

2012s-15

## How Much Do Cartels Typically Overcharge?

*Marcel Boyer, Rachidi Kotchoni*

---

**Série Scientifique**  
*Scientific Series*

---

**Montréal**  
**Mai 2012**

© 2012 *Marcel Boyer, Rachidi Kotchoni*. Tous droits réservés. *All rights reserved*. Reproduction partielle permise avec citation du document source, incluant la notice ©.  
*Short sections may be quoted without explicit permission, if full credit, including © notice, is given to the source.*



Centre interuniversitaire de recherche en analyse des organisations

## **CIRANO**

Le CIRANO est un organisme sans but lucratif constitué en vertu de la Loi des compagnies du Québec. Le financement de son infrastructure et de ses activités de recherche provient des cotisations de ses organisations-membres, d'une subvention d'infrastructure du Ministère du Développement économique et régional et de la Recherche, de même que des subventions et mandats obtenus par ses équipes de recherche.

*CIRANO is a private non-profit organization incorporated under the Québec Companies Act. Its infrastructure and research activities are funded through fees paid by member organizations, an infrastructure grant from the Ministère du Développement économique et régional et de la Recherche, and grants and research mandates obtained by its research teams.*

### **Les partenaires du CIRANO**

#### **Partenaire majeur**

Ministère du Développement économique,  
de l'Innovation et de l'Exportation

#### **Partenaires corporatifs**

Autorité des marchés financiers  
Banque de développement du Canada  
Banque du Canada  
Banque Laurentienne du Canada  
Banque Nationale du Canada  
Banque Royale du Canada  
Banque Scotia  
Bell Canada  
BMO Groupe financier  
Caisse de dépôt et placement du Québec  
Fédération des caisses Desjardins du Québec  
Financière Sun Life, Québec  
Gaz Métro  
Hydro-Québec  
Industrie Canada  
Investissements PSP  
Ministère des Finances du Québec  
Power Corporation du Canada  
Rio Tinto Alcan  
State Street Global Advisors  
Transat A.T.  
Ville de Montréal

#### **Partenaires universitaires**

École Polytechnique de Montréal  
HEC Montréal  
McGill University  
Université Concordia  
Université de Montréal  
Université de Sherbrooke  
Université du Québec  
Université du Québec à Montréal  
Université Laval

Le CIRANO collabore avec de nombreux centres et chaires de recherche universitaires dont on peut consulter la liste sur son site web.

Les cahiers de la série scientifique (CS) visent à rendre accessibles des résultats de recherche effectuée au CIRANO afin de susciter échanges et commentaires. Ces cahiers sont écrits dans le style des publications scientifiques. Les idées et les opinions émises sont sous l'unique responsabilité des auteurs et ne représentent pas nécessairement les positions du CIRANO ou de ses partenaires.

*This paper presents research carried out at CIRANO and aims at encouraging discussion and comment. The observations and viewpoints expressed are the sole responsibility of the authors. They do not necessarily represent positions of CIRANO or its partners.*

ISSN 1198-8177

Partenaire financier

Développement  
économique, Innovation  
et Exportation  
Québec 

# How Much Do Cartels Typically Overcharge? \*

Marcel Boyer<sup>†</sup>, Rachidi Kotchoni<sup>‡</sup>

## Résumé / Abstract

L'estimation des surprix des cartels est au cœur de la politique de lutte aux cartels, car elle est un élément clé de la détermination des pénalités. Connