilon^x$. For $\kappa_x = 0$, $\varepsilon^x_{ft}$ is independently distributed of $\varepsilon^x_{f't'}$ for all $f, f', t, t'$, and the regressor $x$ is strictly exogenous. In this case the FE estimator is unbiased and inference based on the FE estimator is valid. When $\kappa_x \neq 0$, (which is presumably the reason why the regressor $x$ is lagged in the first place), then the FE estimator is subject to bias even though the regressor itself is not endogenous. This bias is stronger for $\kappa_x = 1$ than for $\kappa_x = 0.2$. For each value of $\kappa_x$, we adopt two estimators of $\beta$: the regular FE estimator with Firm fixed effects and the double-filtered IV (DFIV) estimator proposed by Hayakawa et al. (2019) as a possible solution. As with the FE estimator, the DFIV also allows for Firm fixed effects.

Table 2 reports our findings. For each simulation, we report four metrics. The simulated bias is the average deviation of $\hat{\beta}$ from its true value. The root mean square error (RMSE) is an inverse measure of estimation precision and can be seen as a relative cost of implementing an estimator. The 95% confidence interval coverage rates (CR) reports how often the estimated 95% confidence interval of $\hat{\beta}$ covers the true value of $\beta$ and is a measure for the accuracy of inference of the individual estimators. Lastly, we report the aforementioned frequency of statistically significant estimates at the $1$ percent level.[23]

The FE estimator works well only in the case of strict regressor exogeneity. Looking at the experiments with $\kappa_x = 0$ in the upper part of Table 2, we observe no discernible bias. The confidence interval coverage rates are all very close to 95 percent. Similarly, the share of statistically significant FE estimates at the 1% level is close to 0.01, indicating inference is accurate. In contrast, the FE estimator is severely biased in the remaining experiments, even for a relatively small value of $\kappa_x = 0.2$, which corresponds to only 4 percent variance share of $\varepsilon^x_{ft}$ explained by $\varepsilon^y_{ft}$. As predicted by theory, the bias of the FE estimator in experiments with $\kappa_x \neq 0$ does not depend much on $F$, and it declines with $T$. Even though the magnitude of the bias is relatively small for larger $T$ and smaller $\kappa_x$, the inference is highly inaccurate for all sample sizes considered. The 95% CRs all equal zero and the FE estimates of $\hat{\beta}$ are statistically significant at the 1% level in all Monte Carlo replications in experiments with nonzero $\kappa_x$. Thus, even when $\beta = 0$, the identified effect would have led researchers to the wrongful conclusion that the regressor is relevant. This is in line with the theoretical results from Section 2.2 where the requirement for valid inference ($\sqrt{FT}\bar{\theta} \approx \sqrt{F/T} \to 0$) is much stricter than that for consistency ($\bar{\theta} \approx T^{-1} \to 0$). Since $T$ is not

[23]Formally, let $\hat{\beta}^{(r)}$ be the estimate of $\beta$ in the Monte Carlo replication $r = 1, 2, ..., R$, and $CI^{(r)}_{\hat{\beta},95\%}$ be its respective 95 percent confidence interval estimate. Then, the bias of $\hat{\beta}$ is computed as $bias_{\hat{\beta}} = \frac{1}{R}\sum_{r=1}^{R}\left(\hat{\beta}^{(r)} - \beta\right)$, the root mean square error of $\hat{\beta}$ is computed as $RMSE_{\hat{\beta}} = \sqrt{\frac{1}{R}\sum_{r=1}^{R}\left(\hat{\beta}^{(r)} - \beta\right)^2}$, and the 95 percent confidence interval coverage rate is computed as $CR_{\hat{\beta},95\%} = \frac{1}{R}\sum_{r=1}^{R} I\left(\beta \in CI^{(r)}_{\hat{\beta},95\%}\right)$.

large relative to $F$ in our samples, we essentially find a zero probability that the true parameter value will fall in the estimated 95 percent confidence intervals for the FE estimator when $\kappa_x \neq 0$.

Table 2: Simulated bias, root mean square errors (RMSE), 95% confidence interval coverage rates (CR), and share of statistically significant estimates at 1% level of FE and DFIV estimators of $\beta = 0$ in experiments with lagged regressor

| | **Bias** | | **RMSE ($\times 100$)** | | **95% CR ($\times 100$)** | | **Significant $\hat{\beta}$ at 1% level** | |
|---|---|---|---|---|---|---|---|---|
| $F \setminus T$ | 5 | 10 | 5 | 10 | 5 | 10 | 5 | 10 |
| **A. Experiments with $\kappa_x = 0$ (regressor is strictly exogenous)** | | | | | | | | |
| FE estimator | | | | | | | | |
| 4,000 | 0.00 | 0.00 | 3.35 | 1.89 | 95.05 | 93.70 | 0.011 | 0.013 |
| 20,000 | 0.00 | 0.00 | 1.49 | 0.82 | 95.50 | 95.25 | 0.010 | 0.012 |
| DFIV estimator | | | | | | | | |
| 4,000 | 0.00 | 0.00 | 20.44 | 6.40 | 95.05 | 95.00 | 0.011 | 0.010 |
| 20,000 | 0.00 | 0.00 | 9.03 | 2.87 | 94.80 | 95.05 | 0.010 | 0.010 |
| **B. Experiments with $\kappa_x = 0.2$ (weak correlation between $\varepsilon^y_{ft}$ and $\varepsilon^x_{ft}$)** | | | | | | | | |
| FE estimator | | | | | | | | |
| 4,000 | -0.35 | -0.18 | 35.29 | 18.09 | 0.00 | 0.00 | 1.000 | 1.000 |
| 20,000 | -0.35 | -0.18 | 35.23 | 17.98 | 0.00 | 0.00 | 1.000 | 1.000 |
| DFIV estimator | | | | | | | | |
| 4,000 | 0.00 | 0.00 | 19.61 | 6.18 | 95.75 | 95.05 | 0.008 | 0.010 |
| 20,000 | 0.00 | 0.00 | 8.80 | 2.76 | 94.45 | 95.30 | 0.012 | 0.012 |
| **C. Experiments with $\kappa_x = 1$ (moderate correlation between $\varepsilon^y_{ft}$ and $\varepsilon^x_{ft}$)** | | | | | | | | |
| FE estimator | | | | | | | | |
| 4,000 | -0.92 | -0.47 | 91.57 | 46.82 | 0.00 | 0.00 | 1.000 | 1.000 |
| 20,000 | -0.92 | -0.47 | 91.58 | 46.77 | 0.00 | 0.00 | 1.000 | 1.000 |
| DFIV estimator | | | | | | | | |
| 4,000 | 0.00 | 0.00 | 14.71 | 4.51 | 94.60 | 94.90 | 0.014 | 0.013 |
| 20,000 | 0.00 | 0.00 | 6.41 | 2.01 | 95.75 | 95.25 | 0.009 | 0.011 |

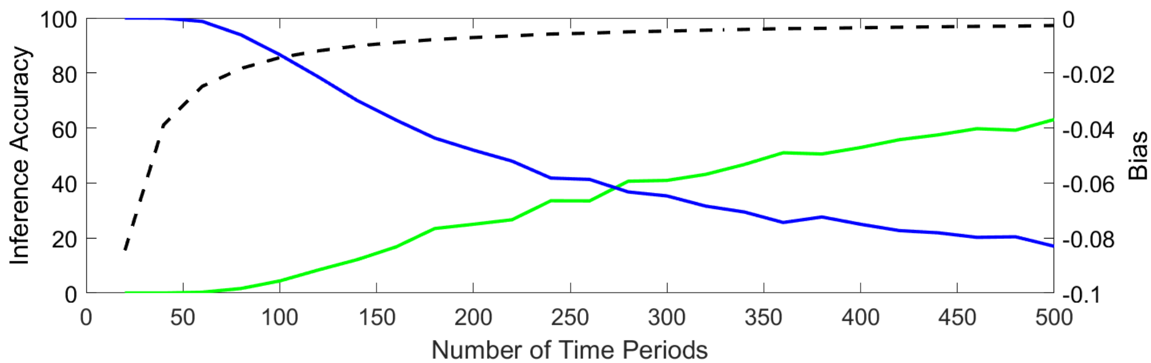
*Note:* The data generating process is given by $y_{ft} = \beta x_{f,t-1} + \varkappa\varepsilon^y_{ft}$, $x_{ft} = 0.8x_{f,t-1} + \varepsilon^x_{ft}$, and $\varepsilon^x_{ft} = \kappa_x\varepsilon^y_{ft} + e^x_{ft}$, for $t = 1, 2, ..., T$, and $f = 1, 2, ..., F$, where $\varepsilon^y_{ft}, e^x_{ft} \sim IIDN\,(0, 1)$ and $\varkappa = 4$. Estimating equation is $y_{ft} = \eta_f + \beta x_{f,t-1} + \varepsilon^y_{ft}$. DFIV is the double filter IV estimator by Hayakawa et al. (2019) using the regressor $x_{f,t-1}$ itself as an instrument. Simulation results are based on $R = 2,000$ Monte Carlo replications.

In contrast to the FE estimator, the DFIV estimator is consistent for all sample sizes considered, regardless of the value of $\kappa_x$. We see no notable bias of the DFIV estimator in Table 2 in any experiment. Confidence interval coverage rates are also close to 95 percent in all cases. These result confirm that the double filtered IV approach by Hayakawa et al. (2019) can be a useful case-specific solution for the lagged regressor construction. Unfortunately, there is also a cost to using the DFIV estimator – a loss in estimation precision. In experiments with a strictly exogenous regressor the RMSE of the DFIV estimator is at least three-fold higher than that of the FE estimator.

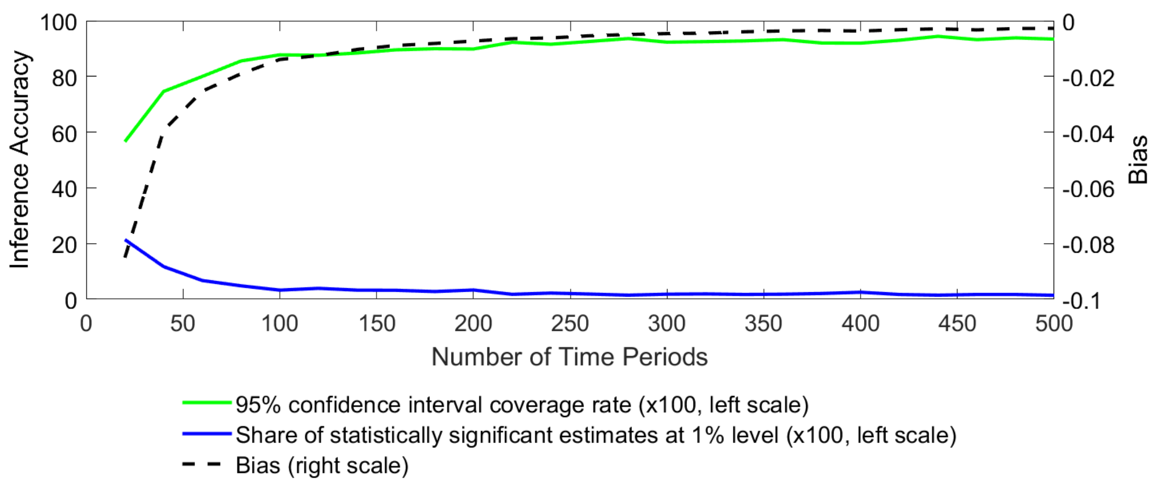
Given the poor performance of the FE estimator on inference, we next ask the question how large the time dimension needs to be for accuracy of inference to sufficiently improve in this de-

Figure 3: Bias, 95 percent confidence interval coverage rates and the share of statistically significant estimates at 1 percent level of FE estimator of $\beta = 0$ in experiments with lagged regressor and $T = 20, 40, ..., 500$.

(a) Experiments with $F = 4,000$ firms




(b) Experiments with $F = 200$ firms




*Note:* Values of $T = 20, 40, ..., 500$ are on X-axis. The data generating process is given by $y_{ft} = \beta x_{f,t-1} + \varkappa \varepsilon^{y}_{ft}$, $x_{ft} = 0.8 x_{f,t-1} + \varepsilon^{x}_{ft}$, and $\varepsilon^{x}_{ft} = \kappa_x \varepsilon^{y}_{ft} + e^{x}_{ft}$, for $t = 1, 2, ..., T$, and $f = 1, 2, ..., F$, where $\varepsilon^{y}_{ft}, e^{x}_{ft} \sim IIDN(0, 1)$ and $\varkappa = 4$. We set $\kappa_x = 0.2$ and $\beta = 0$. Estimating equation is $y_{ft} = \eta_f + \beta x_{f,t-1} + \varepsilon^{y}_{ft}$. Simulation results are based on $R = 2,000$ Monte Carlo replications.

sign. Panel (a) of Figure 3 shows the simulated bias, 95% CR and the share of statistically significant estimates in a sample of $F = 4,000$ firms for increasing values of $T = 20, 40, ..., 500$. As predicted by Corollary 1, the bias diminishes at the rate of $1/T$. The accuracy of inference also improves with an increase in $T$, but is still quite distorted even for $T = 500$ where in about 17 percent of cases the FE estimates are statistically significant at the 1 percent level. Results improve with a smaller sample size of $F = 200$ firms, as it is reported in panel (b). Here, $T = 100$ is sufficiently large for reasonably accurate inference, despite the fact that the magnitude of the bias is more-or-less unaffected by the cross-section dimension. For $F = 200$ and $T = 100$ in panel B, we see 95 confidence interval coverage rate of 87.75 percent and 3.25 percent probability of FE estimates being statistically significant at 1 percent level. Taken together, these results are not encouraging for the idea that the accuracy of inference can be improved by increasing the time dimension. For the vast majority of firm panels in finance research, $F$ is going to be significantly larger than $T$. While considering some additional years (or months) might make the bias of $\hat{\beta}$ less pronounced, even small panels of firms require long time series for each firm for inference to be acceptably accurate. Implementing the DFIV estimator, at least as a robustness analysis, seems to be a more promising way to deal with the issues discussed in this paper.

Since the cost of using the DFIV estimator can be large, we also consider the small sample performance of the Hausman test given in (11). Table 3 shows that the empirical size (the rejection rate of the null hypothesis when the null is true) of the Hausman test is close to the chosen nominal level of 5 percent. The simulated power (the rejection rate of the null hypothesis when the null is false) increases both with sample size and $\kappa_x$. For $\kappa_x = 0.2$, it appears to be sufficient (close to 100) only for larger $F$ and/or $T$. For $\kappa_x = 1$, good power is achieved for all our choices of the sample size. These results suggest that the Hausman test can provide useful guidance, but may lack the power to detect lower correlations between $\varepsilon_{ft}^x$ and $\varepsilon_{ft}^y$ in smaller samples.

Table 3: Empirical size and power of the Hausman test for the weak exogeneity bias in experiments with lagged regressor and $\beta = 0$

| | Size ($\times$100) | | Power ($\times$100) | | | |
|---|---|---|---|---|---|---|
| | $\kappa_x = 0$ | | $\kappa_x = 0.2$ | | $\kappa_x = 1$ | |
| $F \setminus T$ | 5 | 10 | 5 | 10 | 5 | 10 |
| 4,000 | 4.80 | 4.50 | 44.10 | 84.50 | 100 | 100 |
| 20,000 | 4.95 | 4.75 | 98.10 | 100 | 100 | 100 |

*Note:* The Hausman test is given by (11) and it is based on the difference between the FE and DFIV estimators of $\beta$. DFIV is the double filter IV estimator by Hayakawa et al. (2019) using the regressor itself as an instrument. The data generating process is given by $y_{ft} = \beta x_{f,t-1} + \varkappa \varepsilon^y_{ft}$, $x_{ft} = 0.8 x_{f,t-1} + \varepsilon^x_{ft}$, and $\varepsilon^x_{ft} = \kappa_x \varepsilon_{ft} + e^x_{ft}$, for $t = 1, 2, ..., T$, and $f = 1, 2, ..., F$, where $\varepsilon^y_{ft}, e^x_{ft} \sim IIDN\,(0, 1)$ and $\varkappa = 4$. The estimating equation is $y_{ft} = \eta_f + \beta x_{f,t-1} + \varepsilon^y_{ft}$. The reported empirical size is the rejection rate in experiments with $\kappa_x = 0$. Empirical power is the rejection rate in experiments with $\kappa_x = 0.2$ and 1. All experiments are based on $R = 2,000$ Monte Carlo replications.

### 4.2 Experiments with a lagged control variable

In the second set of experiments, we focus on estimation of $\beta$ in regression specifications that augment the regressor ($x$) with a lagged control variable, denoted as $w_{f,t-1}$. Consequently the estimating equation for the FE and DFIV estimators is given by a regression of $y$ on $x$, a lagged value of $w$ and the firm fixed effects (denoted by $\eta_f$):

$$y_{ft} = \beta x_{ft} + \delta w_{f,t-1} + \eta_f + \epsilon_{ft}.$$

We choose to generate data according to the design outlined below. In it, the equation for $y$ features only the strictly exogenous regressor $x_{ft}$ and not the control variable $w_{f,t-1}$. We make this choice to demonstrate that the inclusion of $w_{f,t-1}$ in the estimation can lead to a serious bias in the focal parameter $\beta$ even in the case when the control variable is not necessary. Formally:

**(DGP2)**
$$y_{ft} = \beta x_{ft} + \varkappa \varepsilon^y_{ft}, \tag{14}$$
$$x_{ft} = \rho_x x_{f,t-1} + \varepsilon^x_{ft}, \tag{15}$$
$$w_{ft} = \rho_w w_{f,t-1} + \varepsilon^w_{ft}, \quad \text{and} \quad \varepsilon^w_{ft} = \kappa_w(\varepsilon^y_{ft} - \varepsilon^x_{ft}) + e^w_{ft}, \tag{16}$$

for $f = 1, 2, ..., F$ and $t = 1, 2, ..., T$, where $\varepsilon^y_{ft}, \varepsilon^x_{ft}, e^w_{ft} \sim IIDN\,(0, 1)$ and the initial values are $w_{f,-1} = 0$. We set $\rho_x = \rho_w = 0.8$, and $\varkappa = 4$.

As before we consider $\beta = 0$ below. We vary the key parameter $\kappa_w$ between 0, 0.2 and 1. For $\kappa_w = 0$, the control variable $w$ is strictly exogenous, whereas for $\kappa_w \neq 0$, there is a feedback from $\varepsilon^y_{ft}$ and $\varepsilon^x_{ft}$ to $w_{ft}$. In this case, the contemporaneous values of the control variable are endogenous.

Regardless of the value of $\kappa_w$, the lagged control, $w_{f,t-1}$, is uncorrelated with $\varepsilon^y_{ft}$, which is the typical justification used for lagging control variables.

Table 4 reports the small sample findings for the FE and DFIV estimators. As in the previous experiments, we expect the FE estimator to work well only when $\kappa_w = 0$. This is confirmed in Panel A of Table 4, where we see no discernible bias, coverage rates close to 95 percent and the share of statistically significant estimates at the 1 percent level is in fact close to 1 percent. In Panel B, which features a small value of $\kappa_w = 0.2$, the bias of the FE estimator is non-zero, but small at -0.02 for all sample sizes. However, even a small bias causes inference problems. For $F = 20,000$ and $T = 10$, 41.1 percent of replications feature statistically significant $\hat{\beta}$ at the 1 percent level. As in Table 2, the bias and inference problems of the FE estimator in Table 4 are more serious for $\kappa_w = 1$. In this case, the bias is substantial. It is comparable in size to the bias in the previous experiments with lagged regressor and $\kappa_x = 0.2$.

In contrast to the FE estimator, the DFIV estimator works well for all experiments, regardless of the value of $\kappa_w$. However, similarly to findings in Table 2, the reported RMSE values in Table 4 show a significant cost of using the DFIV estimator when it is not strictly needed.

In Table 5 we see that the Hausman test is correctly sized, but that the power is good only if the magnitude of the bias and the sample size are sufficiently large. Specifically, we see a power of only 10.60 percent when $\kappa_w = 0.2$ and the sample size is $F = 20,000$ and $T = 10$. This shows that it is possible to have a serious problem with the accuracy of inference when using the FE estimator even when the Hausman test does not indicate that there is a problem.

### 4.3 Experiments with endogenous regressor in a Market-Firm-Year panel

Finally, we consider a Monte Carlo design with multiple markets. Let $x_{fmt}$ be the explanatory variable for firm $f = 1, 2, ..., F$, market $m = 1, 2, ....M$, and year $t = 1, 2, ..., T$. We set $M = 3$ and consider the same values for $F$ and $T$ as before. Data is generated according to:

$$\textbf{(DGP3)} \quad y_{fmt} = \beta x_{fmt} + \varkappa \varepsilon^y_{fmt}, \qquad \varepsilon^y_{fmt} = \nu e^x_{fmt} + e^y_{fmt}, \tag{17}$$

$$x_{fmt} = \rho_x x_{f,m,t-1} + \varepsilon^x_{f,m,t}, \quad \text{and} \quad \varepsilon^x_{fmt} = \gamma c_{ft} + e^x_{fmt} \tag{18}$$

where $e^x_{fmt}, e^y_{fmt}, c_{ft} \sim IIDN\,(0, 1)$ and the initial values are $x_{f,m,-1} = 0$. We set $\rho_x = 0.8$, $\nu = 1$, $\varkappa = 4$, and unless stated otherwise, $\gamma = 1$. As before, we set $\beta = 0$.

We identify $\beta$ through the leave-out strategy outlined in Assumption 2. Thus, we use the average of $x$'s for the same firm in the same year, but in other markets. This provides an estimate

Table 4: Simulated bias, root mean square errors (RMSE), 95% confidence interval coverage rates (CR), and share of statistically significant estimates at 1% level of FE and DFIV estimators for estimation of $\beta = 0$ experiments with lagged control variable

| | **Bias** | | **RMSE ($\times 100$)** | | **95% CR ($\times 100$)** | | **Significant $\hat{\beta}$ at 1% level** | |
|---|---|---|---|---|---|---|---|---|
| $F \setminus T$ | 5 | 10 | 5 | 10 | 5 | 10 | 5 | 10 |
| **A. Experiments with $\kappa_w = 0$ (control $w_{f,t-1}$ is strictly exogenous)** | | | | | | | | |
| FE estimator | | | | | | | | |
| 4,000 | 0.00 | 0.00 | 3.37 | 1.78 | 93.80 | 95.05 | 0.011 | 0.010 |
| 20,000 | 0.00 | 0.00 | 1.44 | 0.84 | 95.55 | 94.25 | 0.009 | 0.013 |
| DFIV estimator | | | | | | | | |
| 4,000 | 0.00 | 0.00 | 21.63 | 6.81 | 95.85 | 94.95 | 0.006 | 0.008 |
| 20,000 | 0.00 | 0.00 | 10.08 | 3.04 | 94.75 | 95.05 | 0.012 | 0.006 |
| **B. Experiments with $\kappa_w = 0.2$ (weak correlation between $\varepsilon^w_{ft}$ and $\varepsilon^{x,y}_{ft}$)** | | | | | | | | |
| FE estimator | | | | | | | | |
| 4,000 | -0.02 | -0.02 | 4.03 | 2.69 | 88.40 | 80.95 | 0.030 | 0.068 |
| 20,000 | -0.02 | -0.02 | 2.73 | 2.11 | 65.40 | 34.70 | 0.153 | 0.411 |
| DFIV estimator | | | | | | | | |
| 4,000 | 0.00 | 0.00 | 25.33 | 7.00 | 95.50 | 94.75 | 0.009 | 0.011 |
| 20,000 | 0.00 | 0.00 | 11.14 | 3.08 | 95.70 | 95.05 | 0.011 | 0.008 |
| **C. Experiments with $\kappa_w = 1$ (moderate correlation between $\varepsilon^w_{ft}$ and $\varepsilon^{x,y}_{ft}$)** | | | | | | | | |
| FE estimator | | | | | | | | |
| 4,000 | -0.21 | -0.19 | 21.53 | 19.46 | 0.00 | 0.00 | 1.000 | 1.000 |
| 20,000 | -0.21 | -0.19 | 21.28 | 19.20 | 0.00 | 0.00 | 1.000 | 1.000 |
| DFIV estimator | | | | | | | | |
| 4,000 | 0.00 | 0.00 | 26.20 | 7.07 | 95.55 | 94.70 | 0.009 | 0.010 |
| 20,000 | 0.00 | 0.00 | 11.90 | 3.19 | 94.80 | 94.80 | 0.010 | 0.010 |

*Note:* The data generating process is given by $y_{ft} = \beta x_{ft} + \varkappa \varepsilon^y_{ft}$, $x_{ft} = 0.8 x_{f,t-1} + \varepsilon^x_{ft}$, $w^x_{ft} = 0.8 w_{f,t-1} + \varepsilon^w_{ft}$, and $\varepsilon^w_{ft} = \kappa_w(\varepsilon^y_{ft} - \varepsilon^x_{ft}) + e^w_{ft}$, for $t = 1, 2, ..., T$, and $f = 1, 2, ..., F$, where $\varepsilon^y_{ft}, \varepsilon^x_{ft}, e^w_{ft} \sim IIDN(0, 1)$ and $\varkappa = 4$. Estimating equation is $y_{ft} = \eta_f + \beta x_{ft} + \delta w_{f,t-1} + \varepsilon^y_{ft}$. DFIV is the double filter IV estimator by Hayakawa et al. (2019) using the regressors themselves, $(x_{f,t}, w_{f,t-1})'$, as instruments. All experiments are based on $R = 2,000$ Monte Carlo replications.

of $c_{ft}$ that we then use as an instrument for $x_{fmt}$ when estimating $\beta$ via 2SLS. We refer to these IV estimators as leave-out (LO) estimators. For comparison, we also report on the FE estimator, which will be biased, due to the correlation between the regressor $x_{fmt}$ and the error term $\varepsilon^y_{fmt}$. We explore five forms of fixed effects to illustrate the findings of Proposition 1 and its special cases discussed in Proposition 2. The latter predicts that the bias of LO estimators (if any) depends on the choice of fixed effects.[24]

[24] DGP3 is a special case of the general model (2)-(3) with $i = (f, m, t)$, $v_{f,m,t} = x_{f,m,t}$, and the construction group for the instrument is given by $G(f, m, t) = J(f, m, t) \setminus \{f, m, t\}$ with $J(f, m, t) = \{(f, 1, t), (f, 2, t), ..., (f, M, t)\}$. Hence

Table 5: Empirical size and power of the Hausman test for the weak exogeneity bias in experiments with lagged control variable and $\beta = 0$

| | Size ($\times$100) | | Power ($\times$100) | | | |
|---|---|---|---|---|---|---|
| | $\kappa_w = 0$ | | $\kappa_w = 0.2$ | | $\kappa_w = 1$ | |
| $F \setminus T$ | 5 | 10 | 5 | 10 | 5 | 10 |
| 4,000 | 4.05 | 5.45 | 5.15 | 6.80 | 11.95 | 80.65 |
| 20,000 | 5.40 | 4.60 | 4.80 | 10.60 | 43.05 | 100 |

*Note:* The Hausman test is based on the difference between the FE and DFIV estimators of $\beta$. DFIV is the double filter IV estimator by Hayakawa et al. (2019) using the regressors themselves as instruments. The data generating process is given by $y_{ft} = \beta x_{ft} + \varkappa \varepsilon^y_{ft}$, $x_{ft} = 0.8 x_{f,t-1} + \varepsilon^x_{ft}$, $w^x_{ft} = 0.8 w_{f,t-1} + \varepsilon^w_{ft}$, and $\varepsilon^w_{ft} = \kappa_w(\varepsilon^y_{ft} - \varepsilon^x_{ft}) + e^w_{ft}$, for $t = 1, 2, ..., T$, and $f = 1, 2, ..., F$, where $\varepsilon^y_{ft}, \varepsilon^x_{ft}, e^w_{ft} \sim IIDN(0, 1)$ and $\varkappa = 4$. Estimating equation is $y_{ft} = \eta_f + \beta x_{ft} + \delta w_{f,t-1} + \varepsilon^y_{ft}$. Reported empirical size is the rejection rate in experiments with $\kappa_w = 0$. Empirical power is the rejection rate in experiments with $\kappa_w = 0.2$, and 1. All experiments are based on $R = 2,000$ Monte Carlo replications.

We begin by examining fixed effect structures consistent with Case 3 of Proposition 2. First we consider Firm fixed effects, which results in $|C(f, m, t)| = MT$ and $|G(f, m, t) \cap C(f, m, t)| = M - 1$. Hence, $\bar{\theta}$, as given by (7), is of order $1/(MT)$, and since both $M$ and $T$ are small in our experiments, we can expect a sizable bias, that diminishes with $M$ and $T$ but does not change with $F$. We next look at Year fixed effects. In this case, we have $|C(f, m, t)| = MF$, $|G(f, m, t) \cap C(f, m, t)| = M - 1$, and therefore $\bar{\theta}$ given by (7) is of order $1/(MF)$. Since $F$ is large in our design, we do not expect any sizable bias for this LO estimator. Third, we analyze Firm + Year fixed effects, which, given the discussion above and the results of Online Appendix A, is expected to lead to a similar bias as with Firm fixed effects alone. The fourth type of fixed effects is Firm $\times$ Year and corresponds to Case 1 of Proposition 2. This is the case where the LO estimator will be identical to the FE estimator. Last, we consider Firm $\times$ Market fixed effects, which corresponds to Case 5 of Proposition 2. This case is not expected to have any notable bias.

All of these theoretical predictions are borne out in findings reported in Table 6. All FE estimators are severely biased, regardless of the choice of fixed effects. LO estimators using Firm and Firm + Year fixed effects share similar and sizable bias which in some cases is worse than the bias observed when not instrumenting at all. In contrast, the bias of the LO estimator with Year fixed effects is not discernible. The LO estimator with Firm $\times$ Year fixed effects (Case 1) is numerically identical to the FE estimator with the same fixed effects. Our preferred solution (Case 5) using Firm $\times$ Market fixed effects does not have any discernible bias.

An additional implication of Proposition 2 is that, due to the bias of the first stage regressions, F-statistics can spuriously indicate instrument relevance. Figure 4 reports histograms for first-

$|G(f, m, t)| = M - 1$. The overlap between the set $G(f, m, t)$ and the fixed effects set $C(f, m, t)$ will determine the extent of the bias of LO estimators.

stage F-statistics of the LO estimators using Firm fixed effects only and Firm $\times$ Market fixed effects in experiments with $\gamma = 0$. This means that there is no correlation between the instrument and the regressor and thus the instrument is irrelevant.[25] Despite this irrelevance, the F-statistics for the first stage regressions with Firm fixed effects (Panel (a) of Figure 4) show values of around 17,000, spuriously indicating instrument strength. In contrast, Panel (b) shows that the spurious relevance disappears when the bias is removed.

[25] These experiments feature $F = 4,000$ firms, $M = 3$ markets, and $T = 5$ periods.

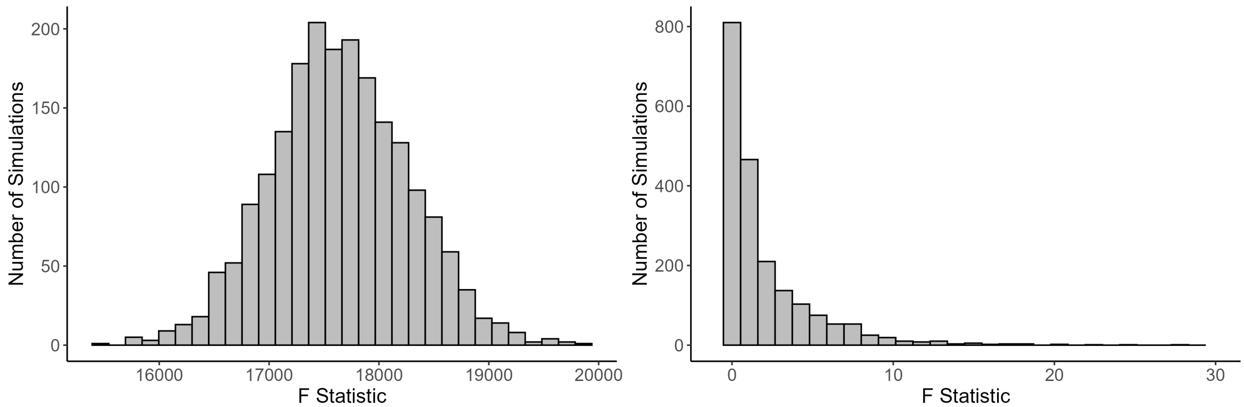
Table 6: Simulated bias, root mean square error (RMSE), 95% confidence interval coverage rates (CR), and share of statistically significant estimates at 1% level of FE and leave-out IV estimators of $\beta = 0$ in Firm-Market-Year panel experiments with regressor endogeneity

| | **Bias** | | **RMSE ($\times 100$)** | | **95% CR ($\times 100$)** | | **Significant $\hat{\beta}$ at 1% level** | |
|---|---|---|---|---|---|---|---|---|
| $F \setminus T$ | 5 | 10 | 5 | 10 | 5 | 10 | 5 | 10 |
| **A. Fixed Effect Estimators** | | | | | | | | |
| Firm Fixed Effects | | | | | | | | |
| 4,000 | 1.39 | 1.11 | 139.26 | 110.85 | 0.00 | 0.00 | 1.000 | 1.000 |
| 20,000 | 1.39 | 1.11 | 139.24 | 110.85 | 0.00 | 0.00 | 1.000 | 1.000 |
| Year Fixed Effects | | | | | | | | |
| 4,000 | 0.90 | 0.81 | 90.37 | 81.12 | 0.00 | 0.00 | 1.000 | 1.000 |
| 20,000 | 0.90 | 0.81 | 90.36 | 81.13 | 0.00 | 0.00 | 1.000 | 1.000 |
| Firm + Year Fixed Effects | | | | | | | | |
| 4,000 | 1.39 | 1.11 | 139.26 | 110.86 | 0.00 | 0.00 | 1.000 | 1.000 |
| 20,000 | 1.39 | 1.11 | 139.24 | 110.86 | 0.00 | 0.00 | 1.000 | 1.000 |
| Firm $\times$ Year Fixed Effects | | | | | | | | |
| 4,000 | 1.81 | 1.62 | 180.76 | 162.24 | 0.00 | 0.00 | 1.000 | 1.000 |
| 20,000 | 1.81 | 1.62 | 180.71 | 162.25 | 0.00 | 0.00 | 1.000 | 1.000 |
| Firm $\times$ Market Fixed Effects | | | | | | | | |
| 4,000 | 1.50 | 1.15 | 150.26 | 114.66 | 0.00 | 0.00 | 1.000 | 1.000 |
| 20,000 | 1.50 | 1.15 | 150.26 | 114.65 | 0.00 | 0.00 | 1.000 | 1.000 |
| **A. Leave-Out Instrumental Variable Estimators** | | | | | | | | |
| Firm Fixed Effects (Case 3) | | | | | | | | |
| 4,000 | -2.87 | -0.57 | 288.02 | 56.78 | 0.00 | 0.00 | 1.000 | 1.000 |
| 20,000 | -2.84 | -0.57 | 284.72 | 56.66 | 0.00 | 0.00 | 1.000 | 1.000 |
| Year Fixed Effects (Case 3) | | | | | | | | |
| 4,000 | 0.00 | 0.00 | 1.98 | 1.32 | 94.60 | 93.75 | 0.011 | 0.014 |
| 20,000 | 0.00 | 0.00 | 0.87 | 0.58 | 94.60 | 94.80 | 0.012 | 0.011 |
| Firm + Year Fixed Effects (Case 3) | | | | | | | | |
| 4,000 | -2.87 | -0.57 | 288.43 | 56.84 | 0.00 | 0.00 | 1.000 | 1.000 |
| 20,000 | -2.85 | -0.57 | 284.80 | 56.68 | 0.00 | 0.00 | 1.000 | 1.000 |
| Firm $\times$ Year Fixed Effects (Case 1) | | | | | | | | |
| 4,000 | 1.81 | 1.62 | 180.76 | 162.24 | 0.00 | 0.00 | 1.000 | 1.000 |
| 20,000 | 1.81 | 1.62 | 180.71 | 162.25 | 0.00 | 0.00 | 1.000 | 1.000 |
| Firm $\times$ Market Fixed Effects (Case 5) | | | | | | | | |
| 4,000 | 0.00 | 0.00 | 3.31 | 1.85 | 95.00 | 94.90 | 0.012 | 0.012 |
| 20,000 | 0.00 | 0.00 | 1.52 | 0.84 | 94.25 | 94.80 | 0.017 | 0.012 |

*Note:* The DGP is given by $y_{fmt} = \beta x_{fmt} + \varkappa \varepsilon^y_{fmt}$, $\varepsilon^y_{fmt} = e^x_{fmt} + e^y_{fmt}$, and $x_{fmt} = 0.8 x_{f,m,t-1} + c_{f,t} + e^x_{fmt}$, for $f = 1, 2, ..., F$, $t = 1, 2, ..., T$, and $m = 1, 2, 3$, where $e^x_{fmt}, e^y_{fmt}, c_{ft} \sim IIDN(0, 1)$, and $\varkappa = 4$. Leave-Out IV estimators are 2SLS estimators that use firm-specific averages of $x_{fmt}$ in other markets as an instrument, given by $z_{fmt} = \sum_{m' \neq m} x_{f,m',t}$. All experiments are based on $R = 2,000$ Monte Carlo replications.

Figure 4: Histograms of F-statistics for Leave-Out (LO) instrument relevance in Firm-Market-Year experiments with irrelevant instrument ($\gamma = 0$), $F = 4,000$, $M = 3$ and $T = 5$.

(a) LO-IV using Firm Fixed Effects (b) LO-IV using Firm $\times$ Market Fixed Effects




*Note:* The DGP is given by $y_{fmt} = \beta x_{fmt} + \varkappa \varepsilon^y_{fmt}$, $\varepsilon^y_{fmt} = e^x_{fmt} + e^y_{fmt}$, and $x_{fmt} = 0.8x_{f,m,t-1} + \gamma c_{f,t} + e^x_{fmt}$, for $f = 1, 2, ..., F$, $t = 1, 2, ..., T$, and $m = 1, 2, 3$, where $e^x_{fmt}, e^y_{fmt}, c_{ft} \sim IIDN\,(0, 1)$, and $\varkappa = 4$. Leave-out instrument is given by $z_{fmt} = \sum_{m' \neq m} x_{f,m',t}$. This instrument is not relevant when $\gamma = 0$. All experiments are based on $R = 2,000$ Monte Carlo replications.

# 5 Empirical Application: Judge Fixed Effects in USPTO Data

There is an active and important research agenda trying to understand the effects of decisions by judges and other officials in questions such as the incarceration of a defendant, the acceptance into a social program, the granting of patents, and other areas. The empirical challenge for such studies is that these decisions are not made at random, and thus, exogenous variation in the independent variable is difficult to find. Beginning with Kling (2006), a common empirical approach has been to use the leniency of randomly assigned decision-makers as an instrument, an approach popularly dubbed as "Judge Fixed Effects" (Frandsen et al., 2023, who also provide a literature overview of applications).

We provide an empirical demonstration of our bias in this application of the leave-out instrument using publicly available data from the U.S. Patent Office (USPTO). The data covers patent applications and unique patent examiner identifiers for 9.23 million patent applications from 1910 until 2014. This data, or data like it, have been used in combination with leave-out instruments to answer a variety of questions in finance, innovation research, and other areas (see, e.g., Farre-Mensa et al., 2020; Feng and Jaravel, 2020; Melero et al., 2020). To illustrate the effects of the bias, we consider a research question that can be answered using only the USPTO data and does not require additional, potentially proprietary or subscription-based, data. We thus focus on the question of whether a successful patent application makes inventors more likely to apply for another patent in the future. Data availability for unsuccessful patent applications requires us to focus on applications starting 2001 (see, e.g., Sampat and Williams, 2019, for detail). We focus on first-time applicants between 2001 and 2009, as identified by their full name on the patent application. They are categorized as a repeating inventor if the same name appears on a later patent application. This definition raises some measurement error concerns. However, we are focused on identifying the estimation bias rather than finding a credible answer to the research question. Discussions of the identification and other details of the analysis can be found in Online Appendix C.

After data restrictions, we consider 1.15 million inventors on 790,000 patent applications. The descriptive statistics of the sample can be seen in Table 7 below. Repeated applications are common, with 53.4% of inventors applying for one or more patents after the first one. We can see that about 64% of first applicants are successful, such that there are sufficiently many treated and untreated inventors. The leave-out instrument to measure examiner generosity is calculated from all examiner decisions in the year of the patent application. Notably, this also includes patent applications by inventors who are not applying for a patent for the first time. In total, we use 2.9 million patent applications for calculating the IV. Unsurprisingly, it has a mean close to 64%.[26] Important for the extent of our bias, we can see that the instrument is based on a median of 77 patent applications. This gives an indication of the magnitude of $|G(i)|$. We combine the analysis with $Art\ Unit \times Year$ fixed effects. With a median value of 19 examiners per art unit, we can further see that $|G(i)|/|C(i)|$ is about 0.05 in our analysis. This provides an idea of the size of the overlap between construction group and fixed effect cell.

To analyze the research question, we first simply regress the success of the application on the measure for an inventor with a repeated application, using $Art\ Unit \times Year$ fixed effects. Column (1) in Table 8 shows that this least squares (LS) analysis renders a positive and highly significant effect of past success on the likelihood of further applications. This estimation, however, suffers from obvious endogeneity problems. In column (2), we use the canonical combination of leave-out instruments and fixed effects to address this. The results show that instrumenting makes the coefficient smaller, but still economically meaningful and highly statistically significant. Furthermore, a five-digit F-statistic promises a strong instrument and valid identification strategy.

However, the estimate in column (2) suffers from bias due to a reintroduction of the focal value by the fixed effects as summarized by Proposition 1 in general and item 2 of Proposition 2 for this case in particular. Not all solutions mentioned in Section 3.1 are feasible here. Since there is no

[26]See also Figure C1 in Online Appendix C for the residualized distribution.

Table 7: Descriptive Statistics for Analysis Sample

| | Mean | St. Dev. | Median | 5% Pctl | 95% Pctl |
|---|---|---|---|---|---|
| Repeated Application | 0.533 | 0.4989 | - | - | - |
| Successful Application | 0.6394 | 0.4802 | - | - | - |
| Examiner Generosity | 0.6515 | 0.2293 | 0.6927 | 0.22 | 0.9667 |
| Examiner Generosity (Outside) | 0.6579 | 0.2349 | 0.7041 | 0.2121 | 0.9762 |
| Patents/ Examiner×Year | 971.8 | 4,609 | 77 | 19 | 340 |
| Patents/ Examiner×Year (Outside) | 378.7 | 1,969 | 49 | 13 | 186 |
| Examiners/ Art unit×Year | 25.85 | 22.08 | 19 | 11 | 82 |

*Note:* The table shows the descriptive statistics for the main analysis of the USPTO data. The first four rows and the last row are based on the 1.15 million inventors or 790 thousand applications in the estimation sample. Inventors have a repeated application if their name appears on at least one more patent application filed after the first one. An application is denoted successful if it has an issue date. Examiner generosity is the leave-out instrument. Outside stands for outside sample and denotes instruments calculated only from applications without any first time applicant on them. Patents/ Examiner×Year values are calculated directly from the 2.9 or 2 million patent applications which are used to calculate the leave-out instrument or outside sample instrument, respectively.

Table 8: Estimation Results USPTO Data

| | | Real Examiners | | Scrambled Examiners | |
|---|---|---|---|---|---|
| | (1) LS | (2) 2SLS | (3) 2SLS | (4) 2SLS | (5) 2SLS |
| Success | 0.085*** | 0.054*** | 0.045*** | 0.086*** | −5.060 |
| | (0.002) | (0.006) | (0.008) | (0.020) | (41.345) |
| 1st Stage F-Stat. | - | 48,046 | 5,712 | 1,309 | 0.01637 |
| Fixed effects | Art unit × Year | Art unit × Year | Art unit × Year | Art unit × Year | Art unit × Year |
| Instrument Group | - | All Applications | Non-first Applications | All Applications | Non-first Applications |
| Examiner | - | Real | Real | Scrambled | Scrambled |
| Clustered st. err. | Art unit | Art unit | Art unit | Art unit | Art unit |
| Observations | 1,146,706 | 1,146,706 | 1,146,706 | 1,146,706 | 1,146,706 |

*Note:* The table displays the results of the different estimations analyzing the effect of a successful patent application on the probability of the inventor applying for at least one additional patent in the data. We consider all inventors who filed for their first patent application in the years 2001 through 2009. 2SLS estimations use a leave-out instrument based on examiner generosity in all patents of that examiner in the year the focal patent is filed (columns (2) and (4)) or all applications without any first-time applicant on them in the same time frame (columns (3) and (5)). The analyses in columns (4) and (5) use examiner IDs scrambled within the fixed effect cell such that the instrument is uninformative. Standard errors are clustered on the level of the Art unit in each estimation. Stars *, **, and *** denote statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

obvious orthogonal dimension to the examiner using a fixed effect on such a dimension in an interaction is not possible here. Further, since most examiners are only working on applications in a single art unit, calculating the instrument outside of the fixed effect cell is also unfeasible.

