Econometric Inference Using Hausman Instruments

Jinyong Hahn[*] UCLA

Zhipeng Liao[†] UCLA

Nan Liu[‡] Xiamen University

Ruoyao Shi[§] UCR

October 14, 2024

Abstract

We examine econometric inferential issues with Hausman instruments. The instrumental variable (IV) estimator based on Hausman instrument has a built-in correlation across observations, which may render the textbook-style standard error invalid. We develop a standard error that is robust to these problems. Clustered standard error is not always valid, but it can be a good pragmatic compromise to deal with the interlinkage problem if Hausman instrument is to be used in econometric models in the tradition of Berry, Levinsohn, and Pakes (1995). Additionally, we find that the Hausman IV is in fact equivalent to the judge IV proposed by Kling (2006), which broadens the implication of our results beyond the industrial organization literature.

JEL Classification: C14, C33, C36

Keywords: BLP, Hausman instrument, Judge instrument, Stable convergence, Uniformly valid inference

1 Introduction

Hausman instrument was first introduced by Hausman (Hausman, Leonard, and Zona, 1994; Hausman, 1996) as a way to address endogeneity of the (log of) price variable in linear demand equations. It was later adopted in the context of nonlinear specifications, following the tradition of Berry, Levinsohn, and Pakes (1995, BLP hereafter), for similar purposes (e.g., Nevo, 2001; Crawford and Yurukoglu, 2012). For a comprehensive discussion and documentation of the Hausman IV in the broader context of Industrial

[*]Department of Economics, UCLA, Los Angeles, CA 90095-1477, USA. Email: hahn@econ.ucla.edu.

[†]Department of Economics, UCLA, Los Angeles, CA 90095-1477, USA. Email: zhipeng.liao@econ.ucla.edu.

[‡]Wang Yanan Institute for Studies in Economics and School of Economics, Xiamen University, Xiamen, Fujian, 361005, China. Email: nanliu@xmu.edu.cn.

[§]Department of Economics, UC Riverside, 900 University Avenue, Riverside, CA, 92521, USA. Email: ruoyao.shi@ucr.edu.

Organization (IO) models, refer to Aguirregabiria (2019). The Hausman IV is also one of the most popular instruments in quantitative marketing and financial market analyses (see, e.g., Crawford and Yurukoglu, 2012; Rossi, 2014; Egan, Hortaçsu, and Matvos, 2017; Scanlon, 2019). Furthermore, as we will demonstrate in Section 5 of the paper, the Hausman IV can be argued to be identical to the commonly used judge IV,[1] which has been extensively applied in empirical research, including studies on pretrial supervision and incarceration (Di Tella and Schargrodsky, 2013; Aizer and Doyle, 2015), foster care (Gross and Baron, 2022), and disability benefits (Dahl, Kostøl, and Mogstad, 2014; Autor, Kostøl, Mogstad, and Setzler, 2019), among others.[2]

All the existing literature applying the Hausman IV focuses on empirical applications, with no theoretical discussion on the asymptotic framework and distribution necessary for econometric inference using the Hausman IV. This paper addresses this gap by introducing various asymptotic frameworks to examine inferential issues associated with the Hausman IV. We utilize a pseudo-panel structure to study the Hausman IV in linear models, where n denotes the number of firms in a given market (market size), and T represents the number of times the market is observed (number of markets).

Our first finding is that the Hausman IV estimator has a built-in correlation across contemporary observations, which concurs with the endogeneity that requires an IV in the first place. This contemporaneous correlation induces a cluster-like dependence in the IV estimator, even when the error terms in both the first-stage and second-stage equations are independent across observations. We demonstrate that the textbook-style standard error formula is valid only under asymptotics where both n and T grow to infinity. If either n or T is fixed, this formula becomes invalid. Additionally, we consider a rescaled version of the textbook-style standard error and a clustered standard error, showing that the former is valid under large n asymptotics, while the latter is valid under large T asymptotics. To overcome these limitations, we develop a uniformly valid standard error that ensure correct asymptotic inference as long as either n or T increases to infinity. This standard error is constructed as an average of the rescaled textbook-style standard error and the clustered standard error.

Our asymptotic analysis is different from the typical results in the econometrics literature. It is because the weak convergence concept, which is a standard tool for asymptotic analysis in econometrics, was not adequate for our asymptotic analysis. We adopted the stable convergence concept, which is rarely

[1] There are different versions of the judge IV depending on how the covariates are handled. However, the leave-one-out versions of the judge IV are all numerically identical to the Hausman IV in the pseudo-panel interpretation that we consider.

[2] For a summary of the applications of the judge IV in recent literature, see Table 1 in Frandsen, Lefgren, and Leslie (2023).

used in econometrics.[3] In this sense, this paper makes a technical contribution as well.

Our results are based on the specification that the underlying model is linear, which superficially rules out the BLP specification. Because our analysis of the failure of the textbook-style standard error is based on the problems with the "numerator" of the standard error in linear models, and because the same issue arises in the " numerator" counterpart of the BLP, it is straightforward to conclude that the textbook-style standard errors are invalid in the BLP specification when Hausman IV is adopted.

Regarding the practical implications for the BLP model with Hausman IV, we concede that this paper does not extend the analysis of the uniformly consistent standard error, which was developed and justified for the linear case. The BLP model involves numerous other components, making it challenging to isolate and focus on the anomalies specifically related to the Hausman IV. While the large T asymptotic results for linear models can be extended to the BLP model with Hausman IV in a straightforward manner—essentially requiring only an adjustment of the "numerator", we are uncertain if the same straightforward extension applies to the large n asymptotic results.[4] Since analyzing the uniformly consistent standard error requires characterizing the asymptotic distributions under conditions where either n or T grows to infinity, we are currently not in a position to establish a uniform inference method. Having said that, we speculate that, in practice, it may be reasonable to use the clustered standard error. Although the applied literature is not explicit on this point, it appears that large T asymptotics are implicitly adopted in many cases.[5]

The remainder of this paper is organized as follows. In Section 2, we introduce the Hausman IV estimator within a benchmark model and provide an intuitive overview of the main findings of this paper. Section 3 derives the asymptotic distribution of the IV estimator in the benchmark model and discusses the textbook-style standard error, clustered standard error, and provides the formula for a consistent standard error. Section 4 extends these results to cases where exogenous regressors are included in the structural equation. In Section 5, we demonstrate that the inference issues observed with the Hausman

[3] Phillips and Ouliaris (1990), Phillips and Sul (2003), Kuersteiner and Prucha (2013), Hahn, Kuersteiner, and Mazzocco (2020), Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008) are some small number of exceptions.

[4] Some of the subtle issues include the following. As in the linear case, the clustered standard error is invalid under the large n asymptotics as considered by Berry, Linton, and Pakes (2004). Having said that, the large n asymptotics is subject to the concern along the line of Armstrong (2016), based on economics (not statistics) consideration. It is not clear to us whether the same concern should be raised about the Hausman IV with large n asymptotics.

[5] For example, while Conlon and Gortmaker (2020) were not explicit about the asymptotics for standard error calculation, it is clear that they adopted the large T asymptotics implicitly due to their reference (p.1123) to Freyberger (2015) for bias correction and standard error adjustment since the latter considered the large T asymptotics.

IV estimator also arise with other IV approaches, such as the judge IV and Bartik IV. Incidentally, we find that the commonly used judge IV can be argued to be identical to the Hausman IV. Therefore, our findings in the current paper have an implication that goes beyond the IO literature. Section 6 concludes the paper. The Appendix contains proofs of the main theoretical results, and the Supplemental Appendix provides auxiliary lemmas used in these proofs.

The following notation will be adopted throughout the paper. We use K to denote a generic strictly positive constant that may vary from one instance to another but remains independent of the panel dimensions n and T. We adopt the convention that a summation over an empty set equals zero. We use $a \equiv b$ to indicate that a is defined as b. For real numbers $a_1, \ldots, a_m$, $(a_j)_{j \leq m} \equiv (a_1, \ldots, a_m)^\top$. For any matrix A, $A^\top$ denotes the transpose of A, and $\|A\|$ denotes the Euclidean norm of A. For any doubly indexed sequence $a_{i,t}$ (where $i = 1, \ldots, n$ and $t = 1, \ldots, T$), we define $\bar{a}_{i,\cdot} \equiv T^{-1} \sum_{t \leq T} a_{i,t}$, $\bar{a}_{\cdot,t} \equiv n^{-1} \sum_{i \leq n} a_{i,t}$ and $\bar{a} \equiv (nT)^{-1} \sum_{t \leq T} \sum_{i \leq n} a_{i,t}$. The summation $\sum_{i' \neq i}$ is taken over all i' except i, which means $\sum_{i' \neq i} a_{i',t} = \sum_{i'=1}^{i-1} a_{i',t} + \sum_{i'=i+1}^{n} a_{i',t}$.

2 Intuitive Overview of the Main Results

The class of models that the Hausman IV is applicable can be written in the linear simultaneous equations model of the form

$$y_{i,t} = \alpha_i + \beta x_{i,t} + u_{i,t}, \tag{1}$$

$$x_{i,t} = \eta_i + \gamma c_t + v_{i,t}, \tag{2}$$

for $i = 1, \ldots, n$ and $t = 1, \ldots, T$, where $y_{i,t}$ is some dependent variable (such as the quantity demanded), $x_{i,t}$ is some endogenous explanatory variable (such as the price), and the c_t can be understood to be the *latent* common shock (such as cost shocks), the residual terms $u_{i,t}$ and $v_{i,t}$ may be correlated, which leads to the endogeneity of $x_{i,t}$.[6] The i denotes a city and the t denotes the time, so the (i,t)-pair indexes the"market"as commonly understood in the IO literature. The model (1) - (2) can be viewed as a result of the more general model, such as (14) presented in Section 4 below, where other included exogenous variables are partialled out.

This linear model is also flexible enough to include the BLP as long as $y_{i,t}$ is understood to be some nonlinear transformation that may depend on some additional parameters. For instance, Nevo (2001)'s

[6]Here, we abstract from the possibility of multiple goods, so in terms of commonly adopted notations, we let $J = 1$. Our econometric analysis goes through even when $J > 1$ as long as it remains finite in the asymptotic framework.

"full model" is similar to our "extended model" in (14) below with $J > 1$, where $y_{i,t}$ is the mean utility from the good, $x_{i,t}$ is the price, $w_{i,t}$ is a vector of product characteristics, and β and θ are the means of individual coefficients β_i and θ_i ($i = 1, \ldots, n$), which are assumed to follow a joint distribution up to additional parameters θ_2 (Nevo, 2001, eq. (3)). Given θ_2, all the $y_{i,t}$, $x_{i,t}$ and $w_{i,t}$, the market share can be calculated from the model by numerical integration, denoted as $s_{i,t}(y, x, w^\top\theta_2)$, so the "dependent variable" $y_{i,t}$ can be obtained by solving the system of equations that matches the model-predicted market share with the observed one from the data (Nevo, 2001, eq. (7)). The additional parameters as well as the transformation complicates notations without shedding any further light on the basic econometric problems, so we abstract away from the BLP-style complication.

Hausman (1996)'s insight is that under some conditions, the $x_{i',t}$ with $i' \neq i$, i.e., the contemporary endogenous regressor from a different city, can serve as the instrument for $x_{i,t}$.[7] Hausman IV can be particularly useful because cost shocks, the preferred instruments for demand estimation, are often unavailable to researchers. When $n \geq 2$, a common practice is to use the average

$$z_{i,t} \equiv (n-1)^{-1} \sum_{i' \neq i} x_{i',t} \tag{3}$$

as the IV for $x_{i,t}$,[8] which means that the IV estimator is numerically equal to

$$\hat{\beta}_{iv} \equiv \frac{\sum_{t \leq T} \sum_{i \leq n} \left(z_{i,t} - \bar{z}_{i,\cdot}\right)\left(y_{i,t} - \bar{y}_{i,\cdot}\right)}{\sum_{t \leq T} \sum_{i \leq n} \left(z_{i,t} - \bar{z}_{i,\cdot}\right)\left(x_{i,t} - \bar{x}_{i,\cdot}\right)} = \frac{\sum_{t \leq T} \sum_{i \leq n} z_{i,t}\left(y_{i,t} - \bar{y}_{i,\cdot}\right)}{\sum_{t \leq T} \sum_{i \leq n} z_{i,t}\left(x_{i,t} - \bar{x}_{i,\cdot}\right)}, \tag{4}$$

where we use the usual partialling out trick to eliminate the fixed effects. Since

$$y_{i,t} - \bar{y}_{i,\cdot} = \left(x_{i,t} - \bar{x}_{i,\cdot}\right)\beta + u_{i,t} - \bar{u}_{i,\cdot} \tag{5}$$

based on the expression for $y_{i,t}$ in (1), applying (5) to (4) gives:

$$\hat{\beta}_{iv} = \frac{\sum_{t \leq T} \sum_{i \leq n} z_{i,t}\left(\left(x_{i,t} - \bar{x}_{i,\cdot}\right)\beta + u_{i,t} - \bar{u}_{i,\cdot}\right)}{\sum_{t \leq T} \sum_{i \leq n} z_{i,t}\left(x_{i,t} - \bar{x}_{i,\cdot}\right)} = \beta + \frac{\sum_{t \leq T} \sum_{i \leq n} z_{i,t}\left(u_{i,t} - \bar{u}_{i,\cdot}\right)}{\sum_{t \leq T} \sum_{i \leq n} z_{i,t}\left(x_{i,t} - \bar{x}_{i,\cdot}\right)}.$$

It is straightforward to show that

$$\sum_{t \leq T} \sum_{i \leq n} z_{i,t}\left(u_{i,t} - \bar{u}_{i,\cdot}\right) = \sum_{i \leq n} \sum_{t \leq T} \gamma u_{i,t}\left(c_t - \bar{c}\right) + \frac{1}{n-1} \underline{\sum_{t \leq T} \sum_{i \leq n} \sum_{i' \neq i} u_{i,t} v_{i',t}} - \frac{T}{n-1} \sum_{i \leq n} \sum_{i' \neq i} \bar{u}_{i,\cdot} \bar{v}_{i',\cdot}, \tag{6}$$

[7] In this paper, we will adopt all his identifying assumptions, and focus on the inferential issues. While Bresnahan and Gordon (2008) have raised questions on the identifying assumptions in Hausman (1996)'s paper, which Nevo (2000) and Aguirregabiria (2019) have also noted, this paper will focus on addressing the inferential issues rather than revisiting the questions of identification.

[8] In the judge IV literature, $z_{i,t}$ is referred to as a leave-out mean or unbiased jackknife instrumental variable, see, e.g., Aizer and Doyle (2015); Autor, Kostøl, Mogstad, and Setzler (2019); Norris, Pecenco, and Weaver (2021).

where $\bar{c} \equiv T^{-1} \sum_{t \leq T} c_t$. Note that the second term on the right (underlined) is a sum over t of the U-statistics $\sum_{i \leq n} \sum_{i' \neq i} u_{i,t} v_{i',t}$. Elementary statistics suggests that the variance of such U-statistic would depend on the covariance between u and v.[9] The potential correlation between the u and v in the model (1) and (2) is the source of endogeneity that would make the OLS inconsistent, and it is the reason why the instrument is sought. Our contribution is to recognize that the U-statistic structure built-in as part of the Hausman IV brings back the endogeneity (i.e., the covariance between u and v) as part of the asymptotic variance.

Having described the intuition, we now summarize the basic theoretical results in the next section. In Theorem 1, we provide the asymptotic distribution of the IV estimator $\hat{\beta}_{iv}$. We consider the asymptotic framework where n and/or T can go to infinity, although we insist that at least one of them should go to infinity by requiring that $nT \to \infty$. It turns out that the asymptotic distribution of $\hat{\beta}_{iv}$ depends on the behavior of n and T in the limit. Theorem 1 is a general result that nests all possible limiting behaviors of n and T. The asymptotic variance may be "random" depending on the limiting behaviors, and in order to accommodate such situations, Theorem 1 presents the asymptotic distribution using the stable convergence concept.

In Lemma 1, we consider the textbook-style standard error derived under the homoscedasticity and independence assumption. The lemma establishes that this standard error is consistent only when both n and T go to infinity; if n is fixed while $T \to \infty$, it ignores the covariance between u and v, and therefore is inconsistent; if T is fixed while $n \to \infty$, it is again inconsistent because it ignores a multiplicative factor that depends on the magnitude of T.

Recall that the concern about the covariance in the U-statistic arose in the decomposition (6) of the "numerator" of $\hat{\beta}_{iv}$. The assumption that $(u_{i,t}, v_{i,t})$ are i.i.d. across i and t implies that the U-statistic $\sum_{i \leq n} \sum_{i' \neq i} u_{i,t} v_{i',t}$ should be independent over t. This suggests that a standard error clustered at the time level might be consistent. However, Lemma 2 demonstrates that such a clustered standard error is consistent only when $T \to \infty$; when T is fixed and $n \to \infty$, it is inconsistent for a reason elaborated in Section 3.2 below.

To address these issues, we develop a new standard error. In Theorem 2, it is shown that the new standard error is consistent in general, and despite the relatively unusual stable convergence framework, it enables asymptotically valid statistical inference, similar to what would be achieved under the usual

[9]For intuition, consider the simple case with $n = 2$, where the U-statistic takes the form $u_{1,t} v_{2,t} + u_{2,t} v_{1,t}$. Obviously the variance of this quantity is equal to $\mathrm{Var}(u_{1,t}) \mathrm{Var}(v_{2,t}) + 2 \mathrm{Cov}(u_{1,t}, v_{1,t}) \mathrm{Cov}(u_{2,t}, v_{2,t}) + \mathrm{Var}(u_{2,t}) \mathrm{Var}(v_{1,t})$, assuming that the u's and the v's have zero means, as well as that $u_{i,t}$ and $v_{i',t'}$ are independent for $i \neq i'$ or $t \neq t'$.

weak convergence.

It is important to emphasize that Theorems 1 and 2, as well as Lemmas 1 and 2, are derived under the assumption that $(u_{i,t}, v_{i,t})$ are i.i.d. across i and t. Therefore, the inconsistency of both the textbook-style standard error and the clustered standard error is not attributable to any cluster structure among the pairs $(u_{i,t}, v_{i,t})$ over i or t.

3 Main Results in the Benchmark Model

In this section, we study the asymptotic properties of the IV estimator $\hat{\beta}_{iv}$ based on the model presented in (1) - (2). We refer to this as the benchmark model because the structural equation (1) does not include any exogenous regressors in equation (1). The IV estimator in an extended model, which includes additional regressors in (1), will be investigated in the next section. Throughout this paper, we consider an asymptotic framework where both n and T are indexed by $m = 1, 2, \ldots$ and both are non-decreasing in m, with $n_m T_m \to \infty$ as $m \to \infty$. For simplicity, the dependence of n_m and T_m on m is suppressed, provided there is no risk of confusion.

Assumption 1 *(i) $(u_{i,t}, v_{i,t})$ are i.i.d. across i and t with $\mathbb{E}[u_{i,t}] = 0$ and $\mathbb{E}[v_{i,t}] = 0$; (ii) c_{t_1} is independent of (u_{i,t_2}, v_{i,t_2}) for any i, and any t_1 and t_2; (iii) $\mathbb{E}[u_{i,t}^4] + \mathbb{E}[v_{i,t}^4] \leq K$ and $\max_t \mathbb{E}[c_t^4] \leq K$, (iv) $\hat{\sigma}_c^2 \equiv T^{-1} \sum_{t=1}^{T} (c_t - \bar{c})^2 \to_p \sigma_c^2$ where $\sigma_c^2 > 0$ almost surely; (v) $n \geq 2$, $T \geq 2$ and $(nT)^{-1} = o(1)$.*

Assumption 1 includes some regularity conditions used for studying the IV estimators. Conditions (i, ii) impose a dependence structure on the unobserved components, i.e., $u_{i,t}$, $v_{i,t}$, and c_t, allowing for correlation between $u_{i,t}$ and $v_{i,t}$. For the common factor c_t, we only require an upper bound on its fourth moment and a lower bound on its "sample variance". The factor c_t may exhibit time-varying distributions, making it non-stationary, and have a general dependence structure over time. Condition (v) requires that both n and T are strictly greater than 1, and nT diverges with the sample size. The restriction $(nT)^{-1} = o(1)$ allows for cases such as: (i) large n and small T; (ii) small n and large T; and (iii) large n and large T.

The independence assumption between the $\{c_t\}$ and $\{(u_{i,t}, v_{i,t})\}$ amounts to a homoscedasticity assumption as well as no apparent cluster structure in the error vector $\{(u_{i,t}, v_{i,t})\}$. We acknowledge that in a typical empirical question where the Hausman IV is applicable, $(u_{i,t}, v_{i,t})$ often has a cluster structure over i or t, or both, but an important and interesting feature of our result is that the asymptotic distribution of the Hausman IV estimator exhibits a clustering problem, introduced by the Hausman IV, even when there is no apparent cluster structure in the original model.

Theorem 1 *Let $\mathcal{F}_0$ denote the sigma-field generated by $\{c_t\}_{t=1}^{T_\infty}$. Under Assumption 1, we have*

$$(nT)^{1/2}(\hat{\beta}_{iv} - \beta) = \frac{(nT)^{-1/2} \sum_{t=1}^{T} \sum_{i=1}^{n} (\gamma u_{i,t}(c_t - \bar{c}) + \varepsilon_{i,t}) + O_p((nT)^{-1/2})}{\gamma^2 \hat{\sigma}_c^2 + O_p((nT)^{-1/2})}, \tag{7}$$

where $\varepsilon_{i,t} \equiv (n-1)^{-1} \sum_{i'=1}^{i-1} (u_{i,t} v_{i',t} + u_{i',t} v_{i,t})$. Moreover, as $m \to \infty$,

$$(nT)^{1/2}(\hat{\beta}_{iv} - \beta) \to \omega_\infty Z \qquad (\mathcal{F}_0\text{-stably}), \tag{8}$$

where $\omega_\infty^2 \equiv (\gamma^2 \sigma_u^2 \sigma_c^2 + (n_\infty - 1)^{-1}(\sigma_u^2 \sigma_v^2 + \sigma_{u,v}^2))/(\gamma^4 \sigma_c^4)$ is independent of $Z \sim N(0,1)$, σ_u^2 and σ_v^2 denote the variances of $u_{i,t}$ and $v_{i,t}$, respectively, and $\sigma_{u,v}$ denotes the covariance between them.[10]

Theorem 1 derives the asymptotic distribution of the IV estimator $\hat{\beta}_{iv}$. The stable limit in (8) is required to address the case of large n and small T, where $\hat{\sigma}_c^2$ does not converge to a non-random constant; instead, its probability limit remains random in such a scenario. In the small n and large T case, it is evident that the covariance $\sigma_{u,v}$, in addition to the variances σ_u^2 and σ_v^2, appears in the "asymptotic variance" ω_∞^2. This arises from the cluster dependence of the Hausman IV $z_{i,t} - \bar{z}_{i,\cdot} = (n-1)^{-1} \sum_{i' \neq i} \left(x_{i',t} - \bar{x}_{i',\cdot}\right)$, which acts through $\varepsilon_{i,t}$, contributing to the U-statistic term in (6). Since $\sigma_{u,v}$ is the source of endogeneity, its appearance in ω_∞^2 highlights the critical importance of accounting for inherent cluster dependence when calculating the standard error for inference on the unknown parameter β.

3.1 Textbook-style Standard Error

In this subsection, we examine the textbook-style standard error for $\hat{\beta}_{iv}$. It turns out that such a standard error is consistent only when both n and T go to infinity, although a minor modification is consistent as long as $n \to \infty$.

Specifically, the textbook-style standard error formula for IV estimators with conditionally homoskedastic residuals is given by:

$$\widehat{\mathrm{SE}}_0(\hat{\beta}_{iv}) \equiv \sqrt{\frac{\left(\sum_{t \le T} \sum_{i \le n} (z_{i,t} - \bar{z}_{i,\cdot})^2\right)\left(\sum_{t \le T} \sum_{i \le n} \hat{u}_{i,t}^2\right)}{nT\left(\sum_{t \le T} \sum_{i \le n} x_{i,t}(z_{i,t} - \bar{z}_{i,\cdot})\right)^2}},$$

where $\hat{u}_{i,t} \equiv y_{i,t} - \bar{y}_{i,\cdot} - \hat{\beta}_{iv}(x_{i,t} - \bar{x}_{i,\cdot})$. The following lemma presents the asymptotic properties of $\widehat{\mathrm{SE}}_0(\hat{\beta}_{iv})$.

[10] The definitions of $\mathcal{G}$-stable convergence and $\mathcal{G}$-mixing convergence can be found in Section A of the Appendix.

Lemma 1 *Under Assumption 1, we have:*

$$\sqrt{nT}\widehat{\mathrm{SE}}_0(\hat{\beta}_{iv}) \to_p \sqrt{\frac{\gamma^2\sigma_u^2\sigma_c^2 + (n_\infty - 1)^{-1}\sigma_u^2\sigma_v^2}{\gamma^4\sigma_c^4}(1 - T_\infty^{-1})}.$$

Lemma 1 shows that the textbook-style standard error is consistent only when both n and T go to infinity. In this scenario the asymptotic variance of the IV estimator in Theorem 1 simplifies to $\sigma_u^2/(\gamma^2\sigma_c^2)$, which is the same as the probability limit of $\widehat{\mathrm{SE}}_0(\hat{\beta}_{iv})$ after scaling.

In the large n and small T case, the textbook-style standard error is inconsistent due to the $1 - T_\infty^{-1}$ factor. However, we can apply a degrees of freedom adjustment to $\widehat{\mathrm{SE}}_0(\hat{\beta}_{iv})$, and show that the adjusted standard error, i.e., $\widehat{\mathrm{SE}}_0(\hat{\beta}_{iv})(1 - T^{-1})^{-1/2}$ is consistent in the large n scenarios.

In the small n and large T case, the covariance term $\sigma_{u,v}$ in the "asymptotic variance" ω_∞^2 is not captured in the probability limit of $\widehat{\mathrm{SE}}_0(\hat{\beta}_{iv})$. This indicates that the textbook-style standard error and its adjusted version are inconsistent as long as $\sigma_{u,v} \neq 0$. Since the endogeneity of $x_{i,t}$ arises from $\sigma_{u,v} \neq 0$, the inconsistency of the textbook-style standard error is a by-product of the necessity of using IV estimation.

3.2 Clustered Standard Error

It is natural to conjecture that the cluster structure induced by the Hausman IV ($\sigma_{u,v}^2$ in Theorem 1) might be intuitively handled by a clustered standard error, which has an additional bonus of providing a protection against potential heteroscedasticity. We now investigate the asymptotic properties of the conventional clustered standard error in this context. We show that the clustered standard error is consistent when $T \to \infty$ but not when T is bounded from above.

Under Assumptions 1(i, ii), it is clear that: $\{u_{i,t}(c_t - \bar{c})\}_{i\leq n, t\leq T}$ are uncorrelated across i and across t; and $u_{i,t}(c_t - \bar{c})$ and $\varepsilon_{i',t'}$ are uncorrelated for any $i, i' \leq n$ and any $t, t' \leq T$. Therefore, the cluster dependence in the estimation error of $\hat{\beta}_{iv}$ is introduced through $\varepsilon_{i,t}$. Indeed, for any $i_1, i_2 \leq n$, $i_1' \leq i_1 - 1$, $i_2' \leq i_2' - 1$ and any $t, t' \leq T$ with $t \neq t'$, we have

$$\mathbb{E}[u_{i_1,t}v_{i_1',t}u_{i_2,t'}v_{i_2',t'}] = \mathbb{E}[u_{i_1,t}v_{i_1',t}]\mathbb{E}[u_{i_2,t'}v_{i_2',t'}] = 0,$$

which implies that

$$\mathbb{E}\left[\varepsilon_{i,t}\varepsilon_{i',t'}\right] = 0$$

as long as $t \neq t'$. Therefore, there is no clustering across t. On the other hand, we notice that

$$\sum_{t\leq T}\sum_{i\leq n}\varepsilon_{i,t} = (n-1)^{-1}\sum_{t\leq T}\sum_{i\leq n}\left(u_{i,t}\sum_{i'\neq i}v_{i',t}\right).$$

For any $t \leq T$ and any $i_1, i_2 \leq n$ with $i_1 \neq i_2$,

$$\mathbb{E}\left[u_{i_1,t}u_{i_2,t}\sum_{i_1'\neq i_1} v_{i_1',t}\sum_{i_2'\neq i_2} v_{i_2',t}\right] = \mathbb{E}\left[u_{i_1,t}u_{i_2,t}v_{i_2,t}v_{i_1,t}\right] = \mathbb{E}\left[u_{i_1,t}v_{i_1,t}\right]\mathbb{E}\left[u_{i_2,t}v_{i_2,t}\right] = \sigma_{u,v}^2,$$

which shows that $\varepsilon_{i,t}$ has an equi-correlation across i, and it arises precisely due to the way the IV is constructed. This motivates the following clustered standard error

$$\widehat{\mathrm{SE}}_1(\hat\beta_{iv}) \equiv \sqrt{\frac{\sum_{t\leq T}\left(\sum_{i\leq n}\hat u_{i,t}\left(z_{i,t}-\bar z_{i,\cdot}\right)\right)^2}{\left(\sum_{t\leq T}\sum_{i\leq n}x_{i,t}\left(z_{i,t}-\bar z_{i,\cdot}\right)\right)^2}}.$$

We next present the asymptotic properties of the clustered standard error.

Lemma 2 *Under Assumption 1, we have*

$$\widehat{\mathrm{SE}}_1(\hat\beta_{iv})^2 = \frac{(nT)^{-1}\sum_{t\leq T}\left(\sum_{i\leq n}\gamma(c_t-\bar c)u_{i,t}-\xi_t\right)^2 + (n-1)^{-1}(\sigma_u^2\sigma_v^2+\sigma_{u,v}^2)}{(nT)\gamma^4\hat\sigma_c^4} + O_p((nT)^{-3/2}), \tag{9}$$

where $\xi_t \equiv n\gamma(c_t-\bar c)(\bar u + \gamma(c_t-\bar c)(\hat\beta_{iv}-\beta))$. *Moreover, if* $T\to\infty$ *as* $m\to\infty$, *then*

$$(nT)\widehat{\mathrm{SE}}_1(\hat\beta_{iv})^2 \to_p \frac{\gamma^2\sigma_u^2\sigma_c^2 + (n_\infty-1)^{-1}(\sigma_u^2\sigma_v^2+\sigma_{u,v}^2)}{\gamma^4\sigma_c^4}. \tag{10}$$

Lemma 2 provides the asymptotic approximation of the clustered standard error. The component denoted as ξ_t in (9) arises from estimating the unknown parameters α_i and β in the structural equation (1). When T approaches infinity, Lemma 2 shows that $(nT)\widehat{\mathrm{SE}}_1(\hat\beta_{iv})^2$ is a consistent estimator of the asymptotic variance of $\hat\beta_{iv}$. On the other hand, if T is bounded from above, the asymptotic approximation in (9) indicates that $\widehat{\mathrm{SE}}_1(\hat\beta_{iv})$ is an inconsistent estimator.

To illustrate this inconsistency, consider the simplified case where $u_{i,t}$ is known, and as a result ξ_t does not present in (9). In this case, the first term in the numerator of the faction on the right hand side of (9), i.e.,

$$(nT)^{-1}\sum_{t\leq T}\left(\sum_{i\leq n}\gamma(c_t-\bar c)u_{i,t}\right)^2 = \gamma^2T^{-1}\sum_{t\leq T}\left(n^{-1/2}\sum_{i\leq n}(c_t-\bar c)u_{i,t}\right)^2,$$

fails to approach to $\gamma^2\sigma_u^2\sigma_c^2$ in large sample, which causes the inconsistency of $\widehat{\mathrm{SE}}_1(\hat\beta_{iv})$.

The intuition underlying this failure is that the stable convergence of $(n^{-1/2}\sum_{i\leq n}(c_t-\bar c)u_{i,t})_{t\leq T}$, combined with the Cramér-Wold device and continuous mapping theorem, would imply that as $m\to\infty$,

$$\gamma^2T^{-1}\sum_{t\leq T}\left(n^{-1/2}\sum_{i\leq n}(c_t-\bar c)u_{i,t}\right)^2 \to \gamma^2\sigma_u^2\sigma_c^2\left(T^{-1}\sum_{t\leq T}Z_t^2\right) \qquad (\mathcal{F}_0\text{-stably}),$$

where $(Z_t)_{t\leq T}$ is a vector of mutually independent standard normal random variables independent of σ_c^2. In other words, the term $(nT)^{-1}\sum_{t\leq T}\left(\sum_{i\leq n}\gamma(c_t-\bar{c})u_{i,t}\right)^2$ converges to a scaled χ^2 random variable when T is small, not the desired non-stochastic component $\gamma^2\sigma_u^2\sigma_c^2$.

3.3 Averaging Textbook-style Standard Error and Clustered Standard Error

We now present a simple consistent standard error combining $\widehat{\mathrm{SE}}_0(\hat{\beta}_{iv})$ and $\widehat{\mathrm{SE}}_1(\hat{\beta}_{iv})$. From Lemma 1, it is clear that $(1-T^{-1})^{-1/2}\widehat{\mathrm{SE}}_0(\hat{\beta}_{iv})$ is a consistent standard error as long as $n\to\infty$. On the other hand, Lemma 2 shows that $\widehat{\mathrm{SE}}_1(\hat{\beta}_{iv})$ is a consistent standard error whenever $T\to\infty$. This motivates the following averaging standard error:

$$\widehat{\mathrm{SE}}_{avg}(\hat{\beta}_{iv})\equiv\frac{n}{T+n}(1-T^{-1})^{-1/2}\widehat{\mathrm{SE}}_0(\hat{\beta}_{iv})+\frac{T}{T+n}\widehat{\mathrm{SE}}_1(\hat{\beta}_{iv}). \tag{11}$$

We show that the averaging standard error is consistent under the general asymptotic framework with $nT\to\infty$ as $m\to\infty$.

Specifically, if both n and T go to infinity, then $(nT)^{1/2}(1-T^{-1})^{-1/2}\widehat{\mathrm{SE}}_0(\hat{\beta}_{iv})$ and $(nT)^{1/2}\widehat{\mathrm{SE}}_1(\hat{\beta}_{iv})$ converge to the same limit ω_∞, and so does $(nT)^{1/2}\widehat{\mathrm{SE}}_{avg}(\hat{\beta}_{iv})$. When n is bounded from above, $\widehat{\mathrm{SE}}_{avg}(\hat{\beta}_{iv})$ is dominated by the second term in (11), which as we have shown in (9) of Lemma 2, is a consistent estimator of ω_∞ after rescaled by $(nT)^{1/2}$. Finally, if T is bounded from above, $\widehat{\mathrm{SE}}_{avg}(\hat{\beta}_{iv})$ is dominated by the first term in (11), which is a consistent estimator of ω_∞ after rescaled by $(nT)^{1/2}$, as indicated by Lemma 1.

As a consequence, we arrive at the following theorem, showing that statistical inference based on the averaging standard error is valid when either n or T approaches infinity.

Theorem 2 *Under Assumption 1, we have* $(nT)^{1/2}\widehat{\mathrm{SE}}_{avg}(\hat{\beta}_{iv})\to_p\omega_\infty$, *and*

$$\frac{\hat{\beta}_{iv}-\beta}{\widehat{\mathrm{SE}}_{avg}(\hat{\beta}_{iv})}\to N(0,1)\qquad(\mathcal{F}_0\text{-mixing}) \tag{12}$$

as $m\to\infty$.

Theorem 2 shows that asymptotically valid inference on β can be conducted using the stable limit stated in (12). For instance, the usual $(1-\alpha)$-confidence interval given by

$$CI_{1-\alpha}=\left[\hat{\beta}_{iv}-z_{\alpha/2}\widehat{\mathrm{SE}}_{avg}(\hat{\beta}_{iv}),\quad\hat{\beta}_{iv}+z_{\alpha/2}\widehat{\mathrm{SE}}_{avg}(\hat{\beta}_{iv})\right] \tag{13}$$

covers β with probability approaching $1-\alpha$ for any $\alpha\in(0,1)$, where $z_{\alpha/2}$ denotes the $(1-\alpha/2)$-quantile of the standard normal distribution.

4 Extended Model with Exogenous Regressors

The extended model incorporates several exogenous regressors, denoted as $w_{i,t}$ into the structural equation. Consequently, equation (1) becomes

$$y_{i,t} = \alpha_i + x_{i,t}\beta + w_{i,t}^\top\theta + u_{i,t}. \tag{14}$$

The additional d_w-dimensional regressors $w_{i,t}$ are allowed to be correlated with the common shock c_t and may exhibit both spatial and time series dependence. Ignoring these regressors could lead to omitted variable bias and/or incorrect standard error for the IV estimator and the inference procedures discussed in the previous section.[11]

To define the IV estimator in the extended model, we introduce the following notation. Let $\hat{\lambda} \equiv \hat{\Sigma}_w^{-1}\hat{\Gamma}_{w,x}$ and define $\hat{x}_{i,t}$ as $\hat{x}_{i,t} \equiv x_{i,t} - \bar{x}_{i,\cdot} - (w_{i,t} - \bar{w}_{i,\cdot})^\top\hat{\lambda}$, where

$$\hat{\Sigma}_w \equiv (nT)^{-1}\sum_{t\leq T}\sum_{i\leq n}(w_{i,t} - \bar{w}_{i,\cdot})w_{i,t}^\top \text{ and } \hat{\Gamma}_{w,x} \equiv (nT)^{-1}\sum_{t\leq T}\sum_{i\leq n}(w_{i,t} - \bar{w}_{i,\cdot})x_{i,t}.$$

Similarly, define $\hat{y}_{i,t}$ as $\hat{y}_{i,t} \equiv y_{i,t} - \bar{y}_{i,\cdot} - (w_{i,t} - \bar{w}_{i,\cdot})^\top\hat{\pi}$, where $\hat{\pi} \equiv \hat{\Sigma}_w^{-1}\hat{\Gamma}_{w,y}$ and $\hat{\Gamma}_{w,y}$ is defined analogously to $\hat{\Gamma}_{w,x}$ with $x_{i,t}$ replaced by $y_{i,t}$. The IV estimator is then given by:

$$\hat{\beta}_{e,iv} \equiv \frac{\sum_{t\leq T}\sum_{i\leq n} z_{i,t}\hat{y}_{i,t}}{\sum_{t\leq T}\sum_{i\leq n} z_{i,t}\hat{x}_{i,t}}. \tag{15}$$

To study the properties of the IV estimator with additional regressors $w_{i,t}$, the following assumption is employed.

Assumption 2 *(i) w_{i_1,t_1} is independent of $(u_{i_2,t_2}, v_{i_2,t_2})$ for any i_1 and i_2, and any t_1 and t_2; (ii) there exist a matrix Σ_w such that $\hat{\Sigma}_w \to_p \Sigma_w$, where the eigenvalues of Σ_w are bounded away from zero almost surely; (iii) there exist a matrix $\Gamma_{w,c}$ such that $\hat{\Gamma}_{w,c} \to_p \Gamma_{w,c}$, where $\hat{\Gamma}_{w,c}$ is defined analogously to $\hat{\Gamma}_{w,x}$ with $x_{i,t}$ replaced by c_t; (iv) $\max_i \max_t \mathbb{E}[||w_{i,t}||^4] \leq K$; (v) $\sigma_{e,c}^2 \equiv \sigma_c^2 - \Gamma_{w,c}^\top\Sigma_w^{-1}\Gamma_{w,c} > 0$ almost surely.*

Condition (i) in Assumption 2 ensures that the regressors $w_{i,t}$ are strictly exogenous, while condition (iv) imposes a uniform upper bound on their fourth moment. Conditions (ii), (iii), and (iv) serve as regularity conditions to ensure that the IV estimator $\hat{\beta}_{e,iv}$ achieves a convergence rate of $(nT)^{-1/2}$. Assumption 2 permits both time series and spatial dependence in $w_{i,t}$, and allows for correlation between $w_{i,t}$ and the common shock c_t. In this scenario, the probability limit of $\hat{\Gamma}_{w,c}$, i.e., $\Gamma_{w,c}$ is a non-zero matrix.

[11] This is particularly important for the Hausman IV because its identification condition is likely to fail unless the advertising and promotional expenditure variables are included in the main regression (Rossi, 2014, p.666, footnote 8).

Theorem 3 *Let $\mathcal{F}_{e,0}$ denote the sigma-field generated by $\{\{c_t\}_{t\leq T_\infty}, \{w_{i,t}\}_{i\leq n_\infty, t\leq T_\infty}\}$. Under Assumptions 1 and 2, we have as $m\to\infty$*

$$(nT)^{1/2}(\hat{\beta}_{e,iv}-\beta)\to \omega_{e,\infty}Z \qquad (\mathcal{F}_{e,0}\text{-stably}), \tag{16}$$

where $\omega^2_{e,\infty}\equiv(\gamma^2\sigma_u^2\sigma_{e,c}^2+(n_\infty-1)^{-1}(\sigma_u^2\sigma_v^2+\sigma_{u,v}^2))/(\gamma^4\sigma_{e,c}^4)$ is independent of $Z\sim N(0,1)$.

We next present the formula for the average standard error, which is based on both the textbook-style standard error and the clustered standard error, defined analogously to their counterparts in the benchmark model. Similar to the benchmark model, neither the textbook-style standard error nor the clustered standard error is consistent within the general asymptotic framework of $nT\to\infty$ employed in this paper. However, these standard errors can be combined to construct a consistent averaging standard error.

Denote the Hausman IV with α_i and $w_{i,t}$ partialled out as:

$$\hat{z}_{i,t}\equiv(n-1)^{-1}\sum_{i'\neq i}\left(x_{i',t}-\bar{x}_{i',\cdot}\right)-(w_{i,t}-\bar{w}_{i,\cdot})^\top\hat{\varphi}, \tag{17}$$

where $\hat{\varphi}\equiv\hat{\Sigma}_w^{-1}\hat{\Gamma}_{w,z}$ and $\hat{\Gamma}_{w,z}\equiv(nT)^{-1}\sum_{t\leq T}\sum_{i\leq n}(w_{i,t}-\bar{w}_{i,\cdot})z_{i,t}$. The textbook-style standard error is defined as

$$\widehat{\mathrm{SE}}_{e,0}(\hat{\beta}_{e,iv})\equiv\sqrt{\frac{\left(\sum_{t\leq T}\sum_{i\leq n}\hat{z}_{i,t}^2\right)\left(\sum_{t\leq T}\sum_{i\leq n}\hat{u}_{e,i,t}^2\right)}{nT\left(\sum_{t\leq T}\sum_{i\leq n}z_{i,t}\hat{x}_{i,t}\right)^2}},$$

where $\hat{u}_{e,i,t}\equiv\hat{y}_{i,t}-\hat{x}_{i,t}\hat{\beta}_{e,iv}$. Under Assumptions 1 and 2, we can show that

$$(nT)\widehat{\mathrm{SE}}_{e,0}(\hat{\beta}_{e,iv})^2\to_p\frac{\gamma^2\sigma_u^2\sigma_{e,c}^2+(n_\infty-1)^{-1}\sigma_u^2\sigma_v^2}{\gamma^4\sigma_{e,c}^4}(1-T_\infty^{-1}), \tag{18}$$

which indicates that the textbook-style standard error does not account for the covariance term $\sigma_{u,v}$, and therefore is inconsistent when n is bounded from above.[12]

Lemma 3 *Under Assumptions 1 and 2, we have*

$$\sqrt{nT}\widehat{\mathrm{SE}}_{e,0}(\hat{\beta}_{iv})\to_p\sqrt{\frac{\gamma^2\sigma_u^2\sigma_c^2+(n_\infty-1)^{-1}\sigma_u^2\sigma_v^2}{\gamma^4\sigma_c^4}(1-T_\infty^{-1})}.$$

The clustered standard error in the extended model is defined as

$$\widehat{\mathrm{SE}}_{e,1}(\hat{\beta}_{e,iv})\equiv\sqrt{\frac{\sum_{t\leq T}\left(\sum_{i\leq n}\hat{z}_{i,t}\hat{u}_{e,i,t}\right)^2}{\left(\sum_{t\leq T}\sum_{i\leq n}z_{i,t}\hat{x}_{i,t}\right)^2}}.$$

[12]See the proof of Lemma 3 in the Appendix for the derivation of (18).

Similar to its counterpart in the benchmark model, the clustered standard error is consistent only when T goes to infinity. Therefore, we can combine the textbook-style standard error and the clustering "robust" standard error to obtain an averaging standard error defined as

$$\widehat{\mathrm{SE}}_{e,avg}(\hat{\beta}_{e,iv}) \equiv \frac{n}{T+n}(1-T^{-1})^{-1/2}\widehat{\mathrm{SE}}_{e,0}(\hat{\beta}_{e,iv}) + \frac{T}{T+n}\widehat{\mathrm{SE}}_{e,1}(\hat{\beta}_{e,iv}), \tag{19}$$

which is consistent, as shown in the lemma below.

Lemma 4 *Under Assumptions 1 and 2, we have* $(nT)^{1/2}\widehat{\mathrm{SE}}_{e,avg}(\hat{\beta}_{e,iv}) \to_p \omega_{e,\infty}$.

By Theorem 3 and Lemmas 3 and 4, $\widehat{\mathrm{SE}}_{e,avg}(\hat{\beta}_{e,iv})$ can be used to construct confidence intervals, as in (13), and to perform statistical inference for the unknown parameter β.

5 Potential Inferential Issues with Other Instruments

The inferential issue attributable to the U-statistic structure can arise in other contexts as well. For example, the judge IV estimator may face similar challenges depending on its usage. As in equation (1) of Kling (2006), we begin with a linear regression model: $Y_i = S_i\gamma_1 + \varepsilon_i$, where Y_i and S_i respectively denote the outcome and the sentencing variable for defendant i. In Kling (2006), the assignment of judge Z_j to case j is supposed to be random, with the instrument $Z_j\pi$ reflecting the judge's leniency. The sentencing variable is given by a first-stage equation: $S_j = Z_j\pi + Q_j^\top\theta + \eta_j$, where Q_j includes a set of indicators for calendar quarter in each district office.

To understand the connection between the judge IV and the Hausman IV, we consider a pseudo-panel representation of the model:

$$Y_{i,t} = S_{i,t}\gamma_1 + w_{i,t}^\top\gamma_2 + u_{i,t}, \tag{20}$$

where $Y_{i,t}$ and $S_{i,t}$ denote the outcome and the treatment variable for defendant i handled by judge t, $w_{i,t}$ denotes the vector of observable characteristics for individuals or cases. Compared to the structural model in (14), equation (20) represents a potentially unbalanced pseudo-panel, as it allows for fixed effects implicitly and permits varying numbers of cases across judges. These differences are superficial in terms of the expression for the Hausman IV estimator $\hat{\beta}_{e,iv}$.

Since the fixed effects are potentially included in $w_{i,t}$, there is no need to partial out α_i explicitly like for (14). We define the following matrices:

$$\tilde{\Sigma}_w \equiv \sum_{t\leq T}\sum_{i\leq n_t} w_{i,t}w_{i,t}^\top \quad \text{and} \quad \tilde{\Gamma}_{w,a} \equiv \sum_{t\leq T}\sum_{i\leq n_t} w_{i,t}a_{i,t}, \ \text{ for } a \in \{u, z, S, Y\},$$

where n_t denotes the number of cases handled by judge t. We then construct the Hausman IV estimator as:

$$\tilde{\beta}_{e,iv} = \frac{\sum_{t \leq T} \sum_{i \leq n_t} z_{i,t} \hat{Y}_{i,t}}{\sum_{t \leq T} \sum_{i \leq n_t} z_{i,t} \hat{S}_{i,t}}, \tag{21}$$

where $\hat{S}_{i,t} \equiv S_{i,t} - w_{i,t}^{\top} \tilde{\Sigma}_w^{-1} \tilde{\Gamma}_{w,S}$ and $\hat{Y}_{i,t} \equiv Y_{i,t} - w_{i,t}^{\top} \tilde{\Sigma}_w^{-1} \tilde{\Gamma}_{w,Y}$, and $z_{i,t}$ is the IV for $S_{i,t}$, which will be discussed below. In the judge IV literature, it is common to use $(z_{i,t}, w_{i,t}^{\top})$ as the IVs for equation (20), and then apply the two-stage least squares (2SLS) estimation to obtain the 2SLS estimators for γ_1 and γ_2. Elementary algebra shows that $\tilde{\beta}_{e,iv}$ is equivalent to the 2SLS estimator of γ_1. Therefore, the Hausman IV estimator is indeed identical to the judge IV estimator.[13]

There are two common ways to construct the (judge) IV for $S_{i,t}$ in the literature, both involving some form of leave-out mean. We focus on the leave-one-out mean, as it is the most straightforward and closely related to the Hausman IV studied in earlier sections of this paper. The first approach is to use the leave-one-out mean of the treatment $S_{i,t}$:[14]

$$z_{i,t} = (n_t - 1)^{-1} \sum_{i' \neq i, i' \leq n_t} S_{i',t}, \tag{22}$$

which is identical to the Hausman IV in (3), with adjustments for heterogeneous n_t. Given the numerical equivalence of $\tilde{\beta}_{e,iv}$ to the 2SLS estimator of γ_1, the main results established in Section 4 apply to the judge IV estimator when n_t is homogeneous across t.

To see that the judge IV estimator $\tilde{\beta}_{e,iv}$ based on IV in (20), shares similar inferential issues as the Hausman IV in general, we abstract Q_t from the first-stage equation of $S_{i,t}$ for now, simplifying it as $S_{i,t} = c_t + v_{i,t}$ in our pseudo-panel representation, where $c_t = Z_t \pi$.[15] Applying the expression for $Y_{i,t}$ in (20) to the definition of $\tilde{\beta}_{e,iv}$, we obtain:

$$\tilde{\beta}_{e,iv} - \gamma_1 = \frac{\sum_{t \leq T} \sum_{i \leq n_t} z_{i,t} u_{i,t} - \tilde{\Gamma}_{w,z}^{\top} \tilde{\Sigma}_w^{-1} \tilde{\Gamma}_{w,u}}{\sum_{t \leq T} \sum_{i \leq n_t} z_{i,t} \hat{S}_{i,t}}. \tag{23}$$

In the numerator of the above equation, the first term is the source of the inferential issue with the judge IV estimator. Using the definition of $z_{i,t}$ in (22) and the first-stage equation of $S_{i,t}$, we can write:

$$\sum_{t \leq T} \sum_{i \leq n_t} z_{i,t} u_{i,t} = \sum_{t \leq T} \sum_{i \leq n_t} c_t u_{i,t} + \sum_{t \leq T} \sum_{i \leq n_t} \sum_{i' \neq i, i' \leq n_t} \frac{u_{i,t} v_{i',t}}{n_t - 1}. \tag{24}$$

[13] Given this equivalence, the main results established in Section 4 an be extended naturally to the judge IV estimator, with appropriate adjustments for the varying cluster sizes n_t across t.

