Constructed Regressors and Overlapping Fixed Effects*

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

April 14, 2025

Abstract

To avoid endogeneity, financial economists often construct alternate regressors using values from other observations, with lagged and leave-out variables being common examples. We examine a bias in these "constructed regressors" that is induced by overlapping fixed effects. We show that inclusion of fixed effects can reintroduce the focal observation's bias through de-meaning. We show generally that the size of the bias is determined by level of overlap and provide an intuitive test for the significance of the bias. We illustrate the bias's magnitude via simulation and with patent examiner data in a Judge FE design. Even when scrambling the patent examiners, thus removing any instrument validity, the bias leads to a first-stage F-statistic over 1,000. We provide general solutions through either adjustment of the fixed effects or construction sets and detail other, case-specific, solutions.

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

Keywords: Lagged Variables · Leave-Out Instruments · Judge Fixed Effects · Hausman Instruments · Jackknife IV (JIVE) · Patents

*We thank Scott Cunningham, Ralf Elsas, Keith Ericson, Paul Goldsmith-Pinkham, Aurobindo Ghosh, Michal Kolesár, Peter Hull, Eli Liebman, 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, 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.

†Federal Reserve Bank of Dallas (alexander.chudik@dal.frb.org)

‡Dept. of Finance, Tippie College of Business, University of Iowa (cameron-ellis@uiowa.edu)

§LMU Munich School of Management, LMU Munich (jaspersen@lmu.de)

1 Introduction

Consider a financial economist who wants to estimate the causal effect of $x_{i,t}$ on $y_{i,t}$, but worries that $x_{i,t}$ is endogenous. A common strategy to overcome this endogeneity is to look for another variable v that correlates with $x_{i,t}$, but doesn't directly impact $y_{i,t}$. Having found v, it is often the case that $v_{i,t}$ remains endogenous to $y_{i,t}$ while other values of v, either along the i or t dimension more plausibly satisfy the exclusion restriction. The economist will use the average of these other values to "construct" a proxy or instrument for $x_{i,t}$. Perhaps the original, and most common, version of this "constructed regressor" strategy is to exploit exogeneity induced by time and use lagged values, either of x itself or some other variable, instead of the contemporaneous value.[1] More recently, econometricians have developed "leave-out" constructions, which average other observations along a data dimension besides time to utilize plausibly exogenous disturbances in the broader data category to circumvent bias within individual observations.

The 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; Nevo, 2001). "Judge fixed effects" are a similar style of leave-out instrument where an individual judge's average rulings on other cases are used as an instrument for the current case (e.g., Aizer and Doyle, 2015; Dobbie et al., 2018). "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).[2] Though largely originating in other fields, leave-out 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).

Even more common in finance research is the use of fixed effects, often Firm and/or Year, to control for unobserved variables that are invariant along some dimension. Leave-out instruments in particular typically require fixed effects to be plausibly exogenous. With Hausman instruments,

[1] A related strategy is the use of "deterministic" contemporaneous variables (whose value in time t is determined only by events prior to t), to motivate the exclusion restriction (e.g., Koijen and Yogo, 2022).

[2] 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 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.

it is important to include (at least) $Firm$ 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 these identification strategies: when constructed regressors are combined with overlapping fixed effects—where observations within a fixed effect share common terms from the constructed regressor—they become mechanically correlated with the exact value they are trying to avoid, leading to bias.[3] This result holds even if the constructed regressor by itself satisfies the exclusion restriction. Furthermore, when used as an instrument, this bias leads to an increase in the apparent "strength" of the first stage, even if the identifying shock does not actually exist, with a larger bias mechanically leading to a larger first stage F-Statistic.

After showing generally that the bias exists when fixed effects overlap, and does not exist when they don't, we develop a resampling-based test for the significance of the bias. Next, we develop general, though potentially restrictive, solutions. We also detail existing, 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. 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 overlapping fixed effects bias induces a first-stage F-statistic of over 1,000.

Constructed regressors, including lags and leave-out instruments, all share a common identification strategy: leveraging plausibly exogenous disturbances from related data categories to avoid bias within individual observations. Since these disturbances cannot be directly observed, they are estimated from other observations. As a result, each individual estimate of the disturbance incorporates both the actual disturbance and the error terms of all the observations used in the construction. However, when overlapping fixed effects are included, the de-meaning process reintroduces the focal observation's error term, thereby reintroducing the exact bias that the identification strategy seeks to eliminate.

The extent of this bias is determined by the structure of the data dimensions used for both the fixed effects and the constructed regressor. The most extreme version of the bias appears in leave-out instruments when fixed effects are included such that every observation used to calculate the

[3]For example, a constructed regressor of $z_{i,t-1}$ overlaps with $Firm(i)$ fixed effects since $z_{i,t}$ is the constructed regressor for $z_{i,t+1}$, which shares the same fixed effect.

fixed effect, except the focal observation, is also used to construct the instrument.[4] In this case, the instrument fully reduces to a multiplicative transformation of the de-meaned focal observation and is equivalent to not instrumenting at all. The intuition for the result is simple. The mean of all leave-out instruments within a perfectly overlapping fixed effect is simply the average of all of the observations within that fixed effect. When this mean is subtracted from each individual leave-out instrument, the only thing remaining is the focal observation.

A more common case in empirical applications is when the observations used in the constructed regressor do not perfectly overlap with the fixed effects. Here, we find that the size of the bias will be proportional to the degree of overlap with more overlap leading to more bias. For example, with T time periods the degree of overlap for a lagged regressor with $Firm$ fixed effects is, approximately, $\frac{1}{T}$ and the number of firms is irrelevant. We can turn to the "within" interpretation for intuition – after projecting out the fixed effect, the lagged value is compared only to that firm's other T observations. The focal value is one of those values and thus, by construction, any bias it contains bleeds over.

We develop a simple and intuitive test for the existence of this bias. The test relies on developing a placebo regressor that shares the same construction strategy, but has no identifying information. This is achieved though scrambling observations along the relevant construction dimension and then re-running the original regression. In the case of judge fixed effects, judges within a court-year are randomly re-sampled. In the case of lagged variables, the lagged values are randomly chosen from a different firm in the same (lagged) year. Because this resampling removes any identifying variation conditional on observables, any significance that remains in the regression must originate from the bias.

We derive general solutions to the bias from the fact that if the observations used in the instrument and those used in the fixed effect do not overlap, then the de-meaning process has no negative consequences for the instrument's validity. This can be achieved in two ways. First, by increasing the fixed effects' granularity through interacting the fixed effect's desired dimension with a separate fixed effect at the level of the group orthogonal to the instrument group. For example, in the Hausman IV, the researcher could include $Market \times Firm$ fixed effects instead of simply $Firm$. The other option is to change the calculation of the instrument by leaving out all observations also used in the fixed effect. In practice, this will often involve an "imputation" ap-

[4]This occurs when the set that defines an observation's fixed effect (also called "cell" below) is equal to, or a strict subset of, the set used for instrument construction (also called "group" below).

proach, estimating the instrument on one part of the sample, and then the primary regressions on a separate part of the sample. However, it may be the case that an "outside-sample" exists that is relevant for the instrument, but not relevant to the primary regression.

There are also case-specific solutions that have been offered in other literatures. In the case of leave-out instruments, the model can be re-framed as a jacknife IV (JIVE) and the researcher can use the IJIVE or UJIVE estimators of Ackerberg and Devereux (2009) and Kolesár (2013), which runs the estimation as a JIVE, but projects out the controlling fixed effects before running the first stage. In the case of lagged independent variables, researchers can turn to "weak exogeneity" literature and use either the half-panel jacknife of Chudik et al. (2018) or the consistent IV estimator of Mikusheva and Sølvsten (2023). Finally, in the case of lagged dependent variables, known as the "Nickell Bias" (Nickell, 1981), researchers can use the dynamic panel estimators of Blundell and Bond (1998) and Arellano and Bond (1991).[5]

Monte Carlo simulations show that the magnitude of the bias can be substantial and, in some cases, worse than just running OLS on the focal observation. Moreover, we demonstrate this in an empirical example using commonly analyzed patent examiner data from the U.S. Patent Office to illustrate the potential magnitude of the bias. We attempt to estimate the causal impact of having your first patent approved on the probability of applying for a second patent. We use two instruments for first-application approval: (1) a traditional leave-out instrument where we average the approval rates of all other patent applications with the same examiner and (2) an "outside-sample" leave-out instrument where we average the approval rates of the non-first-application patents with the same examiner. We combine these instruments with Art Unit - Year fixed effects.[6] Both instruments appear strong, with first-stage F-statistics of 50,000 and 5,000, respectively. We find that the estimated effect using the traditional instrument is 20% larger than when using the outside-sample instrument, with the bias in the direction of OLS.

In order to separate the level of bias from the setting's underlying instrument strength, we perform our suggested test and scramble the patent examiner identifiers within each fixed effect group, effectively removing any actual identifying variation. In this case, the outside-sample instrument correctly shows a first-stage F-statistic near zero and a statistically insignificant second stage. Strikingly, even though the instrument is not valid by construction, the mechanical

[5] Klosin (2024) examines, and provides a solution to, a similar form of bias that they dub "dynamic bias" which occurs when include unit fixed effects in a static panel and the treatment variable is dependent on past values of the outcome variable.

[6] Patent examiners are separated into "art units" which can be viewed as courts. See Sampat and Williams (2019) for a more thorough discussion.

correlation induced by the overlapping fixed effects gives the traditional leave-out instrument a first-stage F-statistic over 1,000 and a statistically significant second stage with an estimated effect almost exactly the same as OLS.

Our primary contribution is to further financial economists' understanding of the potential bias in this extremely common empirical design. 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; Chudik et al., 2018; Kolesár, 2013; Angrist et al., 1999). While that work proves that the bias exists for any 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 is the overlap that determines the bias, and not the general number of fixed effects. Indeed, there are cases where multiplying two fixed effects together can fully remove the bias, even though the number of controls increases dramatically. We also develop an intuitive and simple test for the existence of this bias, which can inform researchers if pursuing potentially costly solutions are necessary.

Lastly, we contribute to the general literature examining biases that may be induced by fixed effects. Such biases have been shown in other settings, such as the incidental parameters problem in maximum likelihood estimation (Neyman and Scott, 1948; Lancaster, 2000) or difference-in-differences with staggered roll-outs (Goodman-Bacon, 2021; Borusyak et al., 2024; Sun and Abraham, 2021). Fixed effects are popular, because they can solve important econometric problems (Petersen, 2008). At the same time, however, they can often lead to problems in a non-intuitive way when combined naïvely with other estimations strategies. We show that such an effect can appear with constructed regressors, but, at the same time, highlight several ways in which a combination of these instruments with fixed effects is still possible.

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:

$$y_i = \beta z_i + \boldsymbol{x}_i'\boldsymbol{\delta} + \eta_{C(i)} + \varepsilon_i, \quad (1)$$

$$z_i = \frac{1}{|G(i)|}\sum_{j\in G(i)} v_j, \quad (2)$$

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 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. $\mathbf{x}_i = (x_{i1}, x_{i2}, ..., x_{ik})'$ 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 (1), the fixed effect for cell $C(i)$ is denoted by $\eta_{C(i)}$. While not typically used in this fashion, this notation subsumes all common panel structures, including those with multiple fixed effects (such as the common *Firm* and *Year* two-way fixed effects). 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 (1) 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 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 a general regularity requirements on errors and regressors given by Assumption 4 in Appendix .1. 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 optimal.

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

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

$$E(v_i \varepsilon_j) = \begin{cases} 0 & \text{for } i \neq j \\ \sigma_{v\varepsilon i} & \text{for } i = j \end{cases} \tag{3}$$

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

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 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. 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 fixed effects (FE) estimator of β in model (1),

$$\hat{\beta} = \left(\tilde{\mathbf{z}}' \mathbf{M}_{\tilde{X}} \tilde{\mathbf{z}}\right) \tilde{\mathbf{z}}' \mathbf{M}_{\tilde{X}} \tilde{\mathbf{y}}, \tag{4}$$

where $\mathbf{M}_{\tilde{X}} = \mathbf{I}_n - \tilde{\mathbf{X}} \left(\tilde{\mathbf{X}}' \tilde{\mathbf{X}}\right)^{-1} \tilde{\mathbf{X}}'$. 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 (1)-(2), and suppose Assumptions 1 and 4 hold. Consider the FE estimator $\hat{\beta}$ given by (4). Then, as $n \to \infty$,*

$$\hat{\beta} \to_p \beta_0 - Q_{zx}^{-1} \Delta_\beta, \tag{5}$$

where

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

in which $c_{ji} = I\left[j \in G(i)\right]$ and $I(.)$ is an indicator function.

All proofs are provided in the appendix.

This proposition shows that the overlap between the sets $C(i)$ and $G(i)$, given by $\sum_{j\in C(i)} c_{ji} = |(i) \cap C(i)|$, can lead to inconsistency of $\hat{\beta}$.[8] For illustrative purposes, assume $\sigma_{v\varepsilon i} = \sigma_{v\varepsilon}$ for all i. Then $\Delta_{\beta,n}$ reduces to

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

where $\bar{\theta} = n^{-1}\sum_{i=1}^{n}\theta_i$, and

$$\theta_i = \frac{|G(i) \cap C(i)|}{|G(i)|\,|C(i)|}, 0 \leq \theta_i \leq 1.$$

Clearly, $\bar{\theta}$ can tend to zero or a positive nonzero value (≤ 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.

In general, $\bar{\theta}$ can be of order $n^{-\alpha}$, for some value of the exponent α in the range $0 \leq \alpha \leq 1$, namely $\bar{\theta} = O\left(n^{-\alpha}\right)$. 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 Monte Carlo section we illustrate how bad inference could be even if the bias of $\hat{\beta}$ is asymptotically negligible (namely when $\hat{\beta} \to_p \beta_0$).

When thinking of Equation (1) as the first stage in an instrumental variable estimation, then important implication of Proposition 1 is that the first stage F-Statistic can be spuriously large when $\beta_0 = 0$. 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.[9]

[8] This bias occurs even if $E(v_i\varepsilon_j) = 0$ for all $i \neq j$. If this condition is relaxed, then the expression for the bias will be more complex.
[9] 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 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 time periods (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

$$y_{ft} = \beta v_{f,t-1} + \eta_f + \varepsilon_{ft}, \tag{6}$$

which is a special case of (1)-(2) with the datapoint i given by the pair (f, t), $\mathbf{x}_i = 0$ (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 (2), we are taking the average over a single observation (formally: $|G(f,t)| = 1$). This is the most common way of including lags in empirical literature, though there are some application in which the average is taken over multiple lags (see, e.g., Gao et al., 2024) 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 $C(f,t) = \{(f,1), (f,2), ..., (f,T)\}$. Hence $|C(f,t)| = T-1$. The following corollary shows 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 (1)-(2) given by (6), 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 β in (6), using a balanced sample on T time periods and F firms. Then, as $F \to \infty$ and T is fixed,

$$\hat{\beta} \to_p \beta_0 - Q_{v,T}^{-1}\Delta_{\beta,T}, \tag{7}$$

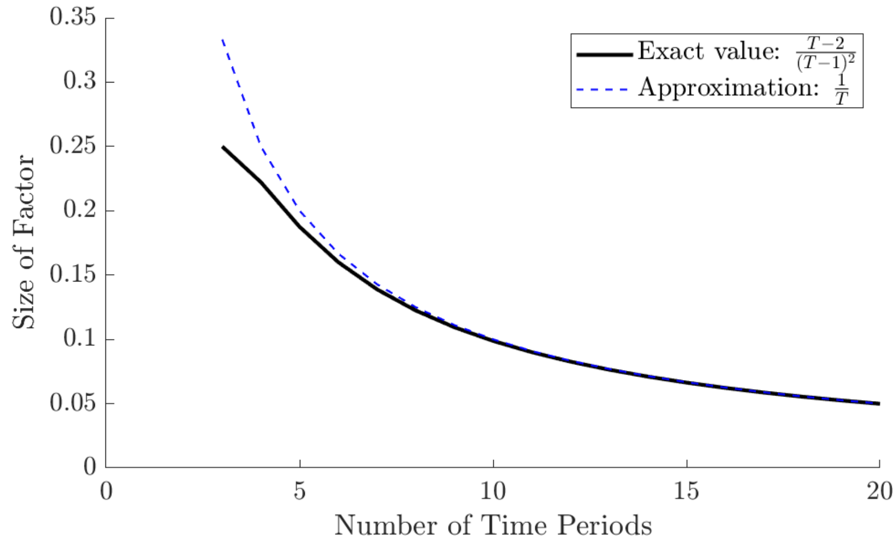
where

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

$\bar{\sigma}_{v\varepsilon} = \lim_{F\to\infty} F^{-1}\sum_{f=1}^{F}\sigma_{v\varepsilon,f}$ 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 specifications 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.[10] As can be seen in Figure 1, the bias can be quite sizeable and is be particularly strong for panels with short T.

Figure 1: Extent of bias in the FE estimation of 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 given number of time periods in the estimation. The blue dashed line shows the approximation $1/T$. The graph starts at $T=3$. With $T=2$ only one period is used in the estimation, thus not allowing for unit-level fixed effects.

The result in 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). 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 β.

Under the usual regularity conditions FE estimator of β in Equation (6) 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 inference will be valid only if F is small relative to T. This rules out many applications where F (the number of firms) is typically quite large in comparison with T (the number of time periods).

[10] 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.

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.*

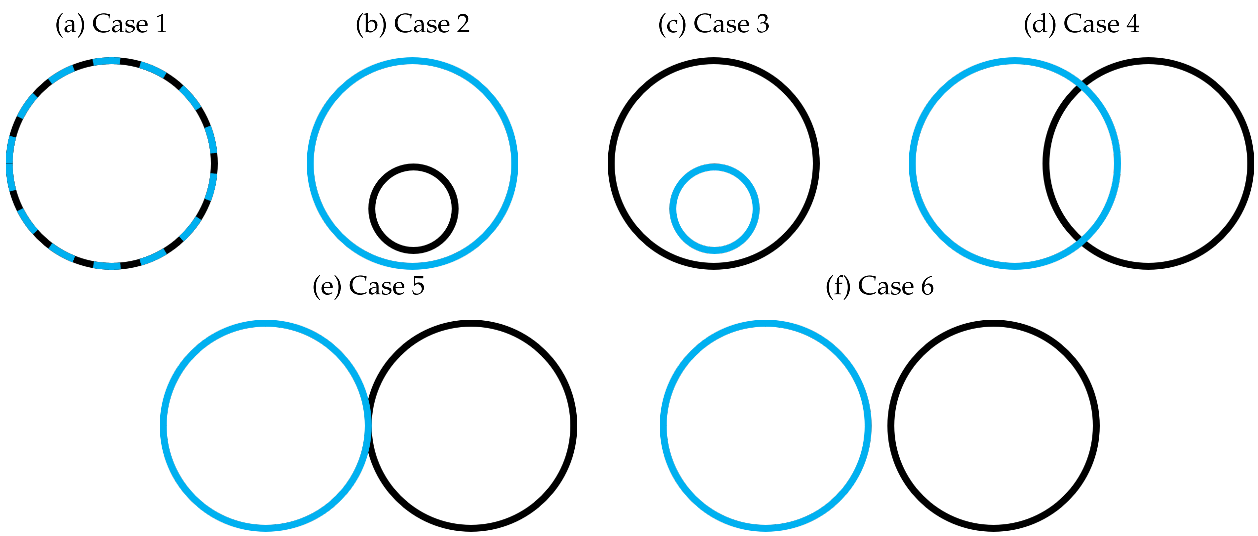
We call $J(i)$ the jackknife set. In the leave-out setting the construction group of observation i is comprised of all observations in $J(i)$ except i itself. Thus, within a jackknife set, the construction groups of two observations differ only regarding a single observation: their own.

The idea is best exemplified by a concrete data application. For this, we borrow the setting of our empirical application described in Section 4. 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. 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.

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 OLS 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 series of small sample properties. 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. 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 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 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 such that $k_1 \geq 2$.*

We can then state the results.

Proposition 2. *Under Assumptions* **??** *through 1 and Assumptions 2 and 3, estimating Equation* (1) *with OLS is equivalent to*

1. *estimating Equation* (1) *with OLS and using*

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

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

2. *estimating Equation* (1) *with OLS and using*

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

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

3. *estimating Equation* (1) *with OLS and using*

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

   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.[11]

[11] 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 $-\frac{1}{k_1 - 1}$.

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 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.[12] 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 persistent 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 is 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 $G(i)$ and $C(i)$, there will be no bias in the estimation as can be seen in Equation (5). 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 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

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

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).

