Journal of Economic Theory and Econometrics, Vol. 35, No. 2, Jun. 2024, 1–10

Identities in Damage Estimation When the Number of Bidders Is Affected by Collusion*

Chirok Han† Jae Nahm‡

Abstract In typical practices of estimating cartel damages, an outcome variable such as winning bid price is regressed on a cartel dummy variable and multiple control variables. When the number of bidders in a bid is influenced by collusion, two approaches can be employed. One approach is to exclude the bidder number variable in the regression analysis. The other approach is to use predicted numbers of bidders for the collusive bids obtained, based on the observed relationship in the non-collusive bids. We show that these two approaches are equivalent in estimating the effects of collusion.

Keywords Damage estimation, cartel, collusion, bid rigging, bidder number, dummy variable approach, forecasting approach.

JEL Classification C10, L40.

*The authors thank the two anonymous reviewers for their helpful comments and suggestions. Han acknowledges support by the Institute of Information & Communications Technology Planning & Evaluation (IITP) Grant funded by the Korean Government (MSIT) (No. 2022-0-00302, Development of global demand forecasting and analysis/prediction system of market/industry trends).

†Corresponding author. Department of Economics, Korea University, 145 Anamro Seongbuk-gu, Seoul, Republic of Korea 02841. Email: chirokhan@korea.ac.kr.

‡Department of Economics, Korea University, 145 Anamro Seongbuk-gu, Seoul, Republic of Korea 02841. Email: shnahm@korea.ac.kr.

Received May 6, 2024, Revised May 28, 2024, Accepted May 29, 2024

1. INTRODUCTION

Damage estimation in bid-rigging is actively pursued in courts to determine the extent of a buyer's damages, typically measured by the degree of price increase resulting from collusion. The magnitude of the price increase is defined as the difference between the price paid in the actual world where collusion occurred and the 'but-for price', also known as the 'hypothetical competitive price,' in the hypothetical situation ("but-for world") where collusion did not take place.

With collusion having taken place, the actual prices paid by buyers are observed, but the but-for prices are not. Estimating the but-for prices requires controlling for other factors that affect competitive prices. This is typically achieved through multiple regression analyses to find the relationship between the factors influencing competitive prices and the resulting competitive prices in scenarios without bid-rigging.

One important variable affecting but-for prices is the number of bidders, as posited by game theory, which suggests Nash equilibrium prices in auctions are influenced by the bidder number. An increase in the number of participants in a bid fosters competition, which drives bid prices lower. Consequently, the coefficient on the bidder number is typically estimated to be negative in regressions (see, Hungria-Gunnelin, 2013).

However, collusion can influence the number of bidders, and controlling for the actual observed bidder number may lead to biased damage estimation. This bias arises because the goal is to measure the 'total' effect of collusion while regressions that control for the bidder number capture only a 'partial' effect. For instance, if a cartel's impact on the bidding price is solely through the reduction of partitipants, the measured partial effect might be minimal, while the total effect of collusion could be substantial.

Other than the often controversial two-stage least squares method, two viable approaches are available to address this issue. The first approach is to simply exclude the bidder number from the control variables, thereby not controlling for the number of bidders. The second approach retains the bidder number variable but substitutes the predicted but-for bidder numbers for the actual ones in collusive bids. The predicted bidder numbers are derived from an auxiliary regression that relates the number of bidders to exogenous control variables using data for non-collusive or benchmark bids.

The purpose of this paper is to establish the equivalence between these two approaches for estimating damages. The equivalence holds for both the 'dummy variable' method and the 'forecasting' method, both of which are widely used in damage estimation (McCrary and Rubinfeld, 2014; Davis and Garcés, 2010).

The former measures the collusion effect by the coefficient on a collusion dummy variable, while the latter compares the actual price to the predicted but-for price. Note that this study focuses on estimating damage with collusion given as an external event, rather than on detecting collusion itself.

In the subsequent sections, we elaborate on the equivalence of the two approaches. Section 2 presents the main arguments and examples for the equivalence. Concluding remarks are provided in Section 3. Mathematical proofs and example Stata code are included in the appendix.

