Estimating the Price Elasticity of Gasoline Demand in Correlated Random Coefficient Models with Endogeneity

Michael Bates* and Seolah Kim†

November 7, 2022

Abstract

We propose a per-cluster instrumental variables approach (PCIV) for estimating linear correlated random coefficient models in the presence of contemporaneous endogeneity and two-way fixed effects. This approach estimates heterogeneous effects and aggregates them to population averages. We demonstrate consistency, showing robustness over standard estimators, and provide analytic standard errors for robust inference. In Monte Carlo simulation, PCIV performs relatively well in finite samples in either dimension. We apply PCIV in estimating the price elasticity of gasoline demand using state fuel taxes as instrumental variables. We find significant elasticity heterogeneity and more elastic gasoline demand on average than with standard estimators. Keywords: instrumental variables, per-cluster estimation, heterogeneous effects, population average effects, local average treatment effects.

*Department of Economics, University of California at Riverside, Riverside, CA 92521, United States of America; michael.bates@ucr.edu

†Department of Economics and Management, Albion College, Albion, MI, 49224, United States of America; skim@albion.edu

[0]We wish to thank Soren Anderson, Otavio Bartalotti, Ozkan Eren, Ruoyao Shi, Gary Solon, Aman Ullah, Jeff Wooldridge and participants at the Midwestern Econometrics Group Annual Meeting and the International Association of Applied Econometrics Summer Meetings for helpful comments. We thank David Ford for excellent research assistance. All errors are our own.

1 Introduction

The price elasticity of the demand for gasoline is an integral parameter for climate mitigation policy, modeling energy and automotive markets, and urban planning (Dahl, 1986; Dahl and Sterner, 1991; Espey, 1998). However, several econometric issues complicate the estimation of this important parameter. First, gasoline prices and the volume purchased likely depend upon each other through the interaction of supply and demand forces. Second, quantities, prices, and taxes trend over time. Finally, there is marked heterogeneity across this market with geography, industry composition, population density, demographic composition, transportation substitutes, economic climates, regulations, and tax rates each varying across states. Assuming homogeneous elasticities across this landscape seems questionable (Wadud et al., 2010; Frondel et al., 2012; Blundell et al., 2012; Hausman and Newey, 2016; Levin et al., 2017), and ignoring this likely heterogeneity when using standard estimators may cause bias. As a result, we develop an estimation strategy that can handle all three complications with a valid instrument. Further, it has the added benefit of providing the distribution of heterogeneous effects rather than only the mean.

The method we introduce to estimate the price elasticity of gasoline demand has broader applicability than our current context.[1] Empirical work often aims to identify the population average effect (PAE)—the average causal relationship between two variables over an entire population of interest. Heterogeneous effects within the population is inherent in these conceptualizations, and often holds economic significance (Heckman and Vytlacil, 1998).[2] These correlated random coefficient (CRC) models, however, add complications for estimation even when regressors are otherwise exogenous (Wooldridge, 2005; Arellano and Bonhomme, 2011;

[1]This method is well suited to settings in which data is grouped into moderately large clusters, endogeneity is plausible, and where heterogeneous effects are likely such as in development and political economy (Miguel et al., 2004), industrial organization and energy markets, (Bushnell et al., 2008), public finance and education (Jackson et al., 2016), labor economics (Kostol and Mogstad, 2014), and trade (Feyrer, 2019).

[2]While Heckman and Vytlacil (1998) coined the helpful phrase "correlated random coefficients models," similar models originate much earlier with examinations appearing in Rubin (1950); Klein (1953); Kuh (1959); Swamy (1971); Mundlak (1978); Raj et al. (1980), and Chamberlain (1992).

Graham and Powell, 2012; Bates et al., 2014). This effect heterogeneity may produce even further complications in so-called two-way fixed effects models with multiple time periods (De Chaisemartin and d'Haultfoeuille, 2020; Borusyak et al., 2021; Callaway and Sant'Anna, 2021; Goodman-Bacon, 2021; Sun and Abraham, 2021; Wooldridge, 2021).

Applied economic researchers, however, often work in settings where some explanatory variables may be correlated with the error term (for instance due to omitted variables or simultaneity). In those cases when a valid instrument is available, it is typical for researchers to turn to pooled two-stage least squares (P2SLS) or fixed effects instrumental variables (FEIV) estimators. Murtazashvili and Wooldridge (2008) provide the conditions under which a general class of FEIV estimators consistently estimate PAEs with endogenous regressors. They find that in CRC models, consistency requires the assumption that the heterogeneous slopes are uncorrelated with the covariance between the instruments and endogenous regressors.[3]

This restriction may not hold in many cases. Consider the local average treatment effect (LATE) framework from Imbens and Angrist (1994). The instrument is not relevant for some individual clusters within the population, and treatment effects differ on average by whether the individual is moved toward treatment by the instrument. Such a setting would prevent FEIV from consistently estimating the average treatment effect (ATE). Even in a population in which all are influenced by a valid instrument, if effect heterogeneity correlates with the strength of the instrument within-cluster, FEIV would continue to fail to estimate the ATE among a population of compliers with the instrument (the LATE).

Fernández-Val and Lee (2013) introduce an approach of estimating cluster-specific slopes and averaging over them.[4] They iteratively apply generalized method of moments (GMM) estimators separately to the time series of each cluster. As the GMM approach is asymptotically biased, they introduce bias corrections. Per-cluster estimation, however, may be taxing on the data, particularly when there are many covariates and relatively short panels.

[3]P2SLS requires a similar restriction for consistency.

[4]Kelejian (1974) also demonstrates that treating these random coefficients as fixed effects is useful for identification in such models with endogenous covariates.

Designating some conditioning variables to have common coefficients may help. Fernández-Val and Lee (2013) preserve degrees of freedom for estimation by netting out the effects of covariates with common coefficients. They estimate this vector of common coefficients in an additional step within each iteration using time averages. Accordingly, parameters on variables that vary only across time cannot be identified, thus ruling out common applications such as two-way fixed effects models.

Our approach, which we term per-cluster instrumental variables (PCIV), in simple settings applies 2SLS estimation on a cluster-by-cluster basis before aggregating them. The general PCIV estimator accommodates differential cluster sizes and covariates with both heterogeneous and homogeneous slopes, allowing the dimension of the vector of covariates to exceed T, as in the presence of two-way fixed effects. We show that the asymptotic distribution of the estimator is normal and centered at the true parameter value without placing restrictions on the correlation between the explanatory variables and random coefficients. In each stage, we do this by first netting out cluster-specific slopes from the left-hand-side variables and variables with homogeneous slopes. Second, we pool over those residuals to consistently estimate homogeneous coefficients using both time-series and cross-sectional variation. We use those estimated common parameters to net out the effects of the covariates from the left-hand-side variables, which we regress on the variables with heterogeneous slopes again by cluster. We aggregate over those cluster-specific slopes using appropriate weights to estimate relevant parameters such as the PAE or LATE. We propose accompanying analytic standard errors that are robust to arbitrary within-cluster correlation. This approach can be viewed as extending an estimator described in Wooldridge (2010 Section 11.7.2) to account for endogeneity.

Due to the long panel and likely effect heterogeneity, the method is ideal for estimating the price elasticity of gasoline demand in the United States. We follow Davis and Kilian (2011); Blundell et al. (2012); Li et al. (2014); Hausman and Newey (2016); Coglianese et al. (2017);

and Hoderlein and Vanhems (2018) in using state fuel taxes as our source of exogenous variation and use them to instrument for average state gasoline prices. Given the variation in state changes to gasoline taxes and the likely heterogeneity in gasoline demand, ignoring such heterogeneity may lead to inconsistent estimates. Accordingly, we perform estimation using standard approaches from the literature and the more robust PCIV approach. As earlier work relies on aging and discontinued data series, we update and bolster the analysis with novel data collection.[5]

This analysis allows us to demonstrate additional advantages of PCIV estimation. First, we find that both the magnitude and significance of the results are sensitive to the methods used. The robust PCIV approach estimates a population average elasticity of the demand for gasoline of -0.4 to -0.6 whereas FEIV estimates range from -0.75 to -0.95. Second, PCIV provides the distribution of state-specific elasticities and first-stage tax-pass-through rates in addition to the mean effects. We see substantial heterogeneity in both. The standard deviation of first- and second-stage state-specific elasticities are respectively 42% and 38% of the means. Third, in seeming violation of the key assumption for consistency of P2SLS and FEIV, we observe a meaningful correlation between the first-stage variation and the state-specific elasticity estimates. Fourth, we find significant divergence between the implicit weighting of FEIV and the natural market share weights we employ with PCIV. Finally, as the instrument is not universally strong, we provide a LATE (-0.56) for the 46 states in which the instrument is strong, which is very close to the PAE though the LATE carries smaller standard errors.

We organize the remainder of the paper as follows. Section 2 introduces the econometric model and the proposed estimator. Section 3 provides the main consistency results. Section 4 contains a Monte Carlo study showing that the PCIV estimator performs comparably to

[5]Along with Davis and Kilian (2011); Li et al. (2014); and Coglianese et al. (2017), we rely on the U.S. Department of Energy, Energy Information Administration, 'Petroleum Marketing Monthly Report: Gasoline Prices by Formulation, Grade, Sales Type' for gasoline prices by state for 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.

standard estimators in the ideal setting but outperforms P2SLS and FEIV when there is correlation between effect heterogeneity and first-stage variation. We estimate the price elasticity of gasoline demand in Section 5 before concluding in Section 6.

2 Model specification and proposed estimator

Consider a correlated random effects model as follows:

$$
\begin{aligned}
y_{ij} &= \mathbf{x}_{1ij}\mathbf{b}_i + \mathbf{x}_{2ij}\boldsymbol{\delta} + e_{ij}, \\
\mathbf{x}_{1ij} &= \mathbf{z}_{ij}\boldsymbol{\gamma}_i + \mathbf{x}_{2ij}\boldsymbol{\eta} + u_{ij},\ i = 1, ..., n; j = 1, ..., T,
\end{aligned}
\tag{1}
$$

where y_{ij} is the dependent variable of the j^{th} unit of cluster i and e_{ij} is an idiosyncratic error. The $1 \times K$ vector, $\mathbf{x}_{1ij}$, includes 1 as well as covariates that may be endogenous and are allowed to vary both between and within clusters. We assume the presence of $\mathbf{z}_{ij}$, a $1 \times L$ ($L \geq K$) vector of instrumental variables, which are excluded from the second-stage equation. A vital 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. We are primarily interested in estimating the PAE, $\boldsymbol{\beta}$. This vector indicates the heterogeneous effects that vary by cluster and may be correlated with $\mathbf{x}_{1ij}$. We allow for a $1 \times H$ vector of exogenous covariates, $\mathbf{x}_{2ij}$, with $\boldsymbol{\delta}$ and $\boldsymbol{\eta}$, $H \times 1$ vectors of homogeneous slopes. We use a standard formulation of the first stage except that we allow the relationship between the instruments and the endogenous variables to vary across clusters, as $\boldsymbol{\gamma}_i = \boldsymbol{\Gamma} + \mathbf{g}_i$ such that $\mathrm{E}(\mathbf{g}_i) = 0$.

2.1 Proposed estimator in a simple model

Before handling the general case, we illustrate the intuition behind our approach by proposing a per-cluster instrumental variables estimator for a simplified setting (PCIVS) in which all clusters are of equal size (such that all the weights are equal) and there are no influential exogenous covariates. PCIVS estimation follows three steps:

1. Use OLS to regress each element of $\mathbf{x}_{1ij}$ on $\mathbf{z}_{ij}$ separately for each cluster i, obtaining the vector of fitted values, $\hat{\mathbf{x}}_{1ij}$.

2. Use OLS to regress y_{ij} on $\hat{\mathbf{x}}_{1ij}$ within each cluster, obtaining the vector of cluster-specific estimates, $\hat{\mathbf{b}}_{i,PCIVS}$.[6]

3. Average over $\hat{\mathbf{b}}_{i,PCIVS}$ for the population to provide the PAE estimate, $\hat{\boldsymbol{\beta}}_{PCIVS}$.

Accordingly, we may write the cluster-specific PCIV estimates as the following:

$$\hat{\mathbf{b}}_{i,PCIVS} = \boldsymbol{\beta} + \mathbf{d}_i + \left(\sum_{j=1}^{T} \mathbf{x}'_{1ij} \mathbf{H}_{\mathbf{z}_{ij}} \mathbf{x}_{1ij} \right)^{-1} \sum_{j=1}^{T} \mathbf{x}'_{1ij} \mathbf{H}_{\mathbf{z}_{ij}} e_{ij}, \tag{2}$$

where $\mathbf{H}_{\mathbf{z}_{ij}} = \mathbf{z}_{ij} \left(\sum_{j=1}^{T} \mathbf{z}'_{ij} \mathbf{z}_{ij} \right)^{-1} \mathbf{z}'_{ij}$. The 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 $\mathbf{x}_{1ij}$ used in the per-cluster regressions, even if it is correlated with both $\mathbf{z}_{ij}$ and $\mathbf{x}_{1ij}$. Maintaining random sampling and homogeneous cluster size, we then average over the estimated cluster-specific slopes to provide our PCIV estimate of $\boldsymbol{\beta}$.

This case restricts the model studied in Murtazashvili and Wooldridge (2008) by disallowing exogenous regressors and time effects. Common parameters pose a significant issue for the simple estimator above and the proposed estimator in Fernández-Val and Lee (2013). In both, estimation of all parameters occurs on a per-cluster basis. Accordingly, each additional parameter significantly reduces the degrees of freedom in each regression and the number of parameters cannot exceed T, disallowing the inclusion of time effects.

Furthermore, the class of FEIV estimators discussed in Murtazashvili and Wooldridge (2008) require additional assumptions for consistency. Consider the following representation

[6] The "statsby" command in Stata 15 (Stata Corp, 2015) allow these first two steps to be completed in a singular line of code.

of the estimate of $\boldsymbol{\beta}$ from 2SLS applied to the time-demeaned covariates and instruments:

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{FEIV}} = \boldsymbol{\beta} + \left(\sum_{i=1}^{n}\sum_{j=1}^{T}\ddot{\mathbf{x}}'_{1ij}\mathbf{H}_{\mathbf{z}}\ddot{\mathbf{x}}_{1ij}\right)^{-1}\left[\sum_{i=1}^{n}\sum_{j=1}^{T}\ddot{\mathbf{x}}'_{1ij}\mathbf{H}_{\mathbf{z}}\ddot{\mathbf{x}}_{1ij}\mathbf{d}_i + \sum_{i=1}^{n}\sum_{j=1}^{T}\ddot{\mathbf{x}}'_{1ij}\mathbf{H}_{\mathbf{z}}\ddot{e}_{ij}\right], \tag{3}$$

where $\mathbf{H}_{\mathbf{z}} = \ddot{\mathbf{z}}_{ij}\left(\sum_{i=1}^{n}\sum_{j=1}^{T}\ddot{\mathbf{z}}'_{ij}\ddot{\mathbf{z}}_{ij}\right)^{-1}\ddot{\mathbf{z}}'_{ij}$. Murtazashvili and Wooldridge (2008) observe that they also "need assumptions such that $\ddot{\mathbf{z}}_{ij}$ is uncorrelated with $\ddot{\mathbf{x}}_{1ij}\mathbf{d}_i$."[7] This is required in a FEIV estimation because FEIV tends to apply relatively more weight to clusters with more identifying first-stage variation.

If the instrument's strength is related to the heterogeneous effects (allowing a LATE to differ from the ATE), $\mathrm{E}[(\ddot{\mathbf{z}}'_{ij}\ddot{\mathbf{x}}_{1ij})^{-1}\ddot{\mathbf{z}}'_{ij}\ddot{\mathbf{x}}_{1ij}\mathbf{d}_i] \neq 0$ and FEIV may be inconsistent. Further, a correlation between the strength of the instrument and the heterogeneous effects at the intensive margin may cause FEIV to fail to estimate even the LATE consistently. It is this assumption that estimators in Fernández-Val and Lee (2013) and our paper avoid without imposing a homogeneous first-stage relationship.[8]

2.2 Proposed estimator in a general model

Empirical contexts are wide-ranging. Clusters are often different sizes and inclusion of time effects and other exogenous covariates is ubiquitous. Here, we introduce an estimator which researchers may apply to the multiplicity of contexts encompassed in equation 1. For ease of exposition, we use matrix notation such that $\mathbf{y}_i$ is $T \times 1$ vector of a dependent variable, $\mathbf{X}_{1i}$ is a $T \times K$ matrix of endogenous variables, $\mathbf{X}_{2i}$ is a $T \times H$ matrix of exogenous covariates, and $\mathbf{Z}_i$ is a $T \times L$ matrix of instruments. We list the steps for estimating the each stage

[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. 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.

[8]Murtazashvili and Wooldridge (2016), and Laage (2019) introduce estimators for CRC models with endogenous regressors, but 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.

of the general PCIV estimator below using OLS. The first-stage estimation is related to the estimators proposed in Pesaran (2006) and more closely follows an estimator described in Wooldridge (2010).

First-stage estimation:

1. Per-cluster, regress $\mathbf{x}_{1ij}$ and $\mathbf{x}_{2ij}$ on $\mathbf{z}_{ij}$, saving the residuals $\tilde{\mathbf{X}}_{1i} = \mathbf{M}_{\mathbf{Z}_i}\mathbf{X}_{1i}$ and $\tilde{\mathbf{X}}_{2i} = \mathbf{M}_{\mathbf{Z}_i}\mathbf{X}_{2i}$ where $\mathbf{M}_{\mathbf{Z}_i} = \mathbf{I}_T - \mathbf{Z}_i(\mathbf{Z}_i'\mathbf{Z}_i)\mathbf{Z}_i'$. This step allows us to eliminate $\boldsymbol{\gamma}_i$ when we estimate $\boldsymbol{\eta}$ in the second step.

This step is akin to the common practice of detrending data to accommodate cluster-specific linear time trends.

2. To consistently estimate $\boldsymbol{\eta}$, regress the residuals $\tilde{\mathbf{x}}_{1ij}$ on $\widetilde{\mathbf{x}}_{2ij}$ pooling over clusters to estimate $\hat{\boldsymbol{\eta}} = (\sum_{i=1}^{n} \mathbf{X}_{2i}'\mathbf{M}_{\mathbf{Z}_i}\mathbf{X}_{2i})^{-1} \sum_{i=1}^{n} \mathbf{X}_{2i}'\mathbf{M}_{\mathbf{Z}_i}\mathbf{X}_{1i}$.

In the analogous example of handling cluster-specific time trends, this step is similar to pooled OLS on detrended data.