2.4 A Test for Existence of the Bias

The formal results and their intuitive explanations make clear that the bias reintroduces the focal value in the estimation mechanically. This interferes with the otherwise informative nature of the construction set. The estimate of coefficient, b, thus features both the informative component and the bias. A simple test for the existence of the bias is thus to remove the informative component from the constructed regressor and then test the estimated coefficient on the newly constructed regressor for statistical significance. If all informative components were removed successfully, a statistically significant coefficient implies the existence of the bias. Specifically, we propose the following two-step test:

Test 1 (for existence of the bias)**.** *The test has two steps*

1. *For each observation i, generate a new construction group $\tilde{G}(i)$ such that $|\tilde{G}(i)| = |G(i)|$, $|\tilde{G}(i) \cap C(i)| = |G(i) \cap C(i)|$, $\tilde{G}(i) \cap G(i) = \varnothing$, and for each $i \in \tilde{G}(i)$ and each $j \in G(i)$ it holds that $Cov(v_i, v_j | \boldsymbol{X}_i, \eta_{C(i)}) = 0$.*
2. *Estimate Equation* (1) *with $z_i = \frac{1}{|\tilde{G}(i)|} \sum_{j \in \tilde{G}(i)} v_j$ and test b for significance.*

The conditions for the new construction set in the first step ensure that the potential bias is of equal magnitude as in the primary estimation since it overlaps with the fixed effect cell to the same degree. The new and original construction sets should not share observations and there should be no correlation between the values of v in the new and the original set when conditioning on control variables and fixed effects.

In practice, $\tilde{G}(i)$ can be generated by scrambling the data on the adequate dimension. In the lagged regressor case, the researcher can, for example, simply choose the lagged value of v for a different firm.[13] In the judge fixed effect case, the researcher can scramble judge ids and then re-estimate the first stage. The procedure is used in Section 4.

[13]This can be achieved easily by scrambling firm ids within a year and then resetting the panel.

2.5 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 data structure can be implemented and we introduce them with the help of examples below. In addition, specific situations may have specific solutions, which we discuss at the end of the section.

If the data is sufficiently rich, one way to ensure that there is no overlap between construction group and fixed effect cell 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 raw $Firm$ 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.[14] Thus, the estimation will be unbiased as long as the other assumptions necessary for leave-out instruments are met.[15]. This is a good example of how adding *more* control variables can actually reduce (or, in this case, eliminate) the bias.

If the data are not rich enough to introduce a fixed effect orthogonal to the construction group, one can 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 three potential ways to do this.

The easiest way to separate construction and fixed effects 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 (1). 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

[14] Note that using $Firm + Market$ fixed effects will not reduce the bias at all.

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

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 4.

When there is no outside-sample instrument available and the identifying variation is a subset of the necessary fixed effect, a split-sample, imputation approach is 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, then the econometrician can limit the estimation to every second period. This way, the data used for regressor construction never appears in the fixed effect cell.

This approach is part of a set of recently developed estimators that take advantage of split samples and separated prediction and estimation procedures to circumvent econometric problems (e.g. Chernozhukov et al., 2018; Wager and Athey, 2018; Borusyak et al., 2024; Abadie et al., 2024). This solution requires the most out of the data, since it must be large enough to support the separate estimation of the instrument and the (potentially 2SLS) regression on individual partitions, but unlike the other solutions, it can used in any data structure.[16]

It is worth noting that some specific data structures with constructed regressors have idiosyncratic solutions available to them. The lagged regressor case can be treated with dynamic panel procedures such as the estimators proposed by Arellano and Bond (1991), Blundell and Bond (1998), and others. Note, however, that such methods can suffer from weak- or many-instrument problems and thus might not be suitable to every data structure (Chudik et al., 2018). The half-panel jackknife estimator of Dhaene and Jochmans (2015) has been applied to the weak exogeneity problem in general by Chudik et al. (2018) and is shown to only have a bias of a magnitude of approximately $\frac{1}{T^2}$, offering another candidate for a (partial) solution. The leave-out case from Section 2.3 can be treated by using improved JIVE estimators, specifically the IJIVE (Ackerberg and Devereux, 2009) and the UJIVE (Kolesár, 2013) in addition to the methods introduced above.

[16]The optimal size of the split will depend on the power of the instrument as well as the power of the primary regression. We leave this calculation as an interesting avenue for future research.

3 Monte Carlo Evidence

To illustrate the magnitude of the potential bias and to show the relative impact of model/data specifications, we employ a Monte Carlo simulation in a realistic setting comparable to our examples above. Our data-generating process is as follows. We label the explanatory hypothesis variable as $x_{f,m,t}$ for firm $f \in F$, market $m \in M$, and year $t \in T$. We also allow a potential firm-level fixed effect η_f in both x and y. We simulate data according to

$$x_{f,m,t} = \alpha \cdot x_{f,m,t-1} + \gamma \cdot c_{f,t} + \nu \cdot q_{f,m,t} + \eta_f + \varepsilon^x_{f,m,t} \tag{8}$$

$$y_{f,m,t} = \beta \cdot x_{f,m,t} + \nu \cdot q_{f,m,t} + \eta_f + \varepsilon^y_{f,m,t} \tag{9}$$

$$\varepsilon^{x,y}_{f,m,t}, c_{f,t}, q_{f,m,t} \sim N(0,1) \tag{10}$$

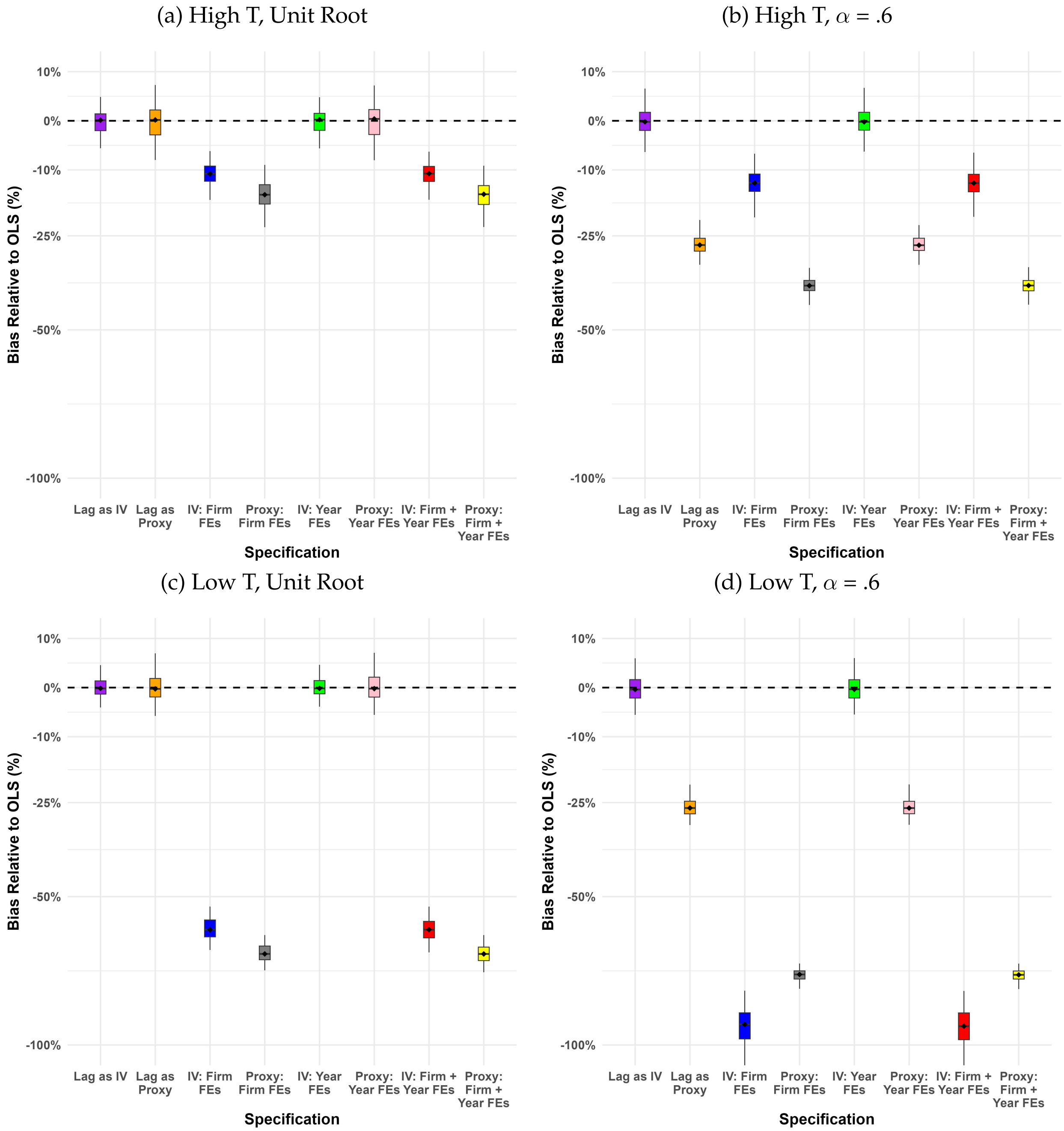
where $y_{f,m,t}$ is the dependent variable of interest, $q_{f,m,t}$ is the firm-market-year quality which is correlated with $y_{f,m,t}$, and $c_{f,t}$ is the firm-year level cost shock which is uncorrelated with $y_{f,m,t}$. $q_{f,m,t}$ and $c_{f,t}$ are unobserved by the econometrician.

Our goal is to recover β. Our first strategy is to exploit that x is auto-correlated and q is not. We approach this through two different methods. First, we take the most common approach and simply proxy for $x_{f,m,t}$ by replacing it in equation (9) with $x_{f,m,t-1}$ and running OLS. Next, we use $x_{f,m,t-1}$ as an instrument for $x_{f,m,t}$ in a 2SLS set up. Our second strategy is to use the average of x's for the same firm in the same year, but in other markets, to estimate $c_{f,t}$, use this as an IV for $x_{f,m,t}$, and then estimate β via 2SLS. We will attempt all of these methods combined with various forms of fixed effects. In order to isolate the bias induced by fixed effects, we initially abstract away from the reason to include fixed effects (time-invariant omitted variables) and set $\eta_f = 0$. Unless stated otherwise, all following regressions feature $\nu = 10$, $\gamma = \alpha = 1$, and $\beta = 0.5$.

Figure 3 presents the results of our first set of Monte Carlo simulations, showing the bias of various lagged estimators, calculated as the difference between the true parameter β and its estimate $\hat{\beta}$, relative to the bias of the OLS estimator. We scale the relative bias by the OLS bias to provide a clear comparison of the different estimators' performance, abstracting from the overall level of endogeneity in the data-generating process.[17] For the lagged-variable construction examples, we utilize 1 $Market$ and treat the data as a $Firm - Year$ panel. In Panels (a) and (b), the simulation features 20 $Years$ and 400 $Firms$. In Panels (c) and (d), the simulation features 4 $Years$ and 1600 $Firms$. Both panels feature 8,000 total observations. Panels (a) and (c) feature a unit-root ($\alpha = 1$), so that $E[x_{t-1}] = x_t$. In panels (b) and (d), the autocorrelation is slightly weaker at $\alpha = .6$.

[17] Specifically, we calculate the bias of a given estimation strategy S as $\text{Bias}_S = (\hat{\beta}_S - \beta)/(\hat{\beta}_{OLS} - \beta)$.

Figure 3: Relative Bias: Lags



Note: Results of a Monte Carlo simulation with the data generating process described in Equations (8) through (10) with $\eta_f = 0$. This figure shows the bias, relative to OLS, of separate models across 500 simulations. Each box represents a separate model specification, and cases are denoted by color. Boxes show the inter-quartile range of the relative bias for each model. The diamond in the middle of each plot denotes the median value. Whiskers indicate variability outside of the inter-quartile range. Note that the y-axis is "log"-scaled both above and below zero for better visual presentation. All simulations have a bias that is ten times larger than the instrument strength. Panel (a) has 20 years, 20 firms, and 20 markets. Panel (b) has 4 years, 100 firms, and 20 markets.

For panels (a) and (c), all estimators without $Firm$ fixed effects are unbiased. This is true even when including year fixed effects, since the construction groups contain values from $t-1$ while the fixed effect cell only contain values from t, so there is no overlap. All of the estimators that include $Firm$ fixed effects (including two-way fixed effects) are biased, with the proxy strategy performing slightly worse than the IV strategy.

For panels (b) and (d), we lower the autocorrelation coefficient slightly, so that $E[x_{t-1}] \neq x_t$. This causes a bias in the proxy variables regardless of fixed effects and the "raw" level of this bias can be see in the 2nd estimation strategy across all panels. With a higher T, in panel (b), the lower autocorrelation has little impact on the IV estimators, since they properly account for the altered relationship in the first stage. However, in a low T setting, the lower autocorrelation magnifies the overlapping fixed effects bias, leading the IV to be worse than the proxy in panel (d).

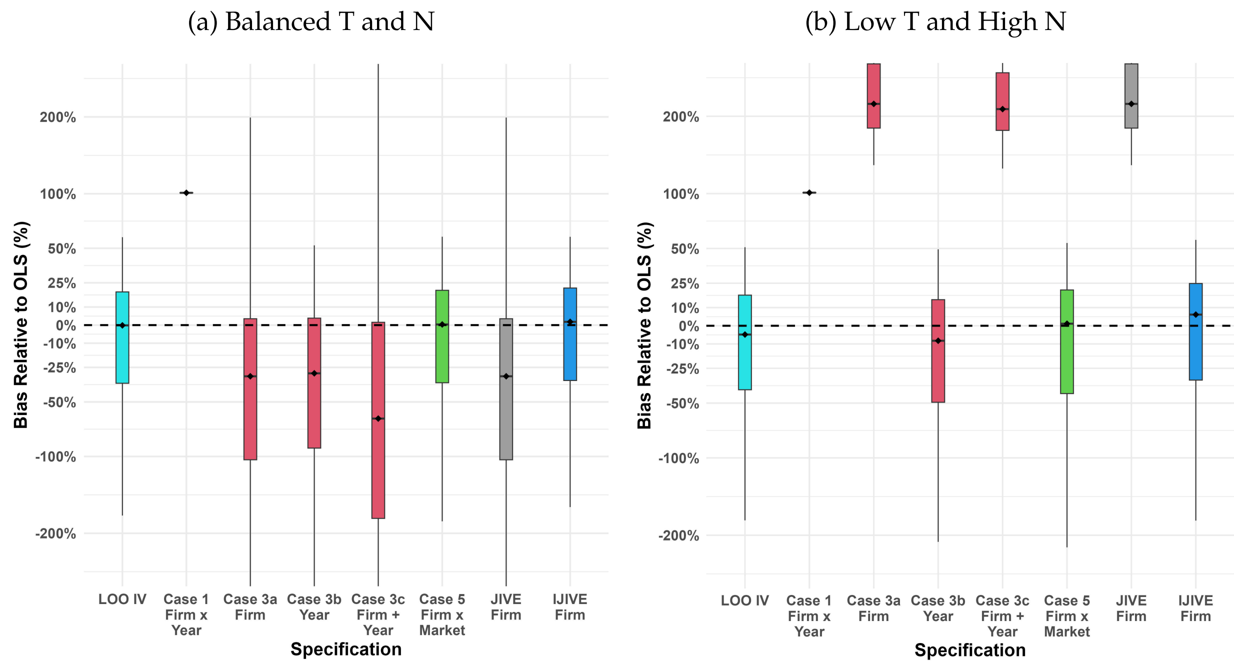
Figure 4 next presents the results of our leave-out Monte Carlo simulations, showing the bias of various leave-out estimators. In Panel (a), the simulation features 20 $Years$, 20 $Firms$, and 20 $Markets$. In Panel (b), the simulation features 4 $Years$, 100 $Firms$, and 20 $Markets$. Both panels feature 8,000 total observations. For Panel (a), the standard leave-out IV estimator, the Case 5 estimator, which combines the leave-out instrument with $Firm \times Market$ fixed effects, and the IJIVE estimator show no bias, on average. As shown in item 1 of Proposition 2, the Case 1 estimator is exactly equivalent to running OLS. The first two Case 3 estimators, which use either $Firm$ fixed effects or $Year$ fixed effects exhibit a median bias of around 20% relative to OLS while the third Case 3 estimator, which uses $Firm$ + $Year$ fixed effects, exhibits double this bias.

In Panel (b), which features the same number of overall observations but different time and firm dimensions, the bias is remarkably different. While the normal IV, Case 5, and the IJIVE remain unbiased, all specifications that involve $Firm$ fixed effects are now worse estimators than simply not instrumenting at all.[18] In contrast, the estimator for Case 3b, which features $Year$ instead of $Firm$ fixed effects, shows almost no bias due to the increasing $Firm$ dimension. This pattern is exactly what was predicted by our theoretical analysis.

We next modify the simulation to include time-invariant characteristics to both x and y by setting $\eta_f \sim N(0,1)$, which necessitates the inclusion of $Firm$ fixed effects. Similar to Figure 4, Panel (a) of Figure 5 shows simulations with a balanced T and N, while Panel (b) shows simulations with a low T and higher N. Both panels feature 8,000 observations, and the bias is now reported relative to an OLS regression that includes $Firm$ fixed effects.

[18] In similar simulations, but with an uninformative instrument, the median first stage F-statistic for the cases with $Firm$ fixed effects is over 20 while all of the rest are all well under 10.

Figure 4: Relative Bias: Leave-outs



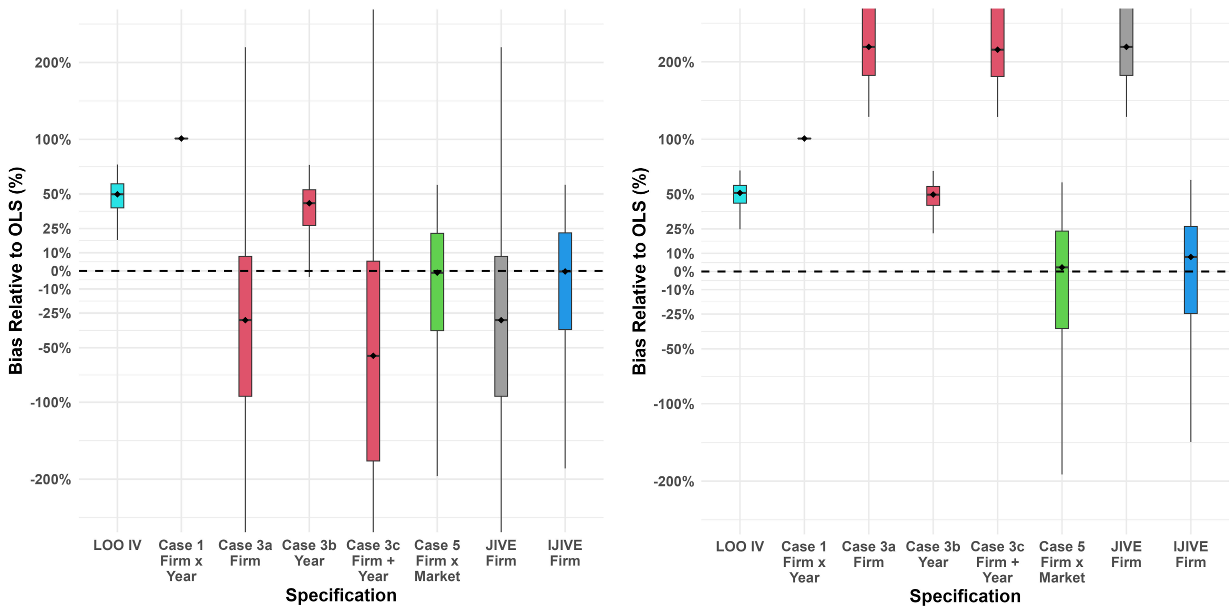
Note: Results of a Monte Carlo simulation with the data generating process described in Equations (8) through (10) with $\eta_f = 0$. This figure shows the bias, relative to OLS, of separate models across 500 simulations. Each box represents a separate model specification, and cases are denoted by color. Boxes show the inter-quartile range of the relative bias for each model. The diamond in the middle of each plot denotes the median value. Whiskers indicate variability outside of the inter-quartile range. Note that the y-axis is "log"-scaled both above and below zero for better visual presentation. All simulations have a bias that is ten times larger than the instrument strength. Panel (a) has 20 years; 20 firms; and 20 markets. Panel (b) has 4 years; 100 firms; and 20 markets.

For Panel (a), the IV is now biased because it does not include firm effects. Cases 3a and 3c feature nearly identical biases to those in Figure 4, because they are accounting for the time-invariant effects, while Cases 5 and the IJIVE remain unbiased. Interestingly, the omitted fixed effect bias and the overlapping fixed effect bias for the specification with $Year$ fixed effects have partially offset, and the bias has flipped signs from Figure 4. However, this offsetting is just a feature of our simulation and, in practice, the biases may be additive.