2. IDENTITIES IN DAMAGE ESTIMATION

Let y_i denote the outcome variable such as the winning bid price and the successful bidding rate, d_i the collusion dummy variable, $\mathbf{x}_i$ the exogenous control variables unaffected by collusion that explain y_i, and n_i the number of participants. The bidder number n_i is determined exogenously for non-collusive bids but can be affected by collusion for the bids with $d_i = 1$. Let $\hat{\pi}$ be the ordinary least squares (OLS) estimator from the regression of n_i on $\mathbf{x}_i$ using the non-collusive bids, and let $\hat{n}_i = \mathbf{x}_i\hat{\pi}$, the predicted bidder number based on $\hat{\pi}$. Let $\bar{n}_i = (1-d_i)n_i + d_i\hat{n}_i$, which is the actual n_i for $d_i = 0$ and the predicted $\hat{n}_i$ for $d_i = 1$. The comparisons in this paper involve the regressions of

(f1) y_i on $\mathbf{x}_i$ using the observations with $d_i = 0$,
(f2) y_i on $\mathbf{x}_i$ and n_i using the observations with $d_i = 0$,
(d1) y_i on d_i and $\mathbf{x}_i$ using all observations, and
(d2) y_i on d_i, $\mathbf{x}_i$, and $\bar{n}_i$ using all observations.

Regressions (f1) and (f2) pertain to the forecasting approach, while regressions (d1) and (d2) are for the dummy variable method.

Let $\hat{\beta}_{f1}$ be the estimated coefficient on $\mathbf{x}_i$ from regression (f1); let $\hat{\beta}_{f2}$ and $\hat{\gamma}_{f2}$ be the estimated coefficients on $\mathbf{x}_i$ and n_i from regression (f2). Let $\hat{\delta}_{d1}$ and $\hat{\beta}_{d1}$ be the estimated coefficients on d_i and $\mathbf{x}_i$ from regression (d1); let $\hat{\delta}_{d2}$, $\hat{\beta}_{d2}$, and $\hat{\gamma}_{d2}$ be the estimated coefficients on d_i, $\mathbf{x}_i$, and $\bar{n}_i$ from regression (d2). The forecasting approach predicts the but-for prices for a collusive bid based on regressions (f1) and (f2), i.e., $\mathbf{x}_i\hat{\beta}_{f1}$ for (f1) and $\mathbf{x}_i\hat{\beta}_{f2} + \hat{\gamma}_{f2}\hat{n}_i$ for (f2). Regressions (d1) and (d2) are associated with the dummy variable approach, where the price overcharge is measured by the estimated coefficients $\hat{\delta}_{d1}$ and $\hat{\delta}_{d2}$ for (d1) and (d2), respectively. Beware that regression (d2) is not a conventional two-stage least squares regression because the predicted bidder number replaces the actual bidder number only for the collusive bids. We have the following identities.

Theorem 1. *(i)* $\hat{\beta}_{f1} = \hat{\beta}_{f2} + \hat{\pi}\hat{\gamma}_{f2}$ *so that* $\mathbf{x}_i\hat{\beta}_{f1} = \mathbf{x}_i\hat{\beta}_{f2} + \hat{\gamma}_{f2}\hat{n}_i$*, i.e., the damage estimators using the forecasting approach based on the two regressions (f1) and (f2) are identical when* n_i *is replaced with* $\hat{n}_i$ *for prediction. (ii)* $\hat{\delta}_{d1} = \hat{\delta}_{d2}$*, i.e., the damage estimators using the dummy-variable approach based on the two regressions (d1) and (d2) are identical.*

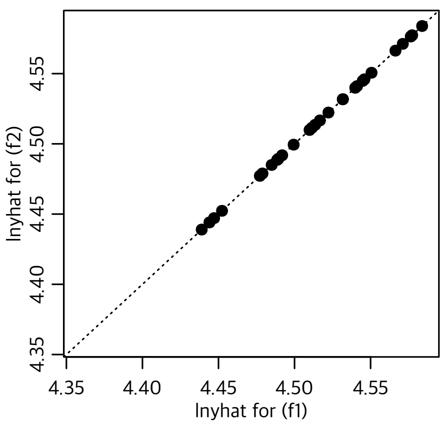
Theorem 1 demonstrates that excluding the bidder number variable is functionally equivalent to including it and replacing the actual number with the predicted number for bids identified as collusive.