This leaves us with the "outside-sample" approach following Sampat and Williams (2019). Here, we calculate the instrument using only patent applications in which none of the inventors are applying for the first time ($n$ = 2 million). Results for this instrument are shown in column (3). The estimated coefficient is 16.1% smaller than that in the canonical estimation and has a 29.6% higher standard error. The standard leave-out IV thus has a bias of 21.6% relative to that of LS.

To illustrate that the bias highlighted in this paper can lead to spuriously large first-stage F-statistics, we randomly scramble the examiner IDs between patents within an $Art\ unit \times Year$ cell. The leave-out instrument calculated on the basis of these scrambled examiner IDs is uninformative and should not lead to a viable identification, given that the applications are now allocated randomly to construction groups. Nevertheless, as can be seen in column (4) of Table 8, a regression with the canonical IV strategy leads to a coefficient that is a) highly statistically significant, b) has a first stage F-statistic above 1000, and c) is about as large as the LS coefficient. At face value, the instrument thus seems reliable and the endogeneity minimal. However, all statistical evidence is generated by the bias. Column (5) shows that using the correct outside-sample instrument from scrambled examiner IDs leads to a statistically insignificant result and a first-stage F-statistic of essentially 0.[27]

The empirical application thus shows that even in data structures that give the *ex-ante* impression that the bias could be small, it can still impact results in a meaningful way. Moreover, if the instrument is completely uninformative, the bias can make it appear informative and valid, instead. In this setting, the bias can also lead to seemingly statistically significant and economically meaningful results when there really are none.[28] On a positive note, the general identification strategy using leave-out instruments based on examiner generosity still seems valid. While first-stage F-statistics decline by over one order of magnitude when the corrected instrument is used, they are still very high.

## 6 Conclusions

In this paper, we have identified a bias that can arise when constructed regressors are combined with fixed effects. Fixed effects de-meaning can reintroduce the focal observation into the con-

[27] We also note that this shows that the bias cannot be generated from the potential covariance structure at the $Art\ unit \times Year$ level (as detailed in Frandsen et al., 2025b), since the outside sample scrambled IV is constructed from the same $Art\ unit \times Year$.

[28] The implications of the results are unaffected by using different definitions of the dependent variable, even if the results themselves are somewhat different. In all analyses, the use of scrambled examiner IDs leads to an informative canonical instrument and an uninformative corrected one. See Online Appendix C for details.

structed regressor, leading to mechanical correlation and biased estimates. We show this result generally and then apply it to two often used econometric specifications: lagged and leave-out regressors. For the latter, we categorize the potential specifications into six cases based on the relationship between the set used for fixed effects (fixed effects cell) and the set used to create the regressor (construction group). Irrespective of the application, bias is most severe when the fixed effect cell overlaps strongly with the construction group.

The bias described in this study may come as a surprise because an exhaustive fixed effect structure is typically seen as a sign of particularly robust identification. This might particularly be the case if the constructed regressors are used as instrumental variables, as is often the case. Here, the typical requirement is that the instrument needs to fulfill the exclusion restriction. In the notation of Equation (2), it is obvious under Assumption 1 that $z_i$ and $\varepsilon_i$ are uncorrelated. Despite this, $z_i - \bar{z}_{C(i)}$ and $\varepsilon_i - \bar{\varepsilon}_{C(i)}$ are mechanically correlated. Thus when examining the validity of an instrument, it is important to consider the entire estimation strategy, instead of defending assumptions piecemeal.[29]

Our findings have important implications for researchers using constructed regressors in finance and economics applications. Careful consideration must be given to the interaction between the fixed effects structure and construction of the regressor to avoid biased estimation and, in the case of instruments, inflated first-stage F-statistics. We illustrate the practical relevance of our results using Monte Carlo simulations and in an empirical application of the judge fixed effects estimation strategy. Combining constructed regressors with fixed effects can generate substantial bias. Guided by our theoretical results, we provide solutions to obtain consistent estimates, and we also point to existing methods in the literature for solutions.

As the use of constructed regressors continues to proliferate, our paper serves as a cautionary note and practical guide. We hope our analysis enables applied researchers to reap the benefits of constructed regressors while avoiding the pitfalls that arise from their interaction with fixed effects.

[29]Implications of the bias identified in this paper for the commonly used specification tests, with a specific focus on using constructed regressors as instruments, is covered in Online Appendix D.

## References


Ackerberg, D. A. and Devereux, P. J. (2009). Improved jive estimators for overidentified linear models with and without heteroskedasticity. The Review of Economics and Statistics, 91(2):351–362.

Ahn, S. C. and Schmidt, P. (1995). Efficient estimation of models for dynamic panel data. Journal of Econometrics, 68:5–27.

Ahn, S. C. and Schmidt, P. (1997). Efficient estimation of models for dynamic panel data models: Alternative assumptions and simplified estimation. Journal of Econometrics, 76:309–321.

Aiello, D. J. (2022). Financially constrained mortgage servicers. Journal of Financial Economics, 144(2):590–610.

Aizer, A. and Doyle, J. J. (2015). Juvenile incarceration, human capital, and future crime: Evidence from randomly assigned judges. The Quarterly Journal of Economics, 130(2):759–803.

Anderson, T. W. and Hsiao, C. (1981). Estimation of dynamic models with error components. Journal of the American Statistical Association, 76:598–606.

Anderson, T. W. and Hsiao, C. (1982). Formulation and estimation of dynamic models using panel data. Journal of Econometrics, 18:47–82.

Angrist, J. D. (2014). The perils of peer effects. Labour Economics, 30:98–108.

Angrist, J. D., Imbens, G. W., and Krueger, A. B. (1999). Jackknife instrumental variables estimation. Journal of Applied Econometrics, 14(1):57–67.

Angrist, J. D. and Krueger, A. B. (1992). The effect of age at school entry on educational attainment: an application of instrumental variables with moments from two samples. Journal of the American statistical Association, 87(418):328–336.

Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. The Review of Economic Studies, 58(2):277–297.

Arellano, M. and Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. Journal of Econometrics, 68:29–51.

Baker, A. C., Larcker, D. F., and Wang, C. C. (2022). How much should we trust staggered difference-in-differences estimates? Journal of Financial Economics, 144(2):370–395.

Benson, A., Li, D., and Shue, K. (2019). Promotions and the Peter Principle. The Quarterly Journal of Economics, 134(4):2085–2134.

Betz, T., Cook, S. J., and Hollenbach, F. M. (2018). On the use and abuse of spatial instruments. Political Analysis, 26(4):474–479.

Blundell, R. and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. Journal of econometrics, 87(1):115–143.

Borusyak, K. and Hull, P. (2023). Nonrandom exposure to exogenous shocks. Econometrica, 91(6):2155–2185.

Butts, K. F. (2025). jive: Jackknife instrumental variables estimation. https://github.com/kylebutts/jive. Git commit 393e4b8. Accessed 30 May 2025.

Chudik, A. and Pesaran, H. M. (2022). An augmented Anderson–Hsiao estimator for dynamic short-T panels. Econometric Reviews, 41(4):416–447.

Chudik, A., Pesaran, M. H., and Yang, J.-C. (2018). Half-panel jackknife fixed-effects estimation of linear panels with weakly exogenous regressors. Journal of Applied Econometrics, 33(6):816–836.

Chyn, E., Frandsen, B., and Leslie, E. (2024). Examiner and judge designs in economics: A practitioner's guide. Journal of Economic Literature, 63(2):401–439.

Coussens, S. and Spiess, J. (2021). Improving inference from simple instruments through compliance estimation. arXiv preprint arXiv:2108.03726.

Dhaene, G. and Jochmans, K. (2015). Split-panel jackknife estimation of fixed-effect models. The Review of Economic Studies, 82(3):991–1030.

Dhaene, G. and Jochmans, K. (2016). Likelihood inference in an autoregression with fixed effects. Econometric Theory, 32:1178–1215.

Dobbie, W., Goldin, J., and Yang, C. S. (2018). The effects of pre-trial detention on conviction, future crime, and employment: Evidence from randomly assigned judges. American Economic Review, 108(2):201–240.

Eisenbach, T. M., Lucca, D. O., and Townsend, R. M. (2022). Resource allocation in bank supervision: Trade-offs and outcomes. The Journal of Finance, 77(3):1685–1736.

Farre-Mensa, J., Hegde, D., and Ljungqvist, A. (2020). What is a patent worth? evidence from the us patent "lottery". The Journal of Finance, 75(2):639–682.

Feng, J. and Jaravel, X. (2020). Crafting intellectual property rights: Implications for patent assertion entities, litigation, and innovation. American Economic Journal: Applied Economics, 12(1):140–181.

Frandsen, B., Lefgren, L., and Leslie, E. (2023). Judging judge fixed effects. American Economic Review, 113(1):253–277.

Frandsen, B., Leslie, E., and McIntyre, S. (2025a). CJIVE: Stata module to perform Cluster Jackknife Instrumental Variables Estimation (CJIVE). Statistical Software Components, Boston College Department of Economics.

Frandsen, B., Leslie, E., and McIntyre, S. (2025b). Cluster jackknife instrumental variables estimation. The Review of Economics and Statistics, forthcoming.

Fruehwirth, J. C., Iyer, S., and Zhang, A. (2019). Religion and depression in adolescence. Journal of Political Economy, 127(3):1178–1209.

Gabaix, X. and Koijen, R. S. (2024). Granular instrumental variables. Journal of Political Economy, 132(7):2274–2303.

Gao, J., Ge, S., Schmidt, L. D., and Tello-Tril, C. (2023). How do health insurance costs affect low- and high-income workers? Working Paper. HKU Jockey Club Enterprise Sustainability Global Research Institute - Archive, revised 24 Apr 2025.

Gormley, T. A. and Matsa, D. A. (2014). Common errors: How to (and not to) control for unobserved heterogeneity. The Review of Financial Studies, 27(2):617–661.

Grieser, W. D. and Hadlock, C. J. (2019). Panel-data estimation in finance: Testable assumptions and parameter (in) consistency. Journal of Financial and Quantitative Analysis, 54(1):1–29.

Han, C. and Phillips, P. C. B. (2010). GMM estimation for dynamic panels with fixed effects and strong instruments at unity. Econometric Theory, 26:119–151.

Han, C., Phillips, P. C. B., and Sul, D. (2014). X-Differencing and dynamic panel model estimation. Econometric Theory, 30:201–251.

Hausman, J. A. (1978). Specification tests in econometrics. Econometrica, 46(6):1251–1271.

Hausman, J. A. (1996). Valuation of new goods under perfect and imperfect competition. In The Economics of New Goods. University of Chicago Press. ISBN 0-226-07415-3.

Hayakawa, K. and Pesaran, M. H. (2015). Robust standard errors in transformed likelihood estimation of dynamic panel data models. Journal of Econometrics, 188:111–134.

Hayakawa, K., Qi, M., and Breitung, J. (2019). Double filter instrumental variable estimation of panel data models with weakly exogenous variables. Econometric Reviews, 38(9):1055–1088.

Honigsberg, C. and Jacob, M. (2021). Deleting misconduct: The expungement of brokercheck records. Journal of Financial Economics, 139(3):800–831.

Hsiao, C., Pesaran, M. H., and Tahmiscioglu, A. K. (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. Journal of Econometrics, 109:107–150.

Inoue, A. and Solon, G. (2010). Two-sample instrumental variables estimators. The Review of Economics and Statistics, 92(3):557–561.

Jiang, W. (2017). Have instrumental variables brought us closer to the truth. Review of Corporate Finance Studies, 6(2):127–140.

Kling, J. R. (2006). Incarceration length, employment, and earnings. American Economic Review, 96(3):863–876.

Kolesár, M. (2013). Estimation in an instrumental variables model with treatment effect heterogeneity. Working Paper.

Lancaster, T. (2002). Orthogonal parameters and panel data. Review of Economic Studies, 69:647–666.

Lavy, V. and Megalokonomou, R. (2024). The short- and the long-run impact of gender-biased teachers. American Economic Journal: Applied Economics, 16(2):176–218.

Ma, Y., Xiao, K., and Zeng, Y. (2022). Mutual fund liquidity transformation and reverse flight to liquidity. The Review of Financial Studies, 35(10):4674–4711.

Melero, E., Palomeras, N., and Wehrheim, D. (2020). The effect of patent protection on inventor mobility. Management Science, 66(12):5485–5504.

Moral-Benito, E. (2013). Likelihood-based estimation of dynamic panels with predetermined regressors. Journal of Business & Economic Statistics, 31:451–472.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. Econometrica, 69(2):307–342.

Nickell, S. (1981). Biases in dynamic models with fixed effects. Econometrica, 9(6):1417–1426.

Paravisini, D., Rappoport, V., and Schnabl, P. (2023). Specialization in bank lending: Evidence from exporting firms. The Journal of Finance, 78(4):2049–2085.

Petersen, M. A. (2008). Estimating standard errors in finance panel data sets: Comparing approaches. The Review of financial studies, 22(1):435–480.

Sampat, B. and Williams, H. L. (2019). How do patents affect follow-on innovation? Evidence from the human genome. American Economic Review, 109(1):203–236.

# Appendix

This appendix is organized in three sections. Section A.1 presents additional regularity assumptions and remarks. Section A.2 presents proofs. Section A.3 outlines DFIV estimator of Hayakawa et al. (2019).

## A.1 Additional Assumptions

**Assumption 4.** *(Existence of FE estimator)* $\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}$ *is invertible, and* $\tilde{\boldsymbol{z}}'\boldsymbol{M}_{\tilde{X}}\tilde{\boldsymbol{z}} > \boldsymbol{0}$.

**Remark 1.** Assumption 4 stipulates the requirements for the existence of the FE estimator. In essence, the assumption rules out perfect collinearity of the constructed and additional regressors.

**Assumption 5.** *(Idiosyncratic errors and regressors) Let* $\lambda_i = z_i\tilde{\varepsilon}_i - E\left(z_i\tilde{\varepsilon}_i\right)$, $\boldsymbol{P}_{\tilde{X}} = \tilde{\boldsymbol{X}}\left(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}\right)^{-1}\tilde{\boldsymbol{X}}'$, *and* $\boldsymbol{M}_{\tilde{X}} = \boldsymbol{I}_n - \boldsymbol{P}_{\tilde{X}}$, *where* $\boldsymbol{I}_n$ *is* $n \times n$ *identity matrix and* $\tilde{\boldsymbol{X}} = \left(\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2, ..., \tilde{\boldsymbol{x}}_n\right)'$. *As* $n \to \infty$,

*(i)* $n^{-1}\sum_{i=1}^{n}\lambda_i \to_p 0$,

*(ii)* $n^{-1}\tilde{\boldsymbol{z}}'\boldsymbol{M}_{\tilde{X}}\tilde{\boldsymbol{z}} \to_p Q_{zx} > 0$, *and*

*(iii)* $n^{-1}\tilde{\boldsymbol{z}}'\boldsymbol{P}_{\tilde{X}}\tilde{\boldsymbol{\varepsilon}} \to_p 0$.

**Remark 2.** Assumption 5 stipulates general high-level requirements on the idiosyncratic errors and regressors required for the consistency of the FE estimator. Condition $(i)$ allows for heteroskedastic and correlated idiosyncratic errors, but rules out strong correlations among errors. Condition $(ii)$ essentially ensures sufficient variation among $z_i$ once the fixed effects and regressors in $\boldsymbol{X}$ were filtered out. Condition $(iii)$ is an exogeneity requirement ensuring that the correlation between the regressors in $\boldsymbol{X}$ and idiosyncratic errors is sufficiently weak for it to not affect the consistency of the FE estimator.