Next, we further examine the dynamics of changing the cardinality of the $Year$ dimension. Panel (a) of Figure 6 shows how the median bias and its inter-quartile range change as the number of time periods $|T|$ increases while the number of $Firms$ (20) and $Markets$ (20) stays constant for select estimators. Due to the low number of observations, the IV estimator without fixed effects exhibits some bias initially, but quickly converges to the true value. For the estimators including fixed effects, we can see that as $|T|$ increases, the absolute value of the bias for all three of them decreases. However, because the number of firms is held steady, the TWFE estimator approaches

Figure 5: Relative Bias: Time-Invariant Effects

(a) Balanced T and N (b) Low T and High N



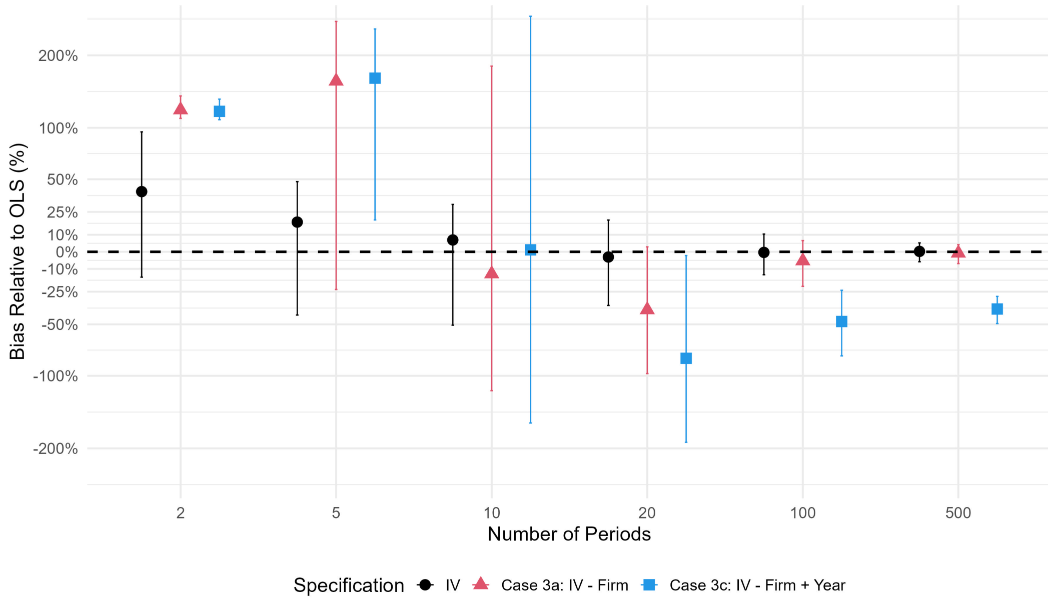
Note: Results of a Monte Carlo simulation with the data generating process described in Equations (8) through (10) with $\eta_f \sim N(0,1)$. This figure shows the bias, relative to an OLS regression that includes $Firm$ fixed effects, of separate models across 500 simulations. Each box represents a separate model specification, and cases are denoted by color. Boxes show the inter-quartile range of the relative bias for each model. The diamond in the middle of each plot denotes the median value. Whiskers indicate variability outside of the inter-quartile range. Note that the y-axis is "log"-scaled both above and below zero for better visual presentation. All simulations have a bias that is ten times larger than the instrument strength. Panel (a) has 20 years; 20 firms; and 20 markets evenly split into 2 regions. Panel (b) has 4 years; 100 firms; and 20 markets.

an asymptote of 30% of the OLS bias even at a $|T| = 500$. The panel further shows that the bias is non-monotonic for all estimators. This is because the equilvalent instrument in item 3 of Proposition 2 includes the true coefficient vector with a positive sign and the additional element with a negative one. When $|T| = 2$ and thus $|C(i)|$ is relatively small compared to $|G(i)|$, the negative sign is so strong that the first stage leads to a negative coefficient, making the second stage coefficient on x larger than the true value (with a mechanism similar as in OLS). As $|T|$ increases, the sign in the first stage eventually flips and now the mechanical correlation (which still has a negative influence on the instrument) biases the coefficient in the second stage downwards. The main take-away is that the sign of the bias in the estimation cannot easily be predicted ex ante.

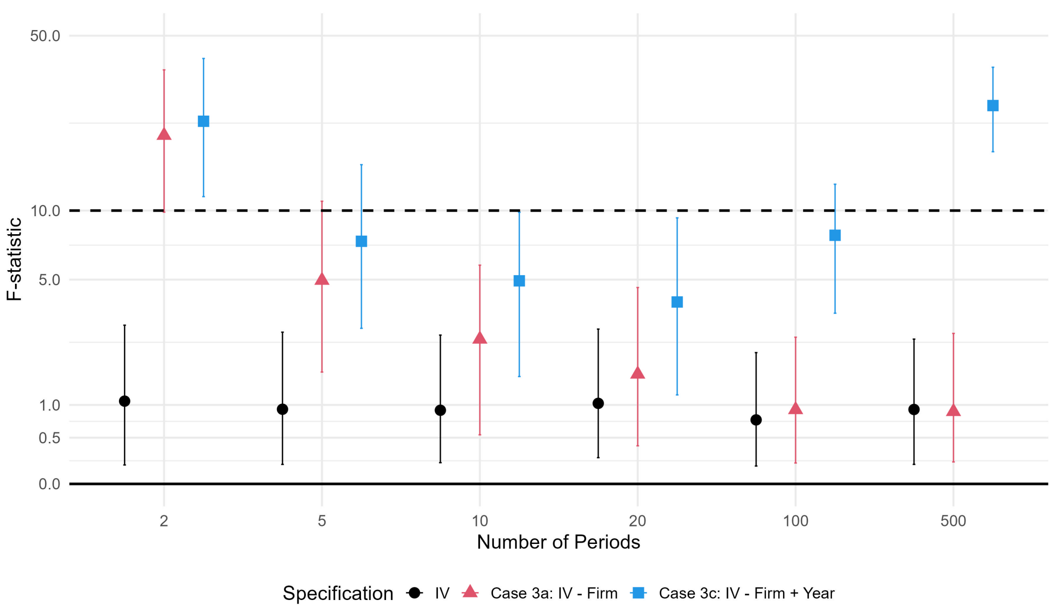
It should be made clear that there is no specific meaning to the values of $|T|$ in Figure 6 or elsewhere in our numerical studies. $\nu = 10$ and $\gamma = 1$ were chosen for expository purposes such that the bias is apparent at lower $|T|$ and $|N|$, but mostly disappears at levels around $|T| = 100$.

Figure 6: Increasing Number of Periods

(a) Relative Bias



(b) First Stage F-Statistic



Note: Results of a Monte Carlo simulation with the data generating process described in Equations (8) through (10) with $\eta_f = 0$. Panel (a) of this figure shows the median bias (relative to the bias of OLS) across 500 simulations for each level of T. We plot the median as well as the inter-quartile range. Panel (b) shows how the bias influences the F-statistic when there is no relevance for the instrument ($\gamma = 0$). Each color/point represents a separate model specification. Note that the y-axis is "log"-scaled both above and below zero for better visual presentation. All simulations have a bias that is ten times larger than the instrument strength, except for the F-statistics, which have no instrument strength; 20 firms; and 20 Markets.

However, in reality, there is nothing special about $|T| = 100$. In practice, the econometrician will not know these values or their relative levels, and as ν increases or γ decreases the level of $|T|$ and $|N|$ where the bias is minimal will increase.

Panel (b) of Figure 6 presents the median first stage F-statistics and the inter-quartile range for a simulation in which the instrument has no strength. For this analysis we set $\gamma = 0$ in Equation (8) and thus the correct F-statistic in the first stage is 0. The standard IV estimator consistently has an F-statistic below 10, correctly indicating a weak instrument (Stock and Yogo, 2002). In contrast, the Case 3 estimators have much higher F-statistics, falsely suggesting a strong instrument. The TWFE estimator exhibits an even larger bias in the F-statistic, with at least 25% of the simulations having a "strong" first stage. Interestingly, the bias in F-Statistics is U-shaped for two-way fixed effects. This is because the bias from the Firm effects is large for small T while, at a large T, the power of the first stage is strong enough to pick up the smaller bias from the Year effects. These simulation results demonstrate the importance of carefully considering the interaction between leave-out instruments and fixed effect structures.

4 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 B.

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 1 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%.[19] 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 2 shows that this OLS 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 applying the instrument makes the coefficient smaller than in OLS, 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 2.5 are feasible here. Since there is no 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

[19]See also Figure B1 in Online Appendix B for the residualized distribution.

Table 1: 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 2: Estimation Results USPTO Data

		Real Examiners		Scrambled Examiners	
	(1)	(2)	(3)	(4)	(5)
	OLS	2SLS	2SLS	2SLS	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.

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 OLS. These numbers are comparable to the numerical simulation, especially when considering that with $|G(i)|/|C(i)|$ at around 0.05, the situation is comparable to that with 20 time periods in Figure 6.

Most importantly, this empirical application allows us to utilize our test for existence of the bias as it is described in Section 2.4. For step 1, 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 2, 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 OLS 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.[20]

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.[21] 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.

[20]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., 2024), since the outside sample scrambled IV is constructed from the same $Art\ unit \times Year$.

[21]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 B for details.

5 Discussion

Our critique of identification via constructed regressors in combination with certain fixed effect structures adds to a growing literature that highlights potential pitfalls of common identification techniques in general and leave-out-related strategies in particular (e.g. Angrist, 2014; Betz et al., 2018; Huber, 2023). In this section, we briefly cover how our bias relates to common assumptions of unbiased identification and discuss certain possible extensions of the model.

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 (1), it is obvious under Assumptions **??** and 1 that z_i and ε_i are uncorrelated. However, they are not uncorrelated conditional on the non-endogenous control variables (i.e. the fixed effects). Nevertheless, that is the necessary condition to fulfill the exclusion restriction (see, e.g., Greene, 2019, chapter 8). Thus when considering the exclusion restriction, it is important to consider the estimation strategy holistically instead of only considering the constructed regressor in isolation.[22]

An common extension of the results in Propositions 1 through 2 is that base variables are often transformed in econometric applications. For example, researchers often take the logs of variables or the variables are normalized into an interval such as [0,1]. How does this affect the problems identified in Section 2?