lny	(1)	(2)	(3)
collusion	0.0653***	0.0653***	0.0927***
	(0.0153)	(0.0129)	(0.0148)
lnbase	0.0077*	0.0028	0.0047
	(0.0039)	(0.0034)	(0.0035)
method=2	0.0501	0.0409	0.0551
	(0.0462)	(0.0391)	(0.0412)
method=3	0.0405	0.0415*	0.0602**
	(0.0253)	(0.0214)	(0.0229)
lnlb	0.1615	-0.1907	-0.0431
	(0.3531)	(0.3044)	(0.3180)
nolb	0.7235	-0.8509	-0.2120
	(1.5508)	(1.3378)	(1.3977)
lncost	0.0628	0.0795**	0.0672*
	(0.0431)	(0.0365)	(0.0385)
$\bar{n}$		-0.0485***	
		(0.0082)	
n			-0.0363***
			(0.0076)
Intercept	3.3016**	4.9901***	4.3307***
	(1.5471)	(1.3387)	(1.3973)
nobs	93	93	93
R-squared	0.5045	0.6500	0.6100
Adj R-sq	0.4637	0.6167	0.5729

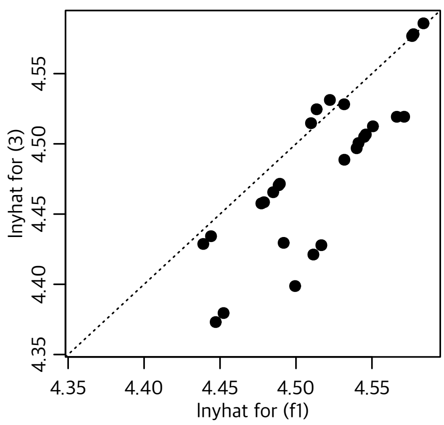
Table 1: EXAMPLE OF DUMMY-VARIABLE REGRESSION ANALYSIS. $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.1$ using the ordinary standard errors.

We illustrate the identities in Theorem 1 with a hypothetical example consisting of 93 observations of auctions: 28 for collusive biddings and 65 for noncollusive biddings. The dependent variable, denoted `lny`, is the logarithm of the successfull bid rate, and the collusion dummy variable is denoted `collusion`. The covariates ($\mathbf{x}_i$) include the logarithm of the base price (`lnbase`), dummy variables for the winner-decision method (one of 1, 2, or 3), the logarithm of the lower bound for the winning bidding rate (`lnlb`, replaced with zero if no lower bound applies), a dummy indicator for no lower bound (`nolb`), and the logarithm of a cost index (`lncost`).

To construct $\hat{n}_i$, n_i is regressed on the covariates $\mathbf{x}_i$ using the 65 noncollusive observations. Then, $\hat{n}_i$ is the predicted values based on this auxiliary regression. The regressions (d1) and (d2) are conducted by regressing `lny` on `collusion` and $\mathbf{x}_i$, and on `collusion`, $\mathbf{x}_i$, and $\bar{n}_i = (1-d_i)n_i + d_i\hat{n}_i$, respectively, using the entire sample. The results of these two regression are presented in columns (1) and (2) of Table 1, where the estimated coefficients of the `collusion` dummy variable are identical, as demonstrated by Theorem 1(ii). Despite the identical estimates, the reported standard errors for `collusion` differ between columns (1) and (2), which is related to the generated regressors problem (Pagan, 1984) inherent in the regression for column (2). Column (3) of Table 1 presents the results from the regression including the actual n_i as a control variable for comparison.



(a) Comparison of (f1) and (f2)



(b) Comparison of (f1) and (3)

Figure 1: COMPARISON OF PREDICTED VALUES FOR THE FORECASTING APPROACH. Model (3) includes n_i as an extra control.

Next, Figure 1 illustrates the predicted values for collusive biddings using the forecasting approach. Figure 1(a) compares the pairwise predicted values obtained from (f1) and (f2), showing that the values are identical for the collusive biddings, thus confirming Theorem 1(i). Figure 1(b) compares (f1) with a third regression that includes the actual n_i instead of $\bar{n}_i$, where the predicted values differ. Stata code for Table 1 and Figure 1 is available in Appendix B.