3. Estimate $\boldsymbol{\gamma}_i$ per-cluster by regressing $(\mathbf{x}_{1ij} - \mathbf{x}_{2ij}\hat{\boldsymbol{\eta}})$ on $\mathbf{z}_{ij}$. Then, we can construct $\hat{\mathbf{X}}_{1i} = \mathbf{Z}_i\hat{\gamma}_i + \mathbf{X}_{2i}\hat{\boldsymbol{\eta}}$, where $\hat{\gamma}_i = (\mathbf{Z}_i'\mathbf{Z}_i)^{-1}\mathbf{Z}_i'(\mathbf{X}_{1i} - \mathbf{X}_{2i}\hat{\boldsymbol{\eta}})$.

Here, we use OLS within each cluster to estimate cluster-specific slops after accounting for covariates with homogeneous slopes.

Second-stage estimation

Second-stage estimation in steps 1 - 3 are directly comparable to the same steps in the first stage, except that y_{ij} is substituted in for $\mathbf{x_{1ij}}$ and $\hat{\mathbf{x}}_{\mathbf{1ij}}$ is substituted in for $\mathbf{z}_{ij}$.

1. For the second stage, we regress y_{ij} and $\mathbf{x}_{2ij}$ on $\hat{\mathbf{x}}_{1ij}$ per-cluster, obtaining the residuals $\dot{\mathbf{y}}_i = \mathbf{M}_{\hat{\mathbf{X}}_{1i}}\mathbf{y}_i$ and $\dot{\mathbf{X}}_{2i} = \mathbf{M}_{\hat{\mathbf{X}}_{1i}}\mathbf{X}_{2i}$, where $\mathbf{M}_{\hat{\mathbf{X}}_{1i}} = \mathbf{I}_T - \hat{\mathbf{X}}_{1i}(\hat{\mathbf{X}}_{1i}'\hat{\mathbf{X}}_{1i})^{-1}\hat{\mathbf{X}}_{1i}'$.

2. Regressing the residuals $\dot{y}_{ij}$ on $\dot{\mathbf{x}}_{2ij}$ pooling over clusters allows us to eliminate $\mathbf{b}_i$ when estimating $\hat{\boldsymbol{\delta}} = (\sum_{i=1}^{n} \mathbf{X}_{2i}'\mathbf{M}_{\hat{\mathbf{X}}_{1i}}\mathbf{X}_{2i})^{-1} \sum_{i=1}^{n} \mathbf{X}_{2i}'\mathbf{M}_{\hat{\mathbf{X}}_{1i}}\mathbf{X}_{1i}$.

3. The heterogeneous slopes $\hat{\mathbf{b}}_i = (\hat{\mathbf{X}}_{1i}'\hat{\mathbf{X}}_{1i})^{-1}\hat{\mathbf{X}}_{1i}'(\mathbf{y}_i - \mathbf{X}_{2i}\hat{\boldsymbol{\delta}})$ can be consistently estimated by regressing $(y_{ij} - \mathbf{x}_{2ij}\hat{\boldsymbol{\delta}})$ on $\hat{\mathbf{x}}_{1ij}$ per cluster.

4. Averaging over $\hat{\mathbf{b}}_i$ obtains $\hat{\boldsymbol{\beta}}_{PCIV} = \sum_{i=1}^{n} w_i \hat{\mathbf{b}}_i$.

The general PCIV estimator allows for flexibly weighting clusters and models where the dimension of covariates may be large, including when $K + H > T$. Researchers often include additional exogenous variables in their models, which they assume to be exogenous. In some instances the instrument may only be exogenous conditional on such covariates. Time effects provide one such example, where we may expect the outcome, endogenous, and instrumental variables to each trend over time. Covariates that vary across and within clusters may also be relevant. However, increasing the dimension of covariates with random slopes in the simple model also entails reducing the degrees of freedom in each cluster-level regression with the estimators presented above.

There are a variety of contexts in which heterogeneously weighting cluster-specific estimates may be attractive. In many panel data settings, clusters represent individuals from the population of interest about whom we have multiple observations. Under random sampling of these individuals we may wish to afford each individual equal weight. However, many popular panel data sets are not random samples of the population.[9]

Such nonrandom sampling schemes would, in turn, lead averaged effects to overweight the populations which are over-represented, obscuring the true PAE. Solon et al. (2015) discuss 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. Researchers can still uncover a consistent estimate of the PAE by taking the weighted average of the estimated cluster-

[9]For instance, the Panel Survey of Income Dynamics over-samples low-income families, and the National Longitudinal Survey of Youth over-samples African American, Hispanic or Latino, military, and economically disadvantaged youth.

specific effects.

Varying cluster sizes may also necessitate weighting for uncovering the PAE. This may be the case in settings with grouped cross-sectional data or state-level panels, as in our application below. When clusters form the population of interest, we have a similar situation to the panel data setting above. However, if we are instead interested in the populations of individuals nested in clusters, we must consider each cluster's size in the population, and w_i may take the relative size of each cluster. Our application falls into this situation, as states vary widely in their relevance to the nationwide market for gasoline.

Efficiency may provide a third rationale for heterogeneously weighting cluster-specific estimates. Swamy (1971) cites this rationale for introducing a precision weighted estimator for CRC models with exogenous regressors and Fernández-Val and Lee (2013) propose an analogous optimal weighting matrix for their GMM estimation. These weighting schemes weight estimates of $\mathbf{b}_i$ by the relative variance of $\mathbf{x}_i$ or covariance between $\mathbf{x}_i$ and $\mathbf{z}_i$. We view the primary advantage of the PCIV estimator to be its robustness, however. Were we to adopt a precision weighting approach similar to Swamy (1971) and Fernández-Val and Lee (2013), 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.

3 Consistency and inference of a proposed estimator

3.1 Consistency

While CRC models are perhaps more plausible representations of many economic contexts, incorporating such heterogeneity comes with greater demands on the data. We enumerate further assumptions below:[10]

(A1) $\{\mathbf{y}_i, \mathbf{X}_{1i}, \mathbf{X}_{2i}, \mathbf{Z}_i, \mathbf{d}_i, \mathbf{g}_i\}_{i=1}^n$ is *i.i.d.* across i.

[10]Note that $\mathrm{E}_i[\cdot]$ refers to the expectation over T for each cluster i.

(A2) $\mathrm{E}[e_{ij} \mid \mathbf{x}_{2ij}, \mathbf{z}_{ij}, \mathbf{d}_i] = 0$, $\mathrm{E}[u_{ij} \mid \mathbf{x}_{2ij}, \mathbf{z}_{ij}] = 0$, $\mathrm{E}[\mathbf{e}_i\mathbf{e}_i'] = \Omega$, $\mathrm{E}[e_{ij}^4] < \infty$. Also, $\mathrm{E}(\mathbf{d}_i) = 0$ where $V(\mathbf{d}_i) = \sigma_{\mathbf{d}}^2 < \infty$, and $\mathrm{E}(\mathbf{g}_i) = 0$ where $V(\mathbf{g}_i) = \sigma_{\mathbf{g}}^2 < \infty$.

(A3) $rank[\mathrm{E}_i(\mathbf{z}_{ij}'\mathbf{x}_{ij})] = K$, $rank[\mathrm{E}_i(\mathbf{z}_{ij}'\mathbf{z}_{ij})] = L$, and $\mathrm{E}[\mathbf{z}_{ij}'\mathbf{z}_{ij}e_{ij}^2]$ is positive definite.

(A4) $\mathrm{E}[\|\mathbf{x}_{2ij}\|^2] < \infty$, $\mathrm{E}[\|\mathbf{z}_{ij}\|^2] < \infty$; $\mathrm{E}[\|\mathbf{z}_{ij}\|^4] < \infty$, and $\mathrm{E}[\|\mathbf{x}_{2ij}\|^4] < \infty$.

(A5) $w_i = O_p(n^{-1})$ where $\sum_{i=1}^n w_i = 1$; As $n, T \to \infty$, $n/T \to 0$.

We assume independence across the cluster i in our model to estimate the population average treatment effect. Assumption (A2) is a standard strict exogeneity assumption and says that once we control for $\mathbf{x}_{2ij}$, $\mathbf{z}_{ij}$, and $\mathbf{d}_i$, ($\mathbf{x}_{2ij}$, $\mathbf{z}_{ij}$) for $s \neq t$ do not explain y_{ij}, and that the variance of the error is well-defined. $\mathrm{E}(\mathbf{d}_i) = 0$ and $\mathrm{E}(\mathbf{g}_i) = 0$ are true by definition when the expectation is taken over the population. The condition that $rank[\mathrm{E}_i(\mathbf{z}_{ij}'\mathbf{x}_{ij})] = K$ is meaningful, and can be viewed as an application of results from Graham and Powell (2012), extended to the endogenous case.[11] It requires 1) variation in the instruments and endogenous regressors within each cluster, and 2) all clusters to be influenced by the instrument. This condition is demanding on the data, but as noted in Graham and Powell (2012), it is a function of the model rather than a particular estimator. Further, the first-stage PCIV regressions demonstrate its plausibility within the sample. Assumption (A4) ensures finite variances in the population model and is needed for the consistency of the estimator's variance. We require an assumption about appropriate weights in (A5), which is needed for estimating the population average treatment effect. The second part of Assumption (A5) is in line with the assumption in Fernández-Val and Lee (2013). Since both Fernández-Val and Lee (2013) and our estimator require estimation at the cluster level after the within-cluster estimation, the assumption on the ratio of n and T convergence is needed. Combined with weighting, this assumption allows our estimator to be consistent both within the cluster and

[11] Assuming $\mathbf{x}_{ij}$ is otherwise exogenous, Graham and Powell (2012) show that regular identification of CRC models may not be possible in many settings, such as if $T \leq K$ or where there is insufficient variation over time in covariates within each cluster regardless of the size of T.

at the population level.

Theorem 1 *Under the assumptions (A1)-(A5),*
$\hat{\boldsymbol{\beta}}_{PCIV} = \boldsymbol{\beta} + \sum_{i=1}^{n} w_i \mathbf{d}_i + \sum_{i=1}^{n} w_i (\mathbf{X}'_{1i}\mathbf{P}_i\mathbf{X}_{1i})^{-1}\mathbf{X}'_{1i}\mathbf{P}_i\mathbf{e}_i \rightarrow \boldsymbol{\beta}$ *as* $n, T \rightarrow \infty$, *where* $\mathbf{P}_i = \mathbf{H}_{\mathbf{Z}_i} + \mathbf{M}_{\mathbf{Z}_i}\mathbf{X}_{2i}(\mathbf{X}'_{2i}\mathbf{M}_{\mathbf{Z}_i}\mathbf{X}_{2i})^{-1}\mathbf{X}'_{2i}\mathbf{M}_{\mathbf{Z}_i}$, *where* $\mathbf{H}_{\mathbf{Z}_i} = \mathbf{Z}_i(\mathbf{Z}'_i\mathbf{Z}_i)^{-1}\mathbf{Z}'_i$.

3.2 Inference

In this section, we derive the asymptotic variance of the general PCIV estimator for inference and describe its estimation. Since we first estimate cluster-level coefficients, $\mathbf{b}_i$, to obtain an estimate of the global-level estimate, $\hat{\boldsymbol{\beta}}_{PCIV}$, we need to account for possible estimation error. We construct a sample variance as follows:

$$\begin{aligned} &\hat{\mathrm{V}}(\hat{\boldsymbol{\beta}}_{PCIV} - \boldsymbol{\beta}) \qquad (4) \\ &= \mathrm{V}\left(\sum_{i=1}^{n} w_i \left[\hat{\mathbf{d}}_i + (\mathbf{X}'_{1i}\mathbf{P}_i\mathbf{X}_{1i})^{-1} \mathbf{X}'_{1i}\mathbf{P}_i\hat{\mathbf{e}}_i\right]\right) \\ &= \sum_{i=1}^{n} w_i^2 \hat{\mathbf{d}}_i\hat{\mathbf{d}}'_i + \sum_{i=1}^{n} w_i^2 (\mathbf{X}'_{1i}\mathbf{P}_i\mathbf{X}_{1i})^{-1} \mathbf{X}'_{1i}\mathbf{P}_i\hat{\mathbf{e}}_i\hat{\mathbf{e}}'_i\mathbf{P}'_i\mathbf{X}_{1i} (\mathbf{X}'_{1i}\mathbf{P}_i\mathbf{X}_{1i})^{-1} \end{aligned}$$

where $\hat{\mathbf{d}}_i = \hat{\mathbf{b}}_i - \hat{\boldsymbol{\beta}}_{PCIV}$, and $\hat{\mathbf{e}}_i = \mathbf{y}_i - \mathbf{X}_{1i}\hat{\mathbf{b}}_i - \mathbf{X}_{2ij}\hat{\boldsymbol{\delta}}$. Throughout this paper, we will apply this estimated variance in the simulation as well as application studies to construct the standard errors. The standard errors from this estimator are robust to heteroskedasticity and arbitrary correlation in the error term within cluster. The proof of Theorem 2 is shown in Appendix B.

Theorem 2 *Under the assumptions (A1)-(A5),* $\sqrt{n}(\hat{\boldsymbol{\beta}}_{PCIV} - \boldsymbol{\beta}) \rightarrow N(0, V(\hat{\boldsymbol{\beta}}_{PCIV}))$, *where* $\hat{V}(\hat{\boldsymbol{\beta}}_{PCIV})$ *is a consistent estimator of* $V(\hat{\boldsymbol{\beta}}_{PCIV})$.

3.3 Finite samples in one dimension

While very large data sets are becoming more commonplace, researchers may not wish to rely on asymptotic arguments with respect to both dimensions of their data in many applications. Consequently, we also consider the properties of the PCIV estimator with a fixed number of clusters (n) and $T \to \infty$, before considering the probability limit more comparable to FEIV with fixed observations per cluster (T) and $n \to \infty$.

We start with fixed n while the number of observations per cluster trends to infinity. While this could be true with panel data, this setting is perhaps more commonplace for clustered cross-sectional data. Accordingly, we take the probability limit of equation 5 as $T \to \infty$, providing the following:

$$\underset{T\to\infty}{\text{plim}}\,(\hat{\boldsymbol{\beta}}_{PCIV} - \boldsymbol{\beta}) = \sum_{i=1}^{n} w_i \mathbf{d}_i + \sum_{i=1}^{n} w_i (\mathrm{E}_i[\mathbf{X}'_{1i}\mathbf{P}_i\mathbf{X}_{1i}])^{-1}\mathrm{E}_i[\mathbf{X}'_{1i}\mathbf{P}_i\mathbf{e}_i] \tag{5}$$

The previously stated rank and validity assumptions applied to each cluster ensure that $\sum_{i=1}^{n} w_i(\mathrm{E}_i[\mathbf{X}'_{1i}\mathbf{P}_i\mathbf{X}_{1i}])^{-1}\mathrm{E}_i[\mathbf{X}'_{1i}\mathbf{P}_i\mathbf{e}_i] = 0$. Thus, the PCIV approach is asymptotically unbiased in estimating the PAE, though consistency requires $n \to \infty$.

Perhaps a more common setting exists when T is fixed and n trends to infinity. This setting is common in applications with panel data. Taking the probability limit of equation 6 with $n \to \infty$ provides the following:

$$\underset{n\to\infty}{\text{plim}}\,(\hat{\boldsymbol{\beta}}_{PCIV} - \boldsymbol{\beta}) = \mathrm{E}[w_i\mathbf{d}_i] + \mathrm{E}[w_i\left(\mathbf{X}'_{1i}\mathbf{P}_i\mathbf{X}_{1i}\right)^{-1}\mathbf{X}'_{1i}\mathbf{P}_i\mathbf{e}_i]. \tag{6}$$

With a fixed number of observations per cluster, additional complications arise. While by definition, $\mathrm{E}(\mathbf{d}_i) = 0$, we first require that T is large enough ($T > K$) to estimate $\boldsymbol{b}_i$.

Second, we must assume $\mathrm{E}[w_i\left(\mathbf{X}'_{1i}\mathbf{P}_i\mathbf{X}_{1i}\right)^{-1}\mathbf{X}'_{1i}\mathbf{P}_i\mathbf{e}_i] = 0$ in order for $\hat{\boldsymbol{\beta}}_{PCIV}$ to consistently estimate the PAE, $\boldsymbol{\beta}$. Unlike in the exogenous case presented in Wooldridge (2010), each estimated $\mathbf{b}_i$ will likely manifest some degree of finite sample bias. The question is whether we may expect the biases to be well-behaved or mean zero in expectation as only the number of clusters approaches infinity. The first way this finite sample bias is well-behaved is apparent from the consistency argument above. The finite sample bias falls as the ratio of observations per cluster to regressors (and instruments) grows.

However, without infinite observations per cluster, we may expect finite sample biases to be nonzero in expectation if the instruments are weak within any clusters, a stronger condition than imposed by (A3). As discussed in Bound et al. (1995) and Staiger and Stock (1997), with weak instruments, the finite sample bias is in the direction of the OLS estimates. If there is also correlation between instrument strength and effect heterogeneity, it would be unreasonable to expect mean-zero finite sample bias. However, this finite sample bias is well-behaved since it falls as the strength of the instrument within each cluster increases. In the extreme, if we have an exogenous instrument that perfectly predicts the regressor, there is no endogeneity issue, and the unbiasedness result from Wooldridge (2005); Arellano and Bonhomme (2011); and Bates et al. (2014) apply.

In cases where some clusters are unmoved by the instrument or the instrument is weak, the PCIV estimator may still recover LATEs, which may be an advantage over competing estimators. As we have modeled the treatment effect and compliance heterogeneity as varying only across individuals, we need no further assumptions regarding monotonicity. That is, while we do not maintain monotonicity across individual clusters, we do maintain monotonicity within each. Unlike with the independent cross-sectional data model discussed in Imbens and Angrist (1994), we observe cluster- (or individual-) specific unbiased estimates of compliance and may perform inference on each. Consequently, we can aggregate our treatment effect estimates according to the cluster-specific, first-stage estimates to get different LATEs, including separate LATEs for compliers (say those who are moved towards treatment by the instrument) and defiers (say those who are moved away from treatment by the instrument). Accordingly, we may exclude never-takers and always-takers or those only weakly moved by the instrument (for instance, those with first-stage F-statistics less than 10, following the rule of thumb from Stock and Yogo (2005)). Thus, PCIV allows us to estimate LATEs over well-defined and identifiable populations for whom finite sample bias is likely to be small. Given the potential for bias in the PCIV approach with finite samples, we use a Monte Carlo study to examine how the per-cluster instrumental variables estimator performs as we vary the number of observations per cluster below.