There are two ways in which transformations are commonly applied to a constructed regressor. The first one is that the variable v, which is underlying the leave-out instrument, is transformed directly. In this case, there is little change to our result Proposition 1, because the transformed version of a variable with a non-zero correlation with the error term is still likely to have a non-zero correlation with the error term. Further, in the special case of the leave-out construction, Proposition 2 will apply unchanged to the transformed variable.

The second way of applying transformations is that the constructed regressor is transformed after construction. In this case, our results may not apply directly, but the way in which they are changed introduces variation in the constructed regressor, which is likely unintended by the econometrician. Again, the transformation will change the correlation structure, but in a way that is unlikely to make the bias disappear. We use Case 1 from Proposition 2 to illustrate this. The

[22]The relationship of our bias with other identification tests, with a specific focus on using constructed regressors as instruments, is covered in Online Appendix E.

proof to this case boils down to

$$z_i - \frac{1}{|C(i)|} \sum_{j \in C(i)} z_j = b \left(v_i - \frac{1}{|C(i)|} \sum_{j \in C(i)} v_j \right). \tag{11}$$

Let the transformation be denoted by the function $\varphi(\cdot)$. If the transformation is linear then it holds that $\varphi(z_i) - \frac{1}{|C(i)|} \sum_{j \in C(i)} \varphi(z_j) = \varphi \left(z_i - \frac{1}{|C(i)|} \sum_{j \in C(i)} z_j \right) = \varphi \left(b \left(v_i - \frac{1}{|C(i)|} \sum_{j \in C(i)} v_j \right) \right)$ $= b \left(\varphi(v_i) - \frac{1}{|C(i)|} \sum_{j \in C(i)} \phi(v_j) \right)$ and the same results as before are obtained. This is, for example, the case if the variable is normalized into [0,1]. If, however, the transformation is non-linear, then $\varphi(z_i) - \frac{1}{|C(i)|} \sum_{j \in C(i)} \varphi(z_j) \neq \varphi \left(z_i - \frac{1}{|C(i)|} \sum_{j \in C(i)} z_j \right)$ and the results will not apply fully. In this case, the extent to which they apply will solely be determined by how close the transformation is to linearity. The log transformation is a good example here, because we know it to be close to linear if the values are close to one another (in relative terms). However, if the relative values are spaced far apart, there will be a larger difference between $\varphi(z_i) - \frac{1}{|C(i)|} \sum_{j \in C(i)} \varphi(z_j)$ and $\varphi \left(z_i - \frac{1}{|C(i)|} \sum_{j \in C(i)} z_j \right)$ and thus the constructed regressor will not fully reduce to the focal value. Nevertheless, all variation that makes using $\varphi(z_i)$ different from using $b\varphi(v_i)$ stems from the non-linearity of the transformation and not from the fact that z_i has become more informative as an regressor.

The theoretical results derived in Proposition 2 in Section 2.3 required the assumption that all construction groups are of equal cardinality. Similarly, our numerical illustrations also used fully balanced panels. In Online Appendix C, we explore numerically whether there are noticeable changes to the results when this assumption does not hold. Technically speaking this is only necessary for Cases 3 and 4 of Proposition 2, because all other cases can be shown to hold in unbalanced panels analytically. However, because these two cases are the empirically most relevant, we explore them numerically. Our findings show that our results are not limited to balanced panels, but also appear when the data is unbalanced. Thus, having instrument groups of different cardinalities does not address the issues raised in Propositions 2 even though these results are only formulated for balanced panels. Additionally, the analysis highlights that unbalanced panels can even exacerbate the bias and lead to more strongly misleading F-statistics associated with the estimators.

Lastly, all our theoretical and numerical analyses consider OLS estimators, independent of the ultimate use of the leave-out construction. In certain empirical literatures, such as demand estimations in industrial organization, it is typical to use methods of moments estimators or other non-linear procedures. While we cannot provide a blanket statement for the applicability of our results to such cases, it is clear from the proofs that if the within transformation for fixed effects is applied before the estimation, then our critique will be valid.

6 Conclusions

In this paper, we have identified a bias that arises when constructed regressors are combined with overlapping fixed effects. The inclusion of these fixed effects acts as a de-meaning process that can reintroduce the focal observation into the constructed 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 and the set used to create the regressor. Irrespective of the application, bias is most severe when the fixed effect cell overlaps strongly with the construction group. Unbiasedness is achieved only when the two sets do not share any observations. We provide a test based on scrambling and resampling of regressor construction groups that indicates whether or not the bias exists in a given data structure.

Our findings have important implications for researchers using constructed regressors in finance applications. Careful consideration must be given to the interaction between the fixed effects structure and construction of the regressor to avoid mechanical correlation and, in the case of instrumental variable estimations, inflated first-stage F-statistics. When possible, researchers should use fine-grained fixed effects such that overlap between the construction group and the fixed effect set is avoided. If this is infeasible, the regressor should be constructed to have minimal overlap with the fixed effects.

We illustrate the practical relevance of our results in numerical studies and an empirical application of the judge fixed effects estimation strategy. Across both of these settings, naively combining constructed regressors with fixed effects can generate substantial bias. Our proposed remedies, guided by our theoretical results, provide solutions to obtain consistent estimates.

As the use of constructed regressors continues to proliferate, our paper serves as a cautionary note and practical guide. Econometric methods leveraging fine-grained variation can be powerful for causal inference in financial research, but their validity depends crucially on understanding how that variation interacts with other dimensions of the empirical specification. We hope our analysis enables applied financial researchers to reap the benefits of constructed regressors while avoiding the pitfalls that arise from their interaction with fixed effects.

References

Abadie, A., Gu, J., and Shen, S. (2024). Instrumental variable estimation with first-stage heterogeneity. Journal of Econometrics, 240(2):105425.

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.

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.

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.

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.

Borusyak, K., Jaravel, X., and Spiess, J. (2024). Revisiting event study designs: Robust and efficient estimation. Review of Economic Studies, page Forthcoming.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. Working Paper.

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. NBER Working Paper, 25528.

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.

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. (2024). Cluster jackknife instrumental variables estimation. Unpublished Working Paper.

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. forthcoming.

Gao, J., Ge, S., Schmidt, L. D., and Tello-Tril, C. (2024). How do health insurance costs affect low-and high-income workers? Working Paper.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. Journal of Econometrics, 225(2):254–277.

Greene, W. H. (2019). Econometric Analysis. Pearson.

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

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

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

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

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

Klosin, S. (2024). Dynamic biases of static panel data estimators. accessed via https://klosins.github.io/Klosin_JMP.pdf.

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

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

Lancaster, T. (2000). The incidental parameter problem since 1948. Journal of Econometrics, 95(2):391–413.

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.

Mikusheva, A. and Sølvsten, M. (2023). Linear regression with weak exogeneity. arXiv preprint arXiv:2308.08958.

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

Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. Econometrica, pages 1–32.

Nickell, S. (1981). Biases in dynamic models with fixed effects. Econometrica: Journal of the econometric society, pages 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.

Słoczyński, T. (2022). Interpreting ols estimands when treatment effects are heterogeneous: smaller groups get larger weights. Review of Economics and Statistics, 104(3):501–509.

Stock, J. H. and Yogo, M. (2002). Testing for weak instruments in linear iv regression.

Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. Journal of Econometrics, 225(2):175–199.

Wager, S. and Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. Journal of the American Statistical Association, 113(523):1228–1242.

This appendix is organized in two sections. Section A.1 presents additional regularity assumptions and remarks, and Section A.2 presents proofs.

.1 Additional Assumptions

Assumption 4. *(Idiosyncratic errors and regressors) Let* $\lambda_i = z_i\tilde{\varepsilon}_i - E(z_i\tilde{\varepsilon}_i)$, $\mathbf{P}_{\tilde{X}} = \tilde{\mathbf{X}}\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1}\tilde{\mathbf{X}}'$, *and* $\mathbf{M}_{\tilde{X}} = \mathbf{I}_n - \mathbf{P}_{\tilde{X}}$, *where* $\mathbf{I}_n$ *is* $n \times n$ *identity matrix and* $\tilde{\mathbf{X}} = (\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, ..., \tilde{\mathbf{x}}_n)'$. *As* $n \to \infty$,

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

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

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

Remark 1. Assumption 4 stipulates general high-level requirements on the idiosyncratic errors and regressors required for the consistency of FE estimator. Condition (i) allows for heteroskedastic and correlated idiosyncratic errors, but rules out strong correlations among errors. [Literature on weak/strong cross section dependence can be cited] Condition (ii) essentially ensures sufficient variation among z_i once the fixed effects and regressors in $\mathbf{X}$ were filtered out. Condition (iii) is an exogeneity requirement ensuring that the correlation between the regressors in $\mathbf{X}$ and idiosyncratic errors is sufficiently weak for it to not affect the consistency of the FE estimator.

.2 Proofs

Proof of Proposition 1. Using $\mathbf{y} = \beta\mathbf{z} + \mathbf{X}\delta + \eta + \varepsilon$, the vector of demeaned variables $\tilde{\mathbf{y}}$ is given by $\tilde{\mathbf{y}} = \beta\tilde{\mathbf{z}} + \tilde{\mathbf{X}}\delta + \tilde{\varepsilon}$. Substituting this expression for $\tilde{\mathbf{y}}$ in (4), and noting that $\mathbf{M}_{\tilde{X}}\tilde{\mathbf{X}} = \mathbf{0}$, we obtain

$$\hat{\beta} - \beta_0 = \left(\frac{\tilde{\mathbf{z}}'\mathbf{M}_{\tilde{X}}\tilde{\mathbf{z}}}{n}\right)^{-1}\frac{\tilde{\mathbf{z}}'\mathbf{M}_{\tilde{X}}\tilde{\varepsilon}}{n}.$$

Under Assumption 4.(ii), $n^{-1}\tilde{\mathbf{z}}'\mathbf{M}_{\tilde{X}}\tilde{\mathbf{z}} \to_p Q_z$, where $Q_z > 0$. Consider $n^{-1}\tilde{\mathbf{z}}'\mathbf{M}_{\tilde{X}}\tilde{\varepsilon}$ next. We have

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

where $n^{-1}\tilde{\mathbf{z}}'\mathbf{P}_{\tilde{X}}\tilde{\varepsilon} \to_p 0$, as $n \to \infty$ under Assumption 4.(iii). Last but not least, consider $n^{-1}\tilde{\mathbf{z}}'\tilde{\varepsilon}$. Let $\mathbf{M_C}$ be the orthogonal projection matrix that filters out the fixed effects, namely $\tilde{\mathbf{z}} = \mathbf{M}_C\mathbf{z}$, and $\tilde{\varepsilon} = \mathbf{M_C}\varepsilon$. Since $\mathbf{M}_C$ is symmetric and idempotent, we have