## A.2 Proofs

*Proof of Proposition 1.* Using $\boldsymbol{y} = \beta\boldsymbol{z} + \boldsymbol{X}\boldsymbol{\delta} + \boldsymbol{\eta} + \boldsymbol{\varepsilon}$, the vector of demeaned variables $\tilde{\boldsymbol{y}}$ is given by $\tilde{\boldsymbol{y}} = \beta\tilde{\boldsymbol{z}} + \tilde{\boldsymbol{X}}\boldsymbol{\delta} + \tilde{\boldsymbol{\varepsilon}}$. Substituting this expression for $\tilde{\boldsymbol{y}}$ in (5), and noting that $\boldsymbol{M}_{\tilde{X}}\tilde{\boldsymbol{X}} = \boldsymbol{0}$, we obtain

$$\hat{\beta} - \beta = \left(\frac{\tilde{\boldsymbol{z}}'\boldsymbol{M}_{\tilde{X}}\tilde{\boldsymbol{z}}}{n}\right)^{-1}\frac{\tilde{\boldsymbol{z}}'\boldsymbol{M}_{\tilde{X}}\tilde{\boldsymbol{\varepsilon}}}{n}.$$

Under Assumption 5.(ii), $n^{-1}\tilde{\boldsymbol{z}}'\boldsymbol{M}_{\tilde{X}}\tilde{\boldsymbol{z}} \to_p Q_{zx}$, where $Q_{zx} > 0$. Consider $n^{-1}\tilde{\boldsymbol{z}}'\boldsymbol{M}_{\tilde{X}}\tilde{\boldsymbol{\varepsilon}}$ next. We have

$$\frac{\tilde{\boldsymbol{z}}'\boldsymbol{M}_{\tilde{X}}\tilde{\boldsymbol{\varepsilon}}}{n} = \frac{\tilde{\boldsymbol{z}}'\tilde{\boldsymbol{\varepsilon}}}{n} - \frac{\tilde{\boldsymbol{z}}'\boldsymbol{P}_{\tilde{X}}\tilde{\boldsymbol{\varepsilon}}}{n},$$

where $n^{-1}\tilde{\boldsymbol{z}}'\boldsymbol{P}_{\tilde{X}}\ \tilde{\boldsymbol{\varepsilon}} \rightarrow\ _p 0$, as $n \to \infty$ under Assumption 5.(iii). Last but not least, consider $n^{-1}\tilde{\boldsymbol{z}}'\tilde{\boldsymbol{\varepsilon}}$. Let $\boldsymbol{M}_C$ be the orthogonal projection matrix that filters out the fixed effects, namely $\tilde{\boldsymbol{z}} = \boldsymbol{M}_C\boldsymbol{z}$, and $\tilde{\boldsymbol{\varepsilon}} = \boldsymbol{M}_C\boldsymbol{\varepsilon}$. Since $\boldsymbol{M}_C$ is symmetric and idempotent, we have

$$\frac{\tilde{\boldsymbol{z}}'\tilde{\boldsymbol{\varepsilon}}}{n} = \frac{\boldsymbol{z}'\tilde{\boldsymbol{\varepsilon}}}{n} = \frac{1}{n}\sum_{i=1}^{n} z_i\tilde{\varepsilon}_i = \frac{1}{n}\sum_{i=1}^{n}\lambda_i + \Delta_{\beta,n},$$

where $\lambda_i = z_i\tilde{\varepsilon}_i - E\left(z_i\tilde{\varepsilon}_i\right)$, and

$$\Delta_{\beta,n} = \frac{1}{n}\sum_{i=1}^{n} E\left(z_i\tilde{\varepsilon}_i\right).$$

$E\left(\lambda_i\right) = 0$ by construction. In addition, under Assumption 5.(i), $\frac{1}{n}\sum_{i=1}^{n}\lambda_i \rightarrow_p 0$ as $n \to \infty$. Noting $\tilde{\varepsilon}_i = \varepsilon_i - \bar{\varepsilon}_{C(i)}$, where $\bar{\varepsilon}_{C(i)} = \left|C\left(i\right)\right|^{-1}\sum_{j\in C(i)}\varepsilon_j$, and using (3) for $z_i$, we can write $\Delta_{\beta,n}$ as

$$\begin{aligned}\Delta_{\beta,n} &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{|G(i)|}\sum_{j\in G(i)} E\left(v_j\varepsilon_i\right)\right)\\ &\quad -\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{|G(i)|\,|C(i)|}\sum_{h\in G(i)}\sum_{j\in C(i)} E\left(v_h\varepsilon_j\right)\right) \qquad (19)\end{aligned}$$

But $E\left(v_j\varepsilon_i\right) = 0$ for $i \neq j$ under Assumption 1. Noting $i \notin G(i)$, it follows $\sum_{j\in G(i)} E\left(v_j\varepsilon_i\right) = 0$. For the second term on the right side of (19), we obtain

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{|G(i)|\,|C(i)|}\sum_{h\in G(i)}\sum_{j\in C(i)} E\left(v_h\varepsilon_j\right)\right) = \frac{1}{n}\sum_{i=1}^{n}\frac{\sum_{j\in C(i)} c_{ji}\sigma_{v\varepsilon j}}{|G(i)|\,|C(i)|},$$

where $c_{ji} = I\left[j \in G(i)\right]$, and $I\left(.\right)$ is indicator function. These results establish

$$\hat{\beta} - \beta + Q_z^{-1}\Delta_{\beta,n} \rightarrow_p 0,$$

as $n \to \infty$, where

$$\Delta_{\beta,n} = \frac{1}{n}\sum_{i=1}^{n}\frac{\sum_{j\in C(i)} c_{ji}\sigma_{v\varepsilon j}}{|G(i)|\,|C(i)|}.$$

Result (6) follows.

□

*Proof of Corollary 1.* Form Proposition 1, we have

$$\Delta_{\beta,n} = \sum_{i=1}^{n}\frac{\sum_{j\in C(i)} c_{ji}\sigma_{v\varepsilon j}}{|G(i)|\,|C(i)|}.$$

But in the special case given by (8), $|G(i)| = 1$, $|C(i)| = T-1$, the summation $\sum_{i=1}^{n}$ becomes $F^{-1}(T-1)^{-1}\sum_{f=1}^{F}\sum_{t=2}^{T}$, $\sigma_{v\varepsilon j}$ becomes $\sigma_{v\varepsilon,f}$, and we obtain

$$\begin{aligned}\Delta_{\beta,(F,T)} &= F^{-1}(T-1)^{-1}\sum_{f=1}^{F}\sum_{t=2}^{T}\frac{I(t<T)\sigma_{v\varepsilon,f}}{T-1}, \\ &= \frac{T-2}{(T-1)^2}\frac{1}{F}\sum_{f=1}^{F}\sigma_{v\varepsilon,f}.\end{aligned}$$

Using $\bar{\sigma}_{v\varepsilon} = \lim_{F\to\infty} F^{-1}\sum_{f=1}^{F}\sigma_{v\varepsilon,f}$, we have

$$\Delta_{\beta,T} = \lim_{F\to\infty}\Delta_{\beta,(F,T)} = \frac{T-2}{(T-1)^2}\bar{\sigma}_{v\varepsilon}$$

In addition, $n^{-1}\tilde{\boldsymbol{z}}'\boldsymbol{M}_{\tilde{X}}\tilde{\boldsymbol{z}}$ reduces to $F^{-1}(T-1)^{-1}\sum_{f=1}^{F}\sum_{t=2}^{T}\tilde{v}_{f,t-1}^{2}$. Let

$$Q_{v,T} = plim_{F\to\infty}F^{-1}(T-1)^{-1}\sum_{f=1}^{F}\sum_{t=2}^{T}\tilde{v}_{f,t-1}^{2},$$

then result (9) follows. □

*Proof of Proposition 2.* We begin by showing item 3 (Case 4). The other items follow immediately from that. From the Frisch-Waugh-Lovell Theorem, estimating Equation (2) is equivalent to estimating

$$y_i^R = \beta z_i^R + \boldsymbol{X}_i^R\gamma + \varepsilon_i^R. \tag{20}$$

We consider $z_i^R = z_i - \frac{1}{|C(i)|}\sum_j \in C(i) z_j$. From Equation (3) and Assumption 2, we know that $z_i = \frac{1}{|G(i)|}\sum_{j\in G(i)} v_j = \frac{1}{|J(i)|-1}\sum_{j\in J(i)\backslash i} v_j$. Substituting, we obtain

$$z_i^R = \frac{1}{|J(i)|-1}\sum_{j\in J(i)\backslash i} v_j - \frac{1}{|C(i)|}\sum_{j\in C(i)}\frac{1}{|J(j)|-1}\sum_{h\in J(j)\backslash j} v_h, \tag{21}$$

$$\begin{aligned}&= \frac{1}{|J(i)|-1}\left(\sum_{j\in\hat{J}(i)} v_j - v_i\right) + \frac{1}{|J(i)|-1}\sum_{j\in\check{J}(i)} v_j \\ &\quad - \frac{1}{|C(i)|}\sum_{j\in C(i)}\frac{1}{|J(j)|-1}\left(\sum_{h\in\hat{J}(j)\backslash j} v_h + \sum_{h\in\check{J}(j)} v_h\right).\end{aligned} \tag{22}$$

By Assumption 3 it holds that $|J(i)| = k_1 \; \forall \; i$. It follows that

$$\begin{aligned}\dot{z}_i =& \frac{1}{|J(i)|-1}\left(\sum_{j\in\hat{J}(i)} v_j - v_i - \sum_{j\in C(i)\setminus\hat{J}(i)} v_j + \sum_{j\in C(i)\setminus\hat{J}(i)} v_j - \frac{1}{|C(i)|}\sum_{j\in C(i)}\sum_{h\in\hat{J}(j)\setminus j} v_h\right) \\ &+ \frac{1}{|J(i)|-1}\left(\sum_{j\in\check{J}(i)} v_j - \frac{1}{|C(i)|}\sum_{j\in C(i)}\sum_{h\in\check{J}(j)} v_h\right). \end{aligned} \tag{23}$$

By Assumptions 2 and 3, $\mathcal{J}$ is a partition of the data and $|\hat{J}(i)| = k_2 \; \forall \; i$. It thus holds that $\sum_{j\in C(j)}\sum_{h\in\hat{J}(j)\setminus j} v_h = (|\hat{J}(j)|-1)\sum_{j\in C(j)} v_j$. Rearranging the terms we obtain

$$\begin{aligned}\dot{z}_i =& \frac{1}{|J(i)|-1}\left(-v_i - \sum_{j\in C(i)\setminus\hat{J}(i)} v_j + \frac{|C(i)|-|\hat{J}(i)|+1}{|C(i)|}\sum_{j\in C(i)} v_j\right) \\ &+ \frac{1}{|J(i)|-1}\left(\sum_{j\in\check{J}(i)} v_j - \frac{1}{|C(i)|}\sum_{j\in C(i)}\sum_{h\in\check{J}(j)} v_h\right). \end{aligned} \tag{24}$$

We reverse the within transformation and obtain item 3 of the proposition. Item 2 follows by observing that in this case $\check{J}(i) = \varnothing$. Item 1 follows from observing that in these cases both $\check{J}(i) = \varnothing$ and $C(i) \setminus \hat{J}(i) = \varnothing$. The observations for Cases 5 and 6 follow from the original statement in Proposition 1 and observing that for both cases $G(i) \cap C(i) = \varnothing$. □

### A.3 Double-Filtered IV estimator

This section provides details on the double-filtered IV (DFIV) estimator of Hayakawa et al. (2019). Consider Firm-Year panel data model with individual observations indexed by $f = 1, 2, ..., F$ and $t = 1, 2, ..., T$. Suppose the dependent variable, denoted as $y_{ft}$, is given by

$$y_{ft} = \eta_f + \boldsymbol{\beta}'\boldsymbol{z}_{ft} + \varepsilon_{ft}, \tag{25}$$

where $\boldsymbol{z}_{ft}$ is a $k \times 1$ vector of regressors, we allow for $k \geq 1$, $\boldsymbol{\beta}$ is a $k \times 1$ vector of the corresponding slope coefficients, $\eta_f$ is the Firm fixed effect, and $\varepsilon_{ft}$ is the idiosyncratic error term. It is assumed $\boldsymbol{z}_{ft}$ is uncorrelated with $\varepsilon_{ft}$, but it can be correlated with future values of $\varepsilon_{ft}$. $\boldsymbol{z}_{ft}$ could contain regressors in form of lags, including the possibility of lags of the dependent variable. It could also include contemporaneous variables, so long as $\boldsymbol{z}_{ft}$ is uncorrelated with $\varepsilon_{ft}$. For future reference, it is convenient to stack (25), for $t = 1, 2, ..., T$, to obtain the model in a matrix form,

$$\boldsymbol{y}_f = \boldsymbol{\tau}_T\eta_f + \boldsymbol{Z}_f\boldsymbol{\beta} + \boldsymbol{\varepsilon}_f, \tag{26}$$

where $\boldsymbol{y}_f = (y_{f1}, y_{f2}, ..., y_{fT})'$, $\boldsymbol{\tau}_T$ is $T \times 1$ vector of ones, $\boldsymbol{Z}_f = (\boldsymbol{z}_{f1}, \boldsymbol{z}_{f2}, ..., \boldsymbol{z}_{fT})'$ and similarly $\boldsymbol{\varepsilon}_f = (\varepsilon_{f1}, \varepsilon_{f2}, ..., \varepsilon_{fT})'$.

