Per-Cluster Instrumental Variables Estimation: Uncovering the Price Elasticity of the Demand for Gasoline

Michael Bates and Seolah Kim *

January 28, 2020

Abstract

We propose a per-cluster instrumental variables estimator (PCIV) for estimating population average effects under correlated random coefficient models in the presence of endogeneity. We demonstrate consistency, showing robustness over standard estimators, and provide analytic standard errors for robust inference. We compare PCIV, fixed-effects instrumental variables, and pooled 2-stage least squares estimators using Monte Carlo simulation verifying that PCIV performs relatively well. We also apply the approaches, examining the monthly responsiveness of gasoline consumption to prices as instrumented by state fuel taxes. We find that US consumers are on average more elastic in their demand for gasoline than previous estimates imply.

*Department of Economics, University of California at Riverside, Riverside, CA 92521, United States of America (email: michael.bates@ucr.edu; skim238@ucr.edu). We wish to thank Lucas Davis, Aman Ullah, Ruoyao Shi, Gary Solon, Jeff Wooldridge, and participants at the Midwest Econometric Group Conference in Madison, WI and at the UC Riverside Brown Bag Seminar for helpful comments. All mistakes are our own.

1 Introduction

The goal of empirical work is often to identify the population average effect (PAE)—the true causal effect of one variable on another that represents the average effect over the entire population of interest. Inherent in these efforts is the idea that these effects may differ across individuals in the population. While the idea of heterogeneous random slopes far predates Heckman and Vytlacil (1998), in allowing the heterogeneous effects to be correlated with explanatory variables, Heckman and Vytlacil (1998) provide the useful term of correlated random coefficient (CRC) models.[1]

The discussion of heterogeneous effects garnered further attention with Imbens and Angrist (1994) introducing the notion of local average treatment effects (LATEs) when using instrumental variables to identify causal effects in cross-sectional settings. They show that with minimal assumptions using instrumental variables, researchers can identify average treatment effects among the population whose "treatment status is influenced by the instrument." Heckman and Vytlacil (2005) and Deaton (2010) comment on the limitations of such LATEs, with Heckman and Vytlacil (2005) demonstrating the use of additional structure to estimate parameters of arguably more interest.

Growing access to large data sets that carry a structure by which observations are related to one another through shared membership in a common cluster may provide hope for estimating PAEs with fewer assumptions. These clusters can take the form of multiple time observations following the same individual as in standard panel data settings, or grouped cross-sectional data where individuals are clustered, for instance into classrooms or establishments or states.

In either setting, knowledge of these groupings of data may be useful to uncover PAEs in CRC models, when the heterogeneous effects enter at the cluster level. The panel data setting is particularly intuitive as we typically think of types of individuals having differences in their responsiveness. Indeed, Wooldridge (2005) provides the conditions under which a

[1] See for instance, Rubin (1950); Klein (1953); Kuh (1959); Swamy (1971); Mundlak (1978); Raj et al. (1980), and Chamberlain (1992), for early examinations of the random coefficients model.

general class of fixed effects estimators is consistent in estimating PAEs in CRC models when the explanatory variables are otherwise exogenous. Bates et al. (2014) notes that the additional assumption necessary for fixed effects to consistently estimate PAEs may not be benign, and proposes a per-cluster estimator, which is unbiased in estimating PAEs even when fixed effects estimation may be inconsistent.

However, applied economic researchers often work in settings where strict exogeneity of the explanatory variables likely does not hold. Murtazashvili and Wooldridge (2008) provides the conditions under which a general class of fixed effects instrumental variables estimators (FEIV) consistently estimate PAEs with endogenous regressors. Specifically, Murtazashvili and Wooldridge (2008) finds that the presence of heterogeneous slopes, requires us to assume that the heterogeneous slopes are uncorrelated with the covariance between the detrended instruments and endogenous regressors.

Unfortunately, this restriction may not hold in important cases. First, consider the case where a LATE exists and is different from the ATE. That is, the instrument is not relevant for some portion of the population, and the effects differ on average between those who are and are not moved by the instrument. Such a setting would prevent FEIV from consistently estimating the ATE. Secondly, consider the application of FEIV on the population who were all influenced by a valid instrument (whom Imbens and Angrist (1994) term "compliers"). If there are cluster-specific heterogeneous effects, and if that heterogeneity is correlated with the strength of the instrument within-cluster, fixed effects would continue to fail to estimate the ATE among the compliers (the LATE). Thus, in some respects, this assumption is stronger than simply assuming that the LATE is no different from the ATE.

Given these limitations of fixed effects instrumental variables, we look to another estimator to identify PAEs. As larger data sets (both in numbers of clusters and numbers of observations per cluster) become available, a natural inclination may be to estimate cluster specific slopes and average over them.[2] We provide the conditions under which such an

[2]For instance, adopting such an approach is alluded to in Solon et al. (2015).

approach is consistent in estimating PAEs. We term this approach per-cluster instrumental variables (PCIV). We compare the finite sample performance of this PCIV approach against standard FEIV and pooled two stage least squares (P2SLS) estimators using Monte Carlo simulation. The PCIV approach performs relatively well even with somewhat small numbers of observations per-cluster. Further, we provide three extensions to address settings common in applied empirical work: additional exogenous covariates, weighting to address heterogeneous cluster sizes, and the exploration of mechanisms using cluster-level covariates.

Finally, we demonstrate the use of per-cluster instrumental variables examining the price elasticity of gasoline demand in the United States. We follow Davis and Kilian (2011), and Coglianese et al. (2017) in instrumenting prices with state fuel taxes. We view this empirical exercise as important in its own right. The average responsiveness of consumers to price increases in gasoline is of particular importance to both economists and policymakers. This population-level parameter is important for gauging macroeconomic effects of gasoline price fluctuations, for speculation in oil markets, for modeling the market for automobiles, urban planning, optimal taxation, and national security (Dahl and Sterner, 1991; Davis and Kilian, 2011; Espey, 1998; Hughes et al., 2008; Coglianese et al., 2017). Perhaps most importantly, in the midst of a growing literature quantifying the costs of global climate change many leading economists are calling for policy intervention to address these externalities.[3] Most notable among these, 27 Nobel Laureate economists, 4 former Chairs of the Federal Reserve, and 15 former Chairs of the Council of Economic Advisers serving under presidents from both major political parties in the United States co-signed a letter calling for immediate national action on climate change (Climate Leadership Council, 2019). In their letter published in the *Wall Street Journal*, they recommend the adoption of a carbon tax as "the most cost-effective lever to reduce carbon emissions at the scale and speed that is necessary." Recommending that it should be increased each year until emissions reduction goals are met, be coupled with a broader carbon adjustment system, and return tax revenue directly to U.S. citizens

[3]For examples, see Tol (2002); Deschênes et al. (2009); Fisher et al. (2012), and Burke et al. (2015).

through equal lump-sum dividends. The efficacy of such a policy relies on consumers' price sensitivity in their demand for carbon-rich products. Petroleum is one such product, the burning of which accounts for 34 percent of the total U.S. anthropogenic green house gas emissions and about 42 percent of total U.S. anthropogenic carbon dioxide emissions (US Energy Information Administration, 2018).

We are concerned that the most convincing existing estimates of this important parameter still seem to rely upon assumptions which may be problematic, most notably homogeneous elasticities of demand for gasoline across states. Accordingly, we estimate the price elasticity of demand for gasoline using the more robust PCIV approach. Both for comparability to Coglianese et al. (2017) and Li et al. (2014) and to demonstrate the performance of this proposed estimator, we also use P2SLS applied to first-differences and FEIV. Furthermore, earlier work rely on aging and discontinued data series, which we update in our analysis with novel data collection.[4] Using data on state-level monthly gasoline sales and prices as instrumented by state taxes from 1989-2018, we find that both the magnitude and significance of the results are sensitive to the methods used. We estimate a larger elasticity of the demand for gasoline (-0.6 to -0.7) than the estimates of monthly responsiveness to prices reported in Coglianese et al. (2017) (-0.3 to -0.4). These larger estimates suggest that carbon tax increases may provide a quicker lever for reaching climate goals than previously thought.

This application allows us to demonstrate additional reasons researchers may find a per-cluster instrumental variables approach useful even when not the primary specification. The procedure provides sample analogues to many of the assumptions made when estimating average or local average effects with fixed effects instrumental variables approaches. First, the state-specific estimates give evidence to the existence of "essential heterogeneity" in elasticities across space. Secondly, analysis of the cluster specific first-stage slopes reveals whether

[4]We along with Davis and Kilian (2011); Li et al. (2014), and Coglianese et al. (2017) rely on US Department of Energy, Energy Information Administration, 'Petroleum Marketing Monthly Report: Gasoline Prices by Formulation, Grade, Sales Type' for gasoline prices by state for the period 1989-2008. However, this series was discontinued in 2011. As a result, we add to this data state-by-month averages of at-the-pump gasoline prices collected from Gasbuddy.com.

the monotonicity assumption necessary for estimating LATEs holds within sample. Third, once cluster specific slopes are estimated, it is simple to find the correlation between the estimated heterogeneous slopes and the within-cluster relevance of the instrument in order to assess the key condition for the consistency of FEIV. Lastly, we compare the natural market share weights we employ with PCIV to the relative strength of the instrument implicitly used by FEIV and P2SLS as weights of state-specific elasticities.

The remainder of the paper is organized as follows. In Section 2, we introduce the econometric model and briefly summarize existing estimators with exogenous and endogenous regressors. In Section 3, we introduce the proposed estimator providing the main consistency results with Section 5 providing possible extensions of the estimator. Section 4 contains a Monte Carlo study that shows that the PCIV estimator outperforms P2SLS and FEIV when the key condition is violated, confirming the results from Section 2. We provide a reexamination of Coglianese et al. (2017) in Section 6 using the approaches laid out in Sections 3 and 5. Section 7 offers concluding remarks.

2 Model specification and previous results

We initially take a standard correlated random effects model for a randomly selected i from the population. The model shown below is similar to those examined in Wooldridge (2005); Arellano and Bonhomme (2011); Graham and Powell (2012) and Laage (2019).

$$y_{ij} = \mathbf{x_{ij}}\mathbf{b_i} + e_{ij}, \quad i = 1, ..., N, \; j = 1, ..., T \tag{1}$$

where y_{ij} is a dependent variable and e_{ij} is an idiosyncratic error. The $1 \times K$ vector, $\mathbf{x_{ij}}$, includes 1 as well as covariates that may be endogenous and are allowed to vary both between and within clusters. A key feature of the model is the $K \times 1$ vector of cluster-specific slopes, $\mathbf{b_i} = \boldsymbol{\beta} + \mathbf{d_i}$, where $E(\mathbf{d_i}) = 0$ by definition. This vector indicates the heterogeneous effects that vary by cluster and may be correlated with $\mathbf{x_{ij}}$.[5]

2.1 Exogenous regressors

Even if $\mathbf{x_{ij}}$ is otherwise strictly exogenous, ignoring the heterogeneous intercepts and coefficients may be problematic for uncovering consistent estimates of parameters of interest. Wooldridge (2005) shows the conditions under which standard fixed effects estimators are consistent in estimating average treatment effects (ATEs). To summarize in this simple case with exogenous regressors and fixed T, the assumptions for consistency, are the following:

$$\mathrm{E}(\mathbf{e_{ij}}|\mathbf{x_{i1}}, ..., \mathbf{x_{ij}}, \mathbf{b_i}) = 0,\ i = 1, ..., N,\ j = 1, ..., T, \tag{2}$$

$$rank\ \mathrm{E}\left(\sum_{j=1}^{T} \ddot{\mathbf{x}}'_{\mathbf{ij}}\ddot{\mathbf{x}}_{\mathbf{ij}}\right) = K, \tag{3}$$

$$\mathrm{E}[\ddot{\mathbf{x}}'_{\mathbf{ij}}\ddot{\mathbf{x}}_{\mathbf{ij}}\mathbf{d_i}] = 0, \tag{4}$$

where $\ddot{\mathbf{x}}_{\mathbf{ij}} = \mathbf{x_{ij}} - T^{-1}\sum_{j=1}^{T}\mathbf{x_{ij}}$. Equation 4 shows that in addition to standard rank and strict exogeneity assumptions, the presence of cluster-specific coefficients necessitates an additional assumption for fixed effects estimators to be consistent.

Though this additional condition is seemingly benign, there are several situations in which equation 4 may not hold. Bates et al. (2014) explores one. As a motivating example, they consider estimation of the effect of socio-economic status (SES) on students academic performance on a standardized mathematics exam. Were the effect of SES to differ across schools in a way that is related to the diversity within schools, fixed effects estimation would be inconsistent, as this setting violates the key condition shown in equation 4.

Bates et al. (2014) proposes using a per-cluster estimator similar to that discussed in Kuh (1959) and Swamy (1971), though the recent work avoids restrictive assumptions made in the earlier work. The per-cluster estimator is unbiased in estimating ATEs without further

[5] In principal $\mathbf{x_{ij}}$ may also include aggregate time variables such that the $\mathbf{b_i}$ allows for cluster-specific flexible time trends as in Wooldridge (2005) and Murtazashvili and Wooldridge (2008).

assumptions, and relaxes the assumption in equation 4 from Wooldridge (2005).[6] To adapt per-cluster estimation to this simple model, we need only two steps. First, estimate $\hat{\mathbf{b}}_i$ for each cluster using OLS on only the within-cluster observations, such that;

$$\hat{\mathbf{b}}_i = \boldsymbol{\beta} + \mathbf{d_i} + \left(\sum_{j=1}^{T} \mathbf{x'_{ij}x_{ij}}\right)^{-1} \left(\sum_{j=1}^{T} \mathbf{x'_{ij}} e_{ij}\right). \tag{5}$$

Second, average $\hat{\mathbf{b}}_\mathbf{i}$ over clusters. Thus,

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{PC}} = \boldsymbol{\beta} + \frac{1}{N}\sum_{i=1}^{N} \mathbf{d_i} + \frac{1}{N}\sum_{i=1}^{N}\left[(\sum_{j=1}^{T} \mathbf{x'_{ij}x_{ij}})^{-1}(\sum_{j=1}^{T} \mathbf{x'_{ij}} e_{ij})\right]. \tag{6}$$

In this simple setting, unbiasedness follows from the rank conditions, the fact that $\mathrm{E}(\mathbf{d_i}) = 0$ by definition, and the strict exogeneity assumption in equation 2.

2.2 Endogenous regressors

We now turn to settings where assumption 2 may not hold, whether the violation originates from measurement error, time-varying omitted variables, or simultaneity. Continuing with a simple model, we introduce $\mathbf{z_{ij}}$, a $1 \times L$ $(L \geq K)$ vector of instrumental variables.

Murtazashvili and Wooldridge (2008) shows the conditions under which a general class of fixed effects instrumental variables estimators is consistent in identifying PAEs. To summarize their findings, consider the following representation of the estimate of β from two stage least squares applied to the demeaned covariates and instruments:

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{FEIV}} = \boldsymbol{\beta} + \left(\sum_{i=1}^{N}\sum_{j=1}^{T} \mathbf{\ddot{x}'_{ij}H_z\ddot{x}_{ij}}\right)^{-1}\left[\sum_{i=1}^{N}\sum_{j=1}^{T} \mathbf{\ddot{x}'_{ij}H_z\ddot{x}_{ij}d_i} + \sum_{i=1}^{N}\sum_{j=1}^{T} \mathbf{\ddot{x}'_{ij}H_z}\ddot{e}_{ij}\right], \tag{7}$$

where $\mathbf{H_z} = \mathbf{\ddot{z}_{ij}}\left(\sum_{i=1}^{N}\sum_{j=1}^{T} \mathbf{\ddot{z}'_{ij}\ddot{z}_{ij}}\right)^{-1}\mathbf{\ddot{z}'_{ij}}$.

[6]The model considered above is a special case of the framework in Bates et al. (2014), which focuses on estimating the effect of a conditionally exogenous covariate that varies only at the cluster level.

Thus, in addition to the standard rank assumptions and the strict exogeneity assumption on the instruments, Murtazashvili and Wooldridge (2008) observe that they also "need assumptions such that $\ddot{\mathbf{z}}_{\mathbf{ij}}$ is uncorrelated with $\ddot{\mathbf{x}}_{\mathbf{ij}}\mathbf{d}_{\mathbf{i}}$." Accordingly, they show that requiring 1) the mean of $\mathbf{b}_{\mathbf{i}}$ to be independent of $\ddot{\mathbf{z}}_{\mathbf{ij}}$ and 2) the covariance of $\ddot{\mathbf{x}}_{\mathbf{ij}}$ and $\mathbf{b}_{\mathbf{i}}$ to not depend on $\ddot{\mathbf{z}}_{\mathbf{ij}}$ is sufficient for consistency of FEIV estimators.[7]

However, a violation of these assumptions occurs if the strength of the instrument is related to the heterogeneous effects. In which case, $\mathrm{E}[(\ddot{\mathbf{z}}_{\mathbf{ij}}'\ddot{\mathbf{x}}_{\mathbf{ij}})^{-1}\ddot{\mathbf{z}}_{\mathbf{ij}}'\ddot{\mathbf{x}}_{\mathbf{ij}}\mathbf{d}_{\mathbf{i}}] \neq 0$. Such a violation of this condition occurs if for those for whom the instrument is relevant (whom we think of as compliers from Imbens and Angrist (1994)) have systematically different responses to the treatment, than does the rest of the population. In other words, in order for FEIV to provide the ATE, we must assume that the LATE does not differ from the ATE. A further implication of this violation is that correlation between the strength of the instrument and the heterogeneous effects at the intensive margin, may cause FEIV to fail to consistently estimate even the LATE.

This or similar restrictions have been questioned by earlier work. Murtazashvili and Wooldridge (2016) propose an estimation method for estimating switching regression models with endogenous variables and switching. In the panel setting with fixed T, in order to allow heterogeneity to be correlated with time-varying explanatory variables, they must linearly model the unobserved heterogeneity using a linear Mundlak (1978) device from the time-averages of the instruments. Laage (2019) investigates the identification of APEs in CRC models also with fixed T panel data. She proposes using a two-step nonparametric estimator in which the first-stages are cross-sectionally estimated by time-period. Both approaches rely on homogeneous first-stage coefficients across individuals, thus ruling out the possibility of compliers, always-takers, and never-takers, usually associated with the LATE.

[7] In the simulation and application we will also consider pooled two-stage least squares and pooled two-stage least squares applied to first differenced data. Clearly, neither estimator avoids making a similar assumption for consistency in estimating the PAE. Given the structure of the data, it is difficult to think of a setting where either would be preferable to FEIV.

3 Proposed estimator

What we term the Per-Cluster Instrumental Variables approach takes much of the same form as the per-cluster approach in Bates et al. (2014), except we utilize the instrument $\mathbf{z_{ij}}$ to estimate the cluster-specific slopes, $\mathbf{b_i}$. Thus,

$$\hat{\mathbf{b}}_{\mathbf{i,PCIV}} = \boldsymbol{\beta} + \mathbf{d_i} + \left(\sum_{j=1}^{T} \mathbf{x'_{ij}} \mathbf{H_{z_i,j}} \mathbf{x_{ij}}\right)^{-1} \sum_{j=1}^{T} \mathbf{x'_{ij}} \mathbf{H_{z_i,j}} e_{ij}, \tag{8}$$

where $\mathbf{H_{z_i}} = \ddot{\mathbf{z}}_{\mathbf{ij}} \left(\sum_{j=1}^{T} \ddot{\mathbf{z}}'_{\mathbf{ij}} \ddot{\mathbf{z}}_{\mathbf{ij}}\right)^{-1} \ddot{\mathbf{z}}'_{\mathbf{ij}}$.

The key intuition here is that because $\mathbf{d_i}$ does not vary within the cluster, it must be mean independent of within-cluster *deviations* of $\mathbf{z_{ij}}$ and $\mathbf{x_{ij}}$ used in the per-cluster regressions, even if it is correlated with both $\mathbf{z_{ij}}$ and $\mathbf{x_{ij}}$. Maintaining random sampling and homogeneous cluster size, we then average over the estimated cluster-specific slopes to provide our PCIV estimate of $\boldsymbol{\beta}$.[8] Thus, our proposed PCIV estimator of $\boldsymbol{\beta}$ can be written as the following:

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{PCIV}} = \frac{1}{N}\sum_{i=1}^{N} \hat{\mathbf{b}}_{\mathbf{i,PCIV}} = \boldsymbol{\beta} + \frac{1}{N}\sum_{i=1}^{N} \mathbf{d_i} + \frac{1}{N}\sum_{i=1}^{N} \left[\left(\sum_{j=1}^{T} \mathbf{x'_{ij}} \mathbf{H_{z_i,j}} \mathbf{x_{ij}}\right)^{-1} \sum_{j=1}^{T} \mathbf{x'_{ij}} \mathbf{H_{z_i,j}} e_{ij}\right]. \tag{9}$$

3.1 Consistency

To find the conditions under which $\hat{\boldsymbol{\beta}}_{\boldsymbol{PCIV}}$ is a consistent estimate of $\boldsymbol{\beta}$, we first consider the ideal setting for this estimator; where both N and T are large. Such settings may have been previously thought to be prohibitive. However, such data is becoming more common from the rise of panels of the human genome to the monotonization of individuals' internet search histories. We take the probability limit of equation 9, as T and $N \to \infty$, providing the following:

$$\operatorname*{plim}_{T,\ N\to\infty} (\hat{\boldsymbol{\beta}}_{\boldsymbol{PCIV}} - \boldsymbol{\beta}) = \mathrm{E}(\mathbf{d_i}) + \mathrm{E}[[\mathrm{E}_i(\mathbf{x'_{ij}} \mathbf{H_{z_i}} \mathbf{x_{ij}})]^{-1} \mathrm{E}_i(\mathbf{x'_{ij}} \mathbf{H_{z_i}} e_{ij})]. \tag{10}$$

[8] In an extension in section 5.1, we discuss the use of weights to uncover the population average effects in the presence of unequal cluster sizes or nonrandom sampling.