In the remainder of this section, we discuss the effects of omitting some control variables and including additional controls in the regression for predicting the number of bidders. If the set of regressors used to predict the number of bidders are different from that in the outcome equation, then the equivalences in Theorem 1 do not hold exactly, but they still hold asymptotically if the models are correctly specified. On one hand, if $\mathbf{x}_i = (\mathbf{x}_{ai}, \mathbf{x}_{bi})$, $\hat{\pi} = (\hat{\pi}_a', \hat{\pi}_b')'$, and $\hat{\pi}_b \xrightarrow{p} 0$, then Theorem 1 still holds with the exact identity replaced by asymptotic equivalence when $\mathbf{x}_{bi}$ is omitted from the bidder-number regression. Note that $\hat{\pi}_b \xrightarrow{p} 0$ means that $\mathbf{x}_{bi}$ is unimportant for determining n_i, so its omission does not cause bias. On the other hand, if extra variables $\mathbf{w}_i$ are included in the bidder-number regression, then again Theorem 1 holds asymptotically, though not exactly, if $\mathbf{w}_i$ is redundant in the determination of y_i conditional on $\mathbf{x}_i$ and $(\mathbf{x}_i, n_i)$ without collusion. Proofs of these claims are provided in the appendix.

3. CONCLUDING REMARKS

This paper establishes an equivalence in cartel damage estimation between two approaches: (i) a regression that omits the bidder number, and (ii) one that replaces the bidder number with its predicted but-for value for collusive bids. This equivalence holds for both dummy-variable and forecasting methods. If the predictors of the bidder number differ from the covariates in the main outcome equation, the equivalence does not hold exactly. Nonetheless, when the models are correctly specified such that the excluded variables are irrelevant to the bidder number and the additional variables are irrelevant to the outcome, the discrepancy between (i) and (ii) is asymptotically negligible.

A. MATHEMATICAL PROOFS

In this appendix, y, X, N, $\bar{N}$, and D are matrices of observations. Let $P_A = A(A'A)^{-1}A'$ and $M_A = I - P_A$. We first have the following identity.

Lemma A.1. *$M_{[X,\bar{N}]}D = M_X D$.*

Proof. Let X_1, $\hat{N}_1$, and D_1 denote the matrices of the relevant variables under bid rigging; let X_0, N_0, and D_0 denote those for the observations without agreement. Note that $\bar{N}_0 = N_0$, $\bar{N}_1 = \hat{N}_1 = X_1\hat{\pi}$, $\hat{\pi} = (X_0'X_0)^{-1}X_0'N_0$, $D_1 = (1,\dots,1)'$, and $D_0 = 0'$. By decomposition,

$$P_{[X,\bar{N}]}D = P_X D + P_{M_X\bar{N}}D. \tag{1}$$

But $X'\bar{N} = X_1'X_1\hat{\pi} + X_0'N_0 = X_1'X_1\hat{\pi} + X_0'X_0\hat{\pi} = X'X\hat{\pi}$, so that

$$M_X\bar{N} = \bar{N} - X(X'X)^{-1}X'\bar{N} = \bar{N} - X\hat{\pi} = \begin{pmatrix} 0 \\ M_{X_0}N_0 \end{pmatrix}. \tag{2}$$

Thus, $\bar{N}'M_X D = 0'D_1 + N_0'M_{X_0}D_0 = 0$, and hence $P_{M_X\bar{N}}D = 0$. This and (1) imply that $P_{[X,\bar{N}]}D = P_X D$, and thus that $M_{[X,\bar{N}]}D = M_X D$. □

Lemma A.2. *(i)* $\hat{\beta}_{f1} = \hat{\beta}_{f2} + \hat{\pi}\hat{\gamma}_{f2}$*; (ii)* $\hat{\delta}_{d1} = \hat{\delta}_{d2}$*; (iii)* $\hat{\beta}_{d1} = \hat{\beta}_{d2} + \hat{\pi}\hat{\gamma}_{d2}$.

Proof of Lemma A.2. (i) Let X_0, N_0, and $\mathbf{y}_0$ be the matrices of $\mathbf{x}_i$, n_i, and y_i for the non-collusive bids. The normal equations for regressions (f1) and (f2) imply $X_0'X_0\hat{\beta}_{f1} = X_0'X_0\hat{\beta}_{f2} + X_0'N_0\hat{\gamma}_{f2}$ because both sides equal $X_0'\mathbf{y}_0$. The result follows immediately. See also Wooldridge (2020) for simple regressions. (ii) By Lemma A.1,

$$\hat{\delta}_{d2} = (D'M_{[X,\bar{N}]}D)^{-1}D'M_{[X,\bar{N}]}\mathbf{y} = (D'M_X D)^{-1}D'M_X\mathbf{y} = \hat{\delta}_{d1}.$$