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

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 4.(i), $\frac{1}{n}\sum_{i=1}^{n}\lambda_i \to_p 0$ 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 (2) for z_i, we can write $\Delta_{\beta,n}$ as

$$\begin{aligned} \Delta_{\beta,n} &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{|G(i)|}\sum_{j\in G(i)} E\left(v_j\varepsilon_i\right)\right) \\ &\quad -\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{|G(i)|\,|C(i)|}\sum_{h\in G(i)}\sum_{j\in C(i)} E\left(v_h\varepsilon_j\right)\right) \end{aligned} \tag{12}$$

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 (12), 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

$$\hat{\beta} - \beta_0 + Q_z^{-1}\Delta_{\beta,n} \to_p 0,$$

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 (5) 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 (6), $|G(i)| = 1$, $|C(i)| = T-1$, 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, $n^{-1}\tilde{\mathbf{z}}'\mathbf{M}_{\tilde{X}}\tilde{\mathbf{z}}$ reduces to $F^{-1}(T-1)^{-1}\sum_{f=1}^{F}\sum_{t=2}^{T}\tilde{v}_{f,t-1}^{2}$. Let $Q_{v,T} = plim_{F\to\infty}F^{-1}(T-1)^{-1}\sum_{f=1}^{F}\sum_{t=}^{T}$
then result (7) follows. □

Proof of Proposition 2. We begin by showing item 3 (Case4). The other items follow immediately from that. From the Frisch-Waugh-Lovell Theorem, estimating Equation (1) is equivalent to estimating

$$y_i^R = \beta z_i^R + \boldsymbol{X}_i^R\gamma + \varepsilon_i^R \tag{13}$$

We consider $z_i^R = z_i - \frac{1}{|C(i)|}\sum_j \in C(i) z_j$. From Equation (2) 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

$$z_i^R = \frac{1}{|J(i)|-1}\sum_{j\in J(i)\backslash i} v_j - \frac{1}{|C(i)|}\sum_{j\in C(i)}\frac{1}{|J(j)|-1}\sum_{h\in J(j)\backslash j} v_h \tag{14}$$

$$\begin{aligned}
&= \frac{1}{|J(i)|-1}\left(\sum_{j\in\hat{J}(i)} v_j - v_i\right) + \frac{1}{|J(i)|-1}\sum_{j\in\check{J}(i)} v_j \\
&\quad - \frac{1}{|C(i)|}\sum_{j\in C(i)}\frac{1}{|J(j)|-1}\left(\sum_{h\in\hat{J}(j)\backslash j} v_h + \sum_{h\in\check{J}(j)} v_h\right)
\end{aligned} \tag{15}$$

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

$$\begin{aligned}\dot{z}_i =& \frac{1}{|J(i)|-1}\left(\sum_{j\in\hat{J}(i)} v_j - v_i - \sum_{j\in C(i)\setminus\hat{J}(i)} v_j + \sum_{j\in C(i)\setminus\hat{J}(i)} v_j - \frac{1}{|C(i)|}\sum_{j\in C(i)}\sum_{h\in\hat{J}(j)\setminus j} v_h\right)\\ &+ \frac{1}{|J(i)|-1}\left(\sum_{j\in\check{J}(i)} v_j - \frac{1}{|C(i)|}\sum_{j\in C(i)}\sum_{h\in\check{J}(j)} v_h\right) \qquad (16)\end{aligned}$$

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

$$\begin{aligned}\dot{z}_i =& \frac{1}{|J(i)|-1}\left(-v_i - \sum_{j\in C(i)\setminus\hat{J}(i)} v_j + \frac{|C(i)|-|\hat{J}(i)|+1}{|C(i)|}\sum_{j\in C(i)} v_j\right)\\ &+ \frac{1}{|J(i)|-1}\left(\sum_{j\in\check{J}(i)} v_j - \frac{1}{|C(i)|}\sum_{j\in C(i)}\sum_{h\in\check{J}(j)} v_h\right). \qquad (17)\end{aligned}$$

We reverse the within transformation and obtain item 3 of the proposition. Item 2 follows by observing that in this case $\check{J}(i) = \varnothing \; i$. Item 1 follows from observing that in these cases both $\check{J}(i) = \varnothing \; i$ 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 $G(i) \cap C(i) = \varnothing$. □

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 **??** through 1, 2, and 3. 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 Ω. We consider a specific balanced panel structure in which for every i the intersection of its two fixed effect cell is equal to the jackknife group $J(i)$. The data structure considered in the numerical simulations of Section 3 is a possible illustration of the case covered here.

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 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 OLS under Assumptions **??** *through 1 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_1 c_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 && \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_1 c_2} \sum_{j \in C(1,i)} \sum_{h \in C_{2j}} v_h \right) && \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_1 c_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_1 c_2} \sum_{j \in C(1,i)} \sum_{h \in C_{2j}} v_h \right) && \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_1 c_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_1 c_2} \sum_{j \in C(1,i)} \sum_{h \in C_{2j}} v_h \right) && \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_1 c_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_1 c_2} \sum_{j \in C(1,i)} \sum_{h \in C_{2j}} v_h \right) && \text{(A9)}
\end{aligned}$$

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

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 Details on USPTO Data Analysis

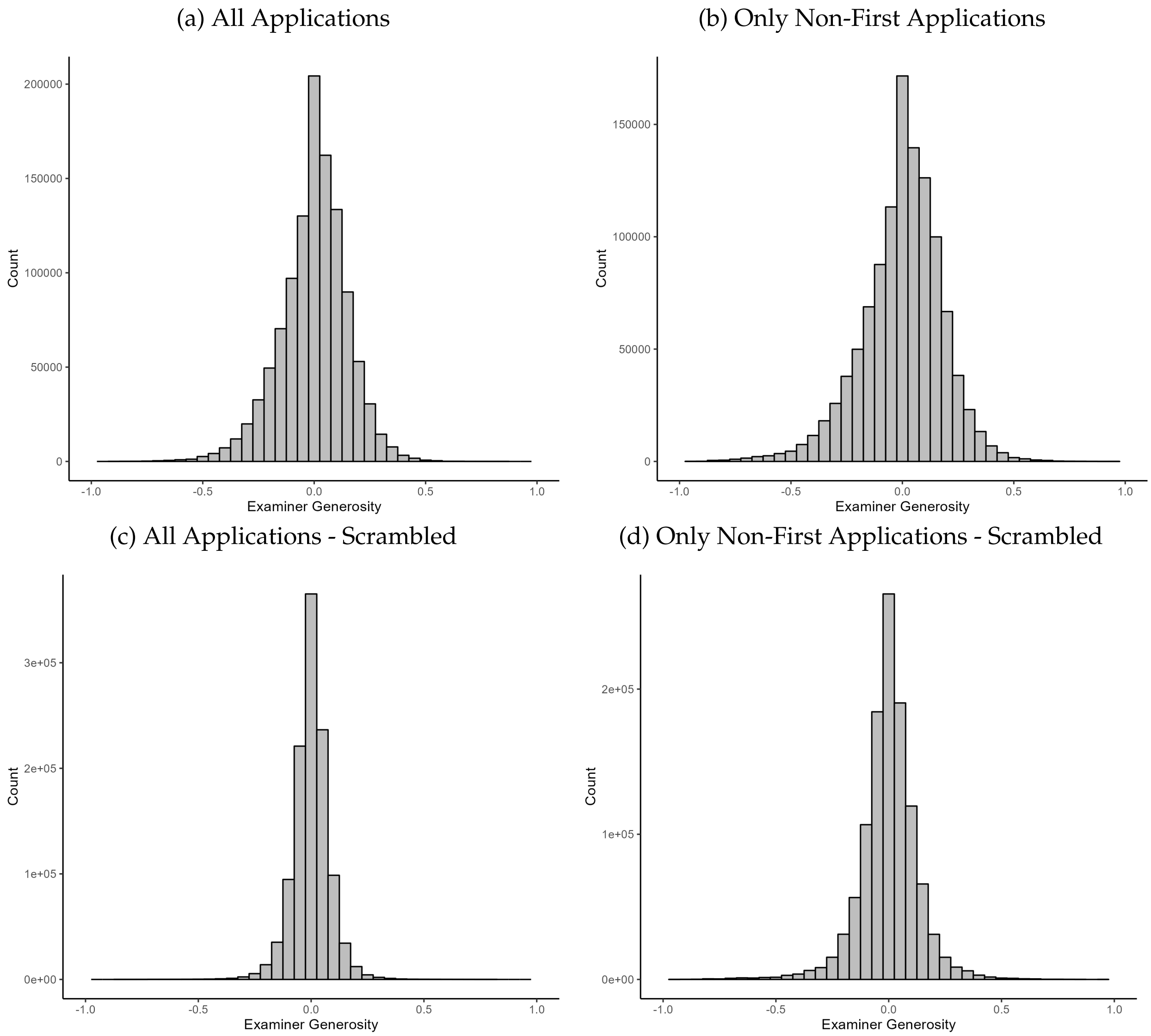
We source data from the 2014 wave of the USPTO Patent Examination Research Dataset (PatEx). We chose this wave because, in contrast to later waves, it provides a unique examiner ID for each patent application, which reduces measurement error in the identification of the instrument group. We combine the 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 29$^{\text{th}}$ 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^{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

$$z_{i,e,t} = \frac{1}{n(e,t)-1}\left(\sum_{j=1}^{n(e,t)} x_{j,e,t} - x_{i,e,t}\right). \tag{B1}$$

Figure B1: 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 (B1)) 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 (B2)). Scrambled instruments result from randomly assigned examiners within a and $Art\ unit \times Year$ cell.

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

$$z_{e,t} = \frac{1}{n(e,t,\neg 1)} \left(\sum_{i=1}^{n(e,t,\neg 1)} x_{i,e,t,\neg 1} \right). \tag{B2}$$

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 B1. 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 (B1) is 0. We do this for both the real examiner IDs and the scrambled ones such that all estimations in Table 2 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 B1 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 B1: Estimation Results with Alternative Definition of Repeated Application

	(1) OLS	(2) 2SLS	(3) 2SLS	(4) 2SLS	(5) 2SLS
Success	0.041***	0.091***	0.106***	0.048***	0.898
	(0.001)	(0.004)	(0.005)	(0.015)	(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 OLS 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 OLS and 2SLS estimates, the results are remarkably similar. The canonical instrument has a 23% relative bias compared to that of OLS. More importantly, the scrambled analysis returns a coefficient about equal to that of the OLS 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 B2: Estimation Results with Alternative Identification of

	(1) OLS	(2) 2SLS	(3) 2SLS	(4) 2SLS	(5) 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 B2 and show the same pattern as in the main analysis. The bias is now more pronounced, it is 54% relative to OLS. 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 OLS, 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 C Unbalanced Panels in the Numerical Study

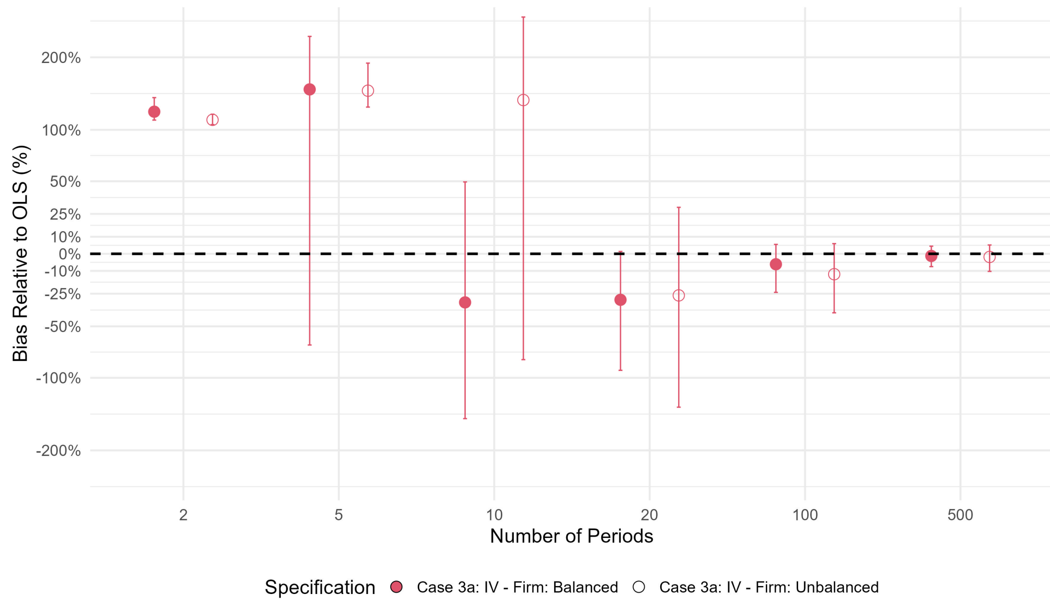
Figure C1 repeats the analysis reported in Figure 6 for Case 3, using an unbalanced panel instead. The simulations are based on the same data-generating process as the main analysis but with two key modifications. First, the number of markets is doubled to increase the potential for unbalancedness. Then, half of the observations are randomly dropped, resulting in an unbalanced panel structure with the same number of observations as the balanced one.

Panel (a) compares the median bias of the estimators between the balanced and unbalanced panel settings. All of our estimators, which use $Firm$ fixed effects, are worse in the unbalanced panel, demonstrating their sensitivity to relative group sizes. Panel (b) shows the first stage F-statistics for the unbalanced panel setting in simulations where the instrument has no strength. All estimators have an upwardly biased F-statistic as the panel loses balance, incorrectly suggesting a stronger instrument.

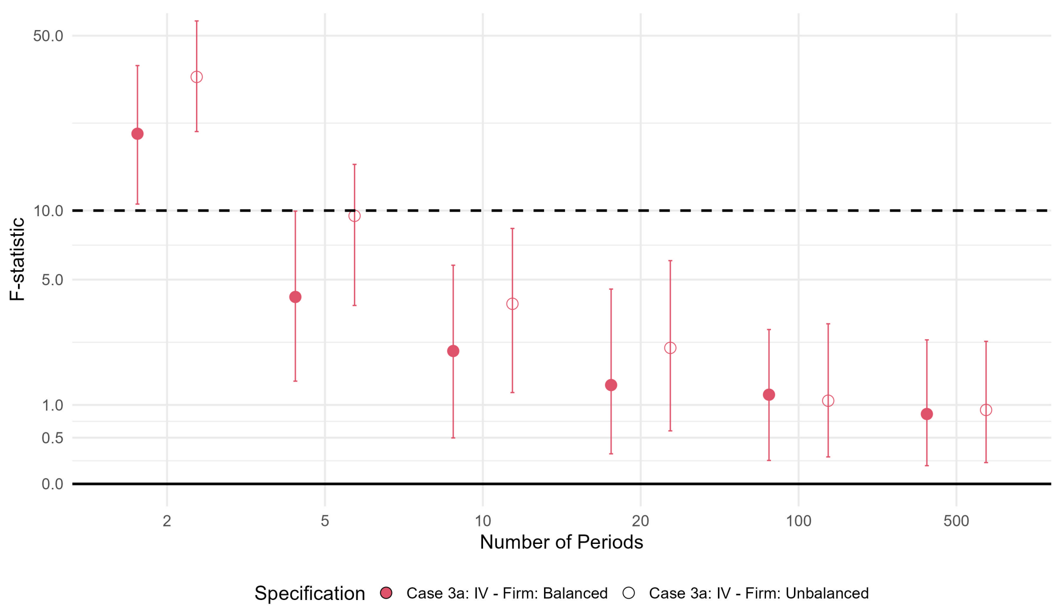
Unbalancedness exacerbates the bias for two potential reasons. Propositions 1 and Online Appendix D show that a smaller number of markets increases the bias. The first potential explanation thus involves how OLS deals with heterogeneous treatment effects. As shown by Słoczyński (2022), when different-sized groups with different treatment effects are combined into one regression, the coefficients will be a weighted average of the two treatment effects, with the smaller group receiving relatively more weight. This intuition carries to our setting. Since the bias is determined by the size of the groups, unbalanced panels will mechanically exhibit different "treatment effects" in the first stage. Since smaller groups induce a larger bias, the groups with the most bias will receive the most weight in estimating the first stage. Second, Online Appendix D also indicates that the absolute value of the bias is decreasing convexly in the number of markets. An average of biases across firms with different numbers of markets will thus be larger than the bias for a firm in the average number of markets.

Figure C1: Unbalanced Panels

(a) Median Bias



(b) First Stage F-Statistic



Note: Panel (a) of this figure shows the median bias of separate models, with and without balanced panels, across 500 simulations for each level of $|T|$. Panel (b) shows how the bias influences the F-statistic when there is no relevance for the instrument ($\gamma = 0$). The data-generating process is summarized in Equations (8) through (10) with $\eta_f = 0$. Panels become unbalanced by doubling the number of markets and then randomly dropping 50% of observations. Inter-quartile ranges are also shown. Each color/point represents a separate model specification. Whiskers indicate the inter-quartile range. Note that the y-axis is "log"-scaled both above and below zero for better visual presentation. All simulations have 20 firms; and 20 Markets.

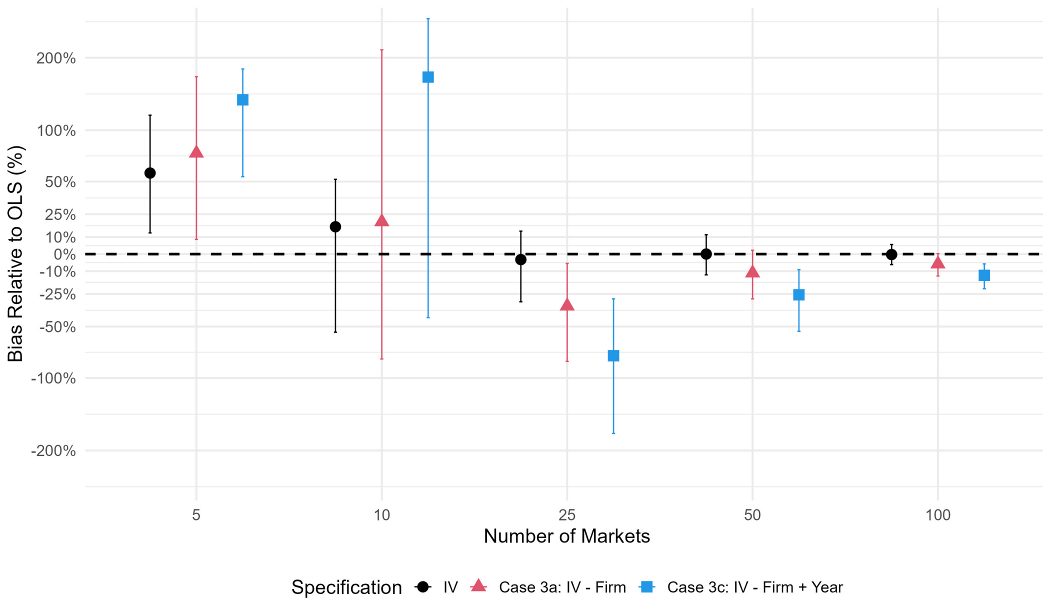
Online Appendix D Varying Number of Markets

Figure D1 depicts the median bias relative to the OLS bias across 500 simulations for varying levels of M, along with the inter-quartile range. Each color and point signifies a distinct model specification. To enhance visual clarity, the y-axis is "log"-scaled both above and below zero. The data-generating process for the simulations is described in Equations (8) through (10) with $\eta_f = 0$.

The standard IV estimator initially exhibits bias due to the limited number of data points but approaches unbiasedness as the number of markets increases and remains unbiased thereafter. In contrast, cases 3a and 3b display a non-monotonic impact of bias, with the median bias being more pronounced when there are 20 markets compared to 10 markets. This finding underscores the complex nature of the bias and its sensitivity to the number of markets in the sample.

Figure D1: Increasing Number of Periods

(a) Relative Bias



Note: Results of a Monte Carlo simulation with the data generating process described in Equations (8) through (10) with $\eta_f = 0$. This figure shows the median bias (relative to the bias of OLS) across 500 simulations for each level of M. We plot the median as well as the inter-quartile range. Each color/point represents a separate model specification. Note that the y-axis is "log"-scaled both above and below zero for better visual presentation. All simulations have a bias that is ten times larger than the instrument strength and contain 20 firms and 20 Years.

Online Appendix E Discussion and Relationship to Judge FE Tests

In this appendix, we will briefly discuss how our argument here relates to other critiques and tests of leave-out instruments.

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 3 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.