Thus, consistency of $\hat{\boldsymbol{\beta}}_{\boldsymbol{PCIV}}$ follows from the assumptions enumerated below:

1. $\mathrm{E}(\mathbf{d_i}) = 0$ by definition,
2. $rank[\mathrm{E}_i(\mathbf{z'_{ij}x_{ij}})] = K$,
3. $rank[\mathrm{E}_i(\mathbf{z'_{ij}z_{ij}})] = L$,
4. $\mathrm{E}_i(\mathbf{z_{ij}}'e_{ij}) = 0$ (The validity of the instrument within each cluster).[9]

The condition that $rank[\mathrm{E}_i(\mathbf{z'_{ij}x_{ij}})] = K$ is meaningful, as it requires, 1) variation in the instrument and endogenous regressor within cluster, and 2) all clusters to be influenced by the instrument. However, whether the sample analogue of this condition is met for each cluster will be apparent when performing per-cluster estimation. Additional assumptions would be required to project out of the sample for which the instrument is relevant. Additionally, unlike FEIV, the PCIV estimator may provide a consistent estimate of $\boldsymbol{\beta}$ for those who comply with the instrument, even when the strength of the instrument is related to the heterogeneous effects.

3.2 Inference

In this section, we derive asymptotic variance of PCIV estimator for inference and describe its estimation. Since we first estimate cluster-level coefficients in order to obtain an estimate of the global-level estimate, $\hat{\boldsymbol{\beta}}_{\boldsymbol{PCIV}}$, we need to get correct standard errors for possible early-stage estimation error. Thus, we need to show the asymptotic variance at each cluster level.

Note that we may write the asymptotic variance of each cluster-level coefficient according

[9]With aggregate time variables included in $\mathbf{x_{ij}}$ and $\mathbf{z_{ij}}$ this is no more restrictive than the exogeneity assumption maintained in Murtazashvili and Wooldridge (2008).

to the following as $T \to \infty$:

$$
\begin{aligned}
&T \cdot \mathrm{V}(\hat{\boldsymbol{b}}_{\boldsymbol{i,PCIV}} - \boldsymbol{b}_{\boldsymbol{i,PCIV}}) \\
=&T \cdot \left(\sum_{j=1}^{T} \mathbf{x'_{ij} H_{z_i} x_{ij}} \right)^{-1} \left(\sum_{j=1}^{T} \mathbf{x'_{ij} H^2_{z_i} x_{ij}} e_{ij}^2 \right) \left(\sum_{j=1}^{T} \mathbf{x'_{ij} H_{z_i} x_{ij}} \right)^{-1} \\
=&\left(\frac{1}{T} \sum_{j=1}^{T} \mathbf{x'_{ij} H_{z_i} x_{ij}} \right)^{-1} \left(\frac{1}{T} \sum_{j=1}^{T} \mathbf{x'_{ij} H^2_{z_i} x_{ij}} e_{ij}^2 \right) \left(\frac{1}{T} \sum_{j=1}^{T} \mathbf{x'_{ij} H_{z_i} x_{ij}} \right)^{-1}
\end{aligned} \tag{11}
$$

We use the cluster-level asymptotic variance to obtain the asymptotic variance of the estimated population-level parameter, $\hat{\boldsymbol{\beta}}_{\boldsymbol{PCIV}}$, as follows.

$$
\begin{aligned}
&NT \cdot \mathrm{V}(\hat{\boldsymbol{\beta}}_{\boldsymbol{PCIV}} - \boldsymbol{\beta}) \\
=&NT \cdot \mathrm{V} \left(\frac{1}{N} \sum_{i=1}^{N} (\boldsymbol{b_i} - \boldsymbol{\beta}) + \frac{1}{N} \sum_{i=1}^{N} (\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b_i}) \right) \\
=&NT \cdot \mathrm{V} \left(\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{d_i} + \frac{1}{N} \sum_{i=1}^{N} (\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b_i}) \right) \\
=&T \cdot \boldsymbol{dd'} + \frac{1}{N} \cdot \sum_{i=1}^{N} \left[\left(\frac{1}{T} \sum_{j=1}^{T} \mathbf{x'_{ij} H_{z_i} x_{ij}} \right)^{-1} \left(\frac{1}{T} \sum_{j=1}^{T} \mathbf{x'_{ij} H^2_{z_i} x_{ij}} e_{ij}^2 \right) \left(\frac{1}{T} \sum_{j=1}^{T} \mathbf{x'_{ij} H_{z_i} x_{ij}} \right)^{-1} \right] \\
&+ \frac{2T}{N} \cdot \sum_{i=1}^{N} \left[\left(\frac{1}{T} \sum_{j=1}^{T} \mathbf{x'_{ij} H_{z_i} x_{ij}} \right)^{-1} \left(\frac{1}{T} \sum_{j=1}^{T} \boldsymbol{d_i} \mathbf{x'_{ij} H_{z_i}} e_{ij} \right) \right] \\
&\longrightarrow T \cdot \sigma_d^2 + Q_1^{-1} \Omega Q_1^{-1} + 2T \cdot Q_1^{-1} Q_3 \quad \text{as } N, T \to \infty,
\end{aligned}
$$

where $Q_1 = \mathrm{E}[\mathbf{x'_{ij} H_{z_i} x_{ij}}], \Omega = \mathrm{E}[\mathbf{x'_{ij} H^2_{z_i} x_{ij}} e_{ij}^2]$, and $Q_3 = \mathrm{E}[\boldsymbol{d_i} \mathbf{x'_{ij} H_{z_i}} e_{ij}]$. Using the obtained asymptotic variance, we can construct a t-statistic within the sample. As the asymptotic variance depends upon $\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b_i} = \left(\sum_{j=1}^{T} \mathbf{x'_{ij} H_{z_i} x_{ij}} \right)^{-1} \left(\sum_{j=1}^{T} \mathbf{x'_{ij} H_{z_i}} e_{ij} \right)$, in the estimated variance we replace e_{ij} with the sample analogue $\hat{e}_{ij}$. Thus, the sample variance is constructed as follows in equation 12.

$$
\begin{aligned}
&\hat{\mathrm{V}}(\hat{\boldsymbol{\beta}}_{\boldsymbol{PCIV}} - \boldsymbol{\beta}) \\
=&\mathrm{V}\left(\frac{1}{N}\sum_{i=1}^{N}\left[\hat{\boldsymbol{d}}_{\boldsymbol{i}} + \left(\sum_{j=1}^{T}\mathbf{x'_{ij}H_{z_i}}e_{ij}\right)\right]\right) \\
=&\frac{1}{N}\sum_{i=1}^{N}(\hat{\boldsymbol{d}}_{\boldsymbol{i}} - \bar{\hat{\boldsymbol{d}}}_{\boldsymbol{i}})^{\mathbf{2}} + \frac{1}{NT}\cdot\left[\sum_{i=1}^{N}\left(\mathbf{x'_{ij}H_{z_i}x_{ij}}\right)^{-1}\left(\frac{1}{T}\sum_{j=1}^{T}\mathbf{x'_{ij}H^2_{z_i}x_{ij}}\hat{e}^2_{ij}\right)\left(\frac{1}{T}\sum_{j=1}^{T}\mathbf{x'_{ij}H_{z_i}x_{ij}}\right)^{-1}\right] \\
&+\frac{2}{N}\cdot\sum_{i=1}^{N}\left[\left(\frac{1}{T}\sum_{j=1}^{T}\mathbf{x'_{ij}H_{z_i}x_{ij}}\right)^{-1}\left(\frac{1}{T}\sum_{j=1}^{T}\hat{\boldsymbol{d}}_{\boldsymbol{i}}\mathbf{x'_{ij}H_{z_i}}\hat{e}_{ij}\right)\right],
\end{aligned}
\tag{12}
$$

where $\hat{\boldsymbol{d}}_{\boldsymbol{i}} = \hat{\boldsymbol{b}}_{\boldsymbol{i}} - \hat{\boldsymbol{\beta}}_{\boldsymbol{PCIV}}$ and $\hat{e}_{ij} = y_{ij} - \boldsymbol{x_{ij}}\hat{\boldsymbol{b}}_{\boldsymbol{i}}$. Throughout this paper, we will apply this estimated variance in the simulation as well as application studies. The consistency of the estimated variance is shown in Appendix A.

For comparison, we use a bootstrap procedure as an alternative to obtain the standard errors for PCIV estimator. Let $C_i = \{\mathbf{y_{ij}}, \mathbf{x_{ij}}, \mathbf{z_{ij}}\}_{j=1}^{T}$, which is a set of observations for each cluster, and $C = \{C_i\}_{i=1}^{N}$. First, we resample N number of C_is from C with replacement, then having a new set of $C_i^* = \{\mathbf{y^*_{ij}}, \mathbf{x^*_{ij}}, \mathbf{z^*_{ij}}\}_{j=1}^{T}$. Note that we do not resample the data within the cluster. By running the regression of $y^*_{ij} = \mathbf{x^*_{ij}b_i} + e_{ij}$, we can obtain the set of $\{\hat{\boldsymbol{\beta}}^*_{\boldsymbol{PCIV,b}}\}_{b=1}^{B}$, where B is the number of bootstrap repetitions. The bootstrap standard errors are given as follows.

$$
s.e.(\hat{\boldsymbol{\beta}}^*_{\boldsymbol{PCIV}}) = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}(\hat{\boldsymbol{\beta}}^*_{\boldsymbol{PCIV,b}} - \bar{\hat{\boldsymbol{\beta}}}^*_{\boldsymbol{PCIV,b}})^2}, \text{where } \bar{\hat{\boldsymbol{\beta}}}^* = \frac{1}{B}\sum_{b=1}^{B}\hat{\boldsymbol{\beta}}^*_{\boldsymbol{PCIV,b}}. \tag{13}
$$

3.3 Finite samples in one dimension

While very large data sets are becoming more common place, in many applications, researchers may not wish to rely on asymptotic arguments with respect to both dimensions of their data. Consequently, we also consider the properties of our estimator with fixed N and T $\rightarrow \infty$, before considering the probability limit more comparable to FEIV with fixed T and N $\rightarrow \infty$.

We start with the number of clusters being fixed while the number of observations per cluster can be very large. While in principal this could be true with panel data, this setting is perhaps more commonplace for clustered cross-sectional data. In such applications, the asymptotic argument makes more sense through the number of observations per-cluster. Accordingly, we take the probability limit of equation 9 as T $\rightarrow \infty$, providing the following:

$$\operatorname*{plim}_{T\rightarrow\infty}(\hat{\boldsymbol{\beta}}_{\boldsymbol{PCIV}}-\boldsymbol{\beta}) = \frac{1}{N}\sum_{i=1}^{N}\mathbf{d_i} + \frac{1}{N}\sum_{i=1}^{N}\left[[\mathrm{E}_i(\mathbf{x'_{ij}H_{z_i}x_{ij}})]^{-1}\mathrm{E}_i[\mathbf{x'_{ij}H_{z_i}}e_{ij}]\right]. \tag{14}$$

The previously stated rank and validity assumptions applied to each cluster ensure that $\frac{1}{N}\sum_{i=1}^{N}[\mathrm{E}_i(\mathbf{x'_{ij}H_{z_i}x_{ij}})]^{-1}\mathrm{E}_i[\mathbf{x'_{ij}H_{z_i}}e_{ij}] = 0$. If researchers observe all clusters or a random sample of clusters in the data, in expectation, $\mathrm{E}(\frac{1}{N}\sum_{i=1}^{N}\mathbf{d_i}) = 0$.[10] Thus, the PCIV approach is asymptotically unbiased in estimating the PAE, though consistency requires $N \rightarrow \infty$.

Next, we consider the robustness of the PCIV approach with fixed T as $N \rightarrow \infty$. Taking the probability limit of equation 9 with N $\rightarrow \infty$ provides the following:

$$\operatorname*{plim}_{N\rightarrow\infty}(\hat{\boldsymbol{\beta}}_{\boldsymbol{PCIV}}-\boldsymbol{\beta}) = \mathrm{E}[\mathbf{d_i}] + \mathrm{E}\left[\left(\sum_{j=1}^{T}\mathbf{x'_{ij}H_{z_i}x_{ij}}\right)^{-1}\sum_{j=1}^{T}\mathbf{x'_{ij}H_{z_i}}e_{ij}\right]. \tag{15}$$

By definition, $\mathrm{E}(\mathbf{d_i}) = 0$. Thus, we must assume $\mathrm{E}[(\sum_{j=1}^{T}\mathbf{x'_{ij}H_{z_i}x_{ij}})^{-1}\sum_{j=1}^{T}\mathbf{x'_{ij}H_{z_i}}e_{ij}] = 0$ in order for $\hat{\boldsymbol{\beta}}_{\boldsymbol{PCIV}}$ to consistently estimate the PAE, $\boldsymbol{\beta}$. With a fixed number of observations per cluster, additional complications arise. First, T must be large enough ($T > K$) to estimate b_{ri}, where $r = 1, ..., K$. Second, in order for this estimator to provide a consistent estimate of $\boldsymbol{\beta}$ without further assumptions, as with the large T case, there must be variation in the instrument, and non-zero covariance between the instrument and endogenous regressors.

Lastly, unlike in the exogenous case presented in Bates et al. (2014), each estimated b_{ri} is

[10]The same holds if researchers observe a random sample of the population, which is nested into clusters.

bound to manifest some degree of finite sample bias. The question is whether we may expect the biases to be mean zero in expectation as only the number of clusters approaches infinity. Bound et al. (1995) summarizes some instances in which we may expect that not to be the case. First, the finite sample bias falls as the ratio of observations per cluster to regressors (and instruments) grows. Second, with weak instruments the the finite sample bias is in the direction of the OLS estimates, violating the mean zero bias condition.[11]

As a result, one of the functions of the Monte Carlo study will be to show how the per-cluster instrumental variables estimator performs against P2SLS and standard FEIV as we vary the number of observations per cluster in Section 4.

4 Simulation Study

We now describe the simulation we use to examine the performance of the P2SLS, FEIV, and PCIV estimators in finite samples. In conducting this simulation study, we consider two types of conditions: where uncorrelated covariance assumption holds, or is violated. We are interested in the bias, efficiency, and asymptotic risk of the the three estimators. We also examine the performance of the analytic standard errors relative to the standard deviation of the bootstrapped estimates. We consider each performance measure under both conditions for a range of cluster sizes and number of clusters.[12]

4.1 Data Generating Process

We generate the data based on our model of interest in Equation 1. Our random intercept d_{0i} and random slopes d_{1i} are drawn from a multivariate normal distribution with zero means and variance-covariance matrix defined by variances $\psi_0 = 0.4^2$, $\psi_1 = 0.25^2$, a correlation of $\rho = 0.5$ giving a covariance of $\psi_{01} = 0.05$.

We generate two types of y_{ij} based on Equation 1 with $\beta_1 = 1$. In the first case, we generate y_{1ij} under uncorrelated covariance assumption holding, and in the second we generate

[11] Bound et al. (1995) results may not directly translate as their discussion focuses on cases where $L-K > 1$.

[12] We use Stata 15 (Stata Corp, 2015) throughout.

y_{2ij} under the violation of the uncorrelated covariance assumption.

$$y_{1ij} = d_{0i} + (\beta_1 + d_{1i})x_{1ij} + o_{ij} + v_{ij}, \qquad v_{1ij} \sim N(0, \sigma_v^2). \tag{16}$$

$$y_{2ij} = d_{0i} + (\beta_1 + d_{2i})x_{2ij} + o_{ij} + v_{2ij}, \qquad v_{2ij} \sim N(0, \sigma_v^2), \tag{17}$$

where v_{1ij} and v_{2ij} are cluster and time varying errors with mean zero and variances of σ_v^2. We generate two types of the individual-level covariate x_{1ij} and x_{2ij} in the following way:

$$x_{1ij} = b_0 d_{0i} + b_1 d_{1i} + b_2 w_j + z_{ij} + a\epsilon_{1ij}, \qquad \epsilon_{1ij} \sim N(0, \sigma_{Xi}^2), \tag{18}$$

$$x_{2ij} = b_0 d_{0i} + b_1 d_{2i} + b_2 w_j + z_{ij} + a\epsilon_{2ij}, \qquad \epsilon_{2ij} \sim N(0, \sigma_{Xi}^2), \tag{19}$$

where w_i is an exogenous cluster level error, ϵ_{ij} is a cluster and time varying error, and

$$a = \sqrt{1 - \psi_0^2 b_0^2 - \psi_1^2 b_1^2 - \sigma_w^2 b_2^2 - 2b_0 b_1 \psi_{01}}. \tag{20}$$

In this set up, coefficients take arbitrary values $b_0 = 1.33$, $b_1 = 2.13$, and $b_2 = 0.20$, which makes both x_{1ij} and x_{2ij} correlated with the random intercept, random slope, and cluster-level covariate. We draw z_{ij} from a normal distribution with mean of zero and variance of σ_{Zi}^2. Transitory endogeneity of our key variables of interest, x_{1ij} and x_{2ij}, enters through the presence of o_{ij} in the generation of y_{1ij} and y_{2ij}, but we imagine is unobserved in the data. We hold that o_{ij} is independent of z_{ij}, but correlated with both x_{1ij} and x_{2ij}.

The key assumption for the consistency of FEIV is that the within-cluster covariance between x_{ij} and z_{ij} is uncorrelated with the random slopes d_i. Accordingly, for uncorrelated covariance, there is no correlation between the heterogeneous slopes d_{1i} and the variances of x_{1ij} and z_{ij}. In contrast, we generate a violation of this condition by having the variances of x_{2ij} and z_{ij} related to d_{2i} with $\sigma_{Zi} = \sigma_{Xi} = \exp(d_{2j})$.

Table 1: Simulated correlations with and without correlated covariance between d and $\ddot{x}\ddot{z}$

Panel A: Uncorrelated Covariance						
	d_0	d_1	x_1	o_1	z	$\ddot{x}_1\ddot{z}$
d_0	1					
d_1	0.0103	1				
x_1	0.3906	0.3633	1			
o_1	0.4012	0.3744	0.3848	1		
z	0.0006	0.0004	0.7886	0.0064	1	
$\ddot{x}_1\ddot{z}$	0.1269	0.0061	0.0532	0.0508	0.0044	1
Panel B: Correlated Covariance						
	d_0	d_2	x_2	o_2	z	$\ddot{x}_2\ddot{z}$
d_0	1					
d_2	0.477	1				
x_2	0.5379	0.5458	1			
o_2	0.5516	0.5603	0.4814	1		
z	0.0006	0.0021	0.7258	0.0073	1	
$\ddot{x}_2\ddot{z}$	0.1269	0.2776	0.1543	0.1573	0.0044	1

Note: $\ddot{x}_1\ddot{z}$ and $\ddot{x}_2\ddot{z}$ stands for the product of the time-demeaned variables of interest and the instrument when the key condition holds and is violated respectively.

The resulting correlations with $N = 250$ and $T = 250$ are shown in Table 1. Under both conditions x is correlated with unobserved heterogeneity d_0, d_1, and d_2 as well as the omitted variable o producing endogeneity. Our variable of interest, x, is also strongly correlated with the instrument z, and z is otherwise orthogonal to the other terms, such that our instrument is both relevant and valid. The key condition hinges on whether the heterogeneous slopes, d_1 or d_2, are correlated with the strength of the instrument within cluster. The fact that the correlations between d_1 and $\ddot{x}_1\ddot{z}$ and between d_2 and $\ddot{x}_2\ddot{z}$ are 0.0061 and 0.2776 respectively reflect two possible states of the key condition.