4 Simulation Study

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

We generate two versions of our outcome: uncorrelated case and correlated case between the first-stage variation and random coefficients. In both cases, we generate the data based on our model in equation 1.

$$
\begin{aligned}
y_{ij} &= x_{ij}\mathbf{b}_i + e_{ij}, \\
x_{ij} &= z_{ij}\boldsymbol{\gamma}_i + u_{ij},
\end{aligned} \tag{7}
$$

where $\mathbf{b}_i = \boldsymbol{\beta} + \mathbf{d}_i = 1 + \mathbf{d}_i$ and $\gamma_i = 1$. An exogenous variable, z_{ij}, is drawn from $N(0, \exp(2\mathbf{d}_i))$. We define composite error terms as $e_{ij} = \alpha_{1i} + o_{ij} + v_{ij}$ and $u_{ij} = \alpha_{2i} + 0.2w_j + 0.32\epsilon_{ij}$, where $\alpha_{2i} = 1.33\alpha_{1i} + 2.13\mathbf{d}_i$ and $v_{ij} \sim N(0, 1.1)$. Both α_{1i} and α_{2i} are

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

random intercepts; and v_{ij}'s and ϵ_{ij}'s are cluster- and time-varying errors respectively[13]. We set $\epsilon_{ij} \sim N(0,1)$ for uncorrelated case and $\epsilon_{ij} \sim N(0, \exp(2\mathbf{d}_i))$ for correlated case[14]. An exogenous cluster-level error, w_j, is drawn from the standard normal distribution. We set $o_{ij} = \xi_{ij} + \eta_{ij}$ with η_{ij} following $N(0,1)$ and with ξ_{ij} being residuals from regressing x_{ij} on z_{ij}. Then endogeneity of our variable of interest, x_{ij}, enters through the presence of o_{ij} in the generation of y_{ij}. Our random intercept, α_{1i}, and random slope, $\mathbf{d}_i$, are drawn as follows:

$$\begin{pmatrix} \alpha_{1i} \\ \mathbf{d}_i \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.4^2 & \rho \\ \rho & 0.25^2 \end{bmatrix} \right),$$

where $\rho = 0$ for uncorrelated case and $\rho = 0.5$ for correlated case. The key assumption for the consistency of P2SLS and FEIV is that the within-cluster covariance between x_{ij} and z_{ij} is uncorrelated with the random slopes, $\mathbf{d}_i$. When there is no violation, as $\mathbf{d}_i$ is not related to the variance of z_{ij} or x_{ij}. In the second case, there is a violation of the key assumption as $\mathbf{d}_i$ acts as a random coefficient on x_{ij} and determines the variance of z_{ij} and x_{ij}.

The resulting correlations with $N = 250$ and $T = 250$ are shown in Table 1. Under both conditions, x_{ij} is correlated with unobserved heterogeneity α_{1i}, $\mathbf{d}_i$, and the omitted variable, o, producing endogeneity. Our variable of interest, x_{ij}, is also strongly correlated with the instrument z_{ij}, and z_{ij} is otherwise orthogonal to the other terms, such that our instrument is both relevant and valid. The key condition hinges on whether the heterogeneous slope, $\mathbf{d}_i$, is correlated with the strength of the instrument within-cluster. The correlations between $\mathbf{d}_i$ and $\ddot{x}\ddot{z}$ are 0.0061 and 0.2776 respectively reflect two possible states of the key condition.

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

[13]Since our parameter of interest is $\mathbf{b}_i$, we combine intercepts into the composite errors.

[14]Having the variance of x_{ij} depend on $\mathbf{d}_i$ ensures that the variance of x_{ij} in both cases are similar, while the variance of x_{ij} is only related to the random coefficients in of x_{ij} on y_{ij}.

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

	Panel A: Uncorrelated Covariance							Panel B: Correlated Covariance					
	α_1	**d**	x	o	z	$\ddot{x}\ddot{z}$		α_1	**d**	x	o	z	$\ddot{x}\ddot{z}$
α_1	1						α_1	1					
d	0.010	1					**d**	0.477	1				
x	0.391	0.363	1				x	0.538	0.546	1			
o	0.401	0.374	0.385	1			o	0.552	0.560	0.481	1		
z	0.001	0.0004	0.789	0.006	1		z	0.001	0.002	0.726	0.007	1	
$\ddot{x}\ddot{z}$	0.127	0.006	0.053	0.051	0.004	1	$\ddot{x}\ddot{z}$	0.127	0.278	0.154	0.157	0.004	1

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

rate from each approach. Each simulation is repeated 500 times. We believe that the fixed T case is most likely to arise in applied research, and focus the simulation discussion on the performance of the three estimators when N is large, and T is fixed. We use different fixed levels of T, ranging from 6 to 250, with $N = 250$ throughout the simulation. For completeness, we include the symmetric examination in which we hold $T = 250$ and vary N in Appendix D. Due to its asymptotic unbiasedness, PCIV generally performs relatively better in cases where N is fixed and T is large.

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. Panel (a) of Figure 1 shows the average bias in the three estimators with $N = 250$ as T increases, maintaining the uncorrelated covariance assumption on the left and violating it on the right. 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 clusters (when clusters have 10 or fewer observations). However, across cluster sizes, PCIV outperforms the other estimators when the uncorrelated covariance assumption is violated. Furthermore, while the estimated biases of FEIV and P2SLS estimators do not show any improvement even with the increasing cluster size, the estimated bias of the PCIV approach gets closer to zero as cluster sizes increase. Table 4 in Appendix C shows that the magnitude of bias in both FEIV and P2SLS is almost identical at approximately 12 percent.

Naturally, researchers are not only interested in the bias of estimators but are also interested in the estimators' precision. We use 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, with a large number of clusters and the number of observations per cluster varying between 6 and 250. Again, we repeat the exercise both with and without the uncorrelated covariance assumption holding.

Panel (b) of Figure 1 reveals that with very small clusters, 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.

We evaluate the performance of the analytic standard errors over cluster size by first depicting the ratio of the mean of the estimated standard errors (SEs) divided by sampling standard deviations (SDs) in panel (c) of Figure 1.

In all cases, the analytic standard errors do reasonably well with large sample sizes along both dimensions. However, they are too small with very small clusters. By nine observations per cluster, the ratio of mean SE to SD for PCIV is comparable to the same ratio for FEIV. The ratio is consistently close to one with T above ten, for all three estimators both with the uncorrelated covariance assumption holding and when it is violated.

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. Panel (d) of Figure 1 shows the evolution of coverage rates for each estimator in the uncorrelated covariance case on the left and correlated covariance on the right with 250 clusters as T grows. In the uncorrelated covariance case, the coverage rates for all three estimators range from 0.89 to 0.98 for all cluster sizes, though they converge close to 0.95 as the cluster size grows.

When there is a correlation between the strength of the instrument and random coefficients, the bias in FEIV and P2SLS is meaningful. Neither standard estimator includes the true value in the 95 percent confidence interval more than 40 percent of the time. Moreover, the rejection rate of the true parameter grows as cluster size grows. 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 cluster size grows. Though PCIV inference performs well overall, the relatively poorer performance when T is small may give researchers reason to adopt a bootstrap approach to standard error estimation with restricted samples.

5 Estimating the price elasticity of demand for gasoline

There is extensive literature estimating the price elasticity of gasoline demand. However, 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 unidirectional causal link. Consequently, we use instrumental variables estimation and follow Davis and Kilian (2011); Blundell et al. (2012); Hausman and Newey (2016); and Coglianese et al. (2017) in using state gasoline taxes to instrument for prices. In many ways, the empirical setting and design are ideal, especially considering that the efficacy of carbon taxes in lowering fuel consumption is one reason why the price elasticity of gasoline demand is of particular interest.

Works such as Davis and Kilian (2011); Levin et al. (2017); and Coglianese et al. (2017)

[14]See Dahl (1986) for a helpful review of early examples and Davis and Kilian (2011); Blundell et al. (2012); Hausman and Newey (2016); Coglianese et al. (2017); Levin et al. (2017); and Hoderlein and Vanhems (2018) for more recent examples.

econometrically model the relationship between gasoline prices and the quantity sold as though responsiveness to prices is homogeneous. However, heterogeneity in elasticities seems likely in this context. Regulation of this market varies across states, as does industry composition, population density, demographic composition, transportation substitutes, macroeconomic climate, and, most importantly, taxes.

Wadud et al. (2010); Frondel et al. (2012); Blundell et al. (2012); and Hausman and Newey (2016) examine how demand elasticities for gasoline may differ across individual characteristics. Of these, only Blundell et al. (2012) and Hausman and Newey (2016) deal with the endogeneity of prices, and both use state gasoline taxes and distance from the Gulf of Mexico as instrumental variables. However, in so doing, each implicitly assumes that the identifying variation (tax changes and tax pass-through) and distance from the gulf homogeneously affect prices. We believe that this assumption may not be benign. For instance, states with developed transportation substitutes may have more elastic demand and raise taxes on gasoline more frequently than states with fewer transportation substitutes and less elastic demand. Further, as tax pass-through rates are theoretically linked to demand and supply elasticities, it would be surprising if the first-stage coefficients were homogeneous while the second-stage elasticities varied across states.

5.1 Model and estimation

As a result, we build both forms of heterogeneity into the econometric model.[15] We allow each state (indexed by i) to differ in both level of gasoline demand, α_{1i}, and in the price elasticity of gasoline demand (depicted by $\mathbf{b}_i = \boldsymbol{\beta} + \mathbf{d}_i$) according to the following CRC

[15]Heterogeneous elasticities may exist at more or less granular levels. However, elasticity heterogeneity only leads to inconsistency in existing estimators when systematically related to heterogeneity in the first stage. The first stage does not vary across individuals but across states, making states an intuitive level to model heterogeneity. As a robustness check in Table 8 in Appendix C, we perform analysis modeling clusters according to the Petroleum Administrative Defense Districts.

model similar in structure to equation 1:

$$\begin{aligned} logsales_{ij} &= \alpha_{1i} + logprice_{ij}\mathbf{b}_i + \mathbf{x}_{ij}\boldsymbol{\delta} + \epsilon_{ij}, \\ logprice_{ij} &= \alpha_{2i} + logtaxes_{ij}\boldsymbol{\gamma}_i + \mathbf{x}_{ij}\boldsymbol{\eta} + u_{ij}. \end{aligned} \tag{8}$$

We are primarily interested in the population average price elasticity of gasoline demand, $\mathrm{E}[\mathbf{b}_i] = \boldsymbol{\beta}$. We use the log of taxes to instrument for the potentially endogenous log of prices. We allow for possible heterogeneity in tax pass-through rates, as denoted by $\boldsymbol{\gamma}_i$.

Given this suspected heterogeneity, our preferred approach is PCIV. Both for comparability to prior elasticity estimates and to benchmark the performance of the PCIV approaches, we apply P2SLS on first differences and FEIV to the data and model above. We use the general PCIV estimator presented in section 2.2 in this setting for two reasons. First, $\mathbf{x}_{ij}$ indicates the presence of exogenous regressors with common parameters in the model. Most notably, there are general time and seasonality trends in gasoline prices, consumption, and taxes, such that our instrument is only exogenous conditional on them. As the inclusion of month-by-year fixed effects is standard in P2SLS and FEIV, we adopt them for all three estimators.[16] Including month-by-year fixed effects is impossible using the simple approach or that described in Fernández-Val and Lee (2013), and allowing for heterogeneous month and year effects by cluster would be extremely taxing on the data.

Second, states vary substantially in size and relevance to the market for gasoline. The inclusion of heterogeneous slopes in the econometric model leads us to pay particular attention to differences in state sizes and volumes of gasoline purchased. 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 differ by state, failing to account for such differences in states' relevance to the market may lead to inconsistent estimates of the parameter over the population. Thus, we incorporate weights representing

[16]We estimate the model with year and month fixed effects and find similar results in Table 7.

the relative average of volume purchased in each state. We use these weights in weighted P2SLS and FEIV regressions, and when aggregating the state-specific coefficients in PCIV estimation. In contrast with PCIV, weighting does not necessarily help recover the PAE with P2SLS and FEIV estimation in the presence of random coefficients Solon et al. (2015). We also provide results from regressions without these weights for comparison.

Coglianese et al. (2017) remark that anticipatory behavior biases the estimates of Davis and Kilian (2011). In appendix E, we examine the issue theoretically and empirically. The anticipatory behavior is primarily a concern only for the P2SLS estimator as it is persistently inconsistent with such violations to strict exogeneity, whereas the inconsistency from this source in FEIV and PCIV converges to zero as the number of time periods grows large. Consistent with theory, Figure 10 shows the robustness of PCIV and FEIV to including additional lags and leads of log prices and log taxes whereas the P2SLS estimator is sensitive to this violation of strict exogeneity. We additionally provide a replication of Davis and Kilian (2011) and Coglianese et al. (2017) in appendix F.

5.2 Data

The data we use carries a fittingly wide scope, containing monthly observations of gasoline prices, taxes, and volume sold from January 1989 through December 2018 throughout the United States. The data provides us with 360 time observations over the 50 states and the District of Columbia.

The data on monthly, statewide, gasoline price averages for 1989 through 2011 comes from the U.S. 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. We add state and federal taxes to the tax-exclusive prices to approximate at-the-pump prices. However, the EIA discontinued this series and the survey on which it relied in 2011, requiring updated subsequent tax changes and significant additional data collection. We supplement this pricing data with average at-the-pump price data from Gasbuddy.com.[17]

We begin with the annual 'Highway Statistics Series' for effective dates of state gasoline taxes and refine these dates against state governmental documentation when possible. Since thirty states have changed their state gasoline taxes in the last decade (some multiple times), the recent data provides useful identifying variation. Following Davis and Kilian (2011), we net out any portion of the state gasoline tax rate due to changes in gasoline prices to avoid building endogeneity back into our estimation.[18] Throughout, we use the EIA, 'Petroleum Marketing Monthly Report: Prime Supplier Sales Volumes by Product and Area' for data on gasoline sales volume.

We add several covariates to check the robustness of our results and possibly improve efficiency. In particular, we include the log of unemployment rates; real per-capita income; population; number of licensed drivers; number of road miles; average, minimum, and maximum temperatures; and precipitation; as well as indicators for missing values of each. We list the summary statistics and sources of control variables in Table 5 in Appendix C.

5.3 Results

We present the elasticity estimates from all three estimators in Table 2. Results without volume weights appear on the left, and estimates incorporating the state-specific gasoline-volume weights appear on the right. For PCIV, the standard errors are estimated as described

[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 month-by-year fixed effects will absorb the constant gap.

[18]We largely use data from Davis and Kilian (2011) during the period from 1989 through 2008 though discovered a few instances in which the taxes they used either failed to capture per-unit taxes fully 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.

in section 3.2. We present the first-stage F-statistics for each estimator along the bottom row of Table 2. We calculate the F-statistics for PCIV using Hotelling's T-squared test.[19]

Table 2: Summary of Results Using Three Estimation Methods

	Without volume weights			Volume weighted		
	P2SLS	FEIV	PCIV	P2SLS	FEIV	PCIV
Log price	-0.724	-0.929	-0.551	-0.463	-0.873	-0.555
	(0.193)	(0.415)	(0.227)	(0.154)	(0.394)	(0.240)
First-stage F-statistic	36.66	79.71	58.35	47.47	63.70	61.16
Controls	N	N	N	N	N	N
Log price	-0.736	-0.828	-0.543	-0.512	-0.760	-0.561
	(0.189)	(0.327)	(0.278)	(0.138)	(0.271)	(0.294)
First-stage F-statistic	36.58	80.92	58.71	46.83	60.26	59.93
Controls	Y	Y	Y	Y	Y	Y

Notes: The sample consists of 18,360 state-by-month observations. First-stage F-statistics for P2SLS and FEIV are obtained from the regression of each endogenous regressor on the exogenous regressors and the instruments. The calculation of the first-stage F-statistics for the PCIV was done using Hotelling's T-squared test. State-clustered standard errors appear in parentheses.

We first note that the log of state gasoline taxes is generally strongly predictive of at-the-pump prices across specifications. The F-statistics range from 50.08 (weighted FEIV) to 90.19 (weighted P2SLS applied to first differences).

The estimates of the elasticity of gasoline demand are consistent in sign, though the magnitude of the estimated elasticities range meaningfully across estimators, and in some instances, across specifications within estimators. FEIV provides the largest point estimates in magnitude, estimating elasticities of -1.03 (unweighted and unconditional) to -0.73 (weighted and conditioning on covariates) with all p-values less than 0.02. These FEIV estimates are double the prevailing estimates in the energy economics literature. The P2SLS on first differences estimator is most efficient, with standard errors often less than half the size of those for FEIV, yet is the most sensitive to including weights for sales volume. While

[19]We present first-stage coefficient estimates in Table 6 in Appendix C.

the unweighted P2SLS elasticity estimate is large at -0.66, the weighted P2SLS estimate is -0.43 to -0.47 (depending on the inclusion of covariates), making the FEIV estimate around 50 percent larger than those from P2SLS. The PCIV estimate of the elasticity of gasoline demand falls between those two with point estimates ranging from -0.52 to -0.55 across all specifications. The p-values associated with those estimates range from 0.020 to 0.065.

5.4 Examination of identifying variation

It is natural to ask which estimates are preferable according to econometric theory. We take multiple approaches to shed light on this question. First, we consider the weighting employed by all three estimators. We examine whether the implicit weighting used in FEIV and P2SLS mimics the weights by sales volume and find significant divergence. Second, we inspect whether there appears to be a relationship between the first-stage variation and the estimated elasticities—an apparent violation of the key condition for the consistency of P2SLS and FEIV. We then explore the heterogeneity in the first- and second-stage relationships between gasoline taxes, prices, and sales volume. In sum, these examinations of the data lead us to prefer the weighted PCIV approach as our decision to model such heterogeneity is borne out by the data. This rich accounting of the identifying variation in the data reveals four states for which the log of gasoline taxes does poorly in predicting changes in prices. As a result, we also estimate a LATE, a weighted average of the state-specific elasticities for the 46 states in which the instrument is strong.

Our population of interest is all gasoline sales within the United States. If there are heterogeneous elasticities across states, we must afford each state weight according to its relevance in the market. In this context, it turns out that the PCIV point estimates are very similar when we do and do not weight for sales volume, though a priori, that result was not obvious. 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 where there is relatively more first-stage variation. We examine the agreement between the two weighting schemes in Figure 2. On the left, we report the negative of the relative first-stage variation in each state from FEIV estimation. Functionally, this is the sum of the products of demeaned and detrended x and z for each state divided by the sum of those state totals. The share of total gasoline sales that occur in the state over the panel is on the right.

Figure 2 shows vast disagreement between the two weighting schemes. The explicit weights we employ are intuitive, with larger gasoline-consuming states receiving influence 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 with just 0.1 percent of the national market.

The FEIV weighting scheme also affords California the highest weight, though it is 63 percent higher than the weights based on sales volume. Texas consumes the second most gasoline in the country (9.5 percent) yet is twenty-second in first-stage variation (1.3 percent). On the other end, there are notable exceptions. With only 1.1 percent of gasoline sales, Connecticut has the fifth largest weight on the standard FEIV estimates. As a further complication for FEIV estimation in this context, both Hawaii and Minnesota receive negative weight in this exercise.