DFIV estimation of $\boldsymbol{\beta}$ is based on the idea of instrumenting $\boldsymbol{z}_{ft}$ with itself, and applying forward filtering to the model and backward filtering to the instruments to eliminate Firm fixed effects. Specifically, define the forward filtering matrix

$$\underset{(T-1)\times T}{\boldsymbol{F}_T} = diag\left(c_1, c_2, ..., c_{T-1}\right) \left[\begin{array}{c|ccccc|c} 1 & \frac{-1}{T-1} & \frac{-1}{T-1} & \frac{-1}{T-1} & \cdots & \frac{-1}{T-1} & \frac{-1}{T-1} \\ \hline 0 & 1 & \frac{-1}{T-2} & \frac{-1}{T-2} & \cdots & \frac{-1}{T-2} & \frac{-1}{T-2} \\ 0 & 0 & 1 & \frac{-1}{T-3} & \cdots & \frac{-1}{T-3} & \frac{-1}{T-3} \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & & \ddots & 1 & \frac{-1}{2} & \frac{-1}{2} \\ 0 & 0 & \cdots & \cdots & 0 & 1 & -1 \end{array}\right],$$

where $c_t = \sqrt{(T-t)/(T-t+1)}$, for $t = 1, 2, ..., T-1$. Multiplying (26) by $\boldsymbol{F}_T$, and noting $\boldsymbol{F}_T\boldsymbol{\tau}_T = 0$, the model becomes

$$\dot{\boldsymbol{y}}_f = \dot{\boldsymbol{Z}}_f\boldsymbol{\beta} + \dot{\boldsymbol{\varepsilon}}_f,$$

where $\dot{\boldsymbol{y}}_f \equiv \boldsymbol{F}_T\boldsymbol{y}_f = (\dot{y}_{f1}, \dot{y}_{f2}, ..., \dot{y}_{f,T-1})'$, $\dot{\boldsymbol{Z}}_f \equiv \boldsymbol{F}_T\boldsymbol{Z}_f = (\dot{\boldsymbol{z}}_{f1}, \dot{\boldsymbol{z}}_{f2}, ..., \dot{\boldsymbol{z}}_{f,T-1})'$, and $\dot{\boldsymbol{\varepsilon}}_f = \boldsymbol{F}_T\boldsymbol{\varepsilon}_f$. Define next the backward filtering matrix,

$$\underset{(T-1)\times T}{\boldsymbol{B}_T} = diag\left(c_{T-1}, c_{T-2}, ..., c_1\right) \begin{bmatrix} -1 & 1 & 0 & 0 & \cdots & \cdots & 0 \\ \frac{-1}{2} & \frac{-1}{2} & 1 & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & & \ddots & \ddots & & \vdots \\ \frac{-1}{T-3} & \frac{-1}{T-3} & \cdots & \frac{-1}{T-3} & 1 & 0 & 0 \\ \frac{-1}{T-2} & \frac{-1}{T-2} & \cdots & \frac{-1}{T-2} & \frac{-1}{T-2} & 1 & 0 \\ \frac{-1}{T-1} & \frac{-1}{T-1} & \cdots & \frac{-1}{T-1} & \frac{-1}{T-1} & \frac{-1}{T-1} & -1 \end{bmatrix},$$

and the backward filtered variables $\ddot{\boldsymbol{Z}}_f \equiv \boldsymbol{B}_T\boldsymbol{Z}_f = (\ddot{\boldsymbol{z}}_{f2}, \ddot{\boldsymbol{z}}_{f3}, ..., \ddot{\boldsymbol{z}}_{f,T})'$, which are used as instruments. DFIV estimator of $\boldsymbol{\beta}$ is then given by

$$\hat{\boldsymbol{\beta}}_{DFIV} = \left(\sum_{f=1}^{F}\sum_{t=2}^{T-1} \ddot{\boldsymbol{z}}_{ft}\dot{\boldsymbol{z}}'_{ft}\right)^{-1} \left(\sum_{f=1}^{F}\sum_{t=2}^{T-1} \ddot{\boldsymbol{z}}_{ft}\dot{y}_{ft}\right). \tag{27}$$

Note, due to application of both forward and backward filters, the smallest time dimension for $\hat{\boldsymbol{\beta}}_{DFIV}$ to be defined is $T = 3$, in contrast to the FE estimator, which is defined for $T \geq 2$.

# Online Appendix A Leave-out Case 3 for Two-way Fixed Effects Estimators

We mention in Section 2.3 that two-way fixed effects estimators do not improve the situation for leave-out constructed regressors in Case 3 when compared to one-way fixed effects estimators. This appendix gives a formal result akin to item 2 in Proposition 2 for two-way fixed effects estimators under Assumptions 1 and 4. We consider a specific balanced panel structure in which for every $i$ the intersection of its two fixed effect cells is equal to the jackknife group $J(i)$. Note that the assumptions on the data structure outlined in the proposition replace assumptions 2 and 3 from Proposition 2. The data structure considered in the Monte Carlo simulations of Section 4.3 is a possible illustration of the case covered here. For the proof, we denote the two fixed effects cells for observation $i$ as $C(1,i)$ and $C(2,i)$ and the set of all observations as $\Omega$.

**Proposition 3.** *Let the data be structured that for all observations $i \in \Omega$ it holds that $J(i) \subset C(1,i)$, $J(i) \subset C(2,i)$ and $C(1,i) \cap C(2,i) = J(i)$. Further assume $|C(1,i)| = c_1 \ \wedge \ |C(2,i)| = c_2 \ \wedge \ |J(i)| = k \ \forall \ i$. Then, under Assumptions 1 and 4, estimating*

$$x_i = \beta z_i + \eta_{C(1,i)} + \eta_{C(2,i)} + \varepsilon_i \text{ with} \tag{A1}$$

$$z_i = \frac{1}{|J(i)|-1} \sum_{j \in J(i) \setminus i} v_j \tag{A2}$$

*with the FE estimator is equivalent to estimating Equation* (A1) *with*

$$z_i = -\frac{1}{|J(i)|-1} \left( v_i + \alpha \sum_{j \in C(1,i) \setminus J(i)} v_j + (1-\alpha) \sum_{j \in C(2,i) \setminus J(i)} v_j \right) \tag{A3}$$

*for any $\alpha \in \mathbb{R}$.*

*Proof.* From the Frisch-Waugh-Lovell Theorem, estimating Equation (A1) is equivalent to estimating

$$x_i - \bar{x}_{C(1,i)} - \bar{x}_{C(2,i)} + \bar{x}_\Omega = \beta(z_i - \bar{z}_{C(1,i)} - \bar{z}_{C(2,i)} + \bar{z}_\Omega) + \varepsilon_i - \bar{\varepsilon}_{C(1,i)} - \bar{\varepsilon}_{C(2,i)} + \bar{\varepsilon}_\Omega. \tag{A4}$$

We denote $z_i^R = z_i - \bar{z}_{C(1,i)} - \bar{z}_{C(2,i)} + \bar{z}_\Omega$ and substitute equation (A2)

$$
\begin{aligned}
z_i^R =& \frac{1}{k-1}\sum_{j\in J(i)\setminus i} v_j - \frac{1}{c_1}\sum_{j\in C(1,i)}\frac{1}{k-1}\sum_{h\in G_j\setminus j} v_h - \frac{1}{c_2}\sum_{j\in C(2,i)}\frac{1}{k-1}\sum_{h\in G_j\setminus j} v_h \\
&+ \frac{1}{c_1c_2}\sum_{j\in C(1,i)}\sum_{h\in C_{2j}}\frac{1}{k-1}\sum_{l\in G_h\setminus h} v_l \qquad \text{(A5)}\\
=& \frac{1}{k-1}\left(\sum_{j\in J(i)} v_j - v_i - \frac{k-1}{c_1}\sum_{j\in C(1,i)} v_j - \frac{k-1}{c_1}\sum_{j\in C(2,i)} v_j + \frac{k-1}{c_1c_2}\sum_{j\in C(1,i)}\sum_{h\in C_{2j}} v_h\right) \qquad \text{(A6)}\\
=& \frac{1}{k-1}\left(-v_i + \frac{1}{c_1}\sum_{j\in C(1,i)} v_j + \frac{1}{c_1}\sum_{j\in C(2,i)} v_j - \frac{1}{c_1c_2}\sum_{j\in C(1,i)}\sum_{h\in C_{2j}} v_h\right)\\
&+ \frac{1}{k-1}\left(\sum_{j\in J(i)} v_j + \alpha\left(\sum_{j\in C(1,i)\setminus J(i)} v_j - \sum_{j\in C(1,i)\setminus J(i)} v_j\right) + (1-\alpha)\left(\sum_{j\in C(2,i)\setminus J(i)} v_j - \sum_{j\in C(2,i)\setminus J(i)} v_j\right)\right.\\
&\left. - \frac{k}{c_1}\sum_{j\in C(1,i)} v_j - \frac{k}{c_1}\sum_{j\in C(2,i)} v_j + \frac{k}{c_1c_2}\sum_{j\in C(1,i)}\sum_{h\in C_{2j}} v_h\right) \qquad \text{(A7)}\\
=& \frac{1}{k-1}\left(-v_i + \frac{1}{c_1}\sum_{j\in C(1,i)} v_j + \frac{1}{c_1}\sum_{j\in C(2,i)} v_j - \frac{1}{c_1c_2}\sum_{j\in C(1,i)}\sum_{h\in C_{2j}} v_h\right)\\
&+ \frac{1}{k-1}\left(-\alpha\sum_{j\in C(1,i)\setminus J(i)} v_j - (1-\alpha)\sum_{j\in C(1,i)\setminus J(i)} v_j + \frac{\alpha c_1 - k}{c_1}\sum_{j\in C(1,i)} v_j\right.\\
&\left. + \frac{(1-\alpha)c_2 - k}{c_2}\sum_{j\in C(2,i)} v_j + \frac{k}{c_1c_2}\sum_{j\in C(1,i)}\sum_{h\in C_{2j}} v_h\right) \qquad \text{(A8)}\\
=& \frac{1}{k-1}\left(-v_i + \frac{1}{c_1}\sum_{j\in C(1,i)} v_j + \frac{1}{c_1}\sum_{j\in C(2,i)} v_j - \frac{1}{c_1c_2}\sum_{j\in C(1,i)}\sum_{h\in C_{2j}} v_h\right)\\
&+ \frac{1}{k-1}\left(-\alpha\sum_{j\in C(1,i)\setminus J(i)} v_j - (1-\alpha)\sum_{j\in C(1,i)\setminus J(i)} v_j + \alpha\frac{c_1 - k}{c_1}\sum_{j\in C(1,i)} v_j\right.\\
&\left. + (1-\alpha)\frac{c_2 - k}{c_2}\sum_{j\in C(2,i)} v_j - \frac{k(\alpha c_1 + (1-\alpha)c_2 - 1)}{c_1c_2}\sum_{j\in C(1,i)}\sum_{h\in C_{2j}} v_h\right) \qquad \text{(A9)}
\end{aligned}
$$

Reversing the within transformation across both clusters $C(1,i)$ and $C(2,i)$ renders the proposition. □

The proposition can be interpreted in different ways. However, the simplest intuition for the applied researcher is that with two-way fixed effects estimators, item 2 of Proposition 2 is true for both fixed effects at the same time. With a standard panel of firm data, when including both a $Firm$ and a $Year$ fixed effect, both $|T|$ and $|F|$ need to become large in order for the bias to disappear.

# Online Appendix B Additional Monte Carlo Findings

This appendix reports an additional Monte Carlo DGP featuring an endogenous regressor that can be instrumented with its lag in a standard Firm-Year panel. Data is generated according to:

$$\textbf{(DGPA1)} \quad y_{ft} = \beta x_{ft} + \varkappa \varepsilon^{y}_{ft}, \qquad \varepsilon^{y}_{ft} = \nu \varepsilon^{x}_{ft} + e^{y}_{ft}, \quad \text{and} \tag{B1}$$

$$x_{ft} = \rho_x x_{f,t-1} + \varepsilon^{x}_{ft}, \tag{B2}$$

for $f = 1, 2, ..., F$ and $t = 1, 2, ..., T$, where $\varepsilon^{x}_{ft}, e^{y}_{ft} \sim IIDN\,(0, 1)$ and the initial values are $x_{f,-1} = 0$. As before, we set $\beta = 0$, $\rho_x = 0.8$, and $\varkappa = 4$.[30] The parameter $\nu$ defines the degree of the endogeneity of the regressor. We consider three experiments based on $\nu_x = 0, 0.2$ and 1. To estimate $\beta$ we use a 2SLS IV estimator featuring firm fixed effects and use $x_{f,t-1}$ as the instrument for $x_{ft}$. We denote this estimator as IV-Lag below. In our DGPA1, this instrument is uncorrelated with $\varepsilon^{y}_{ft}$ and it is strong for all values of $\nu_x$ and all sample sizes considered. What is often overlooked in empirical research is that these two conditions are not sufficient for the IV-Lag estimator to be consistent. As discussed in Section 2.1, we expect bias of order $O(1/T)$, and consequently the IV-Lag estimator will be inconsistent as $F \to \infty$ and $T$ is fixed. In DGPA1, consistency of the IV-Lag estimator under large-$F$ asymptotics is achieved only if $\nu_x = 0$, which is when $x_{ft}$ is strictly exogenous, and instrumenting is not needed to begin with. As a possible solution, we also compute a double filtered IV estimator of Hayakawa et al. (2019) featuring Firm fixed effects and using the lagged regressor as the instrument. We denote this estimator as DFIV-Lag below.

Table B1 confirms our theoretical predictions. The reported biases of the IV-Lag estimator are all very large and the confidence interval coverage rates are all close to zero when $\nu \neq 0$. Unfortunately, it appears to be a common practice in the literature to use the IV-Lag estimator, even if it is clearly expected that the instrument is correlated with past and/or future values of $\varepsilon^{y}_{ft}$, in panel settings with $Firm$ Fixed effects and a relatively small time dimension. In contrast to the IV-Lag estimator, the DFIV-Lag estimator is consistent in this design, regardless of the value of $\nu$. We see no discernible bias of the DFIV-lag estimator in Table B1. Confidence interval coverage rates are also close to 95 percent in all cases. Hence, the DFIV-Lag estimator is a viable solution. Implementing it is again costly in terms of an increased root mean squared error. However, this cost can only be seen in Panel A of Table B1 and in the case shown there, instrumenting would not be necessary in the first place.

[30]The value of $\beta$ does not affect the small sample performance of the individual estimators considered.

Table B1: Simulated bias, root mean square errors (RMSE), 95% confidence interval coverage rates (CR), and share of statistically significant estimates at 1% level of FE, IV-lag and DFIV-lag estimators of $\beta = 0$ in Firm-Year panel experiments with regressor endogeneity

| | **Bias** | | **RMSE** ($\times 100$) | | **95% CR** ($\times 100$) | | **Significant** $\hat{\beta}$ **at 1% level** | |
|---|---|---|---|---|---|---|---|---|
| $F \setminus T$ | 5 | 10 | 5 | 10 | 5 | 10 | 5 | 10 |
| **A. Experiments with** $\nu = 0$ **(regressor is strictly exogenous)** | | | | | | | | |
| FE estimator | | | | | | | | |
| 4,000 | 0.00 | 0.00 | 3.38 | 1.89 | 94.95 | 94.35 | 0.011 | 0.012 |
| 20,000 | 0.00 | 0.00 | 1.53 | 0.84 | 94.30 | 94.90 | 0.010 | 0.014 |
| IV-lag estimator (using lagged regressor as instrument) | | | | | | | | |
| 4,000 | 0.00 | 0.00 | 10.58 | 3.37 | 94.90 | 94.90 | 0.011 | 0.013 |
| 20,000 | 0.00 | 0.00 | 4.64 | 1.51 | 95.30 | 94.95 | 0.009 | 0.012 |
| DFIV-lag estimator (using lagged regressor as instrument) | | | | | | | | |
| 1,000 | -0.01 | 0.00 | 22.45 | 7.42 | 94.95 | 94.45 | 0.009 | 0.009 |
| 4,000 | 0.00 | 0.00 | 9.87 | 3.36 | 95.25 | 94.35 | 0.012 | 0.010 |
| **B. Experiments with** $\nu = 0.2$ **(small degree of endogeneity)** | | | | | | | | |
| FE estimator | | | | | | | | |
| 4,000 | 0.60 | 0.46 | 60.20 | 45.89 | 0.00 | 0.00 | 1.000 | 1.000 |
| 20,000 | 0.60 | 0.46 | 60.17 | 45.86 | 0.00 | 0.00 | 1.000 | 1.000 |
| IV-lag estimator (using lagged regressor as instrument) | | | | | | | | |
| 4,000 | -1.18 | -0.35 | 118.34 | 35.30 | 0.00 | 0.00 | 1.000 | 1.000 |
| 20,000 | -1.18 | -0.35 | 117.78 | 35.09 | 0.00 | 0.00 | 1.000 | 1.000 |
| DFIV estimator | | | | | | | | |
| 1,000 | 0.00 | 0.00 | 23.06 | 7.37 | 94.70 | 95.20 | 0.011 | 0.009 |
| 4,000 | 0.00 | 0.00 | 10.31 | 3.30 | 94.65 | 94.50 | 0.009 | 0.011 |
| **C. Experiments with** $\nu = 1$ **(moderate degree of endogeneity)** | | | | | | | | |
| FE estimator | | | | | | | | |
| 1,000 | 3.01 | 2.29 | 300.57 | 229.34 | 0.00 | 0.00 | 1.000 | 1.000 |
| 4,000 | 3.01 | 2.29 | 300.53 | 229.25 | 0.00 | 0.00 | 1.000 | 1.000 |
| IV-lag estimator (using lagged regressor as instrument) | | | | | | | | |
| 4,000 | -5.89 | -1.75 | 589.49 | 175.60 | 0.00 | 0.00 | 1.000 | 1.000 |
| 20,000 | -5.89 | -1.75 | 589.15 | 175.31 | 0.00 | 0.00 | 1.000 | 1.000 |
| DFIV-lag estimator (using lagged regressor as instrument) | | | | | | | | |
| 1,000 | -0.01 | 0.00 | 32.22 | 10.64 | 94.25 | 94.70 | 0.012 | 0.010 |
| 4,000 | -0.01 | 0.00 | 14.24 | 4.74 | 94.95 | 94.90 | 0.011 | 0.011 |