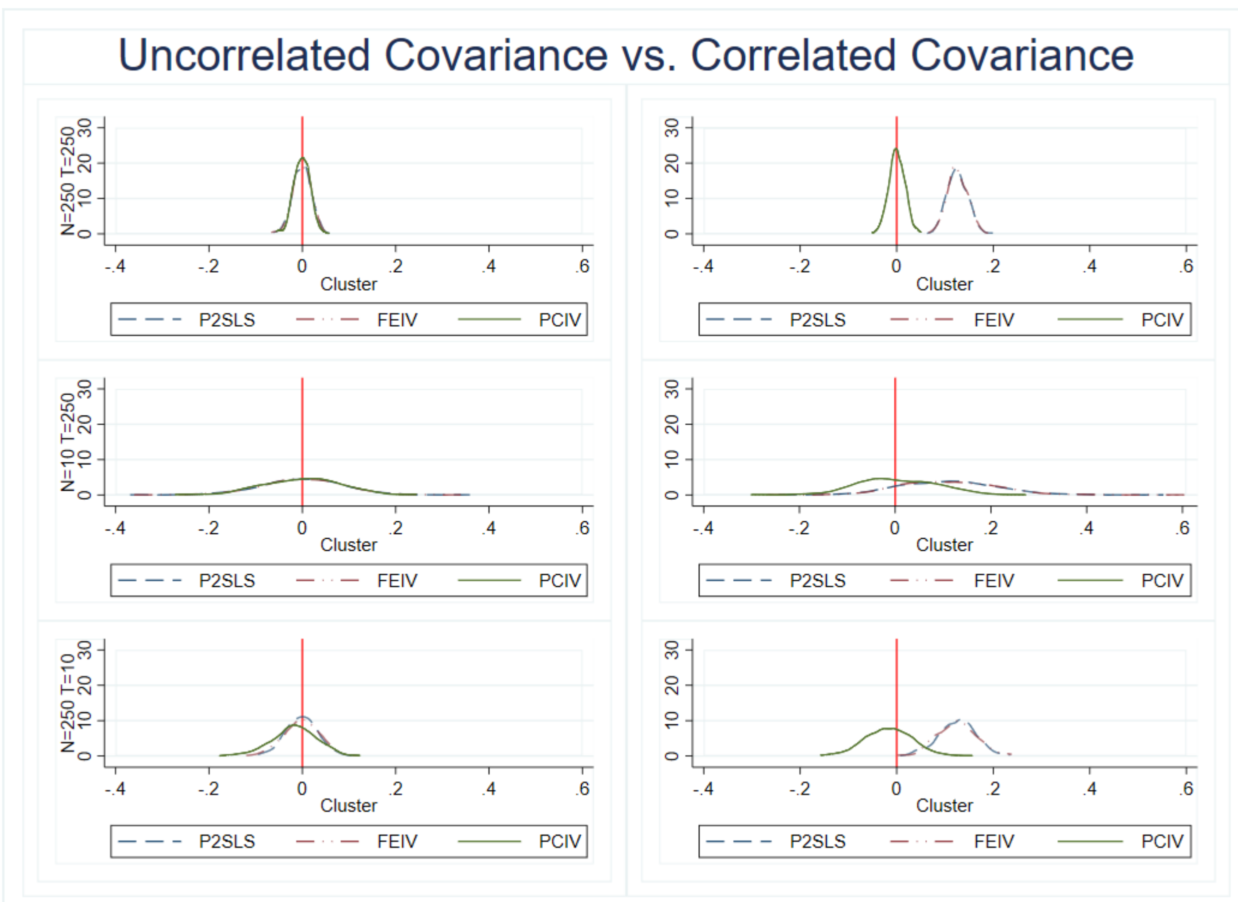
Each simulation is repeated 500 times. Because the asymptotic properties depend on whether T, N, or both go to infinity, we vary both the number of clusters from 6 to 250 and the number of observations per cluster from 6 to 250.

4.2 Results

We evaluate the performance of each method—P2SLS, FEIV, and PCIV—with respect to the bias, asymptotic risk as measured by root mean square error (RMSE), the ratio of mean standard errors by the standard deviations of simulated estimates, and the coverage rate from each approach.

4.2.1 Bias

Figure 1: Kernel Density Plots of Estimation Errors, $\hat{\beta}_1 - \beta_1$

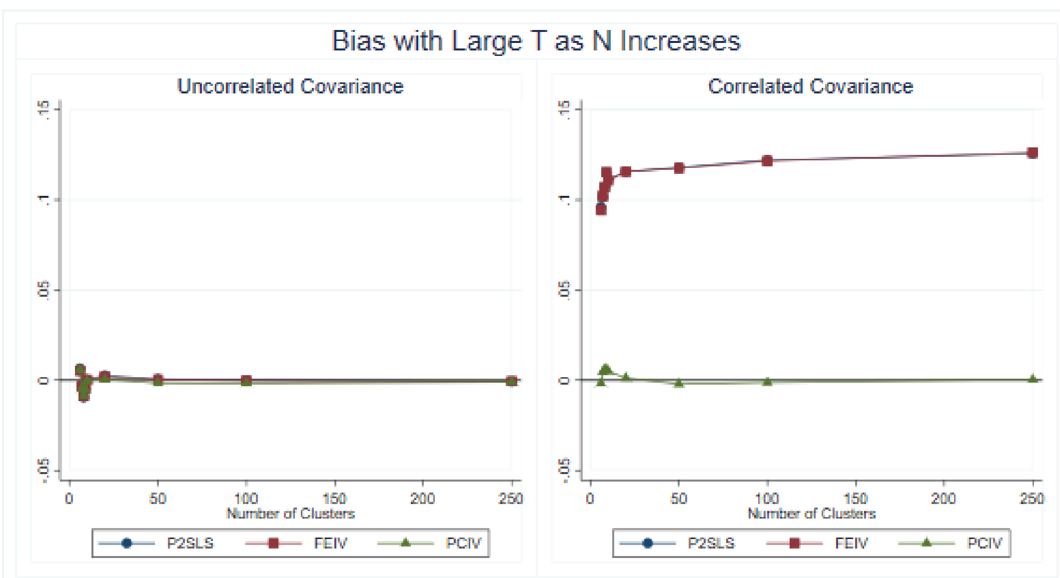


Note: $\hat{\beta}_1$ is the coefficient of x_{ij} across replications for all methods. Left panels are when the uncorrelated covariance assumption holds and right panels are when it is violated. P2SLS = Two-Stage Least Squares; FEIV= Fixed Effects Instrumental Variable; PCIV=Per-Cluster Instrumental Variable.

As the primary potential benefit of PCIV is robustness, we first consider the estimated bias for the coefficient $\hat{\beta}_1$ to assess each estimator's performance. Figure 1 shows the kernel density plot of the bias over each simulation, maintaining the uncorrelated covariance assumption on the left and violating it on the right. Further, we show these density plots first with 250 clusters and 250 observations per cluster at the top, then with only the number of clusters reduced to 10, and finally with 250 clusters and 10 observations per cluster at the bottom.

Figure 1 shows that the distribution of estimation bias with each method tightly centers around zero, when the heterogeneous slopes are uncorrelated with the strength of the instrument and both N and T are large. Naturally, the distribution of estimates spreads with smaller N or T. PCIV is less precise than FEIV or P2SLS in these smaller samples.

Figure 2: Estimated Bias for Coefficient $\beta_1$of x_{ij} Versus number of clusters



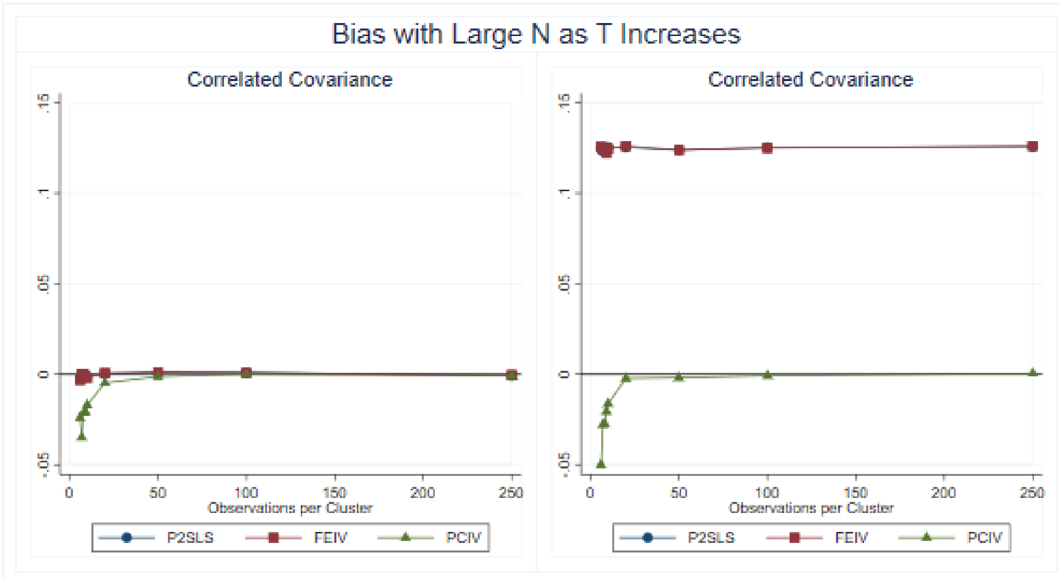
Note: Left panels are when the uncorrelated covariance assumption holds and right panels are when it is violated. P2SLS = Two-Stage Least Squares; FEIV= Fixed Effects Instrumental Variable; PCIV=Per-Cluster Instrumental Variable. Each point represents results from 500 repetitions.

More substantial differences become apparent with a violation of the uncorrelated covariance assumption. With correlation between the strength of the instrument and the heterogeneous slopes, the precision of each of the three estimators falls. However, with large N and T, the distribution of bias in the PCIV estimator remains centered tightly around zero, whereas the the entire distribution of estimates from P2SLS and FEIV lies strictly to the right of zero. With smaller N or T the distributions of estimates overlap (substantially when $N = 10$ and $T = 250$), however only the distribution of PCIV remains centered near zero bias. In contrast, both FEIV and P2SLS are biased when the uncorrelated covariance assumption is violated. From Table 6 in the appendix, the magnitude of bias in both FEIV and P2SLS is almost identical at approximately 12%.

Figure 2 shows the average bias in the three estimators with T fixed at 250 as N increases, again both maintaining and violating the key assumption. With T=250 and the uncorrelated covariance assumption holding, all three estimators show very little mean bias across all numbers of clusters. However, once the uncorrelated covariance assumption is violated, both FEIV and P2SLS demonstrate consistent and significant bias, whereas the PCIV shows little mean bias in its estimates.

Figure 3 repeats the same exercise with N fixed at 250 as T varies. Again, when the uncorrelated covariance assumption holds, FEIV and P2SLS estimators perform consistently well, but demonstrate consistent and significant bias once the uncorrelated covariance assumption is violated. In contrast, the PCIV estimator manifests finite sample bias under both conditions with very small (sized 10 or fewer) clusters. However, across cluster sizes PCIV outperforms the other estimators when the uncorrelated covariance assumption is violated. Furthermore, while estimated bias of FEIV and P2SLS estimators do not show any improvement even with the increasing size of cluster, the estimated bias of the PCIV approach gets closer to zero as cluster sizes increase.

Figure 3: Estimated Bias for Coefficient $\beta_1$of x_{ij} Versus Cluster Size

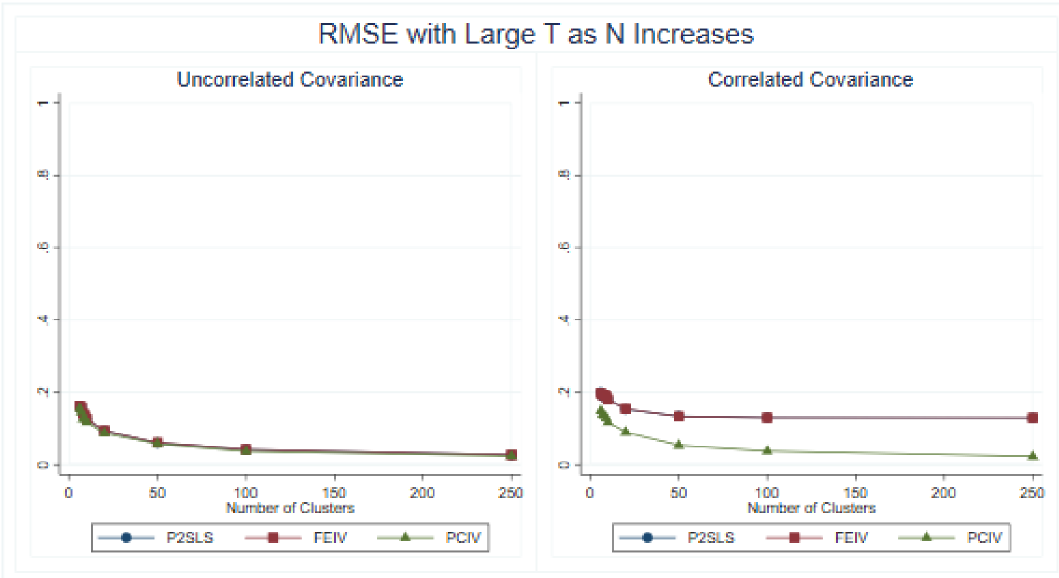


Note: Left panels are when the uncorrelated covariance assumption holds and right panels are when it is violated. P2SLS = Two-Stage Least Squares; FEIV= Fixed Effects Instrumental Variable; PCIV=Per-Cluster Instrumental Variable. Each point represents results from 500 repetitions.

4.2.2 Root mean squared error

Naturally, researchers are not only interested in the bias of estimators, but are also interested in the estimators' precision. As is typical, we use root mean squared errors (RMSEs) which comprise bias and imprecision as a summative measure of performance on both dimensions. As the scale of RMSEs depends on both cluster size and the number of clusters present, we show the RMSE of the estimated coefficient from the three estimators, first, with T=250 and the number of clusters varying between 6 and 250 in Figure 4, and then, with N=250 and cluster sizes varyingin Figure 5. In both figures we repeat the exercise both with and without the uncorrelated covariance assumption holding.

Figure 4: Estimated Root Mean Square Error for Coefficient $\hat{\beta}_1$ of x_{ij} Versus Number of Clusters

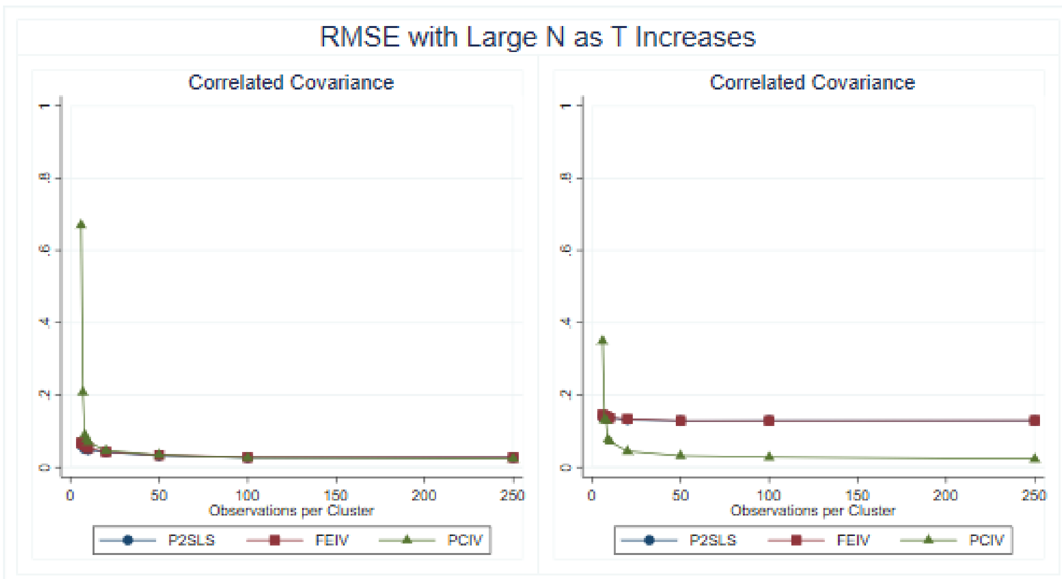


Note: Left panels are when the uncorrelated covariance assumption holds and right panels are when it is violated. P2SLS = Two-Stage Least Squares; FEIV=Fixed Effects Instrumental Variable; PCIV=Per-Cluster Instrumental Variable. Each point represents results from 500 repetitions.

Figure 4 shows that with a large T, all three estimators have similar RSMEs across number of clusters when the uncorrelated covariance assumption holds. In which case, PCIV exhibits only slightly smaller RMSE than FEIV or P2SLS. In the presence of correlation between the strength of the instrument and the heterogeneous effects exists PCIV exhibits smaller RMSE than does either P2SLS or FEIV across all number of clusters considered, indicating that with large number of observations per cluster, PCIV may dominate more standard approaches.

Turning to Figure 5, the first striking pattern is that with very small clusters and $N = 250$, the PCIV approach is prone to large RMSE. However, in this simulation by a cluster size of only 8, the PCIV approach has comparable or lower RMSE than FEIV or P2SLS. We were somewhat surprised by the relative performance of the PCIV estimator at such small cluster sizes as Staiger and Stock (1997) state that the asymptotic distributions provide good approximations on sampling distributions with 10 - 20 observations per instrument.

Figure 5: Estimated Root Mean Square Error for Coefficient $\hat{\beta}_1$ of x_{ij} Versus Cluster Size



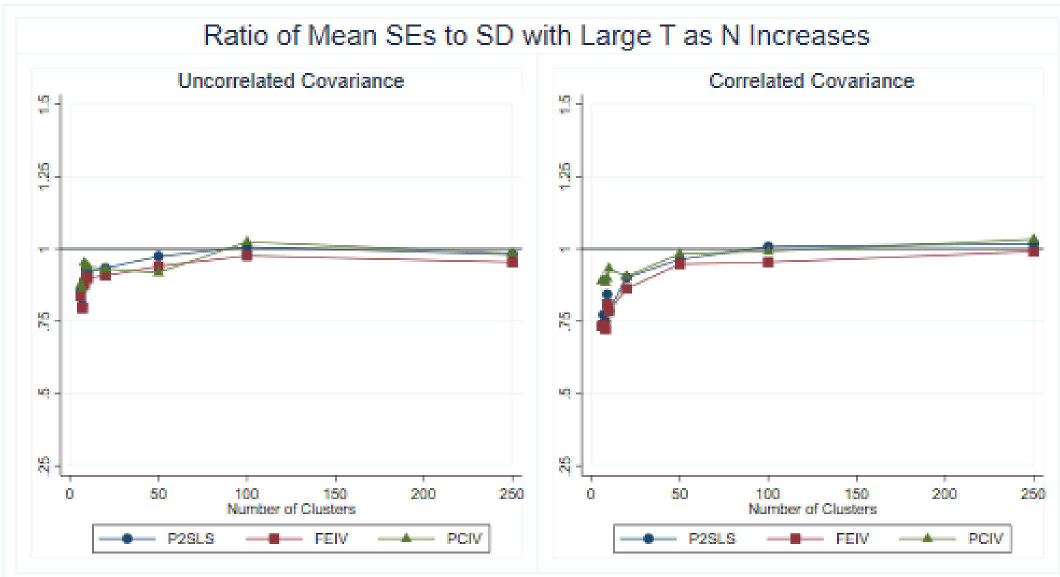
Note: Left panels are when the uncorrelated covariance assumption holds and right panels are when it is violated. P2SLS = Two-Stage Least Squares; FEIV=Fixed Effects Instrumental Variable; PCIV=Per-Cluster Instrumental Variable. Each point represents results from 500 repetitions.

4.2.3 Ratio of Mean SE by SD and Overage Rates

We evaluate of the performance of our analytic standard errors over both numbers of clusters used and size of each cluster. We do so first by depicting the ratio of the mean of the estimated SEs divided by sampling SDs. Figure 6 presents how this ratio changes with the number of clusters and Figure 7 presents the same measure over changes in cluster size. Again the case in which the strength of the instrument is uncorrelated with the heterogeneous effects appears on the left, whereas the case of correlated covariance appears on the right.

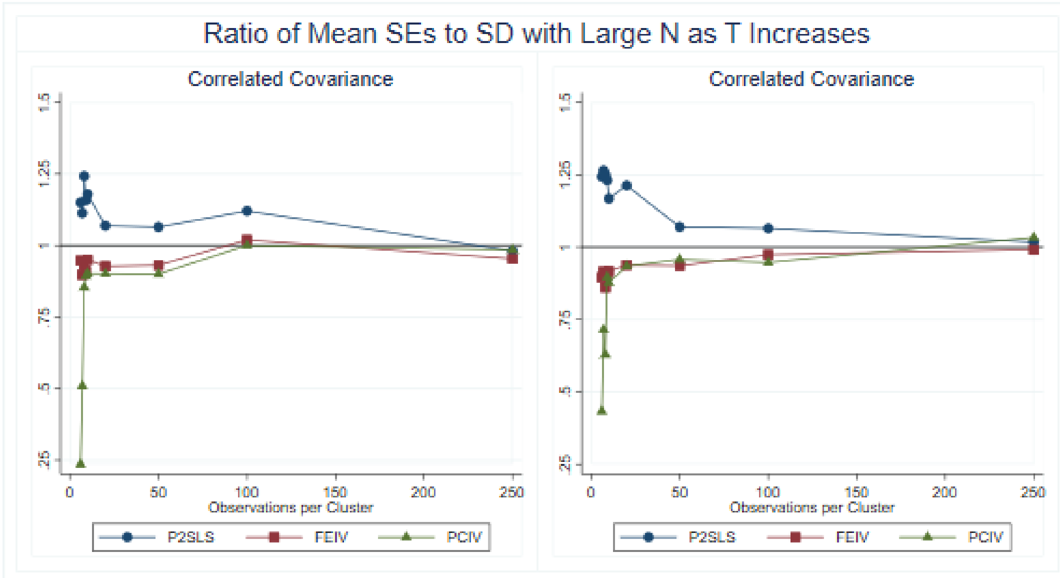
In all cases, the analytic standard errors do reasonably well with large sample sizes along both dimension. The ratio is consistently close to 1 with N and T both above 20 both with the the uncorrelated covariance assumption holding and when it is violated violated.

Figure 6: Ratio of Mean SEs Divided by SDs of Estimates Versus Number of Clusters



Note: Ratio of mean standard errors (SEs) divided by standard deviations (SDs) of the estimates. Left panels are when the uncorrelated covariance assumption holds and right panels are when it is violated. P2SLS = Two-Stage Least Squares; FEIV = Fixed Effects Instrumental Variable; PCIV=Per-Cluster Instrumental Variable. Each point represents results from 500 repetitions.