[14] See, e.g., Di Tella and Schargrodsky (2013); Dahl, Kostøl, and Mogstad (2014); Aizer and Doyle (2015); Autor, Kostøl, Mogstad, and Setzler (2019); Gross and Baron (2022).

[15] Justification of this simplification and the extension of this simplified form are provided later in this section.

It is evident that the second term in this expression shares similar sum-of-U-statistic structure as the second component in (6). Therefore, clustering becomes an issue for econometric inference if the "cluster sizes" n_t are small, as demonstrated in the lemma below.

Lemma 5 *Suppose that c_t has constant variance σ_c^2 and Assumptions 1(i, ii) hold. Then we have:*

$$\mathrm{Var}\left(\sum_{t\leq T}\sum_{i\leq n_t} z_{i,t}u_{i,t}\right) = \sigma_c^2\sigma_u^2\sum_{t\leq T} n_t + (\sigma_u^2\sigma_v^2+\sigma_{u,v}^2)\sum_{t\leq T}\frac{n_t}{n_t-1}. \tag{25}$$

Additionally,

$$\frac{(\sigma_u^2\sigma_v^2+\sigma_{u,v}^2)\sum_{t\leq T}\frac{n_t}{n_t-1}}{\sigma_c^2\sigma_u^2\sum_{t\leq T} n_t} \geq \frac{1}{\sigma_c^2}\frac{\sigma_v^2+\sigma_{u,v}^2/\sigma_u^2}{\sum_{t\leq T} q_t(n_t-1)}, \tag{26}$$

where $q_t \equiv n_t/\sum_{t\leq T} n_t$.

Lemma 5 highlights the significance of cluster dependence introduced by the judge IV in the asymptotic variance of $\tilde{\beta}_{e,iv}$. Specifically, the cluster dependence, captured by $\sigma_{u,v}$, is not asymptotically negligible if the "average" cluster size $\sum_{t\leq T} q_t(n_t-1)$ is bounded from above. Moreover, under this condition, the contribution of this cluster dependence to the asymptotic variance of $\tilde{\beta}_{e,iv}$ can be substantial if σ_c^2 is relatively small. It is clear that Assumption 1(ii) maintained in Lemma 5 aligns with the judge's random assignment design, indicating that even with random assignment, cluster dependence introduced by the judge IV must still be accounted for when conducting valid inference with a small "average" cluster size.

We now turn to the analysis of the judge IV estimator when the instrument is constructed as a leave-out mean of the regression residuals $\hat{S}_{i,t}$:[16]

$$z_{i,t} = (n_t-1)^{-1}\sum_{i'\neq i, i'\leq n_t}\hat{S}_{i',t}. \tag{27}$$

We consider a slightly more complicated first-stage: $S_{i,t} = c_t + w_{i,t}^\top\theta + v_{i,t}$, where the IV in (27), by partialling out the extra control variables, attempts to estimate the IV in (22) derived from the simplified reduced form.[17] The estimation error of the judge IV estimator retains the form presented in (23). Therefore, in order to investigate its inferential issues, it suffices to study the first term in the numerator of (23).

Using the definition of $\hat{S}_{i,t}$ and the expression for $S_{i,t}$, we obtain:

$$z_{i,t} = c_t^e + (n_t-1)^{-1}\sum_{i'\neq i, i'\leq n_t} v_{i,t} - \bar{w}_{-i,t}^\top(\tilde{\theta}-\theta^*),$$

[16] See, e.g., Dobbie, Goldin, and Yang (2018); Norris, Pecenco, and Weaver (2021); Frandsen, Lefgren, and Leslie (2023).

[17] Specifically, we can view the treatment variable in the simplified first-stage as $S_{i,t} - w_{i,t}^\top\theta$, assuming that θ is known. Since the judge IVs take the form of a leave-out mean, partialling out the extra control variables from $S_{i,t}$ is necessary only when $w_{i',t}$ is correlated with $u_{i,t}$. This provides the justification for the IV in (22).

where $c_t^e \equiv c_t - \bar{w}_{-i,t}^\top(\theta^* - \theta)$, $\bar{w}_{-i,t} = (n_t - 1)^{-1} \sum_{i' \neq i, i' \leq n_t} w_{i,t}$ and θ^* denotes the probability limit of $\tilde{\theta} \equiv \tilde{\Sigma}_w^{-1} \tilde{\Gamma}_{w,S}$. This immediately implies that

$$\sum_{t \leq T} \sum_{i \leq n_t} z_{i,t} u_{i,t} = \sum_{t \leq T} \sum_{i \leq n_t} c_t^e u_{i,t} + \sum_{t \leq T} \sum_{i \leq n_t} \sum_{i' \neq i, i' \leq n_t} \frac{u_{i,t} v_{i',t}}{n_t - 1} - \sum_{t \leq T} \sum_{i \leq n_t} u_{i,t} \bar{w}_{-i,t}^\top (\tilde{\theta} - \theta^*). \tag{28}$$

Compared to (23), the expression in (28) contains an extra term $\sum_{t \leq T} \sum_{i \leq n_t} u_{i,t} \bar{w}_{-i,t}^\top (\tilde{\theta} - \theta^*)$, which arises from partialling out $w_{i,t}$. However, this term is of smaller stochastic order than the other two terms under certain regularity conditions, such as Assumptions 1(i, ii) and 2(i-iv). Therefore, the implications of Lemma 5 also apply to the judge IV estimator $\tilde{\beta}_{e,iv}$, when regression residuals are used to construct the IV: even with random assignment, the cluster dependence, captured by $\sigma_{u,v}$, is not asymptotically negligible when the "average" cluster size $\sum_{t \leq T} q_t (n_t - 1)$ is bounded from above, and it could be substantial if the variance of c_t^e is relatively small.

To conclude this section, we would like to emphasize that the U-statistic introduces a challenge due to the unintended interlinkage among observations. This issue can also manifest in the context of Bartik instruments. For example, as noted in Diamond (2016, equation (23)), the instrument is computed using the average log wage in cities within a 25-mile radius of a given city, while excluding the log wage of the city itself. This approach inherently creates an interlinkage problem. In contrast to the Hausman IV or the judge IV estimators, where interlinkage is confined within a cluster, allowing for asymptotic analysis as the number of clusters approaches infinity, the 25-mile radius circles may overlap. This overlapping nature complicates the clustering from an asymptotic analysis perspective. We leave the exploration of this complex issue as a topic for future research.

6 Conclusion

In this paper, we address econometric inferential issues related to Hausman instruments. The IV estimator based on Hausman instruments has a "numerator" that involves U-statistics, naturally introducing a clustering problem even when the errors are independent of each other. The clustering issue can be important depending on the size of the clusters relative to the total sample size. We develop a standard error that is robust to these problems. While clustered standard errors are not always valid, they can serve as a pragmatic compromise for addressing the inter-linkage issue when using Hausman IV in BLP or using the judge IV.

References

AGUIRREGABIRIA, V. (2019): "Empirical industrial organization: models, methods, and applications," *University of Toronto*.

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

ARMSTRONG, T. B. (2016): "Large market asymptotics for differentiated product demand estimators with economic models of supply," *Econometrica*, 84(5), 1961–1980.

AUTOR, D., A. KOSTØL, M. MOGSTAD, AND B. SETZLER (2019): "Disability benefits, consumption insurance, and household labor supply," *American Economic Review*, 109(7), 2613–2654.

BARNDORFF-NIELSEN, O. E., P. R. HANSEN, A. LUNDE, AND N. SHEPHARD (2008): "Designing realized kernels to measure the ex post variation of equity prices in the presence of noise," *Econometrica*, 76(6), 1481–1536.

BERRY, S., J. LEVINSOHN, AND A. PAKES (1995): "Automobile prices in market equilibrium," *Econometrica*, pp. 841–890.

BERRY, S., O. B. LINTON, AND A. PAKES (2004): "Limit theorems for estimating the parameters of differentiated product demand systems," *Review of Economic Studies*, 71(3), 613–654.

BRESNAHAN, T. F., AND R. J. GORDON (2008): *The economics of new goods*, vol. 58. University of Chicago Press.

CONLON, C., AND J. GORTMAKER (2020): "Best practices for differentiated products demand estimation with PyBLP," *The RAND Journal of Economics*, 51(4), 1108–1161.

CRAWFORD, G. S., AND A. YURUKOGLU (2012): "The welfare effects of bundling in multichannel television markets," *American Economic Review*, 102(2), 643–685.

DAHL, G. B., A. R. KOSTØL, AND M. MOGSTAD (2014): "Family welfare cultures," *The Quarterly Journal of Economics*, 129(4), 1711–1752.

DI TELLA, R., AND E. SCHARGRODSKY (2013): "Criminal recidivism after prison and electronic monitoring," *Journal of Political Economy*, 121(1), 28–73.

DIAMOND, R. (2016): "The determinants and welfare implications of US workers' diverging location choices by skill: 1980-2000," *The American Economic Review*, 106(3), 479–524.

DOBBIE, W., J. GOLDIN, AND C. S. YANG (2018): "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges," *The American Economic Review*, 108(2), 201–240.

Egan, M., A. Hortaçsu, and G. Matvos (2017): "Deposit competition and financial fragility: Evidence from the US banking sector," *American Economic Review*, 107(1), 169–216.

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

Freyberger, J. (2015): "Asymptotic theory for differentiated products demand models with many markets," *Journal of Econometrics*, 185(1), 162–181.

Gross, M., and E. J. Baron (2022): "Temporary stays and persistent gains: The causal effects of foster care," *American Economic Journal: Applied Economics*, 14(2), 170–199.

Hahn, J., G. Kuersteiner, and M. Mazzocco (2020): "Estimation with aggregate shocks," *The Review of Economic Studies*, 87(3), 1365–1398.

Häusler, E., and H. Luschgy (2015): *Stable Convergence and Stable Limit Theorems*, vol. 74. Springer.

Hausman, J., G. Leonard, and J. D. Zona (1994): "Competitive analysis with differenciated products," *Annales d'Economie et de Statistique*, pp. 159–180.

Hausman, J. A. (1996): "Valuation of new goods under perfect and imperfect competition," in *The Economics of New Goods*, pp. 207–248. University of Chicago Press.

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

Kuersteiner, G. M., and I. R. Prucha (2013): "Limit theory for panel data models with cross sectional dependence and sequential exogeneity," *Journal of Econometrics*, 174(2), 107–126.

Nevo, A. (2000): "A practitioner's guide to estimation of random-coefficients logit models of demand," *Journal of Economics & Management Strategy*, 9(4), 513–548.

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

Norris, S., M. Pecenco, and J. Weaver (2021): "The effects of parental and sibling incarceration: Evidence from ohio," *American Economic Review*, 111(9), 2926–2963.

Phillips, P. C. B., and S. Ouliaris (1990): "Asymptotic properties of residual based tests for cointegration," *Econometrica*, 58, 165–193.

Phillips, P. C. B., and D. Sul (2003): "Dynamic panel estimation and homogeneity testing under cross section dependence," *The Econometrics Journal*, 6(1), 217–259.

Rossi, P. E. (2014): "Even the rich can make themselves poor: A critical examination of IV methods in marketing applications," *Marketing Science*, 33(5), 655–672.

Scanlon, P. (2019): "New goods and asset prices," *Journal of Financial Economics*, 132(3), 140–157.

Appendix

A Notations and Definitions

We begin by introducing some concepts related to the stable convergence of random variables from the literature, see, e.g., Häusler and Luschgy (2015). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the underlying probability space, and $\mathcal{X}$ be a Polish topological space equipped with its Borel sigma-field $\mathcal{B}(\mathcal{X})$. For a sub-sigma-field $\mathcal{G} \subset \mathcal{F}$, and a $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$-valued random variable X, let $\mathbb{P}^X$ and $\mathbb{P}^{X|\mathcal{G}}$ denote the marginal distribution of X and the conditional distribution of X given $\mathcal{G}$, respectively. Let $\mathcal{C}_b(\mathcal{X})$ denote the space of all continuous, bounded functions $h : \mathcal{X} \longmapsto \mathbb{R}$ equipped with the sup-norm $\|h\|_\infty \equiv \sup_{x\in\mathcal{X}} |h(x)|$.

Definition 1 *Let $\mathcal{G} \subset \mathcal{F}$ be a sub-sigma-field. A sequence $(X_m)_{m\geq 1}$ of $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$-valued random variables is said to converge stably to an $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$-valued random variable X, denoted as*

$$X_m \to X \qquad (\mathcal{G}\text{-stably}),$$

if $\mathbb{P}^{X_m|\mathcal{G}} \to \mathbb{P}^{X|\mathcal{G}}$ weakly as $m \to \infty$. That is

$$\lim_{m\to\infty} \mathbb{E}\left[g\mathbb{E}\left[h(X_m)|\mathcal{G}\right]\right] = \int g \int h(x)\mathbb{P}^{X|\mathcal{G}}(\cdot, dx)d\mathbb{P},$$

for every $\mathcal{G}$-measurable function g with $\mathbb{E}\left[|g|\right] < \infty$ and every $h \in \mathcal{C}_b(\mathcal{X})$. In case that $\mathbb{P}^{X|\mathcal{G}}$ equals $\mathbb{P}^X$ almost surely, then $(X_m)_{m\geq 1}$ is said to converge $\mathcal{G}$-mixing to X, denoted as

$$X_m \to X \qquad (\mathcal{G}\text{-mixing}).$$

The limit $\mathbb{P}^{X|\mathcal{G}}$ in the $\mathcal{G}$-stable convergence is a Markov kernel from $(\Omega, \mathcal{F})$ to $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ such that $\mathbb{P}^{X|\mathcal{G}}(\omega, \cdot)$ is a probability measure on $\mathcal{B}(\mathcal{X})$ for every $\omega \in \Omega$, and $\mathbb{P}^{X|\mathcal{G}}(\cdot, B)$ is $\mathcal{F}$-measurable for every $B \in \mathcal{B}(\mathcal{X})$. For presenting the main results of this paper, we have $\mathcal{X} = \mathbb{R}$ and $\mathbb{P}^{X|\mathcal{G}} = \mathbb{P}^{\eta Z|\mathcal{G}} = N(0, \eta^2)$ for a $\mathcal{G}$-measurable non-negative random variable η, and a standard normal random variable Z which is independent of $\mathcal{G}$, in the stable martingale central limit theorem.

B Proofs of the Main Results

In this section, we provide the proofs of the main results, including Lemmas 1 and 2, and Theorems 1 and 2 from Section 3, as well as Theorem 3, and Lemmas 3 and 4 from Section 4. The auxiliary lemmas used in these proofs are presented in the Supplemental Appendix (hereafter referred to as SA).

Proof of Theorem 1. The expression in (7) follows directly from (4) and Lemmas SA.2 and SA.3 in SA. By Lemma SA.4 in SA,

$$(nT)^{-1/2}\sum_{i\leq n}\sum_{t\leq T}(\gamma u_{i,t}(c_t-\bar{c})+\varepsilon_{i,t})\to\tilde{\omega}_\infty Z \qquad (\mathcal{F}_0\text{-stably}), \tag{29}$$

which together with Assumption 1(iv) and (7) implies that

$$(nT)^{1/2}(\hat{\beta}_{iv}-\beta)=\sum_{i\leq n}\sum_{t\leq T}\frac{\gamma u_{i,t}(c_t-\bar{c})+\varepsilon_{i,t}}{(nT)^{1/2}\gamma^2\hat{\sigma}_c^2}+O_p((nT)^{-1/2}). \tag{30}$$

Since σ_c^2 is $\mathcal{F}_0$-measurable, by Assumption 1(iv) and (29), we can apply Theorem 3.18(b) in Häusler and Luschgy (2015) to get:

$$\left(\sum_{i\leq n}\sum_{t\leq T}\frac{\gamma u_{i,t}(c_t-\bar{c})+\varepsilon_{i,t}}{(nT)^{1/2}},\gamma^2\hat{\sigma}_c^2\right)\to(\tilde{\omega}_\infty Z,\gamma^2\sigma_c^2) \qquad (\mathcal{F}_0\text{-stably}). \tag{31}$$

For any $(x,y)\in\mathbb{R}\times\mathbb{R}$, let

$$g(x,y)\equiv\begin{cases} x/y, & y>0\\ 0, & y\leq 0\end{cases}.$$

Then $g(x,y)$ is Borel-measurable and $\mathbb{P}^{(\tilde{\omega}_\infty Z,\gamma^2\sigma_c^2)}$-continuous almost surely. Therefore by Theorem 3.18(c) in Häusler and Luschgy (2015),

$$g\left(\sum_{i\leq n}\sum_{t\leq T}\frac{\gamma u_{i,t}(c_t-\bar{c})+\varepsilon_{i,t}}{(nT)^{1/2}},\gamma^2\hat{\sigma}_c^2\right)\to g(\tilde{\omega}_\infty Z,\gamma^2\sigma_c^2) \qquad (\mathcal{F}_0\text{-stably}). \tag{32}$$

The claim of the Theorem follows from (32) and the definition of $g(x,y)$. ■

Proof of Lemma 1. By the definition of $\widehat{\mathrm{SE}}_0(\hat{\beta}_{iv})$, we can write

$$\sqrt{nT}\widehat{\mathrm{SE}}_0(\hat{\beta}_{iv})=\sqrt{\frac{(nT)^{-1}\sum_{t\leq T}\sum_{i\leq n}\left(\sum_{i'\neq i}\left(x_{i',t}-\bar{x}_{i',\cdot}\right)\right)^2}{\left((nT)^{-1}\sum_{t\leq T}\sum_{i\leq n}\sum_{i'\neq i}x_{i,t}\left(x_{i',t}-\bar{x}_{i',\cdot}\right)\right)^2}(nT)^{-1}\sum_{t\leq T}\sum_{i\leq n}\hat{u}_{i,t}^2}.$$

The claim of the lemma follows from Assumption 1(iv), and Lemmas SA.2, SA.5 and SA.6 in SA. ■

Proof of Lemma 2. We begin by expressing

$$(nT)(\widehat{\text{SE}}_1(\hat{\beta}_{iv}))^2 = \frac{(nT)^{-1}\sum_{t\leq T}\left(\sum_{i\leq n}\hat{u}_{i,t}\left(z_{i,t}-\bar{z}_{i,\cdot}\right)\right)^2}{\left((nT)^{-1}\sum_{t\leq T}\sum_{i\leq n}x_{i,t}\left(z_{i,t}-\bar{z}_{i,\cdot}\right)\right)^2}. \tag{33}$$

By Lemma SA.2 in SA, the denominator on the right-hand side of (33) satisfies:

$$\left((nT)^{-1}\sum_{t\leq T}\sum_{i\leq n}x_{i,t}\left(z_{i,t}-\bar{z}_{i,\cdot}\right)\right)^2 = \gamma^4\hat{\sigma}_c^4 + O_p((nT)^{-1/2}). \tag{34}$$

Using Lemmas SA.8 and SA.9 in SA, we approximate the numerator on the right-hand side of (33) as

$$\begin{aligned}&(nT)^{-1}\sum_{t\leq T}\left(\sum_{i\leq n}\hat{u}_{i,t}\left(z_{i,t}-\bar{z}_{i,\cdot}\right)\right)^2\\ &= (nT)^{-1}\sum_{t\leq T}\left(\sum_{i\leq n}\gamma(c_t-\bar{c})u_{i,t}-\xi_t\right)^2 + \frac{\sigma_u^2\sigma_v^2+\sigma_{u,v}^2}{n-1} + O_p((nT)^{-1/2}).\end{aligned} \tag{35}$$

By Lemmas SA.10 and SA.11 in SA, we have

$$(nT)^{-1}\sum_{t\leq T}\left(\sum_{i\leq n}\gamma(c_t-\bar{c})u_{i,t}-\xi_t\right)^2 = O_p(1). \tag{36}$$

The claim in (9) follows from Assumption 1(iv) and (33)-(36). In view of (9), to prove (10), it is sufficient to show that as $T\to\infty$,

$$(nT)^{-1}\sum_{t\leq T}\left(\sum_{i\leq n}\gamma(c_t-\bar{c})u_{i,t}-\xi_t\right)^2 \to_p \gamma^2\sigma_u^2\sigma_c^2. \tag{37}$$

Applying Lemmas SA.10 and SA.11 in SA, and applying the Cauchy-Schwarz inequality, we get

$$\begin{aligned}(nT)^{-1}\sum_{t\leq T}\left(\sum_{i\leq n}\gamma(c_t-\bar{c})u_{i,t}-\xi_t\right)^2 &= (nT)^{-1}\sum_{t\leq T}\left(\sum_{i\leq n}\gamma(c_t-\bar{c})u_{i,t}\right)^2 + (nT)^{-1}\sum_{t\leq T}\xi_t^2\\ &\quad - 2\gamma(nT)^{-1}\sum_{t\leq T}\sum_{i\leq n}\gamma(c_t-\bar{c})u_{i,t}\xi_t\\ &= (nT)^{-1}\sum_{t\leq T}\left(\sum_{i\leq n}\gamma(c_t-\bar{c})u_{i,t}\right)^2 + O_p(T^{-1/2})\\ &= \gamma^2\sigma_u^2\hat{\sigma}_c^2 + 2\gamma(nT)^{-1}\sum_{t\leq T}\sum_{i\leq n}(c_t-\bar{c})^2\mu_{i,t} + O_p(T^{-1/2}).\end{aligned} \tag{38}$$

By Assumptions 1(i, ii, iii), we have

$$\mathbb{E}\left[\left|(nT)^{-1}\sum_{t\leq T}\sum_{i\leq n}(c_t-\bar{c})^2\mu_{i,t}\right|^2\right]=(nT)^{-2}\sum_{t\leq T}\sum_{i\leq n}\mathbb{E}[(c_t-\bar{c})^4]\mathbb{E}[\mu_{i,t}^2]\leq KT^{-1},$$

which, together with Markov's inequality, implies that

$$(nT)^{-1}\sum_{t\leq T}\sum_{i\leq n}(c_t-\bar{c})^2\mu_{i,t}=O_p(T^{-1/2}). \tag{39}$$

The desired result in (37) follows from Assumption 1(iv), along with (38), (39) and Slutsky's Theorem. ■

Proof of Theorem 2. First consider the case where both n and T go to infinity. By Lemmas 1 and 2, we have

$$(nT)^{1/2}(1-T^{-1})^{-1/2}\widehat{\mathrm{SE}}_0(\hat{\beta}_{iv})=\frac{\sigma_u}{\gamma\sigma_c}+o_p(1)\quad\text{and}\quad(nT)^{1/2}\widehat{\mathrm{SE}}_1(\hat{\beta}_{iv})=\frac{\sigma_u}{\gamma\sigma_c}+o_p(1),$$

which implies that

$$\begin{aligned}(nT)^{1/2}\widehat{\mathrm{SE}}_{avg}(\hat{\beta}_{iv})=\frac{\sigma_u}{\gamma\sigma_c}&+\frac{n}{T+n}\left((nT)^{1/2}(1-T^{-1})^{-1/2}\widehat{\mathrm{SE}}_0(\hat{\beta}_{iv})-\frac{\sigma_u}{\gamma\sigma_c}\right)\\&+\frac{T}{T+n}\left((nT)^{1/2}\widehat{\mathrm{SE}}_1(\hat{\beta}_{iv})-\frac{\sigma_u}{\gamma\sigma_c}\right)=\frac{\sigma_u}{\gamma\sigma_c}+o_p(1).\end{aligned} \tag{40}$$

Since $\omega_\infty=\sigma_u/(\gamma\sigma_c)$ in this case, from (40) we have $(nT)^{1/2}\widehat{\mathrm{SE}}_{avg}(\hat{\beta}_{iv})\to_p\omega_\infty$. Second, consider the case where n is bounded from above and T approaches infinity. In this scenario,

$$\frac{n}{T+n}=o(1)\quad\text{and}\quad\frac{T}{T+n}=1+o(1). \tag{41}$$

Moreover, Lemma 1 shows that

$$(nT)^{1/2}(1-T^{-1})^{-1/2}\widehat{\mathrm{SE}}_0(\hat{\beta}_{iv})=O_p(1),$$

which together with (10) of Lemma 2 and (41) implies that

$$(nT)^{1/2}\widehat{\mathrm{SE}}_{avg}(\hat{\beta}_{iv})=\frac{T}{T+n}(nT)^{1/2}\widehat{\mathrm{SE}}_1(\hat{\beta}_{iv})+o_p(1)=\omega_\infty+o_p(1).$$

To finish the proof of the first claim of the theorem, consider the last case where T is bounded from above and n approaches infinity. In this scenario,

$$\frac{n}{T+n}=1+o(1)\quad\text{and}\quad\frac{T}{T+n}=o(1). \tag{42}$$

Moreover, by Lemmas SA.10 and SA.11 in SA,

$$(nT)^{1/2}\widehat{\mathrm{SE}}_1(\hat{\beta}_{iv}) = O_p(1),$$

which together with Lemma 1 and (42) implies that

$$(nT)^{1/2}\widehat{\mathrm{SE}}_{avg}(\hat{\beta}_{iv}) = \frac{n}{T+n}(nT)^{1/2}(1-T^{-1})^{-1/2}\widehat{\mathrm{SE}}_0(\hat{\beta}_{iv}) + o_p(1) = \omega_\infty + o_p(1).$$

In sum, we have shown that $(nT)^{1/2}\widehat{\mathrm{SE}}_{avg}(\hat{\beta}_{iv}) \to_p \omega_\infty$ in the asymptotic framework with $nT \to \infty$. It is evident that ω_∞^2 is $\mathcal{F}_0$-measurable. Therefore by the first claim of the lemma, Lemma SA.4 in SA and similar arguments in the proof of Theorem 1, we can show that

$$\frac{\hat{\beta}_{iv} - \beta}{\widehat{\mathrm{SE}}_{avg}(\hat{\beta}_{iv})} \to Z \qquad (\mathcal{F}_0\text{-stably}).$$

Since Z and $\mathcal{F}_0$ are independent, the convergence above is $\mathcal{F}_0$-mixing. ■

Proof of Theorem 3. By Assumptions 1(iv), and 2(ii, iii, v), and Lemmas SB.14 and SB.15 in SA, we have

$$(nT)^{1/2}(\hat{\beta}_{e,iv} - \beta) = (nT)^{-1/2}\sum_{i\leq n}\sum_{t\leq T}\frac{\gamma(c_t - \bar{c} - \hat{\Gamma}_{w,c}^\top\hat{\Sigma}_w^{-1}(w_{i,t} - \bar{w}_{i,\cdot}))u_{i,t} + \varepsilon_{i,t}}{\gamma^2(\hat{\sigma}_c^2 - \hat{\Gamma}_{w,c}^\top\hat{\Sigma}_w^{-1}\hat{\Gamma}_{w,c})} + O_p((nT)^{-1/2}). \quad (43)$$

Since $\sigma_{e,c}^2$ is $\mathcal{F}_{e,0}$-measurable, by Assumptions 1(iv) and 2(ii, iii, v) and (43), we can apply Theorem 3.18(b) in Häusler and Luschgy (2015) to obtain:

$$\begin{pmatrix} \sum_{i\leq n}\sum_{t\leq T}\frac{\gamma(c_t - \bar{c} - \hat{\Gamma}_{w,c}^\top\hat{\Sigma}_w^{-1}(w_{i,t} - \bar{w}_{i,\cdot}))u_{i,t} + \varepsilon_{i,t}}{(nT)^{1/2}} \\ \gamma^2(\hat{\sigma}_c^2 - \hat{\Gamma}_{w,c}^\top\hat{\Sigma}_w^{-1}\hat{\Gamma}_{w,c}) \end{pmatrix} \to \begin{pmatrix} \tilde{\omega}_{e,\infty}Z \\ \gamma^2\sigma_{e,c}^2 \end{pmatrix} \qquad (\mathcal{F}_{e,0}\text{-stably}). \quad (44)$$

The claim of the theorem follows from (44) and the same arguments used in the proof of (32). ■

Proof of Lemma 3. By Lemmas SB.13, SB.16 and SB.17 in SA, we have

$$\begin{aligned}(nT)\widehat{\mathrm{SE}}_{e,0}(\hat{\beta}_{e,iv})^2 &= \frac{\left((nT)^{-1}\sum_{t\leq T}\sum_{i\leq n}\hat{z}_{i,t}^2\right)\times\left((nT)^{-1}\sum_{t\leq T}\sum_{i\leq n}\hat{u}_{e,i,t}^2\right)}{\left((nT)^{-1}\sum_{t\leq T}\sum_{i\leq n}\hat{x}_{i,t}z_{i,t}\right)^2} \\ &= \frac{\gamma^2\hat{\sigma}_{e,c}^2 + \sigma_v^2(n-1)^{-1} + O_p((nT)^{-1/2})}{(\gamma^2\hat{\sigma}_{e,c}^2 + O_p((nT)^{-1/2}))^2}(\sigma_u^2(1-T^{-1}) + O_p((nT)^{-1/2})),\end{aligned}$$

which, together with Assumptions 1(iv) and 2(ii, iii, v), shows that

$$\frac{(nT)\widehat{\mathrm{SE}}_{e,0}(\hat{\beta}_{e,iv})^2}{1-T^{-1}} \to_p \frac{\gamma^2\sigma_{e,c}^2 + \sigma_v^2(n_\infty - 1)^{-1}}{\gamma^4\sigma_{e,c}^4}\sigma_u^2. \quad (45)$$

This implies the claim of the lemma. ■

Proof of Lemma 4. The proof follows from (45), Lemmas SB.13 and SB.24 in SA, and similar arguments to those used in the proof of Theorem 2. Therefore, it is omitted. ■

Proof of Lemma 5. Under the maintained assumptions, we have:

$$\begin{aligned}\mathrm{Var}\left(\sum_{t\leq T}\sum_{i\leq n_t} z_{i,t}u_{i,t}\right) &= \mathrm{Var}\left(\sum_{t\leq T}\sum_{i\leq n_t} c_t u_{i,t}\right) + \mathrm{Var}\left(\sum_{t\leq T}\sum_{i=2}^{n_t}\sum_{i'=1}^{i-1} \frac{u_{i,t}v_{i',t} + u_{i',t}v_{i,t}}{n_t - 1}\right) \\ &= \sigma_c^2\sigma_u^2 \sum_{t\leq T} n_t + (\sigma_u^2\sigma_v^2 + \sigma_{u,v}^2)\sum_{t\leq T}\frac{n_t}{n_t - 1},\end{aligned}$$

which establishes the first claim of the lemma. Furthermore, since $(x-1)^{-1}$ is a convex function for $x \geq 2$, it follows that

$$\frac{\sum_{t\leq T}\frac{n_t}{n_t-1}}{\sum_{t\leq T} n_t} = \sum_{t\leq T}\frac{q_t}{n_t - 1} \geq \frac{1}{\sum_{t\leq T} q_t n_t - 1}.$$

This shows the second claim of the lemma. ■