*Note:* The DGP is given by $y_{ft} = \beta x_{ft} + \varkappa \varepsilon^{y}_{ft}$, $\varepsilon^{y}_{ft} = \nu \varepsilon^{x}_{ft} + e^{y}_{ft}$, and $x_{ft} = 0.8 x_{f,t-1} + \varepsilon^{x}_{ft}$, for $t = 0, 1, 2, ..., T$, $f = 1, 2, ..., F$, where $\varepsilon^{x}_{ft}, e^{y}_{ft}, \sim IIDN\,(0, 1)$ and $\varkappa = 4$. Estimating equation is $y_{ft} = \alpha_f + \beta x_{ft} + \varepsilon^{y}_{ft}$. IV-Lag is two-stage least squares (2SLS) instrumental variable estimator using lagged regressor, $x_{f,t-1}$, as an instrument. DFIV-Lag is the double filter IV estimator by Hayakawa et al. (2019) using lagged regressor, $x_{f,t-1}$, as an instrument. All experiments are based on $R = 2,000$ Monte Carlo replications.

# Online Appendix C Details on USPTO Data Analysis

We source data from the 2014 wave of the USPTO Patent Examination Research Dataset (PatEx). We chose this wave because, in contrast to later waves, it provides a unique examiner ID for each patent application, which reduces measurement error in the identification of the instrument group. We combine the data on applications with that on inventors. A unique inventor is identified by their first, middle, and last name. With this procedure, it is likely that in certain cases, multiple inventors are combined into a single person. We cannot address this measurement problem here, because we do not want to make arbitrary further data restrictions (such as focusing on a specific type of invention) or use additional data for our empirical demonstration. As a result, the research question is probably not identified to a satisfactory degree. However, we are interested in identifying the bias and not the particular research question. We further remove all entries in the data without an examiner ID and inventors with an undefined application date. This process mostly removes older entries in the data which are not of relevance to our analysis.

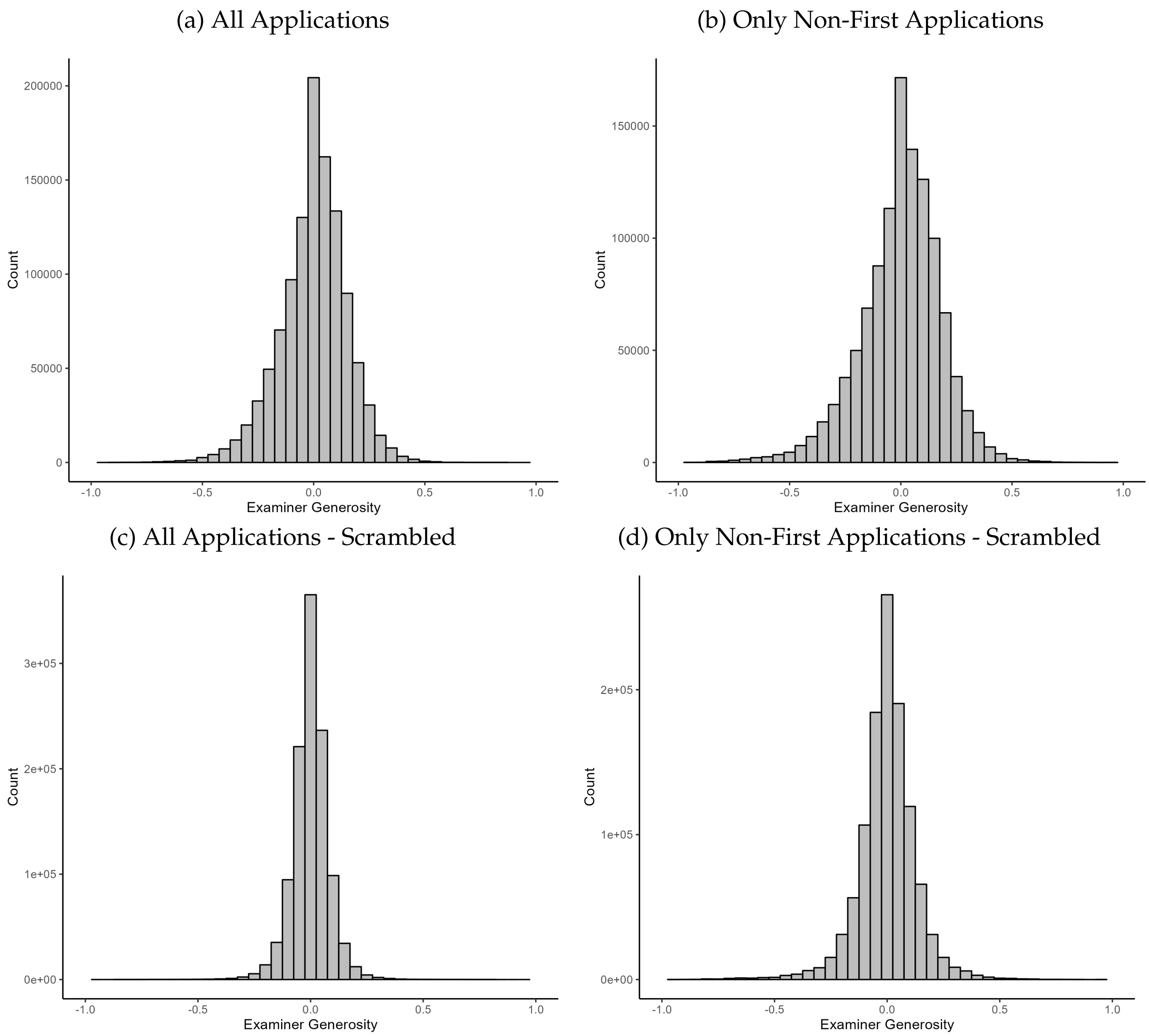
In the regression analyses, we focus on the success of the first application by an inventor. Because failed applications are only reliably documented in the data for applications filed on or after November 29th 2000 (Sampat and Williams, 2019), we focus on applications from 2001 onwards. An application is categorized as the first application if the inventor's name has not appeared on any previous application in the data. If inventors have multiple first applications on the same day, they are deleted from the data. An application is deemed successful if the patent has an issue date in the data. The dependent variable indicates whether the inventors are repeat applicants. They are classified as such if they file for at least one more patent after having filed for the first. We provide a robustness check for this definition below. Because we only observe data up to the year 2014 inclusively, we limit our analysis to the years 2001 through 2009, such that later applicants have sufficient time for a second application.

To calculate the leave-out instrument, we consider all decisions on applications by an examiner in a given year. This includes decisions on applications that are not by first-time applicants. Let $x_{i,e,t}$ be the $i^{\text{th}}$ decision by examiner $e$ in year $t$ and $n(e,t)$ the number of decisions by examiner $e$ in year $t$, then the canonical leave out instrument is given by

$$z_{i,e,t} = \frac{1}{n(e,t)-1}\left(\sum_{j=1}^{n(e,t)} x_{j,e,t} - x_{i,e,t}\right). \tag{C1}$$

The outside sample instrument instead considers all decisions on patents which do not have at least one first-time applicant on them. Denoting these decisions as $x_{i,e,t,\neg 1}$ and their total number

Figure C1: Histograms of Leave-out Instruments (Real and Scrambled Examiner IDs)




*Note:* The graphs show the histograms of the residualized leave-out instruments over the 1.15 million observations used in the analysis of the USPTO data. The instrument for all applications is defined over all decisions of the examiner in the year of the patent application (Equation (C1)) The non-first application instrument only considers examiner decisions on applications by inventors who do not apply for a patent for the first time (Equation (C2)). Scrambled instruments result from randomly assigned examiners within a and $\mathit{Art\ unit} \times \mathit{Year}$ cell.

for examiner $e$ in year $t$ as $n(e,t,\neg 1)$, the outside sample instrument is given by

$$z_{e,t} = \frac{1}{n(e,t,\neg 1)} \left( \sum_{i=1}^{n(e,t,\neg 1)} x_{i,e,t,\neg 1} \right). \tag{C2}$$

Note that in this case, all first-time applicants to examiner $e$ in year $t$ have the same value for the instrument. Distributions for both instruments are given in Figure C1. The figure also shows distributions for the scrambled examiner IDs. Note that we delete all observations from the data in which the denominator in Equation (C1) is 0. We do this for both the real examiner IDs and the scrambled ones such that all estimations in Table 8 are made on the same dataset.

One possible concern for the identification of the dependent variable could be that in the definition of the main analysis, we do not consider the timing of the decision. Rather, we say that any second application by an inventor after the first one constitutes a repeated application, irrespective of whether a decision on the first application has been rendered or not. To provide a robustness check for this identification, we repeat the analysis from the paper with a different definition of the dependent variable. Table C1 describes results when inventors are only categorized as repeated applicants if they apply after the decision on the first patent has been made (that is, after the issue date or the abandonment date).

Table C1: Estimation Results with Alternative Definition of Repeated Application

| | (1)<br>LS | (2)<br>2SLS | (3)<br>2SLS | (4)<br>2SLS | (5)<br>2SLS |
|---|---|---|---|---|---|
| Success | $0.041^{***}$<br>(0.001) | $0.091^{***}$<br>(0.004) | $0.106^{***}$<br>(0.005) | $0.048^{***}$<br>(0.015) | 0.898<br>(8.074) |
| 1st Stage F-Stat. | - | 36,191 | 4,901 | 1,260 | 0.02183 |
| Fixed effects | Art unit × Year | Art unit × Year | Art unit × Year | Art unit × Year | Art unit × Year |
| Instrument Group | - | All Applications | Non-first Applications | All Applications | Non-first Applications |
| Examiner | - | Real | Real | Scrambled | Scrambled |
| Clustered st. err. | Art unit | Art unit | Art unit | Art unit | Art unit |
| Observations | 1,112,915 | 1,112,915 | 1,112,915 | 1,112,915 | 1,112,915 |

*Note:* The table displays the results of the different estimations analyzing the effect of a successful patent application on the probability of the inventor applying for at least one additional patent in the data after the decision on the initial application has been rendered. We consider all inventors who filed for their first patent application in the years 2001 through 2009. 2SLS estimations use a leave-out instrument based on examiner generosity in all patents of that examiner in the year the focal patent is filed (columns (2) and (4)) or all applications without any first-time applicant on them in the same time frame (columns (3) and (5)). The analyses in columns (4) and (5) use examiner IDs scrambled within the fixed effect cell such that the instrument is uninformative. Standard errors are clustered on the level of the Art unit in each estimation. Stars *, **, and *** denote statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

Results are comparable to the main analysis with one important difference. While the main analysis shows the 2SLS coefficient to be smaller than the LS one, the results of the robustness analysis show the opposite pattern. This shows that on the one hand, the identification of the dependent variable is difficult given the present data, on the other hand, there could be two competing endogenous effects. The main analysis shows a quality effect in the sense that inventors with higher-quality patents are both more likely to have a successful application and more likely to have multiple applications. The current analysis implies that higher-quality patents are more likely to succeed in their application and, at the same time, might require fewer follow-up patents because they are already comprehensive. Which of these effects dominates then depends on the exact measurement of the dependent variable. Aside from the difference between LS and 2SLS estimates, the results are remarkably similar. The canonical instrument has a 23% relative bias compared to that of LS. More importantly, the scrambled analysis returns a coefficient about equal to that of the LS estimation with the canonical instrument and an uninformative instrument otherwise. As before, the canonical instrument seems highly valid in the scrambled analysis, with an F-statistic above 1000.

One way to make sure that fewer different inventors are combined in a single person for the sake of the estimation is to instead look for unique inventor-name-by-art-unit combinations. Such a process leads to a significantly increased sample size with roughly 3.86 million such combinations.

Table C2: Estimation Results with Alternative Identification of Individual Inventors

| | (1)<br>LS | (2)<br>2SLS | (3)<br>2SLS | (4)<br>2SLS | (5)<br>2SLS |
|---|---|---|---|---|---|
| Success | 0.072*** | 0.063*** | 0.053*** | 0.079*** | −2.097 |
| | (0.002) | (0.006) | (0.009) | (0.011) | (2.461) |
| 1st Stage F-Stat. | - | 75,512 | 1,525 | 2,214 | 0.999 |
| Fixed effects | Art unit × Year | Art unit × Year | Art unit × Year | Art unit × Year | Art unit × Year |
| Instrument Group | - | All Applications | Non-first Applications | All Applications | Non-first Applications |
| Examiner | - | Real | Real | Scrambled | Scrambled |
| Clustered st. err. | Art unit | Art unit | Art unit | Art unit | Art unit |
| Observations | 3,861,718 | 3,861,718 | 3,861,718 | 3,861,718 | 3,861,718 |

*Note:* The table displays the results of the different estimations analyzing the effect of a successful patent application in a given Art unit on the probability of the inventor applying for at least one additional patent in the same Art unit. We consider all inventors who filed for their first patent application in the years 2001 through 2009. 2SLS estimations use a leave-out instrument based on examiner generosity in all patents of that examiner in the year the focal patent is filed (columns (2) and (4)) or all applications without any first-time applicant on them in the same time frame (columns (3) and (5)). The analyses in columns (4) and (5) use examiner IDs scrambled within the fixed effect cell such that the instrument is uninformative. Standard errors are clustered on the level of the Art unit in each estimation. Stars *, **, and *** denote statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

Results are given in Table C2 and show the same pattern as in the main analysis. The bias is now more pronounced, it is 54% relative to LS. The analysis with scrambled examiner IDs is remarkably consistent between all three analyses. The coefficient when using the canonical leave-out instrument is close to that of LS, and the instrument has a first-stage F-statistic above 1000. However, when the outside-sample instrument is used, the F-statistic is essentially 0 and the weak instrument leads to a statistically insignificant coefficient.

# Online Appendix D Discussion and Relationship to Judge FE Tests

In this appendix, we will briefly discuss how our argument regarding leave-out means and their use as instrumental variables from Section 2.3 relates to other critiques and tests of leave-out instruments from the recent literature.