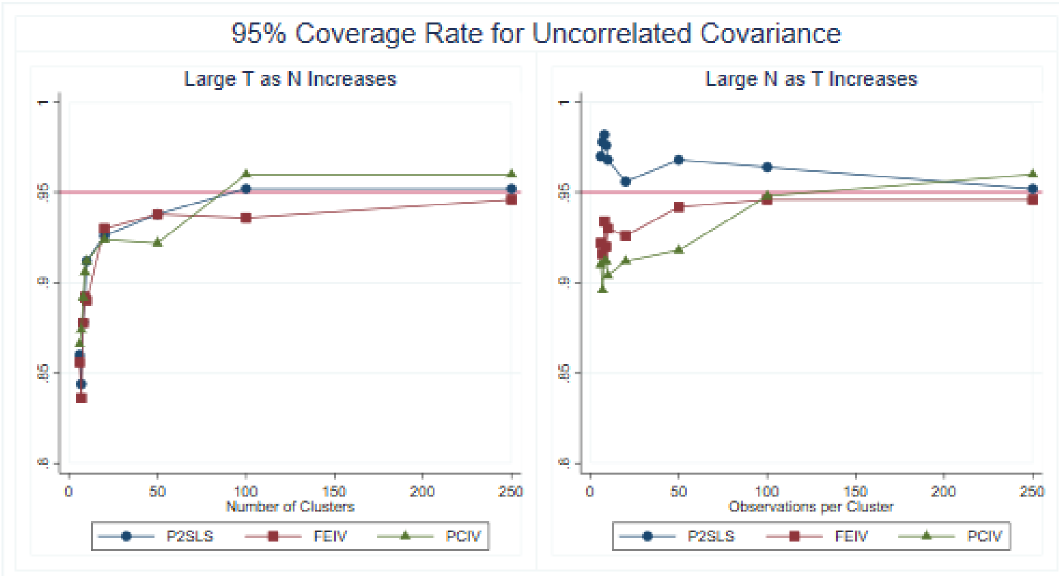
Figure 7: Ratio of Mean SEs Divided by SDs of Estimates Versus Cluster Size



Note: Ratio of mean standard errors (SEs) divided by standard deviations (SDs) of the estimates. Left panels are when the uncorrelated covariance assumption holds and right panels are when it is violated. P2SLS = Two-Stage Least Squares; FEIV= Fixed Effects Instrumental Variable; PCIV=Per-Cluster Instrumental Variable. Each point represents results from 500 repetitions.

Secondly, we show the rate at which the 95 percent confidence interval constructed from our estimated standard errors include the true value of the parameter. Naturally, this should occur 95 percent of the time. In figure 8, we show the evolution of coverage rates for each estimator in the uncorrelated covariance case with 250 clusters as T grows on the left hand side, and with cluster sizes of 250 as N grows on the right hand side. With 250 clusters, the coverage rates for all three estimators range from 0.89 to 0.98 for all cluster sizes, though they converge to 0.95 as the cluster size grows. With 250 observations per cluster, all three estimators reject the true value at a higher rate, but quickly converge to a 95 percent coverage rate as the number of clusters grow.

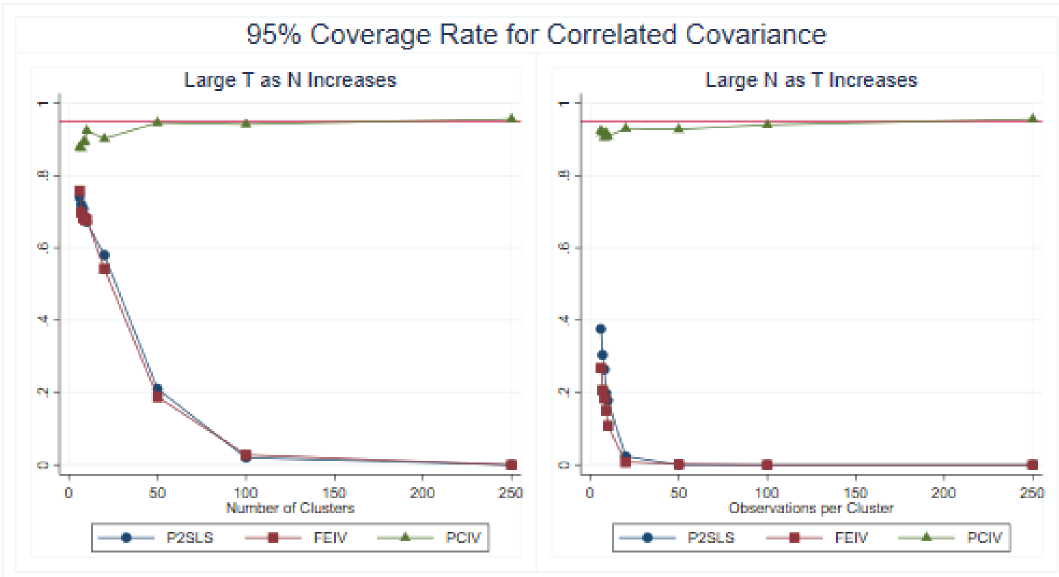
Figure 8: Coverage Rate for the Uncorrelated Case



Note: The horizontal line is to denote the exact 95% coverage rate. Left panels are when the uncorrelated covariance assumption holds and right panels are when it is violated. P2SLS = Two-Stage Least Squares; FEIV= Fixed Effects Instrumental Variable; PCIV=Per-Cluster Instrumental Variable. Each point represents results from 500 repetitions.

In figure 9, we show the evolution of coverage rates for each estimator in the case where the strength of the instrument may be correlated with the heterogeneous coefficients, again with 250 clusters as cluster size grows on the left hand side, and with cluster sizes of 250 as the number of clusters growing on the right hand side. Figure 9 shows that the bias in FEIV and P2SLS is meaningful as each of two estimators reject the true value more than 20 percent of the time. Moreover, the rejection rate of the true parameter grows as the number of clusters and the cluster sizes grow. In contrast, the PCIV rejection rate at the 95 percent confidence level is never more than 6 percentage points from 95 percent and again, it converges to 95 percent as the number of clusters or cluster sizes grow. The relatively poorer performance when either N or T is very small may give reason to researchers to adopt a bootstrap approach to standard error estimation. In the application of these methods to real data we initially show the estimated standard errors with both approaches.

Figure 9: Coverage Rate for the Correlated Case



Note: The horizontal line is to denote the exact 95% coverage rate. Left panels are when the uncorrelated covariance assumption holds and right panels are when it is violated. P2SLS = Two-Stage Least Squares; FEIV = Fixed Effects Instrumental Variable; PCIV=Per-Cluster Instrumental Variable. Each point represents results from 500 repetitions.

5 Extensions

5.1 Defining the population and weighting

In the discussion of uncovering the population average effects we have so far neglected to define the population of interest. In the panel data setting, clusters represent individuals from the population of interest about whom we have multiple observations. Random sampling of these individuals is an important assumption underpinning the above results. However, many popular panel data sets are not random samples of the population. For instance, the Panel Survey of Income Dynamics oversamples low-income families and the National Longitudinal Survey of Youth oversamples African American, Hispanic or Latino, military, and economically disadvantaged youth.

Such nonrandom sampling schemes would in turn lead averaged effects to overweight the populations which are overrepresented, obscuring the true PAE. Solon et al. (2015) discusses overcoming such nonrandom sampling as one possible rationale for when empirical researchers should use weights. Fortunately, researchers may still uncover PAEs using per-cluster approaches with relative ease. In this case the PAE can be identified using the inverse of the probability of selection as a cluster-level weight. By finding the weighted average of the estimated cluster-specific effects in the final step using these weights, researchers can still uncover a consistent estimate of the PAE.

A second issue related to sample-weighting arises in applications where clusters vary in size. This may be the case in settings where we have grouped cross-sectional data or where we have state-level panels as in our application below. In these settings, the population of interest may be either the clusters or the individuals nested within clusters. When clusters form the population of interest, we have a similar situation to the panel data setting. However, if we are instead interested in the populations of individuals, then we must consider the size of each cluster in the population.

Efficiency may provide a third rationale for weighting. With a burdensome set of assumptions including that the variables of interest and the random coefficients are exogenous, Swamy (1971) proposes a consistent and asymptotically efficient estimator of the PAE given below.

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{s}} = \sum_{i=1}^{N} \widetilde{w}_i \hat{\mathbf{b}}_{\mathbf{i}}, \tag{21}$$

where, $\widetilde{w}_i = \left(\sum_{j=1}^{N} \left[\hat{\Delta} + s_{jj}(\mathbf{x_j}'\mathbf{x_j})^{-1} \right]^{-1} \right) \left[\hat{\Delta} + s_{ii}(\mathbf{x_i}'\mathbf{x_i})^{-1} \right]^{-1}$, $s_{ii} = (T-K)^{-1}\hat{\mathbf{e}}_{\mathbf{i}}'\hat{\mathbf{e}}_{\mathbf{i}}$, and $\hat{\Delta} = (N-1)^{-1} \left[\sum_{i=1}^{N} \hat{\mathbf{b}}_{\mathbf{i}}\hat{\mathbf{b}}_{\mathbf{i}}' - N^{-1}\sum_{i=1}^{N} \hat{\mathbf{b}}_{\mathbf{i}} \sum_{i=1}^{N} \hat{\mathbf{b}}_{\mathbf{i}}' \right] - N^{-1}\sum_{i=1}^{N} s_{ii}(\mathbf{x_i}'\mathbf{x_i})^{-1}$. Here, $\hat{\mathbf{e}}_{\mathbf{i}}$ is the vector of least squared residuals from the linear projection of y_i on $\mathbf{x_i}$. Note that with this estimator, the cluster-specific estimated slopes are weighted according to the relative estimated variance in $\mathbf{x_i}$ and the inverse of the relative estimated variance of the cluster-level residuals.

The advantage of the PCIV estimator is its robustness to violations of many of the assumptions on which Swamy's efficient estimator relies. Were we to adopt a similar weighting approach, in which our PCIV estimates of $\mathbf{b_i}$ are weighted by the relative covariance between $\mathbf{z_i}$ and $\mathbf{x_i}$ and the inverse of the relative estimated variance of the cluster-level residuals, we would need to impose the same uncorrelated covariance assumption underpinning FEIV estimation. Consequently, we only pursue weighting schemes that do not impose such restrictive assumptions.

5.2 Additional exogenous covariates

The model used in the simulation is parsimonious. Researchers often include other assumed to be exogenous regressor in instrumental variables regression to improve the efficiency of their estimated coefficients. Year effects provide one such example, though covariates that vary over cluster and time may also be of issue. However, in the PCIV approach increasing the dimension of $\mathbf{x_{ij}}$ also entails reducing the degrees of freedom in each cluster-level regression.

However, by using the residuals from a pooled regression, we can hope to obtain these efficiency gains without cutting into the degrees of freedom on the per-cluster regressions. Murtazashvili and Wooldridge (2016) and Laage (2019) each use similar strategies to handle exogenous additional covariates in their respective approaches. Consider the slightly richer model presented below:

$$y_{ij} = \mathbf{x_{1ij}}\mathbf{b_i} + \mathbf{x_{2ij}}\boldsymbol{\delta} + e_{ij},\ j = 1, ..., T, \tag{22}$$

where $\mathbf{x_{1ij}}$ is a $1 \times K$ vector of potentially endogenous regressors, and $\mathbf{b_i} = \boldsymbol{\beta} + \mathbf{d_i}$ is again a $K \times 1$ vector of cluster-specific slopes, where $E(\mathbf{d_i}) = 0$ by definition. We further allow for a $1 \times H$ vector of exogenous covariates, $\mathbf{x_{2ij}}$, with $\boldsymbol{\delta}$ an $H \times 1$ vector of homogeneous slopes. Again, we do not assume $\mathbf{x_{1ij}}$ to be exogenous, and thus require a $1 \times L$ vector of instruments, $\mathbf{z_{ij}}$, to identify $\boldsymbol{\beta}$.

Distinguishing between $\mathbf{x_{1ij}}$ and $\mathbf{x_{2ij}}$ requires researchers to impose structure on the data–

ideally motivated by contextual knowledge of the environment. While some may be reluctant to make these assumptions, the efficiency gains from doing so may be large, particularly with small T. Each element of $\mathbf{b_i}$ which must estimated per-cluster provides one less degree of freedom for estimation within cluster. By maintaining that $\mathbf{x_{2ij}}$ has homogeneous slopes we can account for correlations between $\mathbf{x_{2ij}}$ and $\mathbf{x_{1ij}}$ by applying the Frisch-Waugh-Lowell Theorem. We accordingly can run pooled regressions of y_{ij}, $\mathbf{z_{ij}}$, and $\mathbf{x_{1ij}}$ on $\mathbf{x_{2ij}}$ keeping the residuals of these regressions $\dddot{y}_{ij}$, $\dddot{\mathbf{z}}_{\mathbf{ij}}$, and $\dddot{\mathbf{x}}_{\mathbf{1ij}}$ respectively.

Accordingly, researchers may subsequently perform the same analysis that is presented in section 3 with the residualized data. In which case the PCIV estimator may be written as the following:

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{\boldsymbol{PCIV}} =&\boldsymbol{\beta} + \frac{1}{N}\sum_{i=1}^{N}\mathbf{d_i} + \frac{1}{N}\sum_{i=1}^{N}\Bigg[\Bigg((\sum_{j=1}^{T}\dddot{\mathbf{z}}_{\mathbf{ij}}'\dddot{\mathbf{x}}_{\mathbf{1ij}})(\sum_{j=1}^{T}\dddot{\mathbf{z}}_{\mathbf{ij}}'\dddot{\mathbf{z}}_{\mathbf{1ij}})^{-1}(\sum_{j=1}^{T}\dddot{\mathbf{z}}_{\mathbf{ij}}'\dddot{\mathbf{x}}_{\mathbf{1ij}})\Bigg)^{-1} \\ &\times \Bigg((\sum_{j=1}^{T}\dddot{\mathbf{z}}_{\mathbf{ij}}'\dddot{\mathbf{x}}_{\mathbf{1ij}})(\sum_{j=1}^{T}\dddot{\mathbf{z}}_{\mathbf{ij}}'\dddot{\mathbf{z}}_{\mathbf{1ij}})^{-1}(\sum_{j=1}^{T}\dddot{\mathbf{z}}_{\mathbf{ij}}'\dddot{e}_{ij})\Bigg)\Bigg].\end{aligned} \tag{23}$$

Equation 23 displays the conditions necessary for $\hat{\boldsymbol{\beta}}_{\boldsymbol{PCIV}}$ to consistently estimate $\boldsymbol{\beta}$. Again with N and T going to infinity, we require no other assumptions for consistency as shown in equation 24.

$$\begin{aligned}\operatorname*{plim}_{N,T\to\infty}(\hat{\boldsymbol{\beta}}_{\boldsymbol{PCIV}} - \boldsymbol{\beta}) =&\mathrm{E}(\mathbf{d_i}) + \mathrm{E}[[\mathrm{E}_i(\dddot{\mathbf{z}}_{\mathbf{ij}}'\dddot{\mathbf{x}}_{\mathbf{1ij}})(\mathrm{E}_i(\dddot{\mathbf{z}}_{\mathbf{ij}}'\dddot{\mathbf{z}}_{\mathbf{1ij}}))^{-1}\mathrm{E}_i(\dddot{\mathbf{z}}_{\mathbf{ij}}'\dddot{\mathbf{x}}_{\mathbf{1ij}})]^{-1} \\ &\times\ \mathrm{E}_i(\dddot{\mathbf{z}}_{\mathbf{ij}}'\dddot{\mathbf{x}}_{\mathbf{1ij}})(\mathrm{E}_i(\dddot{\mathbf{z}}_{\mathbf{ij}}'\dddot{\mathbf{z}}_{\mathbf{1ij}}))^{-1}\mathrm{E}_i(\dddot{\mathbf{z}}_{\mathbf{ij}}'\dddot{e}_{ij})].\end{aligned} \tag{24}$$

Equation 24 makes transparent that we require sufficient all-else-equal variation in the instrument and endogenous regressor within each cluster in addition to requiring $\mathrm{E}(\mathbf{d_i}) = 0$ and $\mathrm{E}_i(\dddot{\mathbf{z}}_{\mathbf{ij}}'\dddot{e}_{ij}) = 0$).[13]

[13] As in section 2, we get asymptotic unbiasedness with fixed N as $T \to \infty$. With Fixed T and $N \to \infty$ we also require within each cluster, we must first have large enough T to estimate $b_{1i}, .., b_{Ki}$, and with additional excluded instruments we must now have $T > L \geq K$. Further, we must again assume that the finite sample bias in estimating each b_i is zero in expectation. As is shown in Bound et al. (1995) and Staiger and Stock (1997) finite sample bias in P2SLS with more than two excluded instruments is a function of the number of

5.3 Estimation of cluster-level parameters and sources of heterogeneity

As a final extension, we wish to discuss one last possible benefit of the PCIV approach. As Bates et al. (2014) shows in the exogenous case, with a small change in the last stage of estimation, the PCIV approach allows researchers to uncover effects of exogenous variables that vary only between clusters including possible mechanisms for the differences between the cluster-level slopes. Here, we extend the approach to cases where $\mathbf{x_{ij}}$ may be endogenous.

Consider the original model in equation 1, where again $\mathbf{b_i} = \boldsymbol{\beta} + \mathbf{d_i}$. We can model $\mathbf{d_i}$ as a function of a $1 \times J$ vector of cluster-level covariates $\mathbf{w_i} = (w_{1i}, ..., w_{Ji})$. We present this often called multi-level model below:

$$\begin{aligned} y_{ij} &= \mathbf{x_{ij}}\mathbf{b_i} + e_{ij}, \\ \mathbf{b_i} &= \boldsymbol{\beta} + \mathbf{w_i}\gamma + \mathbf{u_i}, \end{aligned} \tag{25}$$

where $\mathbf{u_i}$ is a $K \times 1$ vector of cluster-level error terms.

In this environment, rather than averaging over the estimates $\hat{\mathbf{b}}_{\mathbf{i,PCIV}}$, we regress each element of $\hat{\mathbf{b}}_{\mathbf{i,PCIV}}$ on one and the vector, $\mathbf{w_i}$, using OLS in the final stage of PCIV.[14] We obtain the estimates of the PAEs, $\boldsymbol{\beta}$, from the estimated intercepts in these regressions. For ease of exposition, I add the subscript $r = 1, ..., K$ to denote specific coefficients, and let the first element of $\mathbf{b_{ri}}$, b_{1i}, be the cluster-specific intercept (the coefficient on 1 in $\mathbf{x_{ij}}$). The estimated coefficients on $\mathbf{w_i}$ from the regression of $\hat{b}_{1i}$ on one and the vector, $\mathbf{w_i}$, provides the estimates of the effects of $\mathbf{w_i}$ on y_{ij}. Finally, the coefficients on $\mathbf{w_i}$ from the the regressions of the cluster-specific slopes, $\widehat{b_{ri}}$, $r = 1, ..., K$, on one and $\mathbf{w_i}$ provide estimates of the effects of the interaction terms, which may be interpreted as the sources of heterogeneous effects of $\mathbf{x_{ij}}$.

Defining $\mathbf{W_i} \equiv (1, w_{1i}, , , w_{Ji})$ and $\boldsymbol{\delta_r} \equiv (\boldsymbol{\beta_r}, \boldsymbol{\gamma_{r1}}, , , \boldsymbol{\gamma_{rJ}})$, we can follow Bates et al. (2014)

excluded instruments and the strength of those instruments.

[14]If weighting is necessary in order to obtain the PAE as discussed in section 5.1, researchers should use weighted least squares instead of OLS in this final stage of estimation.

in writing the corresponding estimator of $\boldsymbol{\delta_r}$ as the following:

$$\hat{\boldsymbol{\delta}}_{\boldsymbol{r},\boldsymbol{PCIV}} = \boldsymbol{\delta_r} + \left(\sum_{i=1}^{N} \mathbf{W_i'W_i}\right)^{-1} \sum_{i=1}^{N} \mathbf{W_i'}(\mathbf{u_i} + \hat{\mathbf{b}}_{\mathbf{ri,PCIV}} - \mathbf{b_{ri}}). \quad (26)$$

Note that maintaining the consistency of $\hat{\mathbf{b}}_{\mathbf{ri,PCIV}}$ is insufficient for the consistency of estimating $\boldsymbol{\delta_r}$. To consistently estimate $\boldsymbol{\delta_r}$, we must impose three additional conditions. First, that $rank[\mathrm{E}(\mathbf{W_i'W_i})] = J$. Second, we must assume that $\mathbf{W_i}$ is exogenous ($\mathrm{E}(\mathbf{W_i'u_i}) = 0$). Third, we must assume that the estimation error in $\hat{\mathbf{b}}_{\mathbf{ri,PCIV}}$ is unrelated to $\mathbf{W_i}$.

It is also worth noting that in undertaking this analysis, these additional assumptions are necessary to consistently estimate all parameters, including $\boldsymbol{\beta}$ which may be consistently estimated under fewer assumptions. However, there may be significant benefits to understanding why the same intervention has large effects for some entity and small effects in others. As is often the case, revelation of these deeper mechanisms come at a cost of less robust estimates.

6 Estimating the price elasticity of demand for gasoline

The use of P2SLS, FEIV, and related estimators is widespread. Even though in may of the applications of these estimators researchers would likely not maintain homogeneous effects across the population, correlation between this “essential heterogeneity” and the strength of the instrument is not usually discussed.[15] The estimation of the price elasticity of demand for gasoline provides a nice illustration of the PCIV approach applied to a question where the population parameter is of primary interest. As described above, the average responsiveness of consumers to price increases in gasoline is of particular importance to both economists and policymakers with uses ranging from market prediction and urban planning to climate change and national security. Consequently, the accuracy of our best estimates may be important.

[15]Phrasing from Heckman et al. (2006).

Estimating this important parameter is not straightforward. Gasoline prices and the volume purchased likely depend upon each other through the interaction of supply and demand forces. This simultaneity issue requires a source of exogenous variation to establish a uni-directional causal link. We follow Davis and Kilian (2011) in estimating this parameter using state gasoline sales taxes as an instrumental variable for changes in prices. In many ways, the empirical setting and design are ideal, especially considering that the efficacy of carbon taxes in lowering fuel consumption is a primary reason why the price elasticity of demand for gasoline is of particular interest.

However, it is not immediately obvious that P2SLS applied to first differenced data as done in Davis and Kilian (2011) and Coglianese et al. (2017) or FEIV estimate the population average parameter, which is of most interest. The existing literature econometrically models the relationship between gasoline prices and the quantity sold as though responsiveness to prices is homogeneous. However, heterogeneous slopes seem likely in this context. Regulation of this market varies across states, as does the industry composition, population density, transportation substitutes, and the macroeconomic climate. Accordingly, we build such "essential heterogeneity" into the econometric model, allowing each state to differ in the price elasticity of gasoline demand. Our PCIV approach provides a natural way to investigate whether these state-specific slopes are homogeneous or vary across states. Below we describe the process by which we estimate each state-specific price elasticity by PCIV.

The data we use carries a fittingly large scope, containing monthly observations of gasoline prices, taxes, and volume sold from January, 1989 through December, 2018 throughout the United States. This provides us with 360 time observations over all 50 states in addition to the District of Columbia. The data covering January, 1989 through March, 2008 largely reflect that used by Davis and Kilian (2011) and Coglianese et al. (2017), however discontinuation of data series and updating subsequent tax changes required significant additional data collection.[16]

[16]We wish to express gratitude to Lucas Davis for publicly releasing the data and code used for both papers on his website. We provide a direct replication of many of the results from Coglianese et al. (2017)

The data on monthly, statewide, gasoline price averages for 1989 through 2011 comes from om US Department of Energy, Energy Information Administration (EIA), 'Petroleum Marketing Monthly Report: Gasoline Prices by Formulation, Grade, Sales Type.' It measures tax-exclusive prices to end users. To this series, we build in additional taxes to approximate at the pump prices. However, this series and the survey on which it relies was discontinued in 2011 by the EIA. As a result, we supplement this pricing data with average at-the-pump price data from Gasbuddy.com.[17]

We use the annual 'Highway Statistics Series' for exact effective dates of state gasoline taxes and as the base levels of the taxes and checked these dates against state governmental documentation when possible. Since 30 states have changed their state gasoline taxes in the last decade (some multiple times), the recent data provides useful identifying variation. As noted in Davis and Kilian (2011) and Coglianese et al. (2017), some gasoline taxes fail the exclusion restriction to serve as an instrumental variable for prices. Taxes (such as sales tax) which are a function of gasoline prices cannot be used, as they suffer from the same simultaneity problem as do the raw prices. As a result, we net out any portion of the state gasoline tax rate, which is due to changes in gasoline prices.[18] The data on gasoline sales volume used throughout is taken from the EIA, 'Petroleum Marketing Monthly Report: Prime Supplier Sales Volumes by Product and Area.'

In order to address any seasonality and time-trends within the data, we first regress the log of sales volume, the log of prices, the log of state gasoline tax rates, and the unemployment

in Appendix C.

[17]This data was retrieved as the maximum 10-year (January, 2009 - January, 2019) charts for each state plus the District of Columbia from `https://www.gasbuddy.com/Charts` on March 22, 2019 and digitized into four or five daily price averages using `https://automeris.io/WebPlotDigitizer/`. We then averaged over these prices to form a monthly average. There is a level shift between the at-the-pump prices from Gasbuddy.com and the series from EIA. However, for the three years in which the data is overlapping, the trends and fluctuations in prices move in concert. As a result, we believe the constant gap will be absorbed by the month-by-year fixed effects.

[18]Davis and Kilian (2011) did the same for taxes during the period from 1989 through 2008. We largely use their data for this time period though discovered a few instances in which the taxes they used either failed to fully capture per-unit taxes or had price changes built into them. We document these tax changes in the Stata do-file (titled newtaxes.do) and provide within that file links to the documentation for each tax change.

rate on time indicators, keeping the residuals, as described in Section 5.2. We then use these detrended data (denoted by tilde) to estimate the following using 2SLS state by state:

$$\widetilde{logsales}_{ij} = \alpha_{1i} + b_i \widehat{logprice}_{ij} + \delta_i \widetilde{x_{ij}} + \widetilde{\epsilon_{ij}}. \tag{27}$$

The state-specific constant terms absorb any unobserved time-invariant differences between states. We control for state-and-time-varying macroeconomic conditions using $\widetilde{x_{ij}}$, the monthly unemployment rate in the state. Here, $\widehat{logprice}_{ij}$ is the fitted values from the OLS regressions of equation 28.

$$\widetilde{logprice}_{ij} = \alpha_{2i} + \boldsymbol{\gamma_{1i}} \widetilde{\mathbf{logtax}}_{\mathbf{ij}} + \gamma_{2i} \widetilde{x_{ij}} + \widetilde{\epsilon_{ij}}. \tag{28}$$

We include both the just-identified specification, where the contemporaneous price is instrumented by the contemporaneous gasoline log tax, and in order to better predict the first stage, we also include an additional lead and lag of the log of taxes as excluded instruments.

The inclusion of heterogeneous slopes in the econometric model also leads us to pay particular attention to the differences in state sizes and volume of gasoline purchased, in contrast to the earlier literature. Were the responsiveness to prices homogeneous across states, such differences in size may only influence the efficiency of the estimates. However, if the price-elasticities of gasoline demand do differ by state, failing to account for such differences in states' relevance to the market may lead to inconsistent estimates of the price elasticity of gasoline over the population. In our preferred specifications, we weight the state-specific estimates by the time-average of volume purchased in the state. Specifically, after estimating the state-specific elasticities, we estimate the population average effect as the following:

$$\hat{\beta} = \sum_{i=1}^{N} w_i \hat{b}_{i2SLS}, \tag{29}$$

where $\hat{b}_{i2SLS}$ are the state-specific elasticity estimates from equation 27 and w_i are the state-specific weights equalling the share of gasoline sold in state i over the panel. We also provide the raw (unweighted) average for comparison.

We compare these PCIV estimates to estimates of the same parameter using P2SLS applied to first differences as used by Davis and Kilian (2011) and Coglianese et al. (2017) and the benchmark FEIV estimator. We present the results from these estimators in Table 2 both equally weighting state by month observations and incorporating the state-specific gasoline-volume weights. Panel A presents estimates from the just-identified model without weights, Panel B the just-identified model with weights, and Panels C and D present the unweighted and weighted estimates from the over-identified model with contemporaneous log prices instrumented by contemporaneous, lead, and lagged log taxes. For PCIV, the analytic standard errors are estimated as described in section 3.2. We present the Cragg and Donald (1993) F-statistics for the excluded instruments for each estimator in the final row of each panel with the average of state-specific F-statistics listed for PCIV.[19]

Across estimators, the first-stage predictive power of taxes on prices are generally strong with only the unweighted, over-identified P2SLS estimation revealing a concerningly small F-statistic of 5.8, though in other specification P2SLS has persistently smaller first-stage F-statistics. FEIV has particularly strong first-stages with F-statistics ranging from 206.9 to 956.9. PCIV also has a high average of state-specific F-statistics, which range from 40.2 to 190.7.

Turning to the coefficient estimates, the results are generally consistent, though the magnitude and statistical significance of the estimated elasticities range meaningfully. A priori, our preferred approaches are weighted PCIV in the final column as they weight state-specific slopes by volume and are robust to correlation between the strength of the instrument and heterogeneity in demand elasticities. However, whether or not additional leads and lags of taxes are used to instrument for prices seems to matter greatly in this context. Whereas in the over-identified specification in Panel D, the point estimate implies that a 10 percent

[19]We present the first-stage coefficient estimates in Table 7 in the online appendix.

increase in prices decreases gasoline consumption by 9 percent (p-value = 0.015), in the just identified model the point estimate implies that the same 10 percent increase in prices leads to less than a three percent reduction in consumption, which cannot statistically be distinguished from zero (p-value = 0.780).

Table 2: Summary of Results Using Three Estimation Methods

	Just-Identified Contemporaneous Price Specification					
	Panel A (Unweighted)			Panel B (Weighted)		
	P2SLS	FEIV	PCIV	P2SLS	FEIV	PCIV
Log price (contemporaneous)	-0.724 (0.193)*** [0.195]***	-0.925 (0.414)** [0.395]***	-0.575 (1.728) [1.792]	-0.462 (0.153)*** [0.163]***	-0.848 (0.399)** [0.465]*	-0.283 (1.010) [1.017]
F-statistic	17.43	635.5	119.3	40.85	956.9	190.7
	Over-Identified Contemporaneous Price Specification					
	Panel C (Unweighted)			Panel D (Weighted)		
Log price (contemporaneous)	-0.712 (0.192)*** [0.200]***	-0.922 (0.415)** [0.395]***	-0.839 (0.401)** [0.464]*	-0.432 (0.139)*** [0.158]***	-0.854 (0.403)** [0.467]*	-0.920 (0.410)** [0.437]**
F-statistic	5.79	206.9	40.17	17.30	321.3	63.26

Note: The sample consists of 18,360 state-by-month observations. Just-identified specifications use contemporaneous values of log taxes to instrument for contemporaneous prices. Over-identified specifications use a lead, a lag, and contemporaneous values of log taxes to instrument for contemporaneous prices. Weights are constructed from state shares of gasoline volume purchased over the sample period. Analytic standard errors appear in parentheses and state-level cluster-bootstrap standard errors appear in brackets below. Asterisks denote statistical significance at the ***1%, **5%, and *10% level.

There are a few important patterns to note in these results. First, the PCIV analytic clustered standard errors and clustered bootstrap standard errors are close approximations of each other. The ratio of the analytic to bootstrap standard errors range from 0.86 (Panel C) to 0.99 (Panel B). For comparison, the same ratio of the standard FEIV analytic state-clustered standard errors to the state-clustered bootstrap standard errors is from 0.84 (Panels B and D) to 1.05 (Panels A and C). Secondly, the just-identified model illustrates the potential cost in efficiency that may accompany using the more robust PCIV estimator. In the just-identified model the PCIV point estimates both are more sensitive to weighting and are accompanied by standard errors that are two to four times larger than those using FEIV. However, in the over-identified specification the PCIV standard errors are comparable to and occasionally smaller than those from FEIV. As a result, we mostly focus on over-identified specifications throughout the remainder of the paper.

Third, P2SLS applied to first differences is the most efficient estimator, however this efficiency gain may also come at a cost. The divergence between FEIV and P2SLS on first-differences implies a violation of the strict exogeneity assumption of the instrument. However, as long as the dependence between lag or lead values of the instrument are only weakly related to the error term, in a homogeneous coefficient model the inconsistency in FEIV from this dependence converges to zero as T grows large (Wooldridge, 2010). In contrast, P2SLS on first-differences, does not enjoy this same result. Given that we are flexibly detrending the data, strong dependence seems unlikely. However, we further investigate this possible dependence by estimating the following equation using FEIV:

$$\widetilde{logsales}_{ij} = \alpha_{1i} + \beta_1 \widehat{logprice}_{ij} + \beta_2 \widetilde{x_{ij}} + \gamma_1 \widehat{logprice}_{ij+1} + \gamma_2 \widetilde{x_{ij+1}} + \widetilde{\epsilon_{ij}}. \tag{30}$$

We then test the hypothesis that $\gamma_1 = 0$ and $\gamma_2 = 0$. We fall far short of rejecting this null hypothesis (p-value=0.995). Such is not the case with P2SLS.

Indeed in revisiting the analysis of Davis and Kilian (2011), Coglianese et al. (2017) notes evidence of the aforementioned violation of strict exogeneity of taxes as an instrument for prices. Coglianese et al. (2017) shows that anticipatory behavior may be problematic for estimating monthly price elasticities. This is particularly important when using first differences in taxes, prices, and quantities, as in Davis and Kilian (2011). Coglianese et al. (2017)

accordingly uses leads and lags of prices and taxes to address this anticipatory behavior of large consumers on their estimated price elasticity of demand for gasoline.

Table 3: Accounting for anticipatory behavior

Panel A: One lag of log prices as additional endogenous regressor						
	(Unweighted)			(Weighted)		
	P2SLS	FEIV	PCIV	P2SLS	FEIV	PCIV
Cumulative effect	-0.544	-0.918	-0.764	-0.313	-0.855	-0.793
of log prices	(0.136)***	(0.422)**	(0.590)	(0.114)***	(0.404)**	(0.421)*
Panel B: One lead of log prices as additional endogenous regressor						
Cumulative effect	-0.365	-0.923	-0.793	-0.189	-0.856	-0.819
of log prices	(0.167)**	(0.423)**	(0.610)	(0.112)	(0.406)**	(0.414)**
Panel C: One lag and one lead of log prices as additional endogenous regressors						
Cumulative effect	-0.178	-0.918	-0.639	0.001	-0.855	-0.693
of log prices	(0.144)	(0.420)**	(0.515)	(0.113)	(0.406)**	(0.473)

Note: Cumulative effects are the sum of contemporaneous, lead, and/or lag values of prices, instrumented by taxes. All results from over-identified models with an additional lead and lag of log taxes used as excluded instruments. Analytic state-clustered standard errors appear in parentheses. Asterisks denote statistical significance at the ***1%, **5%, and *10% level.

We perform similar analysis, reporting the results from all three estimators in Table 3. For brevity, we list the sum of the price coefficients to summarize the effect of prices as instrumented by state taxes on gasoline purchasing again with and without weighting for volume sold in the state.[20] In Panel A, we first introduce a lag of log prices in addition to contemporaneous log prices. In Panel B we include contemporaneous values and lead of log prices. Finally in Panel C, we include contemporaneous, lead, and lag values of log prices as endogenous regressors. In each case, we use over-identified specifications with an additional lead and lag of log taxes serving as excluded instruments, such that in the final specification

[20]We include the individual coefficient estimates of contemporaneous, lead, and lag prices in Table 9 in the online appendix.

we use the contemporaneous value as well as two leads and two lags of the log of taxes to instrument for the three endogenous variables.

Table 3 reveals further divergence in the estimates when accounting for anticipatory consumer behavior. The estimated cumulative effects are varied with the estimated elasticities ranging from zero (weighted P2SLS) to -0.92 (unweighted FEIV under multiple specifications). P2SLS applied to first differenced data is particularly sensitive to the inclusion of leads and lags. Moving from Table 2 to Table 3, the P2SLS estimates of the cumulative effect fall dramatically and even switch sign from around -0.724 to 0.001. In contrast, the FEIV estimates seem impervious to the inclusion of leads and lags with point estimates changing by a mere 0.002 across specifications. The PCIV estimates serve as a middle ground, as the cumulative effect fall slightly with additional leads and lags, but not dramatically. When including either one additional lead or one additional lag of prices we estimate an approximate -0.8 price elasticity of the demand for gasoline (p-values ranging from 0.048 to 0.060). When both are employed we estimate a cumulative effect about -0.7 (p-value of 0.143).

What do we make of this conglomeration of evidence? The volatility of P2SLS and relative robustness of FEIV suggest that we should privilege FEIV in this context over P2SLS. In order to examine FEIV against PCIV, we explore 1) the presence of heterogeneous elasticities, 2) heterogeneity in the strength of the instruments, 3) correlation between the strength of the instruments and state-specific elasticity estimates, and 4) whether the weighting used in FEIV mimics the weights by sales volume.

The PCIV estimates of b_i provide a natural way to examine heterogeneous elasticities. We present these estimates from the just-identified, contemporaneous specification along the y-axis in the top row of Figure 10, and from the over-identified, contemporaneous specification along the bottom row.[21] The t-statistics on the $\widetilde{logtax}_{ij}$ provide a useful measure of the strength and sign of the instrument on the endogenous regressor, $\widetilde{logprice}_{ij}$. Conveniently, these statistics also shed light on whether there is sufficient variation in the instrument,

[21]With more endogenous regressors the figures become more complicated. We present these in Figure 12 in the online appendix.

and whether the time series is sufficiently long, or whether the PCIV estimates may be encumbered by finite sample bias. Each circle represents a state with the size scaled by the the state's relevance to the gasoline market.

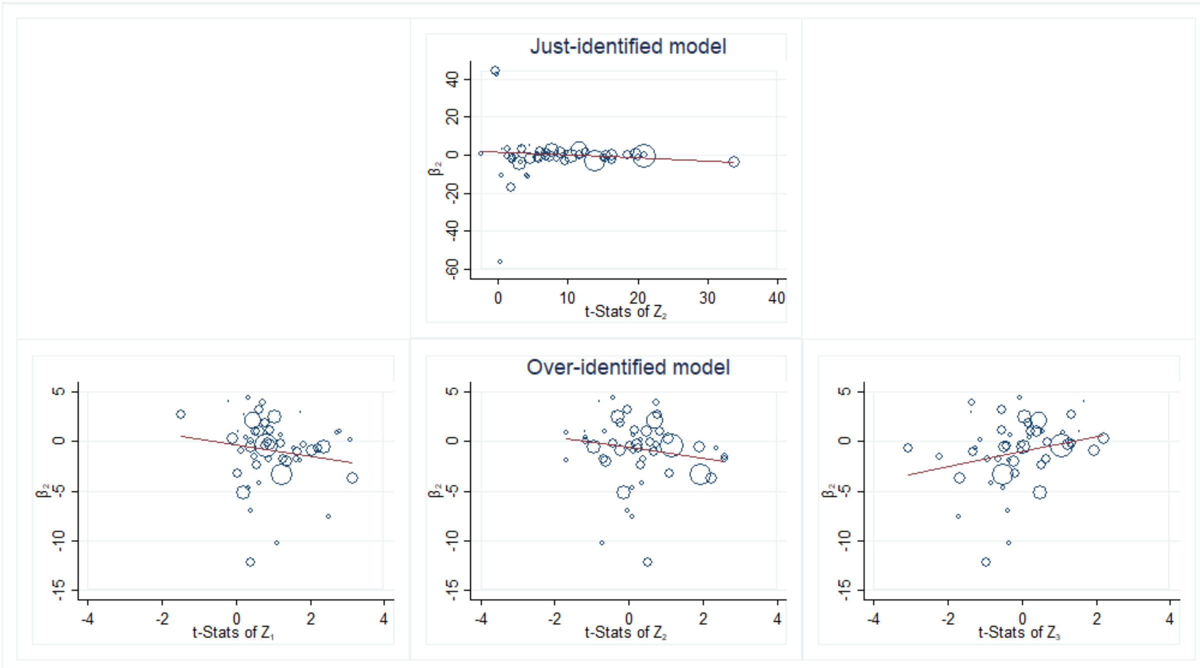
Figure 10 reveals several features of the data that are important for interpreting the estimates presented above, as well as in earlier work. We first consider the just-identified model. Indeed, there is significant heterogeneity in the estimated slopes which range from -56.4 to 44.1. Further, there is also significant heterogeneity in the strength of the instrument as the t-statistics on the first stage range from -2.54 to 33.9. It seems that the extreme outliers are connected to weak instruments as Nevada ($\hat{b_{NV}} = 42.2$), Minnesota ($\hat{b_{MN}} = 44.1$), and New Hampshire ($\hat{b_{NH}} = -56.4$) have first stage t-statistics on the log of gas taxes of -0.25, -0.41, and 0.23 respectively. [22]

Overall, we assess the relationship between the strength of the instrument and heterogeneous slopes by regressing $\hat{b_i}$ on the first-stage t-statistics and find a regression coefficient of -0.16 though it is not statistically significant. Turning to the three graphs along the bottom row from the over identified model, we see a few patterns. First, the heterogeneity in $\hat{b_i}$ remains significant though less dramatic than above, as the stat-specific elasticity estimates range from -12.2 to 4.5. Similarly, the ranges on the t-statistics for each instrument are similarly more compact from -3.1 to 3.1. The three graphs along the bottom row of Figure 10 also reveal steeper slopes in the relationship between the estimated elasticities and strength of the instrument within each state. Indeed the coefficient estimates on the first-stage t-statistics range from -0.58 to 0.85. The relationship between $\hat{b_i}$ and the t-statistic on a 1-month lag in log taxes is statistically significant (p-value = 0.022).[23]

[22]The heterogeneity in elasticity estimates is not purely due to weak instruments. We conduct regional analysis employing PCIV at the EIA's Petroleum Administration for Defense District level and uncover estimates that range from 0.6 to -2.0 with first-stage F-statistics ranging from 24.0 to 630.7. We report the results in Table 8 in Appendix B.

[23]We show the relationships between the state-specific slopes and first stage t-statistics in Tables 11 and 12 in the online appendix.