(iii) Let $e = \mathbf{y} - X\hat{\beta}_{d2} - \bar{N}\hat{\gamma}_{d2} - D\hat{\delta}_{d2}$. Then $X'e = 0$ so that

$$\begin{aligned} X'X\hat{\beta}_{d2} &= X'(\mathbf{y} - D\hat{\delta}_{d2}) - X'\bar{N}\hat{\gamma}_{d2} = X'(\mathbf{y} - D\hat{\delta}_{d1}) - X'\bar{N}\hat{\gamma}_{d2} \\ &= X'X\hat{\beta}_{d1} - X'\bar{N}\hat{\gamma}_{d2}, \end{aligned}$$

where the second identity holds due to part (ii) and the third because of the normal equations $X'\mathbf{y} = X'X\hat{\beta}_{d1} + X'D\hat{\delta}_{d1}$ for (d1). But $X'\bar{N} = X'X\hat{\pi}$ as shown in the proof of Lemma A.1 so that $X'X\hat{\beta} = X'X\tilde{\beta} - X'X\hat{\pi}\hat{\gamma}$. The result follows given that $X'X$ is invertible. □

Proof of Theorem 1. (i) Lemma A.2(i) implies that $\mathbf{x}_i\hat{\beta}_{f1} = \mathbf{x}_i\hat{\beta}_{f2} + \mathbf{x}_i\hat{\pi}\hat{\gamma}_{f2} = \mathbf{x}_i \cdot \hat{\beta}_{f2} + \hat{n}_i\hat{\gamma}_{f2}$ for $d_i = 1$. (ii) Obvious from Lemma A.2(ii). □

THE CASE THAT SOME CONTROL VARIABLES ARE OMITTED

We next consider the case that n_i is regressed on a subset $\mathbf{x}_{ai}$ of $\mathbf{x}_i$, where $\mathbf{x}_i = (\mathbf{x}_{ai}, \mathbf{x}_{bi})$. Consider the two regressions (f1) and (f2). Let $\tilde{n}_i = \mathbf{x}_{ai}\tilde{\pi}_a$, where $\tilde{\pi}_a$ is the OLS estimator from the regression of n_i on $\mathbf{x}_{ai}$ using the non-collusive bids. Then $\tilde{\pi}_a = \hat{\pi}_a + C_0\hat{\pi}_b$ with $C_0 = (X_{0a}'X_{0a})^{-1}X_{0a}'X_{0b}$, where $\hat{\pi} = (\hat{\pi}_a', \hat{\pi}_b')'$, and X_{0a} and X_{0b} are the matrices of $\mathbf{x}_{ai}$ and $\mathbf{x}_{bi}$ for $d_i = 0$. Therefore,

$$\tilde{n}_i - \hat{n}_i = \mathbf{x}_{ai}\tilde{\pi}_a - \mathbf{x}_i\hat{\pi} = \mathbf{x}_{ai}\tilde{\pi}_a - \mathbf{x}_{ai}\hat{\pi}_a - \mathbf{x}_{bi}\hat{\pi}_b = -(\mathbf{x}_{bi} - \mathbf{x}_{ai}C_0)\hat{\pi}_b = -\mathbf{r}_{bi}\hat{\pi}_b.$$

If $\hat{\pi}_b \xrightarrow{p} 0$, then $\tilde{n}_i - \hat{n}_i \xrightarrow{p} 0$ for each i, and thus

$$\mathbf{x}_i\hat{\beta}_{f1} - (\mathbf{x}_i\hat{\beta}_{f2} + \hat{\gamma}_{f2}\tilde{n}_i) = \hat{\gamma}_{f2}(\hat{n}_i - \tilde{n}_i) \xrightarrow{p} 0$$

for each i. The arguments thus far establishes the asymptotic equivalence for the forecasting approach.

For the dummy variable approach, let

$$\ddot{n}_i = (1-d_i)n_i + d_i\tilde{n}_i = \bar{n}_i + d_i(\tilde{n}_i - \hat{n}_i) = \bar{n}_i - d_i\mathbf{r}_{bi}\hat{\pi}_b,$$

and let $\ddot{N}$ be the corresponding full observation vector. We have $X'\ddot{N} = X'\bar{N} + X_1'R_{b1}\hat{\pi}_b$, where R_{b1} is the matrix of $\mathbf{r}_{bi}$ for the collusive bids in the sample. Thus, $\ddot{N}'M_X D = -\hat{\pi}_b'R_{b1}'X_1(X'X)^{-1}X_1'D_1$ is negligible if $\hat{\pi}_b \xrightarrow{p} 0$ ('negligible' in the sense that it converges in probability to zero if divided by the sample size). The asymptotic equivalence follows after some tedius but straightforward algebra.

THE CASE THAT EXTRA CONTROLS ARE INCLUDED

Now consider the case that n_i is regressed on $\mathbf{x}_i$ and some extra $\mathbf{w}_i$ (using the non-collusive bids). For the notations, let $\check{\pi}$ and $\check{\varphi}$ be the OLS estimators from the regression of n_i on $\mathbf{x}_i$ and $\mathbf{w}_i$ using the non-collusive bids. Let $\check{n}_i = \mathbf{x}_i\check{\pi} + \mathbf{w}_i\check{\varphi}$. Introduce the following two regressions for the forecasting approach: (f3) the regression of y_i on $\mathbf{x}_i$ and $\mathbf{w}_i$ using the observations with $d_i = 0$; (f4) the regression of y_i on $\mathbf{x}_i$, $\mathbf{w}_i$, and n_i using the observations with $d_i = 0$. Let the coefficients on $\mathbf{w}_i$ be denoted as α and let us use the '$f3$' and '$f4$' subscripts to denote the estimated coefficients for the corresponding regressions. Then Theorem 1 implies

$$\mathbf{x}_i\hat{\beta}_{f3} + \mathbf{z}_i\hat{\alpha}_{f3} = \mathbf{x}_i\hat{\beta}_{f4} + \mathbf{z}_i\hat{\alpha}_{f4} + \hat{\gamma}_{f4}\check{n}_i.$$

Also, $\hat{\beta}_{f1} = \hat{\beta}_{f3} + (X_0'X_0)^{-1}X_0'W_0\hat{\alpha}_{f3}$ and

$$\begin{pmatrix} \hat{\beta}_{f2} \\ \hat{\gamma}_{f2} \end{pmatrix} = \begin{pmatrix} \hat{\beta}_{f4} \\ \hat{\gamma}_{f4} \end{pmatrix} + \begin{pmatrix} X_0'X_0 & X_0'N_0 \\ N_0'X_0 & N_0'N_0 \end{pmatrix}^{-1} \begin{pmatrix} X_0'W_0 \\ N_0'W_0 \end{pmatrix} \hat{\alpha}_{f4}.$$

When $\hat{\alpha}_{f3} \xrightarrow{p} 0$ and $\hat{\alpha}_{f4} \xrightarrow{p} 0$, we have $\hat{\beta}_{f1} - \hat{\beta}_{f3} \xrightarrow{p} 0$, $\hat{\beta}_{f2} - \hat{\beta}_{f4} \xrightarrow{p} 0$, and $\hat{\gamma}_{f2} - \hat{\gamma}_{f4} \xrightarrow{p} 0$. All these results imply the claimed asymptotic equivalence.

B. STATA CODE

The Stata code for the example applications in Section 2 is provided as follows.

```
use data, clear
global X lnbase i.method lnlb nolb lncost
reg n ${X} if collusion==0
predict nhat
gen nbar = n
replace nbar = nhat if collusion==1
* Dummy variable method
reg lny collusion ${X}
reg lny collusion ${X} nbar
reg lny collusion ${X} n
* Forecasting approach
reg lny ${X} if collusion==0
predict lnyhat1
reg lny ${X} nbar if collusion==0
predict lnyhat2
reg lny ${X} n if collusion==0
predict lnyhat3
twoway scatter lnyhat2 lnyhat1 if collusion==1
twoway scatter lnyhat3 lnyhat1 if collusion==1
```

REFERENCES

Davis, P. and E. Garcés (2010). *Quantitative techniques for competition and antitrust analysis*, Princeton University Press.

Hungria-Gunnelin, R. (2013). "Impact of number of bidders on sale price of auctioned condominium apartments in Stockholm," *International Real Estate Review*, 16, 274–295.

McCrary, J. and D. Rubinfeld (2014). "Measuring benchmark damages in antitrust litigation," *Journal of Econometric Methods*, 3, 63–74.

Pagan, A. (1984). “Econometric issues in the analysis of regressions with generated regressors,” *International Economic Review*, 25, 221–247.

Wooldridge, J. W. (2020). *Introductory econometrics: A modern approach,* 7th ed. Cengage.