Some applications of instrumental variable estimation employ a test in which control variables are added step-wise to the first stage estimation and the econometrician observes whether the coefficient on the instrument changes in the process. This could, in theory, detect the problems identified in this paper, but in practice, it likely is not a reliable test. When considering a wholly uninformative constructed regressor as an instrument that contains mechanical correlation, the instrumented variables (call it $x$) is essentially instrumented with itself and noise. Introducing an independent control variable that is correlated with $x$ can change the coefficient on the instrument, because the control variable is correlated with both $x$ and the mechanical correlation part in the instrument. However, as can be seen in the numerical results of Section 4 the bias can affect the estimated coefficient both positively and negatively and it is unclear which of these effects will be affected the newly included independent variable. Thus, the problem might exist but, at the same time, might not be detected by the test. The test has an even higher chance to become unreliable if the instrument is informative, because if the included control variable is also correlated with the shock that the instrument is measuring, the net effect might again not be noticeable. Further, if $v \neq x$, then such a test would be equivalent to including control variables in order to detect endogeneity of an independent variable, which is not a valid exercise.

There has been extensive work on leave-out constructions as instruments, already. Focusing on the use of spatial instruments and spillovers, Betz et al. (2018) and Huber (2023) consider various necessary conditions for the identification strategy through leave-out instruments to hold. These important conditions are complimentary to our analysis. That is, even if their conditions are met, the instruments can be biased mechanically through the presence of fixed effects and, vice versa, even if the fixed effects are specified correctly, the conditions recounted in these studies still need to be met. As such, these two studies also act as a stand-in for a larger set of conditions which may apply in a given research application (see, e.g, Angrist, 2014, for analyses of peer effects). Because most conditions focus on the instrument itself and not on the broader estimation strategy (including the choice of fixed effects), the results derived here are typically complementary to other necessary conditions.

Given the extension of our results to judge fixed effects, it is natural to ask how the critique impacts the test designed by Frandsen et al. (2023) for such situations. The two aspects are, however, also completely complementary to each other. The test in Frandsen et al. (2023) tests the general admissibility of using judges as identifying variation with regards to the exclusion restriction and monotonicity of the instrument. The test does not consider the typical leave-out construction in the identification strategy but uses means of each judge's propensity for a specific decision directly. As such, an empirical strategy can pass the test of Frandsen et al. (2023) and still use a problematic combination of leave-out instrument and fixed effects as detailed in this paper. Vice versa, a correct construction of the data structure in our sense has no impact on whether the Frandsen et al. (2023) test gets rejected or not.

## Online Appendix E Structured Literature Review

We conducted a structured literature review of the issues 75, 76, and 77 (years 2020 through 2022) of the *Journal of Finance*. We differentiated papers into four possible categories: empirical, theoretical, structural, and other. The last category being comprised of replications/ corrigenda, methods papers, and similar types of research.

We included all papers that had the potential for the bias. This means, the paper featured a constructed regressor (like a lag structure or a leave-out) and a fixed effect that leads to potential overlap of the construction set and the set used for de-meaning. We excluded cases in which the overlap was too small. These were mostly cases with time fixed effects in which the number of firms or comparable entities was large. For all papers listed below, we cannot say for certain, how large the bias will be. We simply listed those papers in which the bias could potentially appear.

An overview of the results of the literature review is given in Table 1. Throughout the review, we focused on empirical and structured papers which made causal claims based on their data analysis. This excluded most asset pricing papers, because here the goal commonly is to predict asset prices rather than to make causal claims. This, however, does not mean that panel data analyses with fixed effects and constructed regressors in asset pricing are without problems. Lagged regressors in particular are common in asset pricing and combining them with fixed effects can lead the contemporaneous values to bleed through into the lagged values. This leads to the potential use of contemporaneous (and future) values in exercises to determine correlations, with the intent for them to be used in predictive modeling.

We list all 33 identified papers in Table E1. Here we also indicate what type of construction the paper uses, whether it is utilized in an instrumental variable estimation and whether the constructed regressor appears in the hypothesis variable or in control variables. We also give an example of where in the paper the estimation results can be found, but these examples are obviously not always limited to the places named in the table.

Some papers listed in Table E1 need additional clarification and are marked with footnotes. These footnotes refer to the aspects listed below.

1. Also contains Nickell (1981) Bias.
2. Structural Paper.
3. Overlap smaller than in Section 5; data allows for easy solution.
4. Bias appears in model calibration.
5. Bias appears only in robustness check.
6. Does test for strict vs. weak exogeneity (using Grieser and Hadlock (2019)), but fails the test.
7. Definition of hypothesis variable unclear and seemingly similar to leave-out variable.

Table E1: Papers with Potential Bias Identified in the 2020 to 2022 Volumes of the *Journal of Finance*

| Paper | Type | IV | Use | Example in Paper |
|---|---|---|---|---|
| Cookson and Niessner (2020)[1] | Lags | - | Control | Table 7 |
| Frydman and Wang (2020) | Lags | - | Hypothesis | Table 3 |
| Farre-Mensa et al. (2020)[3] | Leave-Out | Yes | Hypothesis | Table 2 |
| Hendershott et al. (2020) | Lags | - | Hypothesis | Table 5 |
| Sialm and Zhang (2020) | Lags | - | Hypothesis | Table 3 |
| Mian et al. (2020)[1] | Lags | - | Hypothesis | Table 9 |
| Akbas et al. (2020) | Lags | - | Hypothesis | Table 7 |
| Jagolinzer et al. (2020)[1] | Lags | - | Hypothesis | Table 6 |
| Bennedsen et al. (2020) | Lags | - | Control | Table 4 |
| Schiantarelli et al. (2020) | Leave-Out | Yes | Hypothesis | Table 5 |
| Bräuning and Ivashina (2020) | Lags | Yes | Hypothesis | Table 2 |
| Huang et al. (2020)[1] | Lags | - | Control | Table 5 |
| Dou et al. (2021) | Lags | - | Hypothesis | Table 2 |
| Greenwald et al. (2021)[1,2,4] | Lags | - | Hypothesis | Equation 42 |
| Daniel et al. (2021) | Lags | - | Hypothesis | Table 4 |
| Brown et al. (2021) | Lags | - | Control | Table 2 |
| Shue and Townsend (2021) | Lags | - | Hypothesis | Table 2 |
| Baghai et al. (2021) | Lags | - | Control | Table 4 |
| Cenedese et al. (2021) | Lags | - | Control | Table 2 |
| Falato et al. (2021) | Leave-Out | - | Hypothesis | Table 2 |
| Lee (2021) | Lags | - | Hypothesis | Table 2 |
| Tuzel and Zhang (2021)[5] | Lags | - | Hypothesis | Table 11 |
| Ge (2022)[6] | Lags | Yes | Hypothesis | Table 3 |
| Boguth et al. (2022) | Lags | - | Hypothesis | Table 10 |
| Koijen and Yogo (2022) | Lags | Yes | Hypothesis | Table 4 |
| Birru et al. (2022) | Lags | - | Control | Table 5 |
| Eisenbach et al. (2022) | Leave-Out | Yes | Hypothesis | Table 2 |
| Chang et al. (2022) | Lags | - | Control | Table 3 |
| Gürkaynak et al. (2022) | Lags | - | Hypothesis | Table 5 |
| Wang et al. (2022)[1,2] | Lags | - | Control | Table 8 |
| Bizjak et al. (2022) | Lags | - | Control | Table 5 |
| Schoenherr and Starmans (2022)[7] | Lags | - | Control | Table 3 |
| Meeuwis et al. (2022)[8] | Lags | - | Control | Table 3 |

*Note:* The table includes all papers identified in the structured literature review. The type of constructed regressor categorizes according to the examples provided in Section 2. IV identifies whether the constructed regressor is part of an instrumental variable estimation. Use denotes whether the constructed regressor appears in a hypothesis regressor or a control variable. Note that papers are categorized as *Hypothesis*, when constructed regressors appear in both both hypothesis and control variables. The example in paper is one place of possibly multiple where the estimation results subject to the bias discussed in this paper can be identified. Footnotes on the papers are explained in the text.

8. Note that the control variables in question are proxies for "alternative explanations".

Table E2 shows some additional analyses showing where the potential bias appears in the papers identified in our structural literature review. We can see that lagged regressor constructions are much more common in finance research than leave-outs, but the latter are nevertheless present. We also see that only about 18% of the identified papers use the constructed regressor in an IV estimation. Lastly, it is common for the constructed regressor to appear in the hypothesis variable. It is limited to control variables for only about a third of the identified papers.

Table E2: Possibly Biased Papers by Category

| Type | Count | Relative |
|---|---|---|
| Leave-Out | 4 | 12,12% |
| Lags | 29 | 87,88% |
| Use of IV | 6 | 18,18% |
| Hypothesis | 21 | 63,64% |
| Control | 12 | 36,36% |
| Total | 33 | |

*Note:* Table provides descriptive statistics of the prevalence of the Type, IV utilization and Use of the constructed regressors identified in the structured literature review.

## References for Online Appendix


Akbas, F., Jiang, C., and Koch, P. D. (2020). Insider investment horizon. The Journal of Finance, 75(3):1579–1627.

Baghai, R. P., Silva, R. C., Thell, V., and Vig, V. (2021). Talent in distressed firms: Investigating the labor costs of financial distress. The Journal of Finance, 76(6):2907–2961.

Bennedsen, M., Pérez-González, F., and Wolfenzon, D. (2020). Do ceos matter? evidence from hospitalization events. The Journal of Finance, 75(4):1877–1911.

Betz, T., Cook, S. J., and Hollenbach, F. M. (2018). On the use and abuse of spatial instruments. Political Analysis, 26(4):474–479.

Birru, J., Gokkaya, S., Liu, X., and Stulz, R. M. (2022). Are analyst short-term trade ideas valuable? The Journal of Finance, 77(3):1829–1875.

Bizjak, J. M., Kalpathy, S. L., Mihov, V. T., and Ren, J. (2022). Ceo political leanings and store-level economic activity during the covid-19 crisis: effects on shareholder value and public health. The Journal of Finance, 77(5):2949–2986.

Boguth, O., Duchin, R., and Simutin, M. (2022). Dissecting conglomerate valuations. The Journal of Finance, 77(2):1097–1131.

Bräuning, F. and Ivashina, V. (2020). Monetary policy and global banking. The Journal of Finance, 75(6):3055–3095.

Brown, J. R., Gustafson, M. T., and Ivanov, I. T. (2021). Weathering cash flow shocks. The Journal of Finance, 76(4):1731–1772.

Cenedese, G., Della Corte, P., and Wang, T. (2021). Currency mispricing and dealer balance sheets. The Journal of Finance, 76(6):2763–2803.

Chang, Y.-C., Hsiao, P.-J., Ljungqvist, A., and Tseng, K. (2022). Testing disagreement models. The Journal of Finance, 77(4):2239–2285.

Cookson, J. A. and Niessner, M. (2020). Why don't we agree? evidence from a social network of investors. The Journal of Finance, 75(1):173–228.

Daniel, K., Garlappi, L., and Xiao, K. (2021). Monetary policy and reaching for income. The Journal of Finance, 76(3):1145–1193.

Dou, W. W., Ji, Y., Reibstein, D., and Wu, W. (2021). Inalienable customer capital, corporate liquidity, and stock returns. The Journal of Finance, 76(1):211–265.

Eisenbach, T. M., Lucca, D. O., and Townsend, R. M. (2022). Resource allocation in bank supervision: Trade-offs and outcomes. The Journal of Finance, 77(3):1685–1736.

Falato, A., Hortacsu, A., Li, D., and Shin, C. (2021). Fire-sale spillovers in debt markets. The Journal of Finance, 76(6):3055–3102.

Farre-Mensa, J., Hegde, D., and Ljungqvist, A. (2020). What is a patent worth? evidence from the us patent "lottery". The Journal of Finance, 75(2):639–682.

Frandsen, B., Lefgren, L., and Leslie, E. (2023). Judging judge fixed effects. American Economic Review, 113(1):253–277.

Frydman, C. and Wang, B. (2020). The impact of salience on investor behavior: Evidence from a natural experiment. The Journal of Finance, 75(1):229–276.

Ge, S. (2022). How do financial constraints affect product pricing? evidence from weather and life insurance premiums. The Journal of Finance, 77(1):449–503.

Greenwald, D. L., Landvoigt, T., and Van Nieuwerburgh, S. (2021). Financial fragility with sam? The Journal of Finance, 76(2):651–706.

Gürkaynak, R., Karasoy-Can, H. G., and Lee, S. S. (2022). Stock market's assessment of monetary policy transmission: The cash flow effect. The Journal of Finance, 77(4):2375–2421.

Hayakawa, K., Qi, M., and Breitung, J. (2019). Double filter instrumental variable estimation of panel data models with weakly exogenous variables. Econometric Reviews, 38(9):1055–1088.

Hendershott, T., Li, D., Livdan, D., and Schürhoff, N. (2020). Relationship trading in over-the-counter markets. The Journal of Finance, 75(2):683–734.

Huang, Y., Pagano, M., and Panizza, U. (2020). Local crowding-out in china. The Journal of Finance, 75(6):2855–2898.

Huber, K. (2023). Estimating general equilibrium spillovers of large-scale shocks. The Review of Financial Studies, 36(4):1548–1584.

Jagolinzer, A. D., Larcker, D. F., Ormazabal, G., and Taylor, D. J. (2020). Political connections and the informativeness of insider trades. The Journal of Finance, 75(4):1833–1876.

Koijen, R. S. and Yogo, M. (2022). The fragility of market risk insurance. The Journal of Finance, 77(2):815–862.

Lee, J. (2021). Information asymmetry, mispricing, and security issuance. The Journal of Finance, 76(6):3401–3446.

Meeuwis, M., Parker, J. A., Schoar, A., and Simester, D. (2022). Belief disagreement and portfolio choice. The Journal of Finance, 77(6):3191–3247.

Mian, A., Sufi, A., and Verner, E. (2020). How does credit supply expansion affect the real economy? the productive capacity and household demand channels. The Journal of Finance, 75(2):949–994.

Nickell, S. (1981). Biases in dynamic models with fixed effects. Econometrica, 9(6):1417–1426.

Sampat, B. and Williams, H. L. (2019). How do patents affect follow-on innovation? Evidence from the human genome. American Economic Review, 109(1):203–236.

Schiantarelli, F., Stacchini, M., and Strahan, P. E. (2020). Bank quality, judicial efficiency, and loan repayment delays in italy. The Journal of Finance, 75(4):2139–2178.

Schoenherr, D. and Starmans, J. (2022). When should bankruptcy law be creditor-or debtor-friendly? theory and evidence. The Journal of Finance, 77(5):2669–2717.

Shue, K. and Townsend, R. R. (2021). Can the market multiply and divide? non-proportional thinking in financial markets. The Journal of Finance, 76(5):2307–2357.

Sialm, C. and Zhang, H. (2020). Tax-efficient asset management: evidence from equity mutual funds. The Journal of Finance, 75(2):735–777.

Tuzel, S. and Zhang, M. B. (2021). Economic stimulus at the expense of routine-task jobs. The Journal of Finance, 76(6):3347–3399.

Wang, Y., Whited, T. M., Wu, Y., and Xiao, K. (2022). Bank market power and monetary policy transmission: Evidence from a structural estimation. The Journal of Finance, 77(4):2093–2141.