Figure 10: Heterogeneous Elasticities and Instrument Strength



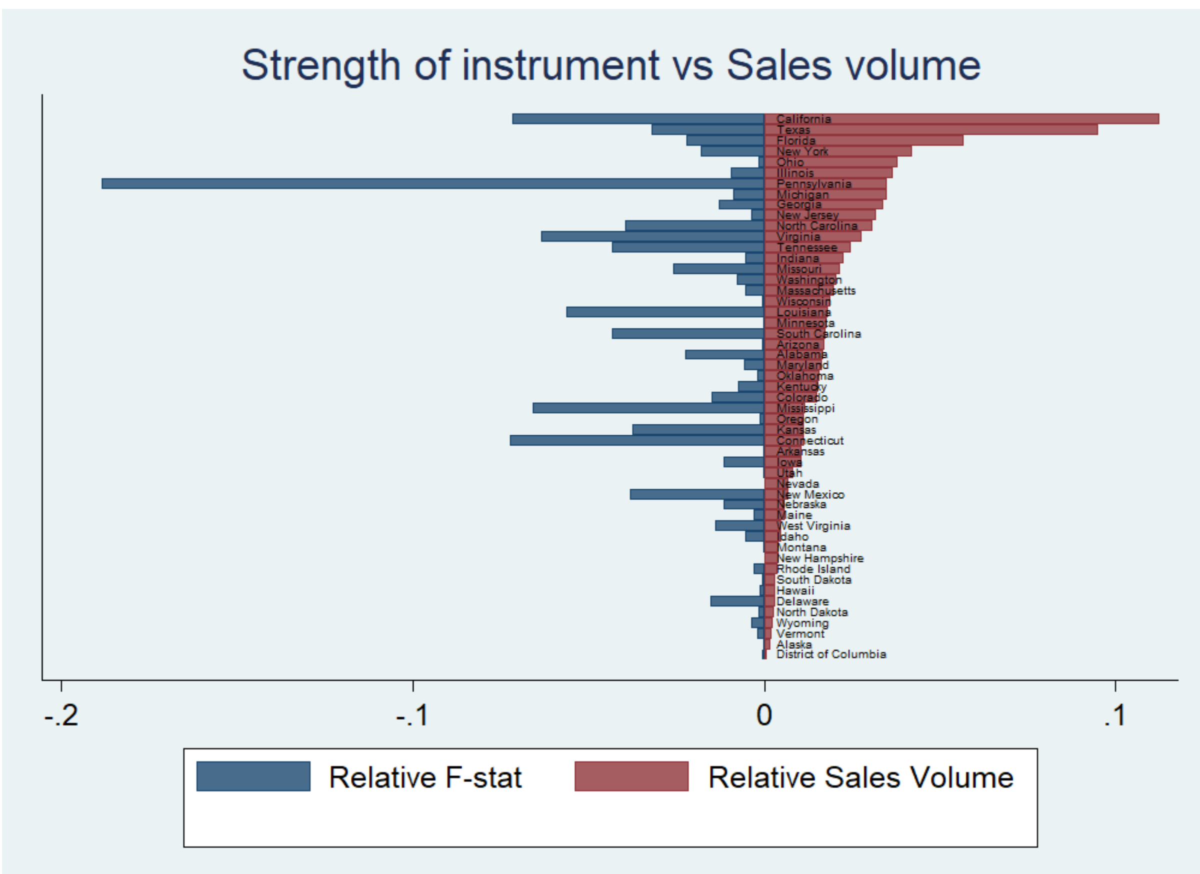
Note: X_2 refers to the contemporaneous Log price. Z_1, Z_2, and Z_3 are one-month lead, contemporaneous, and one-month lag Log tax, respectively.

Finally, we consider the issue of weighting (both implicitly and explicitly) and its role in uncovering a population estimate of the average price elasticity of gasoline demand. As shown in Murtazashvili and Wooldridge (2008) and discussed in Section 2 above, P2SLS and FEIV each implicitly apply additional weight to state-specific coefficients when the instrument is strong. This may, but does not necessarily reflect the natural weighting scheme we employ with PCIV in which we explicitly weight elasticity estimates by the fraction of the market represented in the state. We examine the agreement between the two weighting schemes in Figure 11. One the left we report the relative first-stage F-statistics of each state and on the right we report the share of total gasoline sales that occur in the state over the panel.

Figure 11 shows vast disagreement between the two weighting schemes. The explicit weights we employ are intuitive with larger gasoline consuming states receiving influence on the estimate commensurate with their relevance in the market. California has the highest sales volume, consuming 11.3 percent of the gasoline in the country. The District of Columbia has the lowest sales volume, composed of just 0.1 percent of the national market.

To depict the variation is the strength of the instrument across states we sum the state-specific F-statistics and report the negative of each state's share of the cumulative F-statistics. Pennsylvania leads all states with 18.9 percent of the cumulative F-statistics whereas the bottom 50 percent of states cumulatively comprise under 5 percent of the cumulative F-statistics. Accordingly, Connecticut with only 1.1 percent of gasoline sales (though a state-specific first-stage F-statistic of 439.6) has a larger influence on the standard P2SLS and FEIV estimates than does Ohio, in which 3.8 percent of the nation's gasoline is sold, as its first-stage F-statistic is only 8.7.

Figure 11: Weighting under standard and PCIV approaches



Might the use of weighted least squares with FEIV rectify this issue? In short, not generally or in this context. First, conditional on the appropriate weight, FEIV will continue to overweight states for which the instrument is strong. Secondly, as shown in Solon et al. (2015) weighted least squares in the presence of heterogeneous effects generally does not consistently estimate the population average effect, nor does it always dominate unweighted least squares. In fact, after remarking on the shortcomings of weighted and unweighted least squares, Solon et al. (2015) alludes to a PC estimating approach as a potential way one *could* estimate a population average effect.

Accordingly, the evidence seems to contradict the assumptions necessary for P2SLS or FEIV to consistently estimate the population average elasticity of gasoline demand. Though PCIV is robust to violations of these assumptions, it seems that in this case we lack a long enough time series on each state (or sufficient variation in taxes within each state) to maintain that each state-specific estimate closely approximates the true state-specific parameter.

As a result, we redo the analysis on the sub-sample of states for which the instrument is strong. Naturally, we will no longer be able to claim that we are estimating the PAE for the United States, however, we will be able to find a consistent estimate of the average price elasticity of gasoline demand on a specified and well-defined group of compliers, namely, states that we can identify in which the instrument is strong. Recall that in the presence of heterogeneous slopes, a violation of the key condition prohibits P2SLS and FEIV from consistently estimating even a LATE, whereas PCIV is robust to correlated random slopes. Consequently, we use use the rule of thumb proposed by Staiger and Stock (1997) and formally justified in Stock and Yogo (2005) and keep only the 37 states in which the first stage F-statistic is above ten.[24] We then apply all three estimators to this data and report the results in Table 4. We perform the analysis both on the contemporaneous specification and using a lead, a lag, and both of log prices as endogenous variables.

[24]We use F-statistics from the just-identified contemporaneous specification for ease. The results are similar, if we use F-statistics from the over-identified contemporaneous specification to compose the sample. The state-specific minimum eigenvalues from Cragg and Donald (1993) are generally much smaller with multiple endogenous regressors.

Table 4: Estimated elasticities among states in which the instrument is strong (LATEs)

	Panel A: Just-identified contemporaneous specification					
	(Unweighted)			(Weighted)		
	P2SLS	FEIV	PCIV	P2SLS	FEIV	PCIV
Log price (contemporaneous)	-0.575 (0.186)***	-0.762 (0.418)*	-0.791 (0.526)	-0.425 (0.150)***	-0.757 (0.379)**	-0.681 (0.430)
	Panel B: Over-identified contemporaneous specification					
Log price (contemporaneous)	-0.554 (0.184)***	-0.762 (0.423)*	-0.697 (0.446)	-0.394 (0.133)***	-0.762 (0.383)**	-0.649 (0.433)
	Panel C: Over-Identified with additional lag of log prices					
Log price (cumulative)	-0.488 (0.139)***	-0.762 (0.384)**	-0.669 (0.550)	-0.272 (0.110)***	-0.762 (0.426)*	-0.636 (0.417)
	Panel D: Over-Identified with additional lead of log prices					
Log price (cumulative)	-0.212 (0.154)	-0.762 (0.385)**	-0.662 (0.518)	-0.146 (0.104)	-0.762 (0.428)*	-0.636 (0.368)*
	Panel E: Over-Identified with additional lag and lead of log prices					
Log price (cumulative)	-0.146 (0.140)	-0.762 (0.386)**	-0.647 (0.368)*	0.041 (0.108)	-0.762 (0.431)*	-0.623 (0.316)**

Note: Sample composed of 37 states with first-stage F-statistics from the just-identified specification above 10. Cumulative effects are the sum of contemporaneous, lead, and/or lag values of prices, instrumented by taxes. Analytic state-clustered standard errors appear in parentheses. Asterisks denote statistical significance at the ***1%, **5%, and *10% level.

Among states for which the instrument is strong, the PCIV estimates are much more stable, though they frequently fall on the margin of conventional thresholds of statistical significance. Regardless of whether lag or lead values of log price are included in the model, the weighted PCIV estimates all fall between -0.62 and -0.65 with p-values ranging from 0.048 to 0.134. Even in the just-identified model the weighted estimate is -0.68 with a standard error in line with that from over-identified specifications. As before, the P2SLS estimates applied to first differences are volatile ranging from -0.55 to 0.04, while the FEIV estimates are remarkably stable at -0.76. The stability of the FEIV estimates should not be confused for reliability. Even estimated on this sample of compliers, FEIV does not consistently estimate the LATE; it continues to over-weight states for which the first-stage is particularly strong. As an illustration, if we exclude Pennsylvania, the magnitudes of the FEIV point estimates fall to -0.38 (unweighted) and -0.44 (weighted).

Considering the totality of evidence our best estimates of the price elasticity of demand for gasoline would place the parameter between -0.6 and -0.7, though the confidence intervals for these estimates are wide. These estimates imply that consumers are more sensitive to price hikes than would be implied by Coglianese et al. (2017), Hughes et al. (2008), and Levin et al. (2017). As discussed the small estimates in Coglianese et al. (2017) are due to the use of P2SLS applied to first differences, which is particularly afflicted by anticipatory behavior and improperly weights state-specific elasticities. Hughes et al. (2008) uses crude oil disruptions to instrument for prices in regressions using nationwide monthly data to estimate an elasticity of around -0.03 to -0.22. Levin et al. (2017) uses high-frequency gasoline transaction data to estimate daily monthly demand elasticities at the city, state and national level, employing time and cross-sectional fixed effects in an effort to eliminate the endogeneity of prices. They estimate an elasticity of -0.21 to -0.30 at the state-month level. However, neither of these approaches addresses the endogeneity issue that arises if some price increases are caused by increases in the quantity of gasoline demanded. Consequently, it is unsurprising that the elasticity estimates from these approaches would be smaller.

These high elasticity estimates also reflect that price increases due to taxes increases are more permanent and salient to consumers than are most other gasoline price increases. Indeed Davis and Kilian (2011); Scott (2012); Baranzini and Weber (2013); Li et al. (2014) and Coglianese et al. (2017) each make this argument. Li et al. (2014) in particular demonstrates both points. Regarding permanence, Li et al. (2014) shows that state gasoline taxes have higher AR(1) coefficients than tax exclusive prices. Secondly, Li et al. (2014) also demonstrates that tax changes receive more print and television news coverage than do other price changes of similar magnitude. As a result, these estimates of the price elasticity of demand for gasoline must be viewed as pertaining to permanent changes in prices for relatively informed consumers. However, the responsiveness to permanent and salient price changes is often the parameter of interest in modeling gasoline dependent industries and many policy discussions.

Regarding the efficacy of gasoline taxes in reducing carbon emissions, we may be interested in consumers' responsiveness to the taxes themselves, because the pass-through rate of the tax to consumers may be less than complete.[25] Accordingly, we provide reduced-form estimates of the monthly elasticity of gasoline consumption to gasoline taxes using pooled OLS applied to first differences (POLS), fixed effects (FE), and a per-cluster estimator similar to that used in Bates et al. (2014) (PC). We think of the reduced-form estimation as a correlated random coefficient model with exogenous regressors. Wooldridge (2005) shows that consistency of FE relies on the assumption that the within-state variation in taxes is uncorrelated with state-specific responsiveness to taxes. The same result can be easily extended to POLS. In contrast, Bates et al. (2014) shows that without further assumptions the PC approach is unbiased even under a violation of this assumption. Furthermore, POLS and FE implicitly assign higher weight to the coefficients for states in which there is relatively more within-state variation in taxes and as shown by Solon et al. (2015), appplying WLS does not properly correct for heterogeneous state sizes. We show the relative within-state variance in taxes compared against the relative sales volume of each state in Figure 13 in the appendix, which show extensive disagreement between the two weighting schemes.

[25]Li et al. (2014) also estimates the elasticity of gasoline demand to taxes, but under the assumption of complete pass-through of gasoline taxes to consumers. Our first-stage coefficients in Table 7 contradict that assumption.

Table 5: Reduced-Form Estimation Results Using Three Estimation Methods

	Panel A: Just-Identified					
	(Unweighted)			(Weighted)		
	POLS	FE	PC	POLS	FE	PC
Log tax (Contemporaneous)	-0.148 (0.053)***	-0.228 (0.109)**	-0.259 (0.129)**	-0.094 (0.038)**	-0.263 (0.112)**	-0.336 (0.176)*
	Panel B: Over-Identified					
Log tax (Cumulative Effect)	-0.045 (0.030)	-0.228 (0.112)**	-0.253 (0.132)*	-0.011 (0.022)	-0.264 (0.114)**	-0.337 (0.179)*

Note: Reduced form effect of log state taxes on gasoline sales volume. State-clustered standard errors appear in parentheses. Cumulative effect in Panel B is the sum of the coefficient from the lead, lag, and contemporaneous values of the log of the state gasoline tax. Individual coefficient estimates appear in Table 14 in the appendix. Asterisks denote statistical significance at the ***1%, **5%, and *10% level.

We provide reduced form estimates from all three estimators with and without weighting, both contemporaneous and inclusive of leads and lags in Table 5. Once we account for weighting the PC approach provide a stable estimate -0.34 for the monthly gasoline tax elasticity. We find the same result for the cumulative effect when we also add a lead and lag of gas taxes. Taking this point value literally, a $0.05 per gallon increase in the gas tax, which corresponds to a 9 percent increase in average taxes, leads to a 3.0 percent decrease in the monthly volume of gasoline consumed. For comparison Li et al. (2014) finds that the same $0.05 per gallon increase in taxes leads to a 0.9 to 2.3 percent decrease in annual gasoline sales. Though the 90 percent confidence interval includes the entire range reported in Li et al. (2014), this estimate suggests that consumers are more responsive at least in the short term to gasoline taxes.

7 Concluding Remarks

Whether the purpose of empirical work is to inform and evaluate theory, uncover and explain phenomena, or inform practitioners of best practices, population average effects are generally of high interest.[26] We present the conditions under which per-cluster instrumental variables is consistent in estimating such generally representative parameters and show that it has robustness properties beyond more standard approaches such as pooled two-stage least squares and fixed effects instrumental variables, specifically in the presence of heterogeneous responsiveness to treatment. We develop the theory behind this and demonstrate the performance of per-cluster instrumental variables in simulation.

We then take all three methods to data providing an important, novel examination of the price elasticity of the demand for gasoline in the United States. In this examination, we highlight whether the assumptions necessary for consistent estimation using existing methods are likely to hold in this context, and also the data limitations that persist with per-cluster estimation; both are made possible through the implementation of per-cluster estimation. Though we present consistent estimates of the price elasticity of the demand for gasoline in the United States, we worry the even these estimates may be subject to finite sample bias that originates from weak instruments in some states. As a result, through per-cluster instrumental variables we provide a local average of the price elasticity of demand for gasoline for the states for which the instrument is "strong enough." This is something existing estimators cannot achieve in the presence of heterogeneous state-size and state-specific elasticities. We arrive at an estimate of approximately -0.6, suggesting consumers are more responsive than the previous literature implies. Noting similar issues may afflict reduced form estimates of the average tax elasticity of demand for gasoline in the United States (and elsewhere), we apply per-cluster estimation to this policy-relevant parameter. We find an elasticity of approximately -0.34, which is larger in magnitude than less-robust FE estimates and earlier estimates also using FE.

[26]These useful classifications come from Roth (1986).

Finally, we wish to highlight that use of per-cluster estimation not only adds robustness, but also provides more clarity regarding what parameter is being estimated using Pooled 2-Stage Least Squares or Fixed Effects Instrumental Variables. Even in settings where the panel may lack sufficient length to rely on the consistency of PCIV, we suggest that researchers should employ it to investigate whether key assumptions hold such as monotonicity, effect homogeneity, size homogeneity, and zero correlation between the strength of the instrument and heterogeneous effects, which may inhibit estimation of population average effects with standard estimators.

References

Arellano, M. and S. Bonhomme (2011). Identifying distributional characteristics in random coefficients panel data models. *The Review of Economic Studies 79*(3), 987–1020.

Baranzini, A. and S. Weber (2013). Elasticities of gasoline demand in switzerland. *Energy policy 63*, 674–680.

Bates, M. D., K. E. Castellano, S. Rabe-Hesketh, and A. Skrondal (2014). Handling correlations between covariates and random slopes in multilevel models. *Journal of Educational and Behavioral Statistics 39*(6), 524–549.

Bound, J., D. A. Jaeger, and R. M. Baker (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American statistical association 90*(430), 443–450.

Burke, M., S. M. Hsiang, and E. Miguel (2015). Global non-linear effect of temperature on economic production. *Nature 527*(7577), 235.

Chamberlain, G. (1992). Efficiency bounds for semiparametric regression. *Econometrica: Journal of the Econometric Society*, 567–596.

Climate Leadership Council (2019, January). Economists' statement on carbon dividends. *The Wall Street Journal*. Accessed: February, 2019.

Coglianese, J., L. W. Davis, L. Kilian, and J. H. Stock (2017). Anticipation, tax avoidance, and the price elasticity of gasoline demand. *Journal of Applied Econometrics 32*(1), 1–15.

Cragg, J. G. and S. G. Donald (1993). Testing identifiability and specification in instrumental variable models. *Econometric Theory 9*(2), 222–240.

Dahl, C. and T. Sterner (1991). Analysing gasoline demand elasticities: a survey. *Energy economics 13*(3), 203–210.

Davis, L. W. and L. Kilian (2011). Estimating the effect of a gasoline tax on carbon emissions. *Journal of Applied Econometrics 26*(7), 1187–1214.

Deaton, A. (2010). Instruments, randomization, and learning about development. *Journal of Economic Literature 48*(2), 424–55.

Deschênes, O., M. Greenstone, and J. Guryan (2009). Climate change and birth weight. *American Economic Review 99*(2), 211–17.

Espey, M. (1998). Gasoline demand revisited: an international meta-analysis of elasticities. *Energy Economics 20*(3), 273–295.

Fisher, A. C., W. M. Hanemann, M. J. Roberts, and W. Schlenker (2012). The economic impacts of climate change: evidence from agricultural output and random fluctuations in weather: comment. *American Economic Review 102*(7), 3749–60.

Graham, B. S. and J. L. Powell (2012). Identification and estimation of average partial effects in âirregularâ correlated random coefficient panel data models. *Econometrica 80*(5), 2105–2152.

Heckman, J. and E. Vytlacil (1998). Instrumental variables methods for the correlated random coefficient model: Estimating the average rate of return to schooling when the return is correlated with schooling. *Journal of Human Resources*, 974–987.

Heckman, J. J., S. Urzua, and E. Vytlacil (2006). Understanding instrumental variables in models with essential heterogeneity. *The Review of Economics and Statistics 88*(3), 389–432.

Heckman, J. J. and E. Vytlacil (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica 73*(3), 669–738.

Hughes, J., C. Knittel, and D. Sperling (2008). Evidence of a shift in the short-run price elasticity of gasoline demand. *The Energy Journal 29*(1).

Imbens, G. W. and J. D. Angrist (1994). Identification and estimation of local average treatment effects. *Econometrica 62*(2), 467–475.

Klein, L. R. (1953). *A Textbook of Econometrics.* Evanston: Row, Peterson.

Kuh, E. (1959). The validity of cross-sectionally estimated behavior equations in time series applications. *Econometrica: Journal of the Econometric Society*, 197–214.

Laage, L. (2019). A correlated random coefficient panel model with time-varying endogeneity.

Levin, L., M. S. Lewis, and F. A. Wolak (2017). High frequency evidence on the demand for gasoline. *American Economic Journal: Economic Policy 9*(3), 314–47.

Li, S., J. Linn, and E. Muehlegger (2014). Gasoline taxes and consumer behavior. *American Economic Journal: Economic Policy 6*(4), 302–42.

Mundlak, Y. (1978). Models with variable coefficients: integration and extension. In *Annales de l'INSEE*, pp. 483–509. JSTOR.

Murtazashvili, I. and J. M. Wooldridge (2008). Fixed effects instrumental variables estimation in correlated random coefficient panel data models. *Journal of Econometrics 142*(1), 539–552.

Murtazashvili, I. and J. M. Wooldridge (2016). A control function approach to estimating switching regression models with endogenous explanatory variables and endogenous switching. *Journal of econometrics 190*(2), 252–266.

Raj, B., V. Srivastava, and A. Ullah (1980). Generalized two stage least squares estimators for a structural equation with both fixed and random coefficients. *International Economic Review*, 171–183.

Roth, A. E. (1986). Laboratory experimentation in economics. *Economics & Philosophy 2*(2), 245–273.

Rubin, H. (1950). Note on random coefficients. *Statistical inference in dynamic economic models*, 419–421.

Scott, K. R. (2012). Rational habits in gasoline demand. *Energy Economics 34*(5), 1713–1723.

Solon, G., S. J. Haider, and J. M. Wooldridge (2015). What are we weighting for? *Journal of Human resources 50*(2), 301–316.