Might using weights with FEIV rectify the mismatch of its implicit weights? In short, not generally or in this context. First, even with the addition of volume weights, 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 PAE, nor does it always dominate unweighted least squares.

One advantage of the PCIV estimator is revealing the heterogeneity in first- and second-stage estimates across clusters. We observe significant heterogeneity in the location-specific elasticities and tax pass-through rates. Figure 4 reveals this variation.

The right-hand side of Figure 4 shows the state-specific tax pass-through rates. The weighted average pass-through rate from table 6 implies that for every \$0.10 increase in taxes, prices increase by \$0.024. However, in the District of Columbia and Washington state, the pass-through rates are double the average at 0.61 and 0.46, respectively. On the other end, Georgia, Hawaii, Michigan, and Indiana each have estimated first-stage coefficients of 0.08 and lower, with the estimated effect in Georgia defying the average sign. Unsurprisingly, the F-statistics on the log of taxes in these states are each less than 10, failing to surpass the weak instruments rule of thumb from Stock and Yogo (2005). There is sufficient noise in the District of Columbia estimate to also drive its F-statistic below 10. As a result, we should interpret the second-stage estimates from these locations with caution.

The estimates of state-specific gasoline elasticities on the left-hand side also show significant variation.[20] Montana, Arizona, Idaho, and Nevada have relatively inelastic gasoline demand (elasticities at or lower than 0.3 in magnitude). Perhaps inelastic gasoline demand makes sense in these rural states, which are not densely populated and mostly lack robust public transportation substitutes. In contrast, 14 states have estimated elasticities with magnitudes greater than 0.6 and Pennsylvania has an estimated elasticities of -0.96.

Estimating the state-specific second-stage slopes also allows us to examine the key condition for the consistency of FEIV estimation, namely, whether there is any relationship between effect heterogeneity and first-stage variation. We show the relationship between the relative elasticities and the relative first-stage variation in Figure 3. The PCIV estimates of b_i provide a natural way to examine heterogeneous elasticities. We present these estimates relative to the mean along the x-axis. We measure the relative first-stage variation similarly to in Figure 2, scaled to the mean, which we present along the y-axis. Each circle represents a state with the size scaled by the state's relevance to the gasoline market.

Figure 3 reveals a strong upwards slope as the two statistics have a correlation of 0.14

[20]Conducting regional analysis employing PCIV at the EIA's Petroleum Administration for Defense District level, we uncover significant variation in the seven elasticities. These estimates range from -0.88 to -1.35. However, these results assume no heterogeneity in effects within each region and may suffer from some of the same issues as FEIV. We report the results in Table 8 in Appendix C.

across states. In general, this means that states with relatively more exploitable variation in the first stage also have larger estimated elasticities. This is consistent with the idea that states with less developed public transportation infrastructure (and less elastic gasoline demand) raise gasoline taxes less frequently. This relationship would lead to an upwards bias in the elasticity estimate from FEIV. Indeed, the FEIV elasticity estimates of the PAE are 35 to 90 percent larger in magnitude than the PCIV estimates of the same parameter.

Though PCIV is robust to violations of these assumptions, any given data set may lack a long enough time series on each cluster or sufficient variation in each cluster to maintain that each cluster-specific estimate closely approximates the true cluster-specific parameter. Indeed, we find that the predictive power of the instrument is weak in Hawaii (F-stat = 2.30), Indiana (1.84), Georgia (5.81), Michigan (0.23), and Washington D.C. (7.68).[21] As a result, we redo the analysis on the 46 states in which the first-stage F-statistic is above ten, following the rule of thumb proposed by Staiger and Stock (1997) and justified in Stock and Yogo (2005). Though we acknowledge this parameter may not be relevant for the entire population, this exercise provides us with an estimate of the average price elasticity of gasoline demand on a specified and well-defined group of compliers, namely, states in which the f-statistic of the instrument is above ten.

Table 3: Estimated elasticities among states in which the instrument is strong (LATE)

	Without volume weights			Volume weighted		
	P2SLS	FEIV	PCIV	P2SLS	FEIV	PCIV
Log price	-0.724	-0.929	-0.551	-0.463	-0.873	-0.555
	(0.193)	(0.415)	(0.227)	(0.154)	(0.394)	(0.240)
First-stage F-statistic	36.66	79.71	58.35	47.47	63.70	61.16

Notes: Sample composed of all states with first-stage F-statistics above 10, excluding Alaska, Hawaii, Indiana, Georgia and the District of Columbia. Regressions condition on time-by-month fixed effects. State-clustered standard errors appear in parentheses.

[21]See Table 9 in Appendix C for state-specific statistics.

We report the results from weighted and unweighted PCIV on this population of compliers in Table 3. For comparison, we also apply FEIV and P2SLS to the same data. Among states for which the instrument is strong, the PCIV estimates are -0.52 to -0.54, with p-values of 0.010 or 0.016 depending on weighting. Despite the reduction in sample size, the standard errors here are modestly smaller than those in Table 2 using the entire sample as those states with weak first stages had relative extreme second-stage estimates. The P2SLS and FEIV are larger when estimated on this restricted sample. Recall that the apparent violation of the key condition may prohibit P2SLS and FEIV from consistently estimating even a LATE. Indeed, the relationship between estimated elasticities and first-stage variation is just as strong if not stronger within the states in which the instrument is strong.

6 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. We propose per-cluster instrumental variables estimation to provide both population average estimates and effect heterogeneity in correlated random coefficient models when key variables may be endogenous. Given the vast unobserved differences between people and heterogeneity of their behavior, we believe such environments are widespread, as is the applicability of this approach.

We present the conditions under which PCIV 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 PCIV in simulation.

Our use of PCIV adds rigor to investigation of the price elasticity of demand for gasoline. Our cumulative evidence suggests a price elasticity for the demand for gasoline in the United States to fall between -0.5 and -0.55, keeping in mind that the confidence intervals 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), which find elasticities ranging from -.37 to -.03. However, only Coglianese et al. (2017) address the endogeneity issue in prices. We find that the small estimates there are due to the use of P2SLS applied to first differences, which is particularly afflicted by anticipatory behavior and is only identified off short-run (first month) responses.

These high elasticity estimates may also reflect that price increases due to tax 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, with Li et al. (2014) in particular demonstrating both points. As a result, these estimates of the price elasticity of gasoline demand pertain 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.

Our application shows that the flexibility of PCIV estimation provides more transparency about the underlying data. For instance, we can view whether the implicit weighting of standard estimators correspond to the appropriate population weights and whether there appears to be near zero correlation between the strength of the instrument and heterogeneous effects, which may inhibit estimation of population average effects with standard estimators. PCIV also reveals heterogeneity within the first stage including whether monotonicity holds at the cluster level. Many of these may be estimated even with small numbers of observations per cluster. Finally, in cases where there appears to be insufficient first-stage variation within a subset of clusters to produce a reliable cluster-specific estimate, PCIV allows for the estimation of a local average effect in a defined population of clusters in which the instrument is strongly predictive.

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.

Blundell, R., J. L. Horowitz, and M. Parey (2012). Measuring the price responsiveness of gasoline demand: Economic shape restrictions and nonparametric demand estimation. *Quantitative Economics 3*(1), 29–51.

Borusyak, K., X. Jaravel, and J. Spiess (2021). Revisiting event study designs: Robust and efficient estimation. *arXiv preprint arXiv:2108.12419*.

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.

Bushnell, J. B., E. T. Mansur, and C. Saravia (2008). Vertical arrangements, market structure, and competition: An analysis of restructured us electricity markets. *American Economic Review 98*(1), 237–66.

Callaway, B. and P. H. Sant'Anna (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics 225*(2), 200–230.

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

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.

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

Dahl, C. A. (1986). Gasoline demand survey. *The Energy Journal 7*(1).

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.

De Chaisemartin, C. and X. d'Haultfoeuille (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review 110*(9), 2964–96.

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

Fernández-Val, I. and J. Lee (2013). Panel data models with nonadditive unobserved heterogeneity: Estimation and inference. *Quantitative Economics 4*(3), 453–481.

Feyrer, J. (2019). Trade and incomeâexploiting time series in geography. *American Economic Journal: Applied Economics 11*(4), 1–35.

Frondel, M., N. Ritter, and C. Vance (2012). Heterogeneity in the rebound effect: Further evidence for germany. *Energy Economics 34*(2), 461–467.

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

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.
Hausman, J. A. and W. K. Newey (2016). Individual heterogeneity and average welfare. *Econometrica 84*(3), 1225–1248.
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.
Hoderlein, S. and A. Vanhems (2018). Estimating the distribution of welfare effects using quantiles. *Journal of Applied Econometrics 33*(1), 52–72.
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.
Jackson, C. K., R. C. Johnson, and C. Persico (2016). The effects of school spending on educational and economic outcomes: Evidence from school finance reforms. *The Quarterly Journal of Economics 157*, 218.
Kelejian, H. H. (1974). Random parameters in a simultaneous equation framework: identification and estimation. *Econometrica: Journal of the Econometric Society*, 517–527.
Klein, L. R. (1953). *A Textbook of Econometrics*. Evanston: Row, Peterson.
Kostol, A. R. and M. Mogstad (2014). How financial incentives induce disability insurance recipients to return to work. *American Economic Review 104*(2), 624–55.
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.
Miguel, E., S. Satyanath, and E. Sergenti (2004). Economic shocks and civil conflict: An instrumental variables approach. *Journal of political Economy 112*(4), 725–753.
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.
Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica 74*(4), 967–1012.
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.
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.

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

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

Wadud, Z., D. J. Graham, and R. B. Noland (2010). Gasoline demand with heterogeneity in household responses. *The Energy Journal 31*(1).

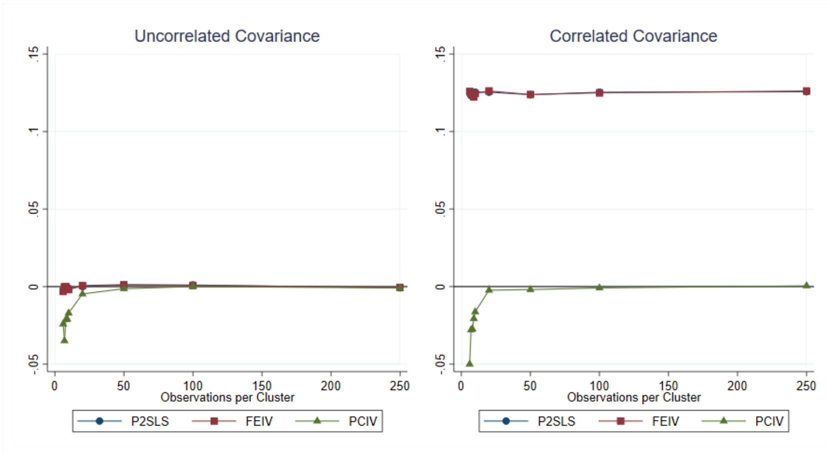
Wooldridge, J. (2021). Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators. *Available at SSRN 3906345*.

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.

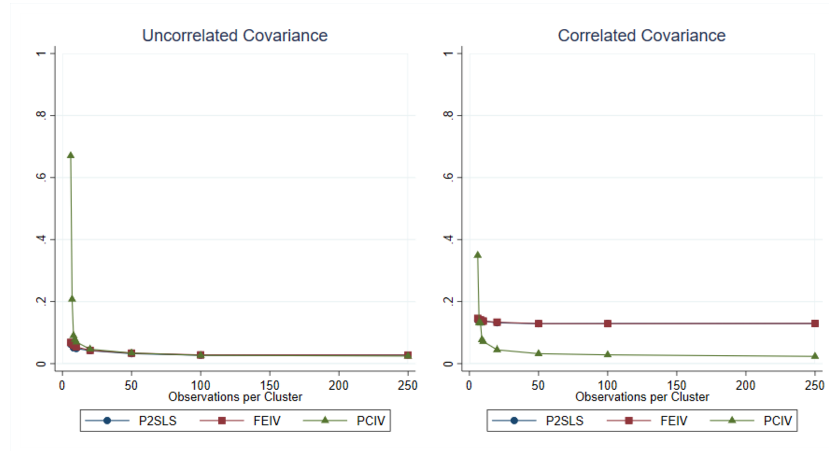
Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

Figure 1: Simulation results across cluster sizes

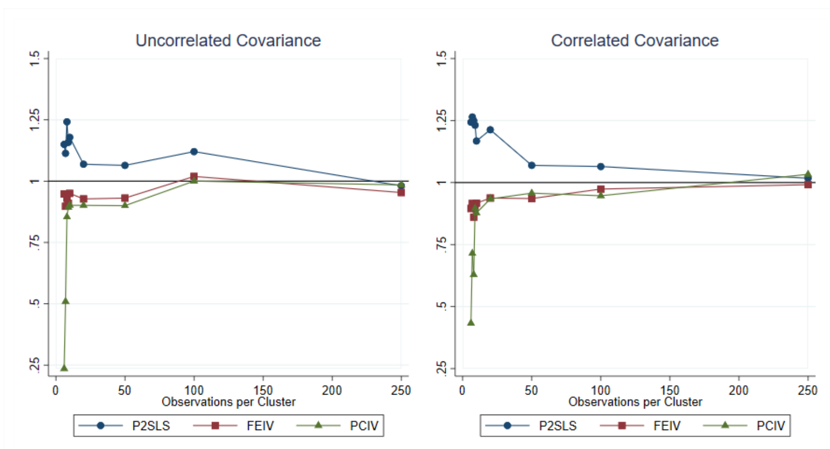
(a) Bias for coefficient β_1 of x_{ij}



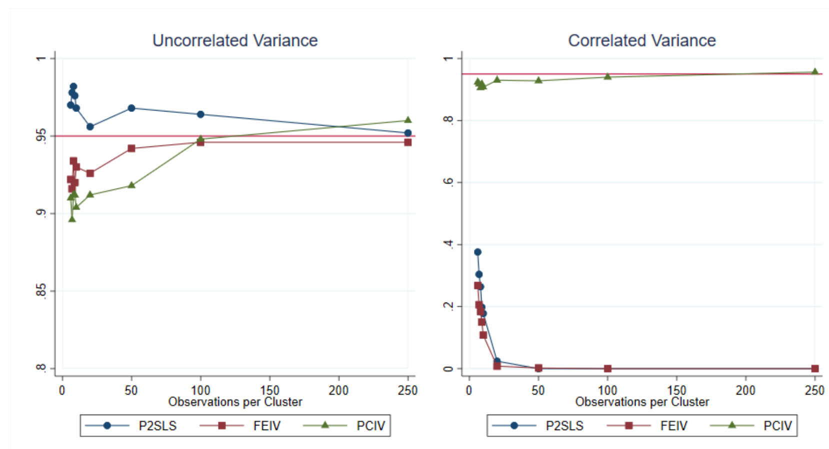
(b) Root mean square error for coefficient $\hat{\beta}_1$ of x_{ij}



(c) Mean standard errors over estimates' standard deviations

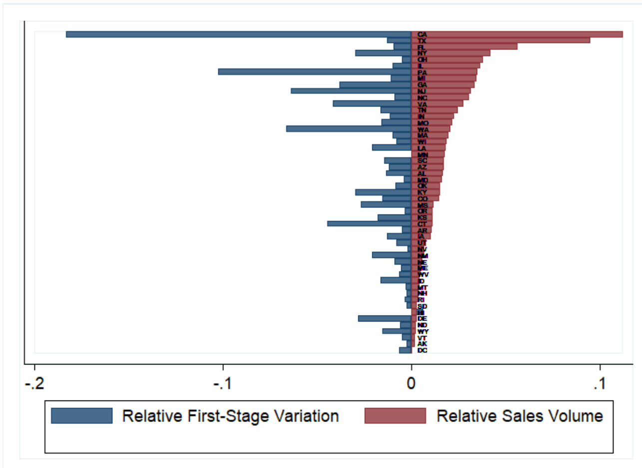


(d) Standard error coverage rates



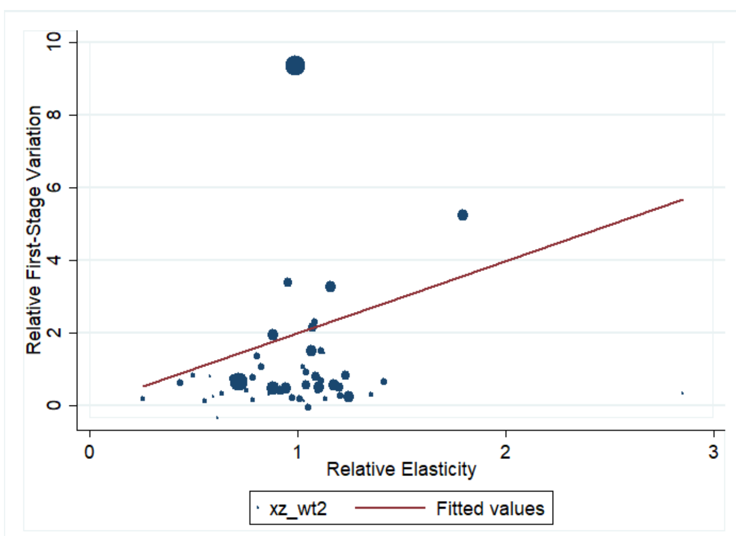
Notes: P2SLS = Pooled Two-Stage Least Squares; FEIV= Fixed Effects Instrumental Variables; PCIV=Per-Cluster Instrumental Variables. Each point represents results from 500 repetitions. Within each panel, figures on the left are when the uncorrelated covariance assumption holds, and figures on the right are when it is violated. Ratio of mean standard errors (SEs) divided by standard deviations (SDs) of the estimates. The horizontal line in (c) denote a ratio of one. The horizontal line in (d) is to denote the exact 95 percent coverage rate. In many instances, FEIV total overlaps with P2SLS, though this is an artifact of the specific DPG in this simulation.

Figure 2: Weighting under FEIV and PCIV approaches



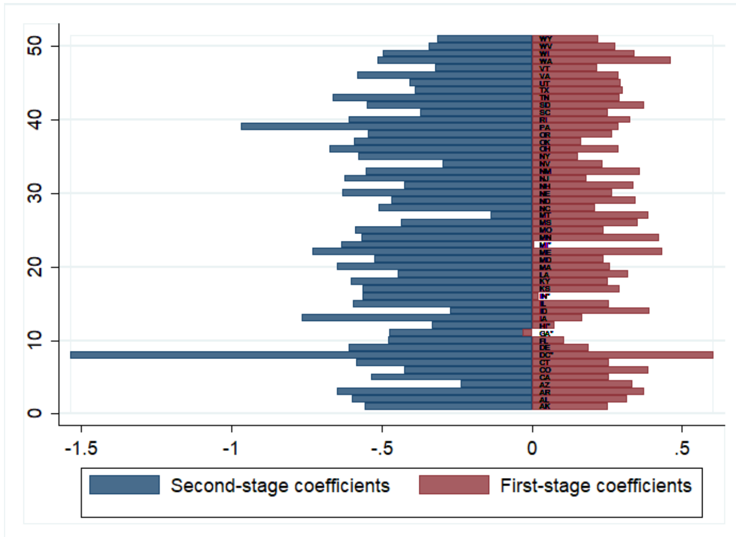
Notes: On the left we report the negative of the relative first-stage variation for FEIV estimation, as measured by the sum of the products of demeaned and detrended *logprice* and *logtax* for each state divided by the sum of those state totals. On the right, we report the share of total gasoline sales that occur in the state over the panel.

Figure 3: Heterogeneous elasticities and first-stage variation



Notes: We present elasticity estimates relative to the mean along the x-axis. Along the y-axis is relative first-stage variation as measured by the sum of the products of demeaned and detrended *logprice* and *logtax* for each state divided by the mean of those state totals. The fitted line has a slope coefficient of 2.000 and a standard error of 1.663.

Figure 4: State-specific first- and second-stage coefficients



Notes: On the left, we report PCIV state-specific elasticity estimates. We report state-specific first-stage tax pass-through estimates on the right. The standard deviation of the state-specific elasticities is 0.204 or 38% of the mean. The standard deviation of the state-specific pass-through rates is 0.115 or 42% of the mean.

A Proof of Theorem 1

$\hat{\boldsymbol{\beta}}_{PCIV}$ is consistent if and only if $\lim\limits_{n,T\to\infty} MSE(\hat{\boldsymbol{\beta}}_{PCIV}) \to 0$. We will show both asymptotic bias and variance will converge to zero asymptotically.

Define $\mathbf{P}_i = \mathbf{H}_{\mathbf{Z}_i} + \mathbf{M}_{\mathbf{Z}_i}\mathbf{X}_{2i}(\mathbf{X}_{2i}'\mathbf{M}_{\mathbf{Z}_i}\mathbf{X}_{2i})^{-1}\mathbf{X}_{2i}'\mathbf{M}_{\mathbf{Z}_i}$, where $\mathbf{H}_{\mathbf{Z}_i} = \mathbf{Z}_i(\mathbf{Z}_i'\mathbf{Z}_i)^{-1}\mathbf{Z}_i'$ and $\mathbf{M}_{\mathbf{Z}_i} = \mathbf{I}_T - \mathbf{H}_{\mathbf{Z}_i}$. Given the model as $\mathbf{y}_i = \mathbf{X}_{1i}\mathbf{b}_i + \mathbf{X}_{2i}\boldsymbol{\delta} + \mathbf{e}_i$ and the estimated $\hat{\boldsymbol{\delta}}$, per-cluster estimator $\hat{\mathbf{b}}_i$ can be written as follows as derived in Section 2.2:

$$\begin{aligned}
\hat{\mathbf{b}}_i &= (\mathbf{X}_{1i}'\mathbf{P}_i\mathbf{X}_{1i})^{-1}\mathbf{X}_{1i}'\mathbf{P}_i(\mathbf{y}_i - \mathbf{X}_{2i}\hat{\boldsymbol{\delta}}) \\
&= (\mathbf{X}_{1i}'\mathbf{P}_i\mathbf{X}_{1i})^{-1}\mathbf{X}_{1i}'\mathbf{P}_i(\mathbf{X}_{1i}\mathbf{b}_i + \mathbf{e}_i) \\
&= \mathbf{b}_i + (\mathbf{X}_{1i}'\mathbf{P}_i\mathbf{X}_{1i})^{-1}\mathbf{X}_{1i}'\mathbf{P}_i\mathbf{e}_i
\end{aligned}$$

Then our estimator of main interest, $\hat{\boldsymbol{\beta}}_{PCIV}$ can be written as follows:

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{PCIV} &= \sum_{i=1}^{n} w_i \hat{\mathbf{b}}_i \\
&= \sum_{i=1}^{n} w_i (\boldsymbol{\beta} + \mathbf{d}_i + (\mathbf{X}_{1i}' \mathbf{P}_i \mathbf{X}_{1i})^{-1} \mathbf{X}_{1i}' \mathbf{P}_i \mathbf{e}_i) \\
&= \boldsymbol{\beta} + \sum_{i=1}^{n} w_i \mathbf{d}_i + \sum_{i=1}^{n} w_i (\mathbf{X}_{1i}' \mathbf{P}_i \mathbf{X}_{1i})^{-1} \mathbf{X}_{1i}' \mathbf{P}_i \mathbf{e}_i
\end{aligned}$$

As $MSE(\hat{\boldsymbol{\beta}}_{PCIV}) = (Bias(\hat{\boldsymbol{\beta}}_{PCIV}))^2 + V(\hat{\boldsymbol{\beta}}_{PCIV})$, we will show each term converges to zero asymptotically. Note that the weight, w_i, is known and satisfies the condition as in Assumption (A6).

$$\begin{aligned}
Bias(\hat{\boldsymbol{\beta}}_{PCIV}) &= \mathrm{E}[\hat{\boldsymbol{\beta}}_{PCIV} - \boldsymbol{\beta}] \\
&= \mathrm{E}\left[\sum_{i=1}^{n} w_i (\hat{\mathbf{b}}_i - \boldsymbol{\beta})\right] \\
&= \mathrm{E}\left[\sum_{i=1}^{n} w_i (\mathbf{b}_i - \boldsymbol{\beta}) + \sum_{i=1}^{n} w_i (\hat{\mathbf{b}}_i - \mathbf{b}_i)\right] \\
&= \sum_{i=1}^{n} w_i \mathrm{E}[\mathbf{d}_i] + \sum_{i=1}^{n} w_i \mathrm{E}[\hat{\mathbf{b}}_i - \mathbf{b}_i]
\end{aligned}$$

From the last equality, $\mathrm{E}[w_i \mathbf{d}_i] = w_i \mathrm{E}[\mathbf{d}_i]$ can hold because w_i's are known constants, which are not correlated with $\mathbf{d}_i$'s. As $\mathrm{E}[\mathbf{d}_i] = 0$ by Assumption (A2), we will simplify $\mathrm{E}[\hat{\mathbf{b}}_i - \mathbf{b}_i]$. By Continuous Mapping Theorem (CMT),

$$\begin{aligned}
\mathrm{E}[\hat{\mathbf{b}}_i - \mathbf{b}_i] &= \mathrm{E}\left[(\mathbf{X}_{1i}' \mathbf{P}_i \mathbf{X}_{1i})^{-1} \mathbf{X}_{1i}' \mathbf{P}_i \mathbf{e}_i\right] \\
&= (\mathrm{E}[\mathbf{X}_{1i}' \mathbf{P}_i \mathbf{X}_{1i}])^{-1} \mathrm{E}\left[\mathbf{X}_{1i}' \mathbf{P}_i \mathbf{e}_i\right]
\end{aligned}$$

Note that $\mathbf{X}_{1i} = \mathbf{Z}_i\boldsymbol{\gamma}_i + \mathbf{X}_{2i}\boldsymbol{\eta} + u_i$ as in equation 1. Then, both terms in the denominator and the numerator can be written as follows:

$$\begin{aligned}
\mathrm{E}[\mathbf{X}_{1i}'\mathbf{P}_i\mathbf{X}_{1i}] &= \mathrm{E}\left[(\mathbf{Z}_i\boldsymbol{\gamma}_i + \mathbf{X}_{2i}\boldsymbol{\eta} + \mathbf{u}_i)'\mathbf{P}_i(\mathbf{Z}_i\boldsymbol{\gamma}_i + \mathbf{X}_{2i}\boldsymbol{\eta} + \mathbf{u}_i)\right] \\
&= \mathrm{E}[\boldsymbol{\gamma}_i'\mathbf{Z}_i'\mathbf{P}_i\mathbf{Z}_i\boldsymbol{\gamma}_i] + \mathrm{E}[\boldsymbol{\eta}'\mathbf{X}_{2i}'\mathbf{P}_i\mathbf{Z}_i\boldsymbol{\gamma}_i] + \mathrm{E}[\mathbf{u}_i'\mathbf{P}_i\mathbf{Z}_i\boldsymbol{\gamma}_i] + \mathrm{E}[\boldsymbol{\gamma}_i'\mathbf{Z}_i'\mathbf{P}_i\mathbf{X}_{2i}\boldsymbol{\eta}] \\
&+ \mathrm{E}[\boldsymbol{\eta}'\mathbf{X}_{2i}'\mathbf{P}_i\mathbf{X}_{2i}\boldsymbol{\eta}] + \mathrm{E}[\mathbf{u}_i'\mathbf{P}_i\mathbf{X}_{2i}\boldsymbol{\eta}] + \mathrm{E}[\boldsymbol{\gamma}_i'\mathbf{Z}_i'\mathbf{P}_i\mathbf{u}_i] + \mathrm{E}[\boldsymbol{\eta}'\mathbf{X}_{2i}'\mathbf{P}_i\mathbf{u}_i] + \mathrm{E}[\mathbf{u}_i'\mathbf{P}_i\mathbf{u}_i]
\end{aligned}$$

Note that $\mathbf{P}_i\mathbf{Z}_i = \mathbf{H}_{\mathbf{Z}_i}\mathbf{Z}_i + \mathbf{M}_{\mathbf{Z}_i}\mathbf{X}_{2i}(\mathbf{X}_{2i}'\mathbf{M}_{\mathbf{Z}_i}\mathbf{X}_{2i})^{-1}\mathbf{X}_{2i}'\mathbf{M}_{\mathbf{Z}_i}\mathbf{Z}_i = \mathbf{Z}_i$, and $\mathbf{X}_{2i}'\mathbf{P}_i\mathbf{X}_{2i} = \mathbf{X}_{2i}'\mathbf{X}_{2i}$. Also, $\mathrm{E}[\mathbf{u}_i'\mathbf{P}_i\mathbf{Z}_i\boldsymbol{\gamma}_i] = 0$, $\mathrm{E}[\mathbf{u}_i'\mathbf{P}_i\mathbf{X}_{2i}\boldsymbol{\eta}] = 0$, $\mathrm{E}[\boldsymbol{\gamma}_i'\mathbf{Z}_i'\mathbf{P}_i\mathbf{u}_i] = 0$, and $\mathrm{E}[\boldsymbol{\eta}'\mathbf{X}_{2i}'\mathbf{P}_i\mathbf{u}_i] = 0$ by Assumption (A2). Also, as $rank(\mathbf{P}_i) = L + H$, we get $\mathrm{E}[\mathbf{u}_i'\mathbf{P}_i\mathbf{u}_i] = (L+H)\cdot\sigma_u^2$. Then,

$$\mathrm{E}\left[\mathbf{X}_{1i}'\mathbf{P}_i\mathbf{X}_{1i}\right] = \boldsymbol{\gamma}_i'\mathbf{Z}_i'\mathbf{Z}_i\boldsymbol{\gamma}_i + \boldsymbol{\eta}'\mathbf{X}_{2i}'\mathbf{Z}_i\boldsymbol{\gamma}_i + \boldsymbol{\gamma}_i'\mathbf{Z}_i'\mathbf{X}_{2i}\boldsymbol{\eta} + \boldsymbol{\eta}'\mathbf{X}_{2i}'\mathbf{X}_{2i}\boldsymbol{\eta} + (L+H)\cdot\sigma_u^2.$$

By WLLN, as $\frac{1}{T}\mathbf{Z}_i'\mathbf{Z}_i \xrightarrow{p} \mathrm{E}[\mathbf{Z}_i'\mathbf{Z}_i] \equiv Q_{ZZ}$, the first term $\boldsymbol{\gamma}_i'\mathbf{Z}_i'\mathbf{Z}_i\boldsymbol{\gamma}_i = O_p(T)$. Likewise, the second, third, and the fourth terms follow the order of T. Next, the numerator in $\mathrm{E}[\hat{\mathbf{b}}_i - \mathbf{b}_i]$ can be simplified as follows. With $\mathbf{M}_{\mathbf{Z}_i}\mathbf{Z}_i = 0$ and (A2), the first and second terms become zero.

$$\begin{aligned}
\mathrm{E}[\mathbf{X}_{1i}'\mathbf{P}_i\mathbf{e}_i] &= \mathrm{E}\left[(\mathbf{Z}_i\boldsymbol{\gamma}_i + \mathbf{X}_{2i}\boldsymbol{\eta} + u_i)'\mathbf{P}_i\mathbf{e}_i\right] \\
&= \mathrm{E}[\boldsymbol{\gamma}_i'\mathbf{Z}_i'\mathbf{P}_i\mathbf{e}_i] + \mathrm{E}[\boldsymbol{\eta}'\mathbf{X}_{2i}'\mathbf{P}_i\mathbf{e}_i] + \mathrm{E}[u_i'\mathbf{P}_i\mathbf{e}_i] \\
&= (L+H)\cdot\sigma_{u,e}
\end{aligned}$$

We write

$$\mathrm{E}[\hat{\mathbf{b}}_i - \mathbf{b}_i] = \frac{(L+H)\cdot\sigma_{u,e}}{\boldsymbol{\gamma}_i'\mathbf{Z}_i'\mathbf{Z}_i\boldsymbol{\gamma}_i + \boldsymbol{\eta}'\mathbf{X}_{2i}'\mathbf{Z}_i\boldsymbol{\gamma}_i + \boldsymbol{\gamma}_i'\mathbf{Z}_i'\mathbf{X}_{2i}\boldsymbol{\eta} + \boldsymbol{\eta}'\mathbf{X}_{2i}'\mathbf{X}_{2i}\boldsymbol{\eta} + (L+H)\cdot\sigma_u^2} = O_p(T^{-1}).$$

Then, $Bias(\hat{\boldsymbol{\beta}}_{PCIV}) = O_p(T^{-1})$.

Now, we will look at the variance of $\hat{\boldsymbol{\beta}}_{PCIV}$. By Assumption (A1),

$$\begin{aligned} V(\hat{\boldsymbol{\beta}}_{PCIV} - \boldsymbol{\beta}) &= V\left(\sum_{i=1}^{n} w_i \mathbf{d}_i + \sum_{i=1}^{n} w_i(\hat{\mathbf{b}}_i - \mathbf{b}_i)\right) \\ &= \sum_{i=1}^{n} w_i^2 V(\mathbf{d}_i) + \sum_{i=1}^{n} w_i^2 V(\hat{\mathbf{b}}_i - \mathbf{b}_i) \end{aligned}$$

First, $V(\mathbf{d}_i) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{d}_i\mathbf{d}_i' \equiv \sigma_{\mathbf{d}}^2 \mathbf{I}_K$. Second,

$$\begin{aligned} V(\hat{\mathbf{b}}_i - \mathbf{b}_i) &= V((\mathbf{X}_{1i}'\mathbf{P}_i\mathbf{X}_{1i})^{-1}\mathbf{X}_{1i}'\mathbf{P}_i\mathbf{e}_i) \\ &= \mathrm{E}[(\mathbf{X}_{1i}'\mathbf{P}_i\mathbf{X}_{1i})^{-1}\mathbf{X}_{1i}'\mathbf{P}_i\mathbf{e}_i\mathbf{e}_i'\mathbf{P}_i\mathbf{X}_{1i}(\mathbf{X}_{1i}'\mathbf{P}_i\mathbf{X}_{1i})^{-1}]. \end{aligned}$$

We will break down by each component of the variance.

$$\begin{aligned} \mathbf{X}_{1i}'\mathbf{P}_i\mathbf{X}_{1i} &= \mathbf{X}_{1i}'\mathbf{H}_{\mathbf{Z}_i}\mathbf{X}_{1i} + \mathbf{X}_{1i}'\mathbf{M}_{\mathbf{Z}_i}\mathbf{X}_{2i}(\mathbf{X}_{2i}'\mathbf{M}_{\mathbf{Z}_i}\mathbf{X}_{2i})^{-1}\mathbf{X}_{2i}'\mathbf{M}_{\mathbf{Z}_i}\mathbf{X}_{1i} \\ &= \mathbf{X}_{1i}'\mathbf{Z}_i(\mathbf{Z}_i'\mathbf{Z}_i)^{-1}\mathbf{Z}_i'\mathbf{X}_{1i} + \mathbf{X}_{1i}'(\mathbf{I}_N - \mathbf{Z}_i(\mathbf{Z}_i'\mathbf{Z}_i)^{-1}\mathbf{Z}_i')\mathbf{X}_{2i}(\mathbf{X}_{2i}'(\mathbf{I}_N - \mathbf{Z}_i(\mathbf{Z}_i'\mathbf{Z}_i)^{-1}\mathbf{Z}_i')\mathbf{X}_{2i})^{-1} \\ &\quad \times \mathbf{X}_{2i}'(\mathbf{I}_N - \mathbf{Z}_i(\mathbf{Z}_i'\mathbf{Z}_i)^{-1}\mathbf{Z}_i')\mathbf{X}_{1i} \end{aligned}$$

By WLLN, $\frac{1}{T}\mathbf{X}_{1i}'\mathbf{Z}_i \xrightarrow{p} \mathrm{E}[\mathbf{X}_{1i}'\mathbf{Z}_i] \equiv Q_{1Z}$. Likewise, we define $Q_{Z1} \equiv \mathrm{E}[\mathbf{Z}_i'\mathbf{X}_{1i}]$, $Q_{ZZ} \equiv \mathrm{E}[\mathbf{Z}_i'\mathbf{Z}_i]$, $Q_{Z2} \equiv \mathrm{E}[\mathbf{Z}_i'\mathbf{X}_{2i}]$, $Q_{2Z} \equiv \mathrm{E}[\mathbf{X}_{2i}'\mathbf{Z}_i]$, $Q_{22} \equiv \mathrm{E}[\mathbf{X}_{2i}'\mathbf{X}_{2i}]$, $Q_{12} \equiv \mathrm{E}[\mathbf{X}_{1i}'\mathbf{X}_{2i}]$, and $Q_{21} \equiv \mathrm{E}[\mathbf{X}_{2i}'\mathbf{X}_{1i}]$. Then, we write $\frac{1}{T}\mathbf{X}_{1i}'\mathbf{P}_i\mathbf{X}_{1i} = Q_{1Z}Q_PQ_{Z1}$ asymptotically, where $Q_P = Q_{ZZ}^{-1} + (Q_{12} - Q_{1Z}Q_{ZZ}^{-1}Q_{Z2})(Q_{22} - Q_{2Z}Q_{ZZ}^{-1}Q_{Z2})^{-1}(Q_{21} - Q_{2Z}Q_{ZZ}^{-1}Q_{Z1})$. Next, the second component of the variance is written as below.