Staiger, D. and J. H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica 65*(3), 557–586.

Stock, J. H. and M. Yogo (2005). Testing for weak instruments in linear iv regression. *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, 80.

Swamy, P. (1971). A vb [1971], statistical inference in random coefficient regres-sion models. *Lecture Notes in Operations Research and Mathematical Systems 55*.

Tol, R. S. (2002). Estimates of the damage costs of climate change. part 1: Benchmark estimates. *Environmental and resource Economics 21*(1), 47–73.

US Energy Information Administration (2018, July). Energy and the environment explained: where green house gases come from. Technical report. Accessed February, 2019.

Wooldridge, J. M. (2005). Fixed-effects and related estimators for correlated random-coefficient and treatment-effect panel data models. *Review of Economics and Statistics 87*(2), 385–390.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

A Consistency of the Estimated Variance for PCIV Estimator

For the consistency of the estimated variance of our PCIV estimator, additional assumptions are required. Also, we set $L = K$ for simplicity.

1. $\mathrm{E}\|\mathbf{x_{ij}}\|^2 < \infty$
2. $\mathrm{E}\|\mathbf{z_{ij}}\|^2 < \infty$
3. $\mathrm{E}[\mathbf{z_{ij}}\mathbf{z'_{ij}}]$ is positive definite
4. $\mathrm{E}\|\mathbf{z_{ij}}\|^4 < \infty$

$$
\begin{aligned}
\hat{\mathrm{V}}(\hat{\boldsymbol{\beta}}_{\boldsymbol{PCIV}} - \boldsymbol{\beta}) &= \mathrm{V}\left(\frac{1}{N}\sum_{i=1}^{N}\left[\hat{\boldsymbol{d}}_{\boldsymbol{i}} + \left(\sum_{j=1}^{T}\mathbf{x_{ij}}'\mathbf{z_{ij}}\right)^{-1}\left(\sum_{j=1}^{T}\mathbf{z'_{ij}}\hat{e}_{ij}\right)\right]\right) \\
&= \frac{1}{N}\mathrm{V}\left(\hat{\boldsymbol{d}}_{\boldsymbol{i}} + \left(\sum_{j=1}^{T}\mathbf{x_{ij}}'\mathbf{z_{ij}}\right)^{-1}\left(\sum_{j=1}^{T}\mathbf{z'_{ij}}\hat{e}_{ij}\right)\right), \\
&= \frac{1}{N}\mathrm{V}(\hat{\boldsymbol{d}}_{\boldsymbol{i}}) + \frac{1}{NT}\hat{\mathrm{V}}(\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b}_{\boldsymbol{i}}) + \frac{2}{N}\widehat{\mathrm{COV}}(\hat{\boldsymbol{d}}_{\boldsymbol{i}}, \hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b}_{\boldsymbol{i}})
\end{aligned}
$$

We will show

a) $\mathrm{V}(\hat{\boldsymbol{d}}_{\boldsymbol{i}}) \xrightarrow{p} \mathrm{V}(\boldsymbol{d}_{\boldsymbol{i}})$

b) $\hat{\mathrm{V}}(\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b}_{\boldsymbol{i}}) \xrightarrow{p} \mathrm{V}(\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b}_{\boldsymbol{i}})$

c) $\widehat{\mathrm{COV}}(\hat{\boldsymbol{d}}_{\boldsymbol{i}}, \hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b}_{\boldsymbol{i}}) \xrightarrow{p} \mathrm{COV}(\boldsymbol{d}_{\boldsymbol{i}}, \hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b}_{\boldsymbol{i}}) \equiv Q_{xz}^{-1}Q_{zed}$

Then, as $n \to \infty$,

$$
\hat{\mathrm{V}}(\hat{\boldsymbol{\beta}}_{\boldsymbol{PCIV}} - \boldsymbol{\beta}) \to \mathrm{V}(\hat{\boldsymbol{\beta}}_{\boldsymbol{PCIV}} - \boldsymbol{\beta})
$$

a) Note that $\mathrm{E}[\hat{\boldsymbol{d}}_{\boldsymbol{i}}] = \mathrm{E}[\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \hat{\boldsymbol{\beta}}] = \mathrm{E}[\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \bar{\hat{\boldsymbol{b}}}_{\boldsymbol{i}}] = 0$.

$$
\begin{aligned}
\mathrm{V}(\hat{\boldsymbol{d}}_{\boldsymbol{i}}) &= \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{d}}_{\boldsymbol{i}}^{2} = \frac{1}{N}\sum_{i=1}^{N}\left(\boldsymbol{d}_{\boldsymbol{i}} + \hat{\boldsymbol{d}}_{\boldsymbol{i}} - \boldsymbol{d}_{\boldsymbol{i}}\right)^2 \\
&= \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{d}_{\boldsymbol{i}}^{2} + \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\boldsymbol{d}}_{\boldsymbol{i}} - \boldsymbol{d}_{\boldsymbol{i}}\right)^2
\end{aligned}
$$

By WLLN,

$$\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{d_i}^2 \xrightarrow{p} \sigma_d^2$$

Likewise,

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{\boldsymbol{d}}_{\boldsymbol{i}}-\boldsymbol{d_i})^2 \xrightarrow{p} \mathrm{E}\left[(\hat{\boldsymbol{d}}_{\boldsymbol{i}}-\boldsymbol{d_i})^2\right]$$

Since $\hat{\boldsymbol{d}}_{\boldsymbol{i}}$ is a consistent estimator of $\boldsymbol{d_i}$, as $n \to \infty$,

$$\mathrm{E}\left[(\hat{\boldsymbol{d}}_{\boldsymbol{i}}-\boldsymbol{d_i})^2\right] \xrightarrow{p} 0$$

b) First, by WLLN,

$$\hat{Q}_{xz} = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{T}\mathbf{x'_{ij}z_{ij}} \xrightarrow{p} \mathrm{E}\left[\mathbf{x'_{ij}z_{ij}}\right] \equiv Q_{xz}$$

Note that $\hat{e}_{ij} \equiv y_{ij} - \mathbf{x_{ij}}\hat{\boldsymbol{b}}_{\boldsymbol{i}} = y_{ij} - \mathbf{x_{ij}}\boldsymbol{b_i} + \mathbf{x_{ij}}\boldsymbol{b_i} - \mathbf{x_{ij}}\hat{\boldsymbol{b}}_{\boldsymbol{i}} = e_{ij} - \mathbf{x_{ij}}(\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b_i})$.

Then, $\hat{e}_{ij}^2 = e_{ij}^2 - 2(\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b_i})'\mathbf{x_{ij}}e_{ij} + (\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b_i})'\mathbf{x_{ij}}\mathbf{x_{ij}}'(\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b_i})$.

$$\begin{aligned}
\hat{\Omega}_{zze} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{T}\mathbf{z_{ij}}\mathbf{z_{ij}}'\hat{e}_{ij}^2 \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{T}\mathbf{z_{ij}}\mathbf{z_{ij}}'\left(e_{ij}^2 - 2(\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b_i})'\mathbf{x_{ij}}e_{ij} + (\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b_i})'\mathbf{x_{ij}}\mathbf{x_{ij}}'(\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b_i})\right) \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{T}\mathbf{z_{ij}}\mathbf{z_{ij}}'e_{ij}^2 - \frac{2}{NT}\sum_{i=1}^{N}\sum_{j=1}^{T}(\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b_i})'\mathbf{x_{ij}}e_{ij}\mathbf{z_{ij}}\mathbf{z_{ij}}' \\
&+ \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{T}\left((\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b_i})'\mathbf{x_{ij}}\right)^2\mathbf{z_{ij}}\mathbf{z_{ij}}' \\
&\equiv \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{T}\mathbf{z_{ij}}\mathbf{z_{ij}}'e_{ij}^2 - 2R_{1n} + R_{2n}
\end{aligned}$$

By WLLN,

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{T}\mathbf{z_{ij}}\mathbf{z_{ij}}'e_{ij}^2 \xrightarrow{p} \mathrm{E}\left[\mathbf{z_{ij}}\mathbf{z_{ij}}'e_{ij}^2\right] \equiv \Omega_{zze}$$

Also,

$$R_{1n} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{T}(\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b_i})'\mathbf{x_{ij}}e_{ij}$$

$$\left\| \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{T} (\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b}_{\boldsymbol{i}})' \mathbf{x_{ij}} e_{ij} \mathbf{z_{ij}} \mathbf{z_{ij}}' \right\| \leq \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{T} \left\| (\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b}_{\boldsymbol{i}})' \mathbf{x_{ij}} e_{ij} \mathbf{z_{ij}} \mathbf{z_{ij}}' \right\|$$

$$= \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{T} tr \left(e_{ij}^2 \left((\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b}_{\boldsymbol{i}})' \mathbf{x_{ij}} \right)^2 \mathbf{z_{ij}} \mathbf{z_{ij}}' \mathbf{z_{ij}} \mathbf{z_{ij}}' \right)^{1/2}$$

$$= \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{T} |e_{ij}| \left| (\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b}_{\boldsymbol{i}})' \mathbf{x_{ij}} \right| \|\mathbf{z_{ij}}\|^2$$

$$\leq \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{T} |e_{ij}| \left\| \hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b}_{\boldsymbol{i}} \right\| \|\mathbf{x_{ij}}\| \|\mathbf{z_{ij}}\|^2$$

$$\|R_{1n}\| \leq \left\| \hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b}_{\boldsymbol{i}} \right\| \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{T} |e_{ij}| \|\mathbf{x_{ij}}\| \|\mathbf{z_{ij}}\|^2$$

By Cauchy-Schwartz inequality,

$$\mathrm{E}\left(|e_{ij}| \|\mathbf{x_{ij}}\| \|\mathbf{z_{ij}}\|^2 \right) \leq \left(\mathrm{E}\,|e_{ij}|^2 \right)^{\frac{1}{2}} \left(\mathrm{E}\,\|\mathbf{x_{ij}}\|^2 \|\mathbf{z_{ij}}\|^4 \right)^{\frac{1}{2}}$$

$$\leq \left(\mathrm{E}\,|e_{ij}|^2 \right)^{\frac{1}{2}} \left(\mathrm{E}\,\|\mathbf{x_{ij}}\|^2 \right)^{\frac{1}{2}} \left(\|\mathbf{z_{ij}}\|^4 \right)^{\frac{1}{2}} < \infty$$

As $\|\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b}_{\boldsymbol{i}}\| \xrightarrow{p} 0$, we have

$$R_{1n} \xrightarrow{p} 0.$$

Lastly,

$$R_{2n} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{T} (\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b}_{\boldsymbol{i}})' \mathbf{x_{ij}} \mathbf{x_{ij}}' (\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b}_{\boldsymbol{i}})$$

$$\left\| \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{T} \left((\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b}_{\boldsymbol{i}})' \mathbf{x_{ij}} \right)^2 \mathbf{z_{ij}} \mathbf{z_{ij}}' \right\| \leq \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{T} \left((\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b}_{\boldsymbol{i}})' \mathbf{x_{ij}} \right)^2 \|\mathbf{z_{ij}}\| \, tr \left(\mathbf{z_{ij}} \mathbf{z_{ij}}' \right)^{1/2}$$

$$= \left\| \hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b}_{\boldsymbol{i}} \right\|^2 \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{T} \|\mathbf{x_{ij}}\|^2 \|\mathbf{z_{ij}}\|^2$$

Following additional assumptions and $\|\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b}_{\boldsymbol{i}}\| \xrightarrow{p} 0$, we have

$$R_{2n} \xrightarrow{p} 0.$$

c)

$$\begin{aligned}\widehat{\text{COV}}(\hat{\boldsymbol{d}}_{\boldsymbol{i}}, \hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b}_{\boldsymbol{i}}) &= \frac{1}{N}\sum_{i=1}^{N} \hat{\boldsymbol{d}}_{\boldsymbol{i}}(\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b}_{\boldsymbol{i}}) \\ &= \frac{1}{N}\sum_{i=1}^{N} \left(\boldsymbol{d}_{\boldsymbol{i}} + \hat{\boldsymbol{d}}_{\boldsymbol{i}} - \boldsymbol{d}_{\boldsymbol{i}}\right)(\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b}_{\boldsymbol{i}}) \\ &= \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{d}_{\boldsymbol{i}}(\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b}_{\boldsymbol{i}}) + \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\boldsymbol{d}}_{\boldsymbol{i}} - \boldsymbol{d}_{\boldsymbol{i}}\right)\left(\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b}_{\boldsymbol{i}}\right)\end{aligned}$$

By WLLN,

$$\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{d}_{\boldsymbol{i}}(\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b}_{\boldsymbol{i}}) \xrightarrow{p} \text{E}\left[\boldsymbol{d}_{\boldsymbol{i}}(\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b}_{\boldsymbol{i}})\right]$$

$$\begin{aligned}\left\|\frac{1}{N}\sum_{i=1}^{N}\left(\hat{\boldsymbol{d}}_{\boldsymbol{i}} - \boldsymbol{d}_{\boldsymbol{i}}\right)\left(\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b}_{\boldsymbol{i}}\right)\right\| &\leq \frac{1}{N}\sum_{i=1}^{N}\left\|\left(\hat{\boldsymbol{d}}_{\boldsymbol{i}} - \boldsymbol{d}_{\boldsymbol{i}}\right)\left(\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b}_{\boldsymbol{i}}\right)\right\| \\ &\leq \frac{1}{N}\sum_{i=1}^{N}\left\|\hat{\boldsymbol{d}}_{\boldsymbol{i}} - \boldsymbol{d}_{\boldsymbol{i}}\right\|\left\|\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b}_{\boldsymbol{i}}\right\|\end{aligned}$$

As $\|\hat{\boldsymbol{d}}_{\boldsymbol{i}} - \boldsymbol{d}_{\boldsymbol{i}}\| \xrightarrow{p} 0$ and $\|\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b}_{\boldsymbol{i}}\| \xrightarrow{p} 0$, we have

$$\frac{1}{N}\sum_{i=1}^{N}\left(\hat{\boldsymbol{d}}_{\boldsymbol{i}} - \boldsymbol{d}_{\boldsymbol{i}}\right)\left(\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b}_{\boldsymbol{i}}\right) \xrightarrow{p} 0$$

Then, we have

$$\widehat{\text{COV}}(\hat{\boldsymbol{d}}_{\boldsymbol{i}}, \hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b}_{\boldsymbol{i}}) \xrightarrow{p} \text{E}\left[\boldsymbol{d}_{\boldsymbol{i}}(\hat{\boldsymbol{b}}_{\boldsymbol{i}} - \boldsymbol{b}_{\boldsymbol{i}})\right] = \text{E}\left[\left(\sum_{j=1}^{T}\mathbf{x}_{\mathbf{ij}}'\mathbf{z}_{\mathbf{ij}}\right)^{-1}\left(\sum_{j=1}^{T}\mathbf{z}_{\mathbf{ij}}' e_{ij}\boldsymbol{d}_{\boldsymbol{i}}\right)\right] \equiv Q_{xz}^{-1}Q_{zed}$$

B Supplementary tables and figures

Table 6: Comparison of Estimates for $\hat{\beta}_1[x_{ij}]$ with Simulated Data

	Uncorrelated Covariance					Correlated Covariance				
Method	Bias	SD	RMSE	$\frac{MeanSE}{SD}$	CR	Bias	SD	RMSE	$\frac{MeanSE}{SD}$	CR
N=250, T=250										
P2SLS	-0.001	0.020	0.028	0.981	0.952	0.126	0.021	0.129	1.017	0.000
FEIV	0.000	0.020	0.028	0.954	0.946	0.126	0.021	0.130	0.991	0.000
PCIV	-0.001	0.017	0.024	0.985	0.960	0.001	0.017	0.023	1.033	0.956
N=10, T=250										
P2SLS	0.000	0.089	0.126	0.920	0.912	0.111	0.102	0.182	0.786	0.672
FEIV	0.000	0.090	0.127	0.899	0.890	0.111	0.101	0.180	0.784	0.678
PCIV	0.000	0.083	0.118	0.939	0.912	0.005	0.083	0.117	0.932	0.924
N=250, T=10										
P2SLS	-0.001	0.035	0.049	1.179	0.968	0.125	0.039	0.137	1.167	0.178
FEIV	-0.002	0.038	0.054	0.950	0.930	0.125	0.041	0.137	0.917	0.108
PCIV	-0.017	0.048	0.070	0.902	0.904	-0.016	0.049	0.071	0.878	0.908

Note: Both bias and RMSE are multiplied by 100. P2SLS=Pooled Two-Stage Least Squares; FEIV=Fixed Effects Instrumental Variable; PCIV=Per-Cluster Instrumental Variable; RMSE= Root Mean Squared Error; $\frac{MeanSE}{SD}$=Ratio of mean standard errors divided by standard deviations; CR=Coverage Rate.

Table 7: First Stage Estimation Results Using Three Estimation Methods

	Just-Identified					
	Panel A (Unweighted)			Panel B (Weighted)		
	P2SLS	FEIV	PCIV	P2SLS	FEIV	PCIV
Log tax (Contemporaneous)	0.204	0.246	0.277	0.204	0.310	0.327
	(0.033)***	(0.027)***	(0.035)***	(0.029)***	(0.044)***	(0.036)***
	[0.033]***	[0.026]***	[0.006]***	[0.031]***	[0.044]***	[0.001]***
F-statistic	17.43	635.5	119.3	40.85	956.9	190.7

	Over-Identified					
	Panel C (Unweighted)			Panel D (Weighted)		
	P2SLS	FEIV	PCIV	P2SLS	FEIV	PCIV
Log tax (1-month lead)	0.008	-0.014	-0.003	0.047	0.040	0.115
	(0.012)	(0.040)	(0.078)	(0.023)**	(0.024)	(0.043)***
	[0.024]	[0.042]	[0.076]	[0.024]**	[0.090]	[0.033]***
Log tax (Contemporaneous)	0.204	0.204	0.273	0.204	0.202	0.233
	(0.033)***	(0.033)***	(0.032)***	(0.030)***	(0.029)***	(0.030)***
	[0.031]***	[0.033]***	[0.031]***	[0.031]***	[0.041]***	[0.029]***
Log tax (1-month lag)	-0.004	0.056	0.010	-0.040	0.066	-0.022
	(0.011)	(0.041)	(0.101)	(0.025)	(0.046)	(0.038)
	[0.025]	[0.044]	[0.096]	[0.025]	[0.044]	[0.041]
F-statistic	5.79	206.9	40.17	17.30	321.3	63.26

Note: Panels A is the estimation without using the weight while Panel B is estimated using the weight. All control for local unemployment rates and de-trend the data by directly (or akin to) including year-by-month fixed effects. Analytic standard errors appear in parentheses and state-level cluster-bootstrap standard errors appear in brackets below. Asterisks denote statistical significance at the ***1%, **5%, and *10% level.

Table 8: Analysis by Petroleum Administration For Defense District (PADD)

PADD Log price	New England	Central Atlantic	Lower Atlantic	Midwest	Gulf Coast	Rocky Mountain	West Coast
	-2.025	-1.880*	0.285	-0.469	-1.146***	0.642	-0.976*
	(1.657)	(1.121)	(0.503)	(0.593)	(0.439)	(1.609)	(0.593)
Observations	2,160	2,160	2,160	5,400	2,160	1,800	2,520
First-stage F-stat	380.5	24	630.7	92.7	143.8	41.2	150.6
Number of states	6	6	6	15	6	5	7

Note: All control for local unemployment rates and de-trend the data by directly (or akin to) including year-by-month fixed effects. State-clustered standard errors appear in parentheses. Asterisks denote statistical significance at the ***1%, **5%, and *10% level.

Table 9: Accounting for anticipatory behavior

	One lag of log prices as additional endogenous regressor					
	Panel A (Unweighted)			Panel B (Weighted)		
	P2SLS	FEIV	PCIV	P2SLS	FEIV	PCIV
Log price	-0.713	-0.879	-0.367	-0.447	-0.778	-0.624
(Contemporaneous)	(0.189)***	(0.438)**	(0.589)	(0.124)***	(0.434)*	(0.421)
Log price	0.169	-0.038	-0.396	0.135	-0.077	-0.168
(1-month lag)	(0.118)	(0.268)	(0.670)	(0.077)*	(0.129)	(0.432)
Cumulative effect	-0.544	-0.918	-0.764	-0.313	-0.855	-0.793
of log price	(0.136)***	(0.422)**	(0.590)	(0.114)***	(0.404)**	(0.421)*

	One lead of log prices as additional endogenous regressor					
	Panel C (Unweighted)			Panel D (Weighted)		
	P2SLS	FEIV	PCIV	P2SLS	FEIV	PCIV
Log price	0.352	-0.174	0.460	0.261	-0.350	-0.081
(1-month lead)	(0.110)***	(0.431)	(0.609)	(0.114)**	(0.394)	(0.414)
Log price	-0.717	-0.749	-1.253	-0.449	-0.506	-0.736
(Contemporaneous)	(0.194)***	(0.266)***	(0.737)*	(0.167)***	(0.186)***	(0.461)
Cumulative effect	-0.365	-0.923	-0.793	-0.189	-0.856	-0.819
of log price	(0.167)**	(0.423)**	(0.610)	(0.112)	(0.406)**	(0.414)**

	One lag and one lead of log prices as additional endogenous regressors					
	Panel E (Unweighted)			Panel F (Weighted)		
	P2SLS	FEIV	PCIV	P2SLS	FEIV	PCIV
Log price	0.371	-0.189	0.784	0.290	-0.387	0.240
(1-month lead)	(0.117)***	(0.426)	(0.514)	(0.119)**	(0.385)	(0.473)
Log price	-0.721	-0.709***	-1.254	-0.469	-0.371	-0.566
(Contemporaneous)	(0.190)***	(0.202)	(0.453)***	(0.148)***	(0.109)***	(0.353)
Log price	0.171	-0.020	-0.169	0.180	-0.097	-0.367
(1-month lag)	(0.111)	(0.272)	(0.603)	(0.061)***	(0.126)	(0.370)
Cumulative effect	-0.178	-0.918	-0.639	0.001	-0.855	-0.693
of log price	(0.144)	(0.420)**	(0.515)	(0.113)	(0.406)**	(0.473)

Note: All results from over-identified models with an additional lead and lag of log taxes used as excluded instruments. Panels A is the estimation without using the weight while Panel B estimated using the weight. Analytic standard errors appear in parentheses and state-level cluster-bootstrap standard errors appear in brackets below. Asterisks denote statistical significance at the ***1%, **5%, and *10% level.

Table 10: Accounting for anticipatory behavior

	One lag of log prices as additional endogenous regressor					
	Panel A (Unweighted)			Panel B (Weighted)		
	P2SLS	FEIV	PCIV	P2SLS	FEIV	PCIV
Cumulative effect of log prices	-0.544 (0.136)*** [0.141]***	-0.918 (0.422)** [0.399]**	-0.764 (0.590) [0.501]	-0.313 (0.114)*** [0.138]**	-0.855 (0.404)** [0.493]*	-0.793 (0.421)* [0.462]*
	One lead of log prices as additional endogenous regressor					
	Panel C (Unweighted)			Panel D (Weighted)		
Cumulative effect of log prices	-0.365 (0.167)** [0.187]*	-0.923 (0.423)** [0.4]**	-0.793 (0.610) [0.491]	-0.189 (0.112) [0.126]	-0.856 (0.406)** [0.495]*	-0.819 (0.414)** [0.46]*
	One lag and one lead of log prices as additional endogenous regressors					
	Panel E (Unweighted)			Panel F (Weighted)		
Cumulative effect of log prices	-0.178 (0.144) [0.149]	-0.918 (0.420)** [0.399]**	-0.639 (0.515) [0.509]	0.001 (0.113) [0.139]	-0.855 (0.406)** [0.473]*	-0.693 (0.473) [0.488]

Note: Cumulative effects are the sum of contemporaneous, lead, and/or lag values of prices, instrumented by taxes. All results from over-identified models with an additional lead and lag of log taxes used as excluded instruments. Analytic standard errors appear in parentheses and state-level cluster-bootstrap standard errors appear in brackets below. Asterisks denote statistical significance at the ***1%, **5%, and *10% level.

Table 11: Regression coefficients of $\widehat{\mathbf{b_i}}$ on cluster-level first-stage t-statistics from Figure 10

T-statistic on Log tax	Panel A	Panel C
Contemporaneous	-0.078	-0.156
	(0.245)	(0.193)
Constant	0.069	1.519
	(2.672)	(3.178)
Observations	51	51
Volume weighted	No	Yes
R-squared	0.002	0.020

Note: Panels A and C correspond to the panels from Table 2. Robust standard errors appear in parentheses. Asterisks denote statistical significance at the ***1%, **5%, and *10% level.

Table 12: Regression coefficients of $\widehat{\mathbf{b_i}}$ on cluster-level first-stage t-statistics from Figure 10

	Panel B			Panel D		
Log Tax						
1-month lead	-0.429			-0.582		
	(0.518)			(0.441)		
Contemporaneous		-0.401		-0.517		
		(0.473)			(0.367)	
1-month lag			0.851			0.748
			(0.427)*			(0.327)**
Constant	-0.407	-0.748	-0.775	-0.334	-0.617	0.999
	(0.703)	(0.483)	(0.457)*	(0.753)	(0.486)	(0.443)**
Observations	51	51	51	51	51	51
R-squared	0.014	0.014	0.075	0.030	0.033	0.080

Note: Panels B and D correspond to the panels from Table 2. Robust standard errors appear in parentheses. Asterisks denote statistical significance at the ***1%, **5%, and *10% level.

Table 13: Estimated elasticities among states in which the instrument is strong (LATEs)

	(Unweighted)			(Weighted)		
	P2SLS	FEIV	PCIV	P2SLS	FEIV	PCIV
Panel A: Just-identified contemporaneous specification						
Log price (contemporaneous)	-0.575	-0.762	-0.791	-0.425	-0.757	-0.681
	(0.186)***	(0.418)*	(0.526)	(0.150)***	(0.379)**	(0.430)
	[0.185]***	[0.409]*	[0.517]	[0.156]**	[0.472]	[0.422]
Panel B: Over-identified contemporaneous specification						
Log price (contemporaneous)	-0.554	-0.762	-0.697	-0.394	-0.762	-0.649
	(0.184)***	(0.423)*	(0.446)	(0.133)***	(0.383)**	(0.433)
	[0.189]***	[0.413]*	[0.439]	[0.148]***	[0.475]	[0.424]
Panel C: Over-Identified with additional lag of log prices						
Log price (cumulative)	-0.488	-0.762	-0.669	-0.272	-0.762	-0.636
	(0.139)***	(0.384)**	(0.550)	(0.110)***	(0.426)*	(0.417)
	[0.145]***	[0.416]*	[0.431]	[0.129]**	[0.477]	[0.426]
Panel D: Over-Identified with additional lead of log prices						
Log price (cumulative)	-0.212	-0.762	-0.662	-0.146	-0.762	-0.636
	(0.154)	(0.385)**	(0.518)	(0.104)	(0.428)*	(0.368)*
	[0.157]	[0.418]*	[0.429]	[0.110]	[0.480]	[0.430]
Panel E: Over-Identified with additional lag and lead of log prices						
Log price (cumulative)	-0.146	-0.762	-0.647	0.041	-0.762	-0.623
	(0.140)	(0.386)**	(0.368)*	(0.108)	(0.431)*	(0.316)**
	[0.147]	[0.421]*	[0.451]	[0.130]	[0.482]	[0.429]

Note: Cumulative effects are the sum of contemporaneous, lead, and/or lag values of prices, instrumented by taxes. Analytic standard errors appear in parentheses and state-level cluster-bootstrap standard errors appear in brackets below. Asterisks denote statistical significance at the ***1%, **5%, and *10% level.

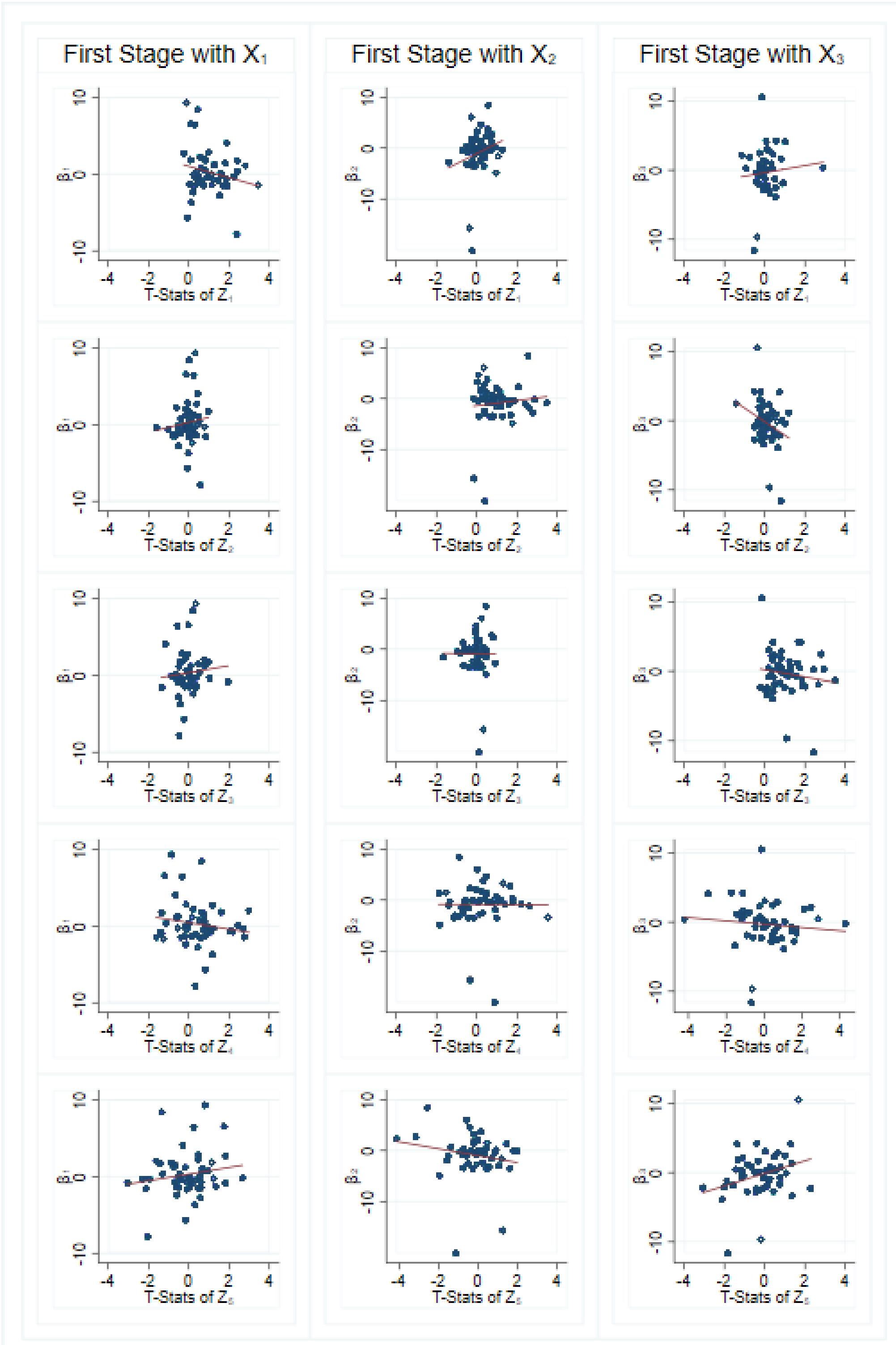
Table 14: Reduced-Form Estimation Results Using Three Estimation Methods

	Just-Identified					
	Panel A (Unweighted)			Panel B (Weighted)		
	P2SLS	FEIV	PCIV	P2SLS	FEIV	PCIV
Log tax	-0.148	-0.228	-0.259	-0.094	-0.263	-0.336
(Contemporaneous)	(0.053)***	(0.109)**	(0.129)**	(0.038)**	(0.112)**	(0.176)*
	[0.055]***	[0.101]**	[0.121]**	[0.040]**	[0.001]***	[0.162]**

	Over-Identified					
	Panel C (Unweighted)			Panel D (Weighted)		
	P2SLS	FEIV	PCIV	P2SLS	FEIV	PCIV
Log tax	0.067	-0.017	-0.040	0.033	-0.077	-0.057
(1-month lead)	(0.026)**	(0.080)	(0.120)	(0.027)	(0.083)	(0.150)
	[0.030]**	[0.076]	[0.111]	[0.030]	[0.090]	[0.141]
Log tax	-0.147	-0.147	-0.221	-0.095	-0.091	-0.072
(Contemporaneous)	(0.053)***	(0.053)***	(0.064)***	(0.037)**	(0.039)**	(0.079)
	[0.039]***	[0.055]***	[0.061]***	[0.039]**	[0.041]**	[0.068]
Log tax	0.035	-0.063	0.008	0.050	-0.096	-0.206
(1-month lag)	(0.028)	(0.052)	(0.099)	(0.013)***	(0.037)**	(0.102)**
	[0.016]	[0.053]	[0.095]	[0.016]***	[0.044]**	[0.087]**
Cumulative Effect	-0.045	-0.228	-0.253	-0.011	-0.264	-0.337
	(0.030)	(0.112)**	(0.132)*	(0.022)	(0.114)**	(0.179)*
	[0.043]	[0.105]**	[0.128]**	[0.041]	[0.125]**	[0.167]**

Note: Panels A is the estimation without using the weight while Panel B is estimated using the weight. Analytic standard errors appear in parentheses and state-level cluster-bootstrap standard errors appear in brackets below. Asterisks denote statistical significance at the ***1%, **5%, and *10% level.

Figure 12: Relationship between state-specific elasticities and first-stage t-statistics in the over-identified specification with leads and lags of prices



Note: X_1, X_2, and X_3 respectively refer to the lag, contemporaneous, and lead values of log prices. Z_1, Z_2, Z_3, Z_4, and Z_5 refer to the 2nd lag, 1st lag, contemporaneous, 1st lead, and 2nd lead of Log tax, respectively.

Figure 13: Reduced form weighting under standard and PC approaches

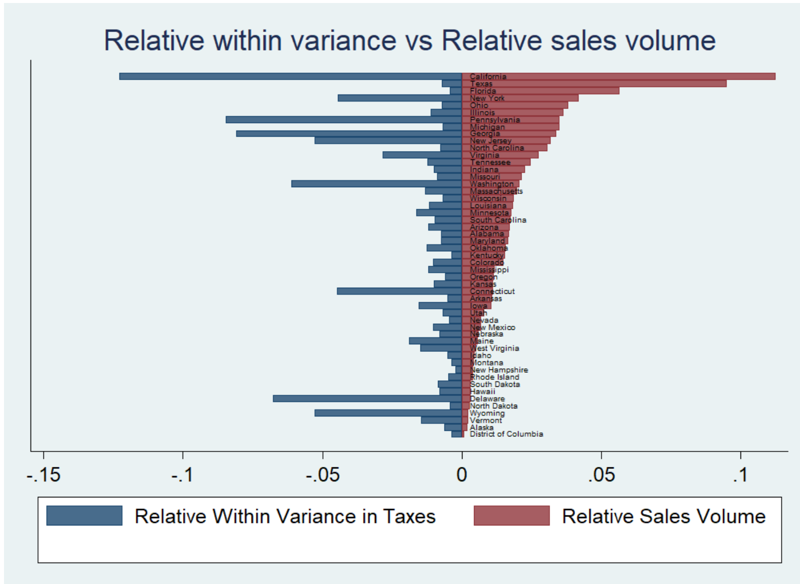
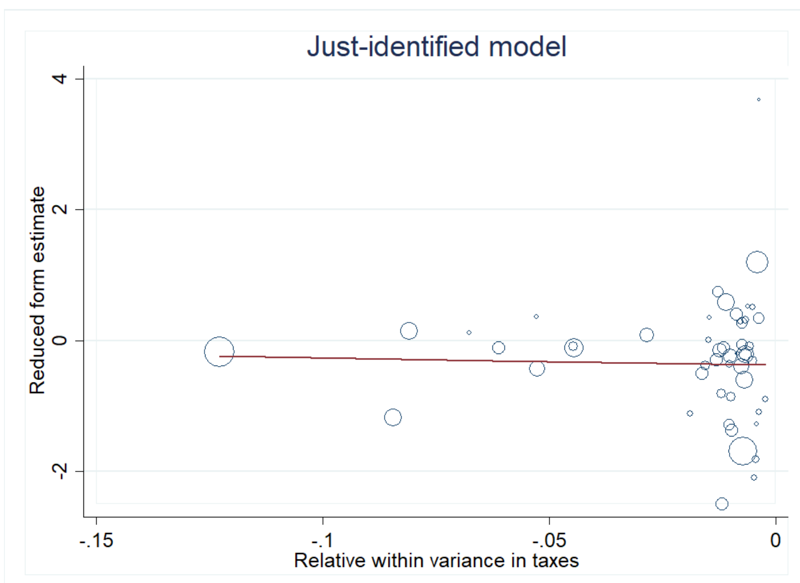


Figure 14: Reduced form weighting under standard and PC approaches



C Replication of Coglianese et al. (2017) adding only new methods

Panel A presents the estimates of contemporaneous deviations in log gasoline prices on log gasoline sales, where log prices are just identified by the contemporaneous deviation in log gasoline taxes. Panel B is similar except that in this case lag and lead taxes are used to instrument for the contemporaneous log prices. Panels C and D repeat the same exercises, except that we weight observations by the time average of the volume of sales in each state. We present both the analytic state-clustered standard errors as well as state-cluster bootstrap standard errors below. For PCIV, the analytic standard errors are estimated as described in section 3.2.

Across specifications, the sign of the estimates is generally consistent. However, the magnitude of the effects and precision with which they are estimated vary substantially with estimated elasticities ranging from -0.69 to -2.97. There are some note worthy patterns. First, the robustness of the PCIV estimator is not free. The additional robustness comes at the cost of efficiency. However, better prediction in the first stage can help substantially. Moving from the just identified specifications in Panels A and C to the overidentified specifications in Panels B and D, the standard errors fall by 23 to 54 percent.

Secondly, weighting brings the estimates from the three approaches much closer together. Whereas in the unweighted regressions shown in Panels A and B, the estimates range from -0.78 to -2.97 and -0.75 to -1.135 respectively, when we weight the point estimate ranges are much tighter from -0.71 to -1.47 in Panel C and -0.69 to -0.79 in Panel D. Finally, the analytic and bootstrap standard errors are close with neither universally smaller or larger than the other, providing reassurance that with a relatively long panel, either estimation approach may be suitable.

Table 15: Summary of Results Using Three Estimation Methods

	just Identified					
	Panel A (Unweighted)			Panel B (Weighted)		
	P2SLS	FEIV	PCIV	P2SLS	FEIV	PCIV
Log price	-1.135	-0.776	-2.970	-0.714	-0.801	-1.472
(contemporaneous)	(0.250)***	(0.39)**	(1.692)*	(0.209)***	(0.326)**	(1.008)
	[0.244]***	[0.410]*	[1.639]*	[0.210]***	[0.380]**	[1.019]

	Over-Identified					
	Panel A (Unweighted)			Panel B (Weighted)		
	P2SLS	FEIV	PCIV	P2SLS	FEIV	PCIV
Log price	-1.135	-0.745	-0.935	-0.743	-0.794	-0.686
(contemporaneous)	(0.243)***	(0.383)*	(0.810)	(0.213)***	(0.327)**	(0.776)
	[0.249]***	[0.403]*	[0.740]	[0.192]***	[0.374]**	[0.746]

Note: These regression results are selected from Coglianese et al. (2017). Panels A is the estimation without using the weight while Panel B is estimated using the weight. Analytic standard errors appear in parentheses and state-level cluster-bootstrap standard errors appear in brackets below. Asterisks denote statistical significance at the ***1%, **5%, and *10% level.

Table 16: Summary of Results Using Three Estimation Methods

	with Lead and Lag					
	Panel A (Unweighted)			Panel B (Weighted)		
	P2SLS	FEIV	PCIV	P2SLS	FEIV	PCIV
Log price	0.540	0.627	-0.135	0.693	0.391	0.678
(1-month lead)	(0.171)***	(0.284)**	(2.124)	(0.203)***	(0.349)	(1.313)
	[0.175]***	[0.330]*	[2.143]	[0.199]***	[0.352]	[1.334]
Log price	-1.152	-1.167	-3.067	-0.693	-0.687	-1.058
(contemporaneous)	(0.250)***	(0.401)***	(2.277)	(0.265)***	(0.247)***	(1.797)
	[0.247]***	[0.259]***	[2.179]	[0.259]***	[0.252]***	[1.667]
Log price	0.244	-0.187	-0.235	0.350	-0.486	-1.004
(1-month lag)	(0.197)	(0.282)	(1.294)	(0.232)	(0.353)	(1.232)
	[0.206]	[0.365]	[1.250]	[0.238]	[0.389]	[1.167]
Cumulative Effect	-0.368	-0.728	-3.437	0.349	-0.781	-1.384
	(0.239)	(0.398)*	(2.153)	(0.529)	(0.334)***	(1.039)
	[0.238]	[0.421]	[2.086]	[0.522]	[0.389]**	[1.045]

	with Additional Leads and Lags					
	Panel A (Unweighted)			Panel B (Weighted)		
	P2SLS	FEIV	PCIV	P2SLS	FEIV	PCIV
Log price	0.553	0.655	-0.333	0.643	0.284	-0.543
(1-month lead)	(0.183)***	(0.319)**	(0.563)	(0.205)***	(0.326)	(0.507)
	[0.184]***	[0.335]*	[0.515]	[0.189]***	[0.317]	[0.515]
Log price	-1.152	-1.155	-1.121	-0.736	-0.643	-0.328
(contemporaneous)	(0.251)***	(0.399)***	(0.548)**	(0.255)***	(0.252)**	(0.438)
	[0.242]***	[0.266]***	[0.536]**	[0.214]***	[0.257]**	[0.419]
Log price	0.239	-0.192	0.377	0.311	-0.420	0.007
(1-month lag)	(0.195)	(0.281)	(0.527)	(0.224)	(0.318)	0.671
	[0.205]	[0.364]	[0.526]	[0.191]	[0.296]	[0.637]
Cumulative Effect	-0.360	-0.691	-1.077	0.219	-0.779	-0.865
	(0.241)	(0.393)*	(0.689)	(0.509)	(0.334)**	(0.719)
	[0.233]	[0.415]	[0.918]	[0.373]	[0.381]**	[0.685]

Note: These regression results are selected from Coglianese et al. (2017). Panels A is the estimation without using the weight while Panel B is estimated using the weight. Analytic standard errors appear in parentheses and state-level cluster-bootstrap standard errors appear in brackets below. Asterisks denote statistical significance at the ***1%, **5%, and *10% level.