$$\begin{aligned} \mathbf{X}_{1i}'\mathbf{P}_i\mathbf{e}_i\mathbf{e}_i'\mathbf{P}_i\mathbf{X}_{1i} &= \mathbf{X}_{1i}'\mathbf{H}_{\mathbf{Z}_i}\mathbf{e}_i\mathbf{e}_i'\mathbf{H}_{\mathbf{Z}_i}\mathbf{X}_{1i} + \mathbf{X}_{1i}'\mathbf{M}_{\mathbf{Z}_i}\mathbf{X}_{2i}(\mathbf{X}_{2i}'\mathbf{M}_{\mathbf{Z}_i}\mathbf{X}_{2i})^{-1}\mathbf{X}_{2i}'\mathbf{M}_{\mathbf{Z}_i}\mathbf{e}_i\mathbf{e}_i'\mathbf{M}_{\mathbf{Z}_i}\mathbf{X}_{2i} \\ &\quad \times (\mathbf{X}_{2i}'\mathbf{M}_{\mathbf{Z}_i}\mathbf{X}_{2i})^{-1}\mathbf{X}_{2i}'\mathbf{M}_{\mathbf{Z}_i}\mathbf{X}_{1i}. \end{aligned}$$

By WLLN, we define $\frac{1}{T}\mathbf{Z}_i'\mathbf{e}_i\mathbf{e}_i'\mathbf{Z}_i \xrightarrow{p} \mathrm{E}[\mathbf{Z}_i'\mathbf{e}_i\mathbf{e}_i'\mathbf{Z}_i] \equiv Q_{Ze}$. Then, we write $\mathbf{X}_{1i}'\mathbf{P}_i\mathbf{e}_i\mathbf{e}_i'\mathbf{P}_i\mathbf{X}_{1i} = Q_{1Z}Q_P\Omega_{Ze}Q_PQ_{Z1}$. By CMT,

$$V(\hat{\mathbf{b}}_i - \mathbf{b}_i) = \frac{1}{T}(Q_{1Z}Q_PQ_{Z1})^{-1}(Q_{1Z}Q_P\Omega_{Ze}Q_PQ_{Z1})(Q_{1Z}Q_PQ_{Z1})^{-1} = O_p(T^{-1})$$

Then, we have $V(\hat{\boldsymbol{\beta}}_{PCIV}) = O_p(1/n) + O_p(1/nT) = O_p(1/n)$. To summarize,

$$\begin{aligned} MSE(\hat{\boldsymbol{\beta}}_{PCIV}) &= (Bias(\hat{\boldsymbol{\beta}}_{PCIV}))^2 + V(\hat{\boldsymbol{\beta}}_{PCIV}) \\ &= O_p(1/T^2) + O_p(1/n). \end{aligned}$$

As we assume Assumption (A5), we can show that $\lim_{n,T\to\infty} MSE(\hat{\boldsymbol{\beta}}_{PCIV}) \xrightarrow{p} 0$. As $n, T \to \infty$,

$$\hat{\boldsymbol{\beta}}_{PCIV} \xrightarrow{p} \boldsymbol{\beta}.$$

B Proof of Theorem 2

First, we will show the consistency of an estimated variance.

$$\hat{V}(\hat{\boldsymbol{\beta}}_{PCIV} - \boldsymbol{\beta}) = \hat{V}\left(\sum_{i=1}^{n} w_i \mathbf{d}_i + \sum_{i=1}^{n} w_i (\mathbf{X}_{1i}'\mathbf{P}_i\mathbf{X}_{1i})^{-1}\mathbf{X}_{1i}'\mathbf{P}_i\mathbf{e}_i\right)$$

By using Assumptions (A1) and (A2),

$$\hat{V}(\hat{\boldsymbol{\beta}}_{PCIV} - \boldsymbol{\beta}) = \sum_{i=1}^{n} w_i^2 \hat{V}(\mathbf{d}_i) + \sum_{i=1}^{n} w_i^2 V((\mathbf{X}_{1i}'\mathbf{P}_i\mathbf{X}_{1i})^{-1}\mathbf{X}_{1i}'\mathbf{P}_i\hat{\mathbf{e}}_i)$$

First, we will show $\hat{V}(\mathbf{d}_i) \to V(\mathbf{d}_i)$.

$$\begin{aligned}\hat{V}(\mathbf{d}_i) &= \frac{1}{n}\sum_{i=1}^{n} \hat{\mathbf{d}}_i\hat{\mathbf{d}}_i' = \frac{1}{n}\sum_{i=1}^{n}(\hat{\mathbf{d}}_i - \mathbf{d}_i + \mathbf{d}_i)(\hat{\mathbf{d}}_i - \mathbf{d}_i + \mathbf{d}_i)' \\ &= \frac{1}{n}\sum_{i=1}^{n} \mathbf{d}_i\mathbf{d}_i' + \frac{1}{n}\sum_{i=1}^{n}(\hat{\mathbf{d}}_i - \mathbf{d}_i)(\hat{\mathbf{d}}_i - \mathbf{d}_i)'\end{aligned}$$

By WLLN, it is easy to show that $\frac{1}{n}\sum_{i=1}^{n} \mathbf{d}_i\mathbf{d}_i' \xrightarrow{p} \sigma_{\mathbf{d}}^2\mathbf{I}_K$ as $\mathrm{E}[\mathbf{d}_i] = 0$. Also, as $\hat{\mathbf{d}}_i \xrightarrow{p} \mathbf{d}_i$ as $n \to \infty$, the second term then converges to zero asymptotically. Therefore, we can show $\hat{V}(\mathbf{d}_i) \to \sigma_{\mathbf{d}}^2\mathbf{I}_K \equiv V(\mathbf{d}_i)$ asymptotically. Next, let's show the convergence of the second term of the sample variance.

$$V((\mathbf{X}_{1i}'\mathbf{P}_i\mathbf{X}_{1i})^{-1}\mathbf{X}_{1i}'\mathbf{P}_i\hat{\mathbf{e}}_i) = \mathrm{E}[(\mathbf{X}_{1i}'\mathbf{P}_i\mathbf{X}_{1i})^{-1}\mathbf{X}_{1i}'\mathbf{P}_i\hat{\mathbf{e}}_i\hat{\mathbf{e}}_i'\mathbf{P}_i\mathbf{X}_{1i}'(\mathbf{X}_{1i}'\mathbf{P}_i\mathbf{X}_{1i})^{-1}]$$

Given that $\mathbf{P}_i = \mathbf{H}_{\mathbf{Z}_i} + \mathbf{M}_{\mathbf{Z}_i}\mathbf{X}_{2i}(\mathbf{X}_{2i}'\mathbf{M}_{\mathbf{Z}_i}\mathbf{X}_{2i})^{-1}\mathbf{X}_{2i}'\mathbf{M}_{\mathbf{Z}_i}$, we will show that $\frac{1}{T}\mathbf{Z}_i'\hat{\mathbf{e}}_i\hat{\mathbf{e}}_i'\mathbf{Z}_i \xrightarrow{p} \mathrm{E}[\mathbf{Z}_i'\mathbf{e}_i\mathbf{e}_i'\mathbf{Z}_i] \equiv \Omega_{Ze}$.

Next, let's show the convergence of the second term in equation 10. Using the above expression, we will show the convergence of the below inequality.

$$\left\|\frac{1}{T}\mathbf{Z}_i'(\hat{\mathbf{e}}_i - \mathbf{e}_i)(\hat{\mathbf{e}}_i - \mathbf{e}_i)'\mathbf{Z}_i\right\| \leq \frac{1}{T}\|\mathbf{Z}_i\|^2\|\hat{\mathbf{e}}_i - \mathbf{e}_i\|^2 \qquad (9)$$

Using $\hat{\mathbf{X}}_{1i} = \mathbf{Z}_i\hat{\boldsymbol{\gamma}}_i + \mathbf{X}_{2i}\hat{\boldsymbol{\eta}}$, we write $\hat{\mathbf{e}}_i = \mathbf{y}_i - \hat{\mathbf{X}}_{1i}\hat{\mathbf{b}}_i - \mathbf{X}_{2i}\hat{\boldsymbol{\delta}} = \mathbf{e}_i - \mathbf{Z}_i(\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i)(\hat{\mathbf{b}}_i - \mathbf{b}_i) - \mathbf{X}_{2i}(\hat{\boldsymbol{\eta}}-\boldsymbol{\eta})(\hat{\mathbf{b}}_i-\mathbf{b}_i)-\mathbf{Z}_i(\hat{\boldsymbol{\gamma}}_i-\boldsymbol{\gamma}_i)\mathbf{b}_i-\mathbf{X}_{2i}(\hat{\boldsymbol{\eta}}-\boldsymbol{\eta})\mathbf{b}_i-\mathbf{Z}_i\boldsymbol{\gamma}_i(\hat{\mathbf{b}}_i-\mathbf{b}_i)-\mathbf{X}_{2i}\boldsymbol{\eta}(\hat{\mathbf{b}}_i-\mathbf{b}_i)-\mathbf{X}_{2i}(\hat{\boldsymbol{\delta}}-\boldsymbol{\delta})$. Then, we write equation 9 as follows.

$$
\begin{aligned}
&\frac{1}{T}\|\mathbf{Z}_i\|^2\|\hat{\mathbf{e}}_i - \mathbf{e}_i\|^2 \\
&\leq \frac{1}{T}\|\hat{\mathbf{b}}_i - \mathbf{b}_i\|^2\|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i\|^2\|\mathbf{Z}_i\|^4 + \frac{1}{T}\|\hat{\mathbf{b}}_i - \mathbf{b}_i\|^2\|\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}\|^2\|\mathbf{X}_{2i}\|^2\|\mathbf{Z}_i\|^2 + \frac{1}{T}\|\mathbf{b}_i\|^2\|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i\|^2\|\mathbf{Z}_i\|^4 \\
&+ \frac{1}{T}\|\mathbf{b}_i\|^2\|\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}\|^2\|\mathbf{X}_{2i}\|^2\|\mathbf{Z}_i\|^2 + \frac{1}{T}\|\hat{\mathbf{b}}_i - \mathbf{b}_i\|^2\|\boldsymbol{\gamma}_i\|^2\|\mathbf{Z}_i\|^2 + \frac{1}{T}\|\hat{\mathbf{b}}_i - \mathbf{b}_i\|^2\|\boldsymbol{\eta}\|^2\|\mathbf{X}_{2i}\|^2\|\mathbf{Z}_i\|^2 \\
&+ \frac{1}{T}\|\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}\|^2\|\mathbf{X}_{2i}\|^2\|\mathbf{Z}_i\|^2 + 2\frac{1}{T}\|\hat{\mathbf{b}}_i - \mathbf{b}_i\|^2\|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i\|\|\mathbf{Z}_i\|^3\|\mathbf{X}_{2i}\|\|\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}\| \\
&+ 2\frac{1}{T}\|\hat{\mathbf{b}}_i - \mathbf{b}_i\|\|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i\|^2\|\mathbf{Z}_i\|^4\|\mathbf{b}_i\| + 2\frac{1}{T}\|\hat{\mathbf{b}}_i - \mathbf{b}_i\|\|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i\|\|\mathbf{Z}_i\|^3\|\mathbf{X}_{2i}\|\|\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}\|\|\mathbf{b}_i\| \\
&+ 2\frac{1}{T}\|\hat{\mathbf{b}}_i - \mathbf{b}_i\|^2\|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i\|\|\mathbf{Z}_i\|^4 + 2\frac{1}{T}\|\hat{\mathbf{b}}_i - \mathbf{b}_i\|^2\|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i\|\|\mathbf{Z}_i\|^3\|\mathbf{X}_{2i}\|\|\boldsymbol{\eta}\|\|\hat{\mathbf{b}}_i - \mathbf{b}_i\| \\
&+ 2\frac{1}{T}\|\hat{\mathbf{b}}_i - \mathbf{b}_i\|\|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i\|\|\mathbf{Z}_i\|^3\|\mathbf{X}_{2i}\|\|\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}\| + 2\frac{1}{T}\|\hat{\mathbf{b}}_i - \mathbf{b}_i\|\|\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}\|\|\mathbf{X}_{2i}\|\|\mathbf{Z}_i\|^3\|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i\|\|\mathbf{b}_i\| \\
&+ 2\frac{1}{T}\|\hat{\mathbf{b}}_i - \mathbf{b}_i\|\|\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}\|^2\|\mathbf{X}_{2i}\|^2\|\mathbf{Z}_i\|^2\|\mathbf{b}_i\| + 2\frac{1}{T}\|\hat{\mathbf{b}}_i - \mathbf{b}_i\|^2\|\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}\|\|\mathbf{X}_{2i}\|\|\mathbf{Z}_i\|^3\|\boldsymbol{\gamma}_i\| \\
&+ 2\frac{1}{T}\|\hat{\mathbf{b}}_i - \mathbf{b}_i\|^2\|\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}\|\|\mathbf{Z}_i\|^2\|\mathbf{X}_{2i}\|^2\|\boldsymbol{\eta}\| + 2\frac{1}{T}\|\hat{\mathbf{b}}_i - \mathbf{b}_i\|\|\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}\|\|\mathbf{X}_{2i}\|\|\mathbf{Z}_i\|^3\|\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}\|\|\mathbf{b}_i\| \\
&+ 2\frac{1}{T}\|\mathbf{b}_i\|^2\|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i\|\|\mathbf{Z}_i\|^3\|\mathbf{X}_{2i}\|\|\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}\| + 2\frac{1}{T}\|\mathbf{b}_i\|\|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i\|\|\mathbf{Z}_i\|^2\|\boldsymbol{\gamma}_i\|\|\hat{\mathbf{b}}_i - \mathbf{b}_i\| \\
&+ 2\frac{1}{T}\|\mathbf{b}_i\|\|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i\|\|\mathbf{Z}_i\|^3\|\mathbf{X}_{2i}\|\|\boldsymbol{\eta}\|\|\hat{\mathbf{b}}_i - \mathbf{b}_i\| + 2\frac{1}{T}\|\mathbf{b}_i\|\|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i\|\|\mathbf{Z}_i\|^3\|\mathbf{X}_{2i}\|\|\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}\| \\
&+ 2\frac{1}{T}\|\mathbf{b}_i\|\|\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}\|\|\mathbf{X}_{2i}\|\|\mathbf{Z}_i\|^3\|\boldsymbol{\gamma}_i\|\|\hat{\mathbf{b}}_i - \mathbf{b}_i\| + 2\frac{1}{T}\|\mathbf{b}_i\|\|\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}\|\|\mathbf{Z}_i\|^2\|\mathbf{X}_{2i}\|^2\|\boldsymbol{\eta}\|\|\hat{\mathbf{b}}_i - \mathbf{b}_i\| \\
&+ 2\frac{1}{T}\|\mathbf{b}_i\|\|\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}\|\|\mathbf{Z}_i\|^2\|\mathbf{X}_{2i}\|^2\|\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}\| + 2\frac{1}{T}\|\hat{\mathbf{b}}_i - \mathbf{b}_i\|^2\|\boldsymbol{\gamma}_i\|\|\mathbf{Z}_i\|^3\|\mathbf{X}_{2i}\|\|\boldsymbol{\eta}\| \\
&+ 2\frac{1}{T}\|\hat{\mathbf{b}}_i - \mathbf{b}_i\|\|\boldsymbol{\gamma}_i\|\|\mathbf{Z}_i\|^3\|\mathbf{X}_{2i}\|\|\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}\| + 2\frac{1}{T}\|\hat{\mathbf{b}}_i - \mathbf{b}_i\|\|\boldsymbol{\eta}\|\|\mathbf{Z}_i\|^2\|\mathbf{X}_{2i}\|^2\|\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}\|
\end{aligned}
$$

The convergence of the above inequality can be shown by the consistency of the three estimators respectively: 1) $\hat{\boldsymbol{\eta}} \xrightarrow{p} \boldsymbol{\eta}$, 2) $\hat{\boldsymbol{\gamma}}_i \xrightarrow{p} \boldsymbol{\gamma}_i$, and 3) $\hat{\boldsymbol{\delta}} \xrightarrow{p} \boldsymbol{\delta}$. In addition, using the Hölder's inequality, we can easily show $\mathrm{E}[\|\mathbf{Z}_i\|^3\|\mathbf{X}_{2i}\|] \leq \left(\mathrm{E}[(\|\mathbf{Z}_i\|^3)^{\frac{4}{3}}]\right)^{\frac{3}{4}} (\mathrm{E}[(\|\mathbf{X}_{2i}\|)^4])^{\frac{1}{4}} < \infty$ under Assumption (A4).

First, $\hat{\boldsymbol{\eta}} = \boldsymbol{\eta} + \left(\sum_{i=1}^n \mathbf{X}_{2i}'\mathbf{M}_{\mathbf{Z}_i}\mathbf{X}_{2i}\right)^{-1}\left(\sum_{i=1}^n \mathbf{X}_{2i}'\mathbf{M}_{\mathbf{Z}_i}\mathbf{u}_i\right)$. By Assumption (A3),

$$\operatorname*{plim}_{n,T\to\infty} \hat{\boldsymbol{\eta}} = \boldsymbol{\eta} + (Q_{2Z}Q_{ZZ}^{-1}Q_{Z2})^{-1}Q_{2Z}Q_{ZZ}^{-1} \operatorname*{plim}_{n,T\to\infty} \frac{1}{nT}\sum_{i=1}^n \mathbf{Z}_i'\mathbf{u}_i = \boldsymbol{\eta}$$

Next, $\hat{\boldsymbol{\gamma}}_i = \boldsymbol{\gamma}_i + (\mathbf{Z}_i'\mathbf{Z}_i)^{-1}\mathbf{Z}_i'\mathbf{u}_i + (\mathbf{Z}_i'\mathbf{Z}_i)^{-1}\mathbf{Z}_i'\mathbf{X}_{2i}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})$. Under Assumption (A3) and the consistency of $\boldsymbol{\eta}$,

$$\operatorname*{plim}_{n,T\to\infty} \hat{\boldsymbol{\gamma}}_i = \boldsymbol{\gamma}_i + Q_{ZZ}^{-1} \operatorname*{plim}_{n,T\to\infty} \frac{1}{T}\mathbf{Z}_i'\mathbf{u}_i + Q_{ZZ}^{-1}Q_{Z2} \operatorname*{plim}_{N,T\to\infty} (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}) = \boldsymbol{\gamma}_i$$

Lastly, $\hat{\boldsymbol{\delta}} = \boldsymbol{\delta} + \left(\sum_{i=1}^n \mathbf{X}_{2i}'\mathbf{M}_{\hat{\mathbf{X}}_{1i}}\mathbf{X}_{2i}\right)^{-1}\left(\sum_{i=1}^n \mathbf{X}_{2i}'\mathbf{M}_{\hat{\mathbf{X}}_{1i}}\mathbf{e}_i\right)$, where $\mathbf{M}_{\hat{\mathbf{X}}_{1i}} = \mathbf{I}_T - \hat{\mathbf{X}}_{1i}(\hat{\mathbf{X}}_{1i}'\hat{\mathbf{X}}_{1i})^{-1}\hat{\mathbf{X}}_{1i}' = \mathbf{I}_T - \mathbf{P}_i\mathbf{X}_{1i}(\mathbf{X}_{1i}'\mathbf{P}_i\mathbf{X}_{1i})^{-1}\mathbf{X}_{1i}'\mathbf{P}_i'$. Define $Q_P = Q_{ZZ}^{-1} + (Q_{12} - Q_{1Z}Q_{ZZ}^{-1}Q_{Z2})(Q_{22} - Q_{2Z}Q_{ZZ}^{-1}Q_{Z2})^{-1}(Q_{21} - Q_{2Z}Q_{ZZ}^{-1}Q_{Z1})$.

$$\begin{aligned}
\operatorname*{plim}_{n,T\to\infty} \hat{\boldsymbol{\delta}} &= \boldsymbol{\delta} + \left(Q_{22} - Q_{2Z}Q_PQ_{Z1}(Q_{1Z}Q_PQ_{Z1})^{-1}Q_{1Z}Q_PQ_{Z2}\right)^{-1} \\
&\quad \times \left[\operatorname*{plim}_{n,T\to\infty} \frac{1}{nT}\sum_{i=1}^n \mathbf{X}_{2i}'\mathbf{e}_i + Q_{2Z}Q_PQ_{Z1}(Q_{1Z}Q_PQ_{Z1})^{-1}Q_{1Z}Q_P \operatorname*{plim}_{n,T\to\infty} \frac{1}{nT}\sum_{i=1}^n \mathbf{Z}_i\mathbf{e}_i\right] \\
&= \boldsymbol{\delta},
\end{aligned}$$

By WLLN, $\frac{1}{T}\mathbf{Z}_i'\mathbf{e}_i\mathbf{e}_i'\mathbf{Z}_i \xrightarrow{p} \mathrm{E}[\mathbf{Z}_i'\mathbf{e}_i\mathbf{e}_i'\mathbf{Z}_i] \equiv \Omega_{Ze}$. As $n, T \to \infty$, and $n/T \to c$ where $0 < c < \infty$,

$$\hat{V}(\hat{\boldsymbol{\beta}}_{PCIV}) \xrightarrow{p} V(\hat{\boldsymbol{\beta}}_{PCIV}).$$

Now, we show the asymptotic normality of $\hat{\boldsymbol{\beta}}_{PCIV}$.

$$\frac{1}{T}\mathbf{Z}_i'\hat{\mathbf{e}}_i\hat{\mathbf{e}}_i'\mathbf{Z}_i = \frac{1}{T}\mathbf{Z}_i'\mathbf{e}_i\mathbf{e}_i'\mathbf{Z}_i + \frac{1}{T}\mathbf{Z}_i'(\hat{\mathbf{e}}_i - \mathbf{e}_i)(\hat{\mathbf{e}}_i - \mathbf{e}_i)'\mathbf{Z}_i \tag{10}$$

Using the Cauchy-Schwarz inequality, we can show that Ω_{Ze} is finite by Assumptions (A2) and (A3). Note that $(\mathrm{E}[e_{ij}^4])^{\frac{1}{4}} = (\mathrm{E}[(y_{ij} - \mathbf{x}_{1ij}\mathbf{b}_i - \mathbf{x}_{2ij}\boldsymbol{\delta})^4])^{\frac{1}{4}} \leq (\mathrm{E}[y_{ij}^4])^{\frac{1}{4}} + \|\mathbf{b}_i\|(\mathrm{E}[\mathbf{x}_{2ij}^4])^{\frac{1}{4}} +$

$\|\boldsymbol{\delta}\|(\mathrm{E}[\mathbf{x}_{2ij}^4])^{\frac{1}{4}} < \infty$ by Assumption (A4). Then,

$$\|\Omega_{Ze}\| \leq \mathrm{E}\left[\|\mathbf{Z}_i\|^2 \|e_i\|^2\right] \leq \left(\mathrm{E}[\|\mathbf{Z}_i\|^4]\right)^{\frac{1}{2}} \left(\mathrm{E}[\|e_i\|^4]\right)^{\frac{1}{2}} < \infty.$$

. We then have $\frac{1}{T}\mathbf{Z}_i'\mathbf{e}_i \xrightarrow{d} N(0, \Omega_{Ze})$.

$$\begin{aligned}\sqrt{n}(\hat{\boldsymbol{\beta}}_{\boldsymbol{PCIV}} - \boldsymbol{\beta}) &= \underbrace{\sqrt{n}\sum_{i=1}^{n} w_i \mathbf{d}_i}_{=O_p(1)} + \underbrace{\sqrt{n}\sum_{i=1}^{n} w_i (\mathbf{X}_{1i}'\mathbf{P}_i\mathbf{X}_{1i})^{-1}\mathbf{X}_{1i}'\mathbf{P}_i\mathbf{e}_i}_{=O_p\left(\frac{\sqrt{n}}{\sqrt{T}}\right)} \\ &\xrightarrow{d} N(0, V(\hat{\boldsymbol{\beta}})),\end{aligned}$$

where $V(\hat{\boldsymbol{\beta}}) = \lambda(\sigma_{\mathbf{d}}^2\mathbf{I}_K + (Q_{1Z}Q_PQ_{Z1})^{-1}Q_{1Z}Q_P\Omega_{Ze}Q_PQ_{Z1}(Q_{1Z}Q_PQ_{Z1})^{-1})$, and $\lambda = O_p(1)$.

C Supplementary tables and figures

Table 4: 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 the mean of standard errors divided by standard deviations; CR=Coverage Rate.

Table 5: Summary Statistics

	Mean	SD	p25	Median	p75
After tax price (1999 USD)	2.69	0.93	1.97	2.34	3.25
State taxes (1999 USD)	0.52	0.09	0.45	0.52	0.58
Unemployment rate	5.54	1.86	4.20	5.20	6.60
Real per-capita income (1999 USD thousands)	39.37	20.42	21.47	35.80	54.19
Population (in millions)	5.88	6.43	1.77	4.13	6.73
Licensed drivers (in millions)	4.05	4.16	1.28	2.85	4.94
Road miles (in thousands)	49.02	55.70	0.37	34.10	88.75
Minimum temperature (degrees F)	40.99	17.05	28.70	41.40	54.70
Average temperature (degrees F)	52.12	17.82	38.90	53.35	66.90
Maximum temperature (degrees F)	63.25	18.80	49.10	65.30	79.00
Rainfall (inches)	3.16	2.03	1.61	2.89	4.34
Sales volume (millions of gallons)	6.96	7.55	1.94	5.22	8.55
Observations	18360				

Notes: The data gasoline prices, taxes, and sales volumes come from the U.S. Department of Energy, Energy Information Administration (EIA). We supplement the price data after 2011 from gasbuddy.com. Population and average income come from the Bureau of Economic Analysis (BEA). The count of licensed drivers and road miles come from the Federal Highway Administration's Highway Statistics. Precipitation and rainfall data come from National Centers for Environmental Information, National Oceanic and Atmospheric Administration.

Table 6: First-Stage Estimation Results

	Without volume weights			Volume weighted		
	P2SLS	FEIV	PCIV	P2SLS	FEIV	PCIV
Log price	0.204	0.248	0.282	0.204	0.311	0.253
	(0.034)	(0.028)	(0.066)	(0.030)	(0.039)	(0.070)
First-stage F-statistic	36.66	79.71	58.35	47.47	63.70	61.16
Controls	N	N	N	N	N	N
Log price	0.204	0.241	0.280	0.205	0.298	0.278
	(0.034)	(0.027)	(0.050)	(0.030)	(0.038)	(0.053)
First-stage F-statistic	36.58	80.92	58.71	46.83	60.26	59.93
Controls	Y	Y	Y	Y	Y	Y

Notes: Standard errors are clustered by state. The calculated first-stage F-statistics for P2SLS and FEIV are obtained from the regression of each endogenous regressor on the exogenous regressors and the instruments. The calculation of the first-stage F-statistics for the PCIV was done using the Hotelling's T-squared test.

Table 7: Estimated elasticities using month and year fixed effect

	Without volume weights			Volume weighted		
	P2SLS	FEIV	PCIV	P2SLS	FEIV	PCIV
Log price	-1.080	-0.884	-0.542	-0.749	-0.843	-0.546
	(0.204)	(0.389)	(0.246)	(0.168)	(0.373)	(0.260)
First-stage F-statistic	65.66	88.16	97.29	21.74	71.79	110.53
Controls	N	N	N	N	N	N

Notes: All regression include month and year fixed effects instead of month-by-year fixed effects. The calculated first-stage F-statistics for P2SLS and FEIV are obtained from the regression of each endogenous regressor on the exogenous regressors and the instruments. The calculation of the first-stage F-statistics for the PCIV was done using the Hotelling's T-squared test.

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

PADD	New England	Central Atlantic	Lower Atlantic	Midwest	Gulf Coast	Rocky Mountains	West Coast
log price	-0.970	-1.160	-0.889	-1.011	-0.977	-0.845	-0.866
	(0.369)	(0.369)	(0.369)	(0.369)	(0.369)	(0.369)	(0.369)
F-statistic	1.83	1.83	1.83	1.83	1.83	1.83	1.83
Number of states	6	6	6	15	6	5	7
Observations	2,160	2,160	2,160	5,400	2,160	1,800	2,520

Notes: All regressions include state FEs, control for local unemployment rates, and detrended data. State-clustered standard errors appear in parentheses.

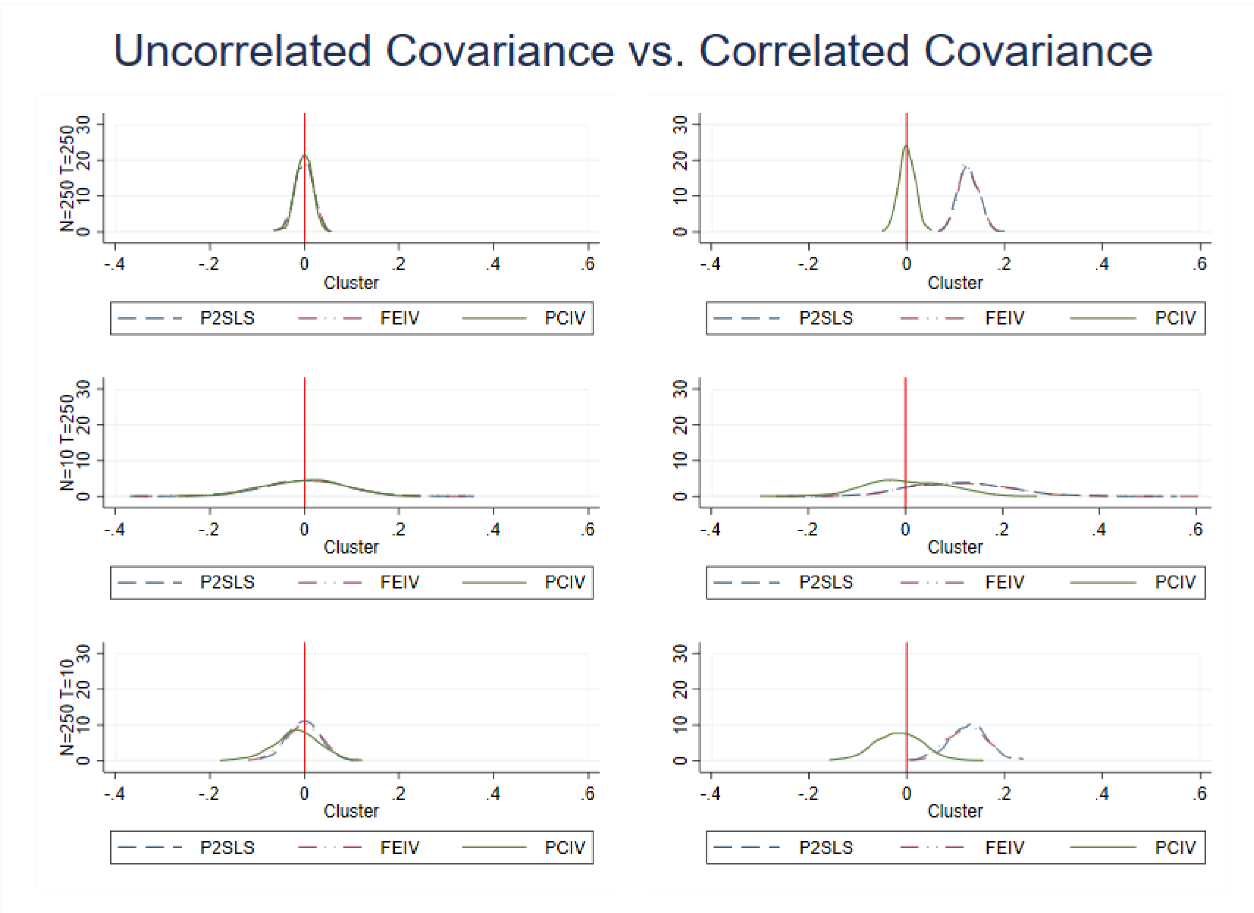
Table 9: State-Specific Statistics

State	First-stage Coefficient	First-stage F-stat	Second-Stage Coefficient	State	First-stage Coefficient	First-stage F-stat	Second-Stage Coefficient
AK	0.32	721.94	-0.61	MT	0.40	66.90	-0.15
AL	0.26	13.92	-0.57	NE	0.27	910.25	-0.64
AZ	0.34	182.59	-0.25	NV	0.25	35.76	-0.31
AR	0.38	22.83	-0.66	NH	0.34	806.14	-0.44
CA	0.26	96.98	-0.55	NJ	0.29	65.23	-0.66
CO	0.39	925.73	-0.44	NM	0.36	1,067.34	-0.56
CT	0.26	1,684.51	-0.60	NY	0.16	350.58	-0.59
DE	0.19	553.34	-0.62	NC	0.27	542.76	-0.51
DC*	0.61	7.81	-1.55	ND	0.35	148.33	-0.48
FL	0.11	65.81	-0.49	OH	0.30	321.95	-0.68
GA*	-0.03	5.16	-0.49	OK	0.17	48.56	-0.60
HI	0.08	2.57	-0.34	OR	0.27	160.17	-0.56
ID	0.40	63.60	-0.28	PA	0.30	645.35	-0.98
IL	0.26	585.79	-0.61	RI	0.33	852.93	-0.62
IN*	0.03	2.92	-0.57	SC	0.26	352.31	-0.38
IA	0.17	202.26	-0.78	SD	0.38	565.64	-0.56
KS	0.30	378.40	-0.57	TN	0.30	854.90	-0.67
KY	0.22	353.23	-0.64	TX	0.31	985.06	-0.40
LA	0.32	1,000.56	-0.46	UT	0.30	149.98	-0.42
ME	0.44	276.81	-0.74	VT	0.22	188.52	-0.33
MD	0.25	1,037.18	-0.54	VA	0.30	1,249.75	-0.59
MA	0.27	654.83	-0.66	WA	0.47	33.15	-0.52
MI*	0.01	0.64	-0.64	WV	0.33	626.07	-0.34
MN	0.43	457.88	-0.58	WI	0.35	572.51	-0.51
MS	0.36	1,207.74	-0.45	WY	0.23	67.84	-0.32
MO	0.24	311.03	-0.60				

Notes: All estimates are from PCIV estimation without covariates. "*" denotes states with first-stage F-statistics lower than 10.

D Additional Simulation Results

Figure 5: Kernel Density Plots of Estimation Errors, $\hat{\beta}_1 - \beta_1$



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

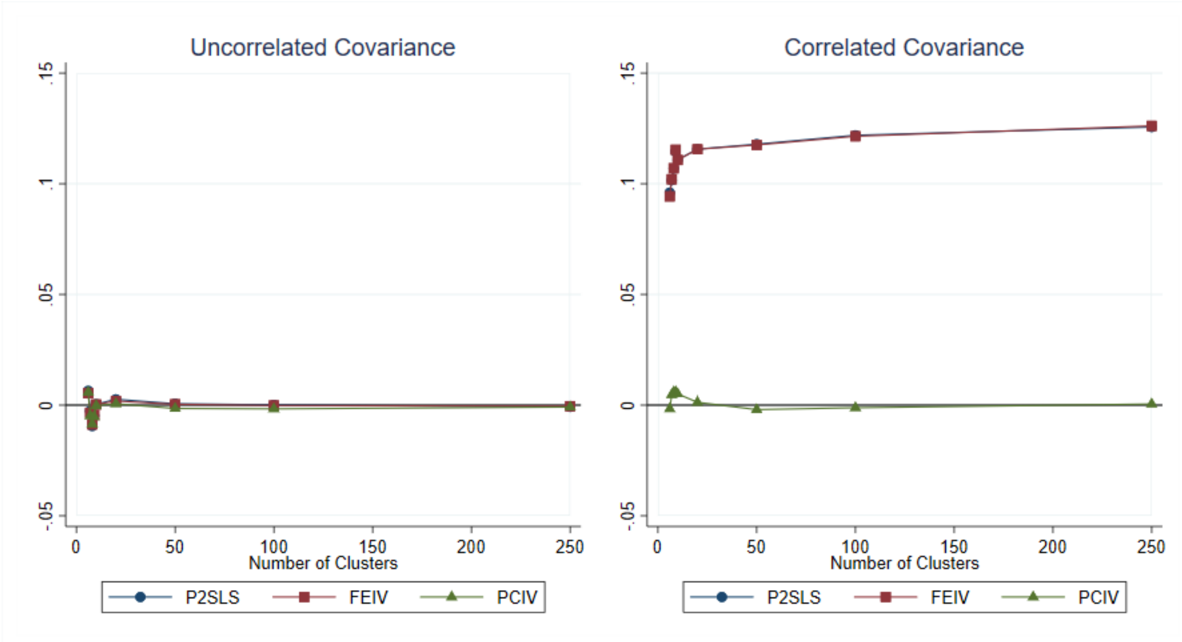
Figure 5 shows the kernel density plot of the bias over each simulation, maintaining the uncorrelated covariance assumption on the left and violating it on the right. Further, we show these density plots first with 250 clusters and 250 observations per cluster at the top, then with only the number of clusters reduced to 10, and finally with 250 clusters and 10 observations per cluster at the bottom.

Figure 5 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.

More substantial differences become apparent with a violation of the uncorrelated covariance assumption. With a 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 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.

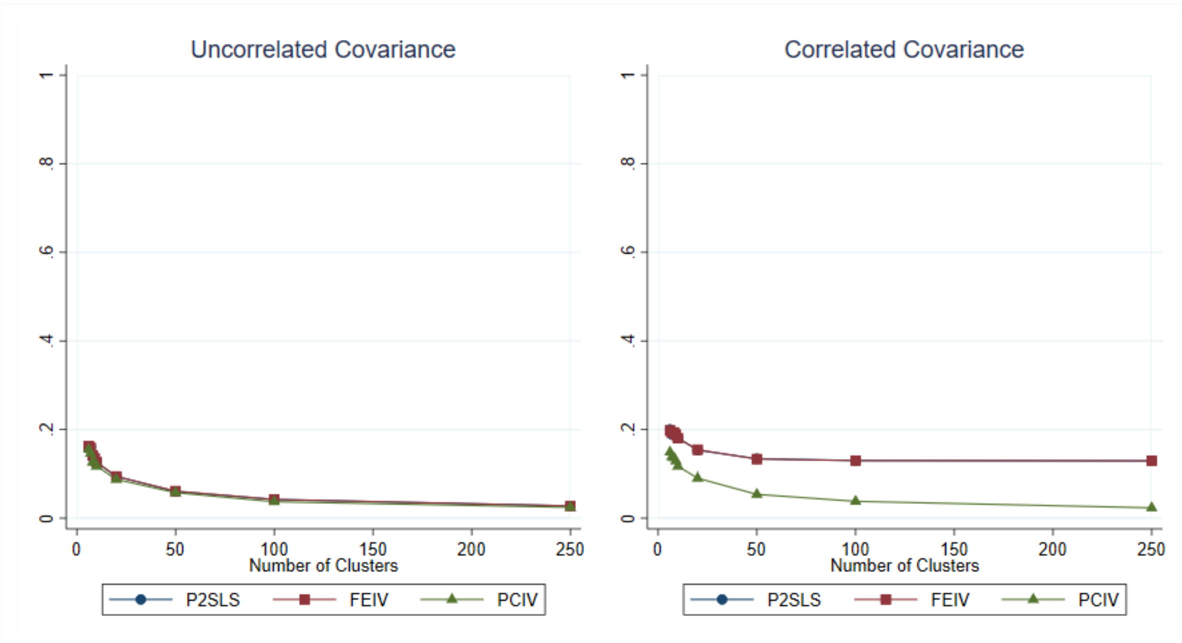
Figure 6 shows the average bias in the three estimators with T fixed at 250 as N increases, again both maintaining and violating the key assumption. With T=250 and the uncorrelated covariance assumption holding, all three estimators show very little mean bias across all numbers of clusters. However, 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 6: Estimated bias for coefficient β_1 of x_{ij} versus the number of clusters



Notes: 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: Estimated root mean square error for coefficient $\hat{\beta}_1$ of x_{ij} versus the number of clusters

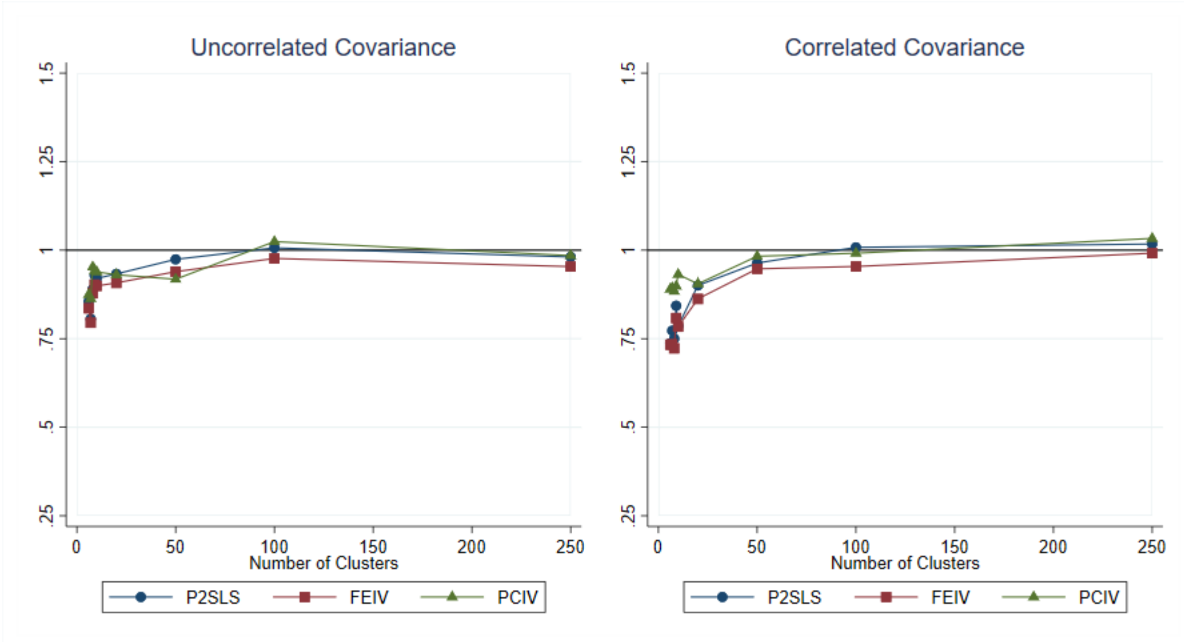


Notes: Left panels are when the uncorrelated covariance assumption holds, and right panels are when it is violated. P2SLS = Two-Stage Least Squares; FEIV=Fixed Effects Instrumental Variable; PCIV=Per-Cluster Instrumental Variable. Each point represents results from 500 repetitions.

Figure 7 shows that with a large T, all three estimators have similar RSMEs across the 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 a large number of observations per cluster, PCIV may dominate more standard approaches.

Figure 8 presents how this ratio changes with the number of clusters. In all cases, the analytic standard errors do reasonably well with large sample sizes along both dimensions. The ratio is consistently close to 1 with N above 20, both with the uncorrelated covariance assumption holding and when it is violated.

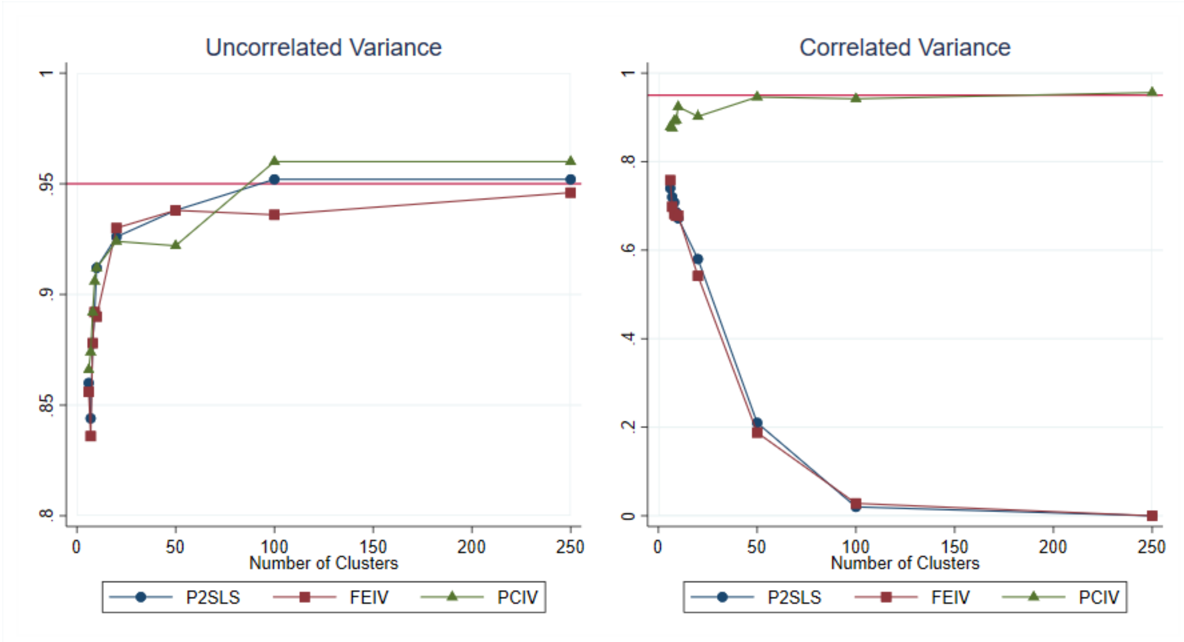
Figure 8: Ratio of mean SEs divided by SDs of estimates versus the number of clusters



Notes: The ratio of the mean standard errors (SEs) divided by standard deviations (SDs) of the estimates describes the performance of the analytic standard errors. The left panels are when the uncorrelated covariance assumption holds, and the right panels depict when it is violated. P2SLS = Two-Stage Least Squares; FEIV = Fixed Effects Instrumental Variable; PCIV=per-cluster instrumental variables. Each point represents results from 500 repetitions.

Figure 9 shows the evolution of coverage rates for each estimator both in cases where the strength of the instrument is uncorrelated with random coefficients and when we violate that assumption. Again, we show the evolution of coverage rates as the number of clusters grows from 7 to 250 with 250 observations per cluster. Figure 9 shows that all three estimators perform similarly when there is no correlation between instrument strength and the random coefficients. However, when we violate the uncorrelated covariance assumption, the right panel shows the bias in FEIV and P2SLS leads these estimators to largely reject the true parameter, and the rejection rate of the true parameter grows as the number of clusters grows. In contrast, the PCIV rejection rate at the 95 percent confidence level remains close to 95 percent throughout.

Figure 9: Coverage rate of the 95 percent confidence interval across the number of clusters



Notes: The horizontal line denotes the exact 95 percent coverage rate. Left panels are when the uncorrelated covariance assumption holds, and the right panels are when it is violated. P2SLS = Two-Stage Least Squares; FEIV = Fixed Effects Instrumental Variable; PCIV=per-cluster instrumental variables. Each point represents results from 500 repetitions.

E Anticipatory behavior

The difference in estimates between P2SLS applied to first differences and FEIV may imply a violation of the strict exogeneity assumption of the instrument. Indeed in revisiting the analysis of Davis and Kilian (2011), Coglianese et al. (2017) note evidence anticipatory behavior providing one such violation. This violation is particularly important when using first differences in taxes, prices, and quantities, as in Davis and Kilian (2011). Coglianese et al. (2017) accordingly use leads and lags of prices and taxes to address the anticipatory behavior of large consumers on their estimated price elasticity of demand for gasoline.

However, as long as the dependence between lag or lead values of the instrument is only weakly related to the error term, 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.[22] Given that we are flexibly detrending the data, strong dependence seems unlikely. We investigate this possible dependence by running reduced form regression of volume purchases on the log of taxes and a lead of log taxes. Significant coefficient estimates on the lead of log taxes would indicate strong dependence.

Table 10: Anticipatory behavior and dependence in reduced-form regressions

	Without volume weights			Volume weighted		
	P2SLS	FEIV	PCIV	P2SLS	FEIV	PCIV
Log tax	-0.150	-0.157	-0.228	-0.105	-0.154	-0.255
	(0.053)	(0.047)	(0.296)	(0.035)	(0.030)	(0.313)
Lead of log tax	0.070	-0.043	-0.012	0.037	-0.073	0.014
	(0.026)	(0.066)	(0.197)	(0.028)	(0.064)	(0.208)
First-stage F-statistic	18.39	44.11	58.62	40.54	31.88	59.04

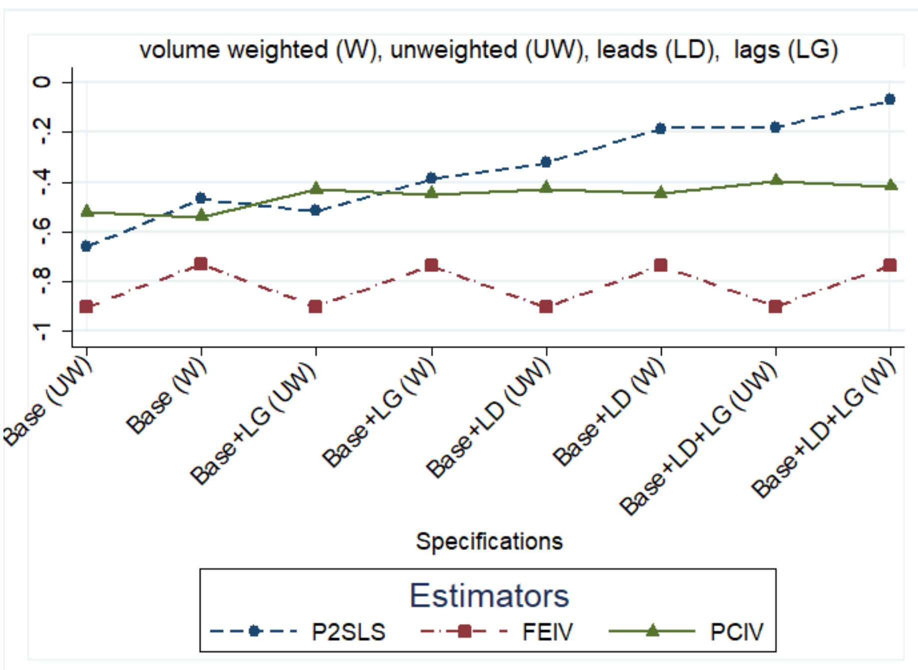
Notes: Headings refer to the estimator for which the reduced form estimates appear. Regressions use the entire sample. The calculation of the first-stage F-statistics for the PCIV was done using Hotelling's T-squared test. State-clustered standard errors appear in parentheses.

We present the results from these regressions in Table 10. Indeed, examining the coeffi-

[22]See Wooldridge (2010) Chapters 10 and 11 for more complete discussions of the issue.

cient estimates of lead taxes on the volume purchased, we see that the lead is only statistically significant with P2SLS applied to first differences. We find no evidence that additional gasoline purchases in anticipation of tax increases drive the FEIV (or PCIV) estimates, as all the coefficients are negative and quantitatively small. This apparent lack of strong dependence leads us to prefer the FEIV estimates over P2SLS applied to first-differences. Despite the small coefficient estimates on the lead of taxes, as a precaution, we follow Coglianese et al. (2017) in examining the sensitivity of elasticity estimates to the inclusion of leads and lags. We report the results from this exercise in Figure 10 in Appendix C, which shows that the magnitude of FEIV and PCIV estimates are robust to the inclusion of leads and lags of log prices and log taxes.

Figure 10: Sensitivity of elasticity estimates across specifications by estimator



Notes: Specification options include volume weighted (W), unweighted (UW), leads (LD), and lags (LG). All elasticity estimates are cumulative as they sum over contemporaneous lead and lag coefficient estimates where applicable.

F Replication of Davis and Kilian (2011) and Coglianese et al. (2017) with additional estimators

We apply all three methods to the data and modelling of Davis and Kilian (2011) and Coglianese et al. (2017) here. Davis and Kilian (2011) provides an early example of instrumenting for gasoline price changes using changes in state gasoline taxes to account for the simultaneity issues in studying prices and quantities. The data comes from the Energy Information Administration and stops in 2008 before the series on state gasoline prices was discontinued. Davis and Kilian (2011) use pooled OLS applied to first-differences in all variables to address unobserved, time-invariant, state-level intercepts.

Coglianese et al. (2017) note apparent anticipatory behavior in gasoline sales to state gasoline taxes changes, violating the strict exogeneity assumption of the instrument. Unlike FEIV and PCIV, this is particularly problematic for first-difference estimation as the inconsistency persists even as the length of the panel grows long (Wooldridge, 2010). As a result, Coglianese et al. (2017) address the issue by examining how the cumulative effect changes when they include a lead and a lag of log prices and instrument for them using a lead and a lag of log taxes. In additional robustness checks, they also include an additional lead and lag of log taxes for over-identified specifications.

We replicate the results from both studies in Table 11 exploring how the results change with using FEIV and PCIV instead of pooled OLS applied to first-differences and when we apply volume weights with each estimator. The furthest left column provides estimates that are directly comparable to Davis and Kilian (2011) in Panel A and to Coglianese et al. (2017) in Panels B and C. As neither Davis and Kilian (2011) and Coglianese et al. (2017) weight by volume of gasoline sold, we refrain from using volume weights for regressions on the left and apply volume weights on the right. Panel A of Table 11 presents the estimates of contemporaneous deviations in log gasoline prices on log gasoline sales, instrumented by the contemporaneous deviation in log gasoline taxes. Panel B of Table 11 accounts for anticipatory behavior following Coglianese et al. (2017) by including a lead and lag of log prices as additional endogenous regressors instrumenting the set with log taxes and well as the lead and lag of log taxes. The cumulative effect is the sum of all three coefficients. Panel C is similar to Panel B except that we follow Coglianese et al. (2017) in including two leads and two lags of taxes used to instrument for the endogenous prices.

The exercise shows a few key points. First, we replicate the results of Davis and Kilian (2011) and Coglianese et al. (2017) in the first column, which is unsurprising given that Lucas Davis graciously provides all data and code on his website. Second, consistent with theory, FEIV and PCIV are less sensitive to the inclusion of leads and lags of log prices than is pooled 2SLS applied to first-differences. Pooled 2SLS applied to first-differences is the most volatile estimator with both the largest and smallest point estimates depending on the specification. In fact, when we account for anticipatory behavior and weight by the states' relevance to the market, the cumulative effect of log prices on volume sold switches sign showing demand increasing with an increase in prices. Finally, the PCIV estimates continue to be markedly smaller (and some would say more plausible) than those from FEIV, insinuating that the correlation between first-stage variation and heterogeneous slopes is meaningful.

Table 11: Replication of Davis and Kilian (2011) and Coglianese et al. (2017)

	Without volume weights			Volume weighted		
	P2SLS	FEIV	PCIV	P2SLS	FEIV	PCIV
Panel A: Contemporaneous log prices, just identified						
	DK2011					
Cumulative Log price	-1.135	-0.777	-0.414	-0.714	-0.801	-0.405
	(0.250)	(0.394)	(0.279)	(0.209)	(0.327)	(0.463)
First-stage F-statistic	246.65	82.69	55.64	76.82	28.69	47.99
Panel B: Additional lead and lag of log prices, just identified						
	CDKS2017					
Cumulative Log price	-0.360	-0.691	-0.355	0.219	-0.779	-0.348
	(0.241)	(0.396)	(0.356)	(0.509)	(0.334)	(0.591)
First-stage F-statistic	31.21	51.21	54.89	54.84	44.61	48.08
Panel C: Additional lead and lag of log prices, over identified						
	CDKS2017					
Cumulative Log price	-0.368	-0.728	-0.367	0.349	-0.781	-0.359
	(0.239)	(0.402)	(0.233)	(0.529)	(0.335)	(0.388)
First-stage F-statistic	52.26	82.99	55.04	73.32	8.49	47.68

Notes: All regressions include month-by-year fixed effects and control for state unemployment rates. DK2011 corresponds to replications of Davis and Kilian (2011) and CDKS2017 corresponds to Coglianese et al. (2017). Cluster-robust standard errors are in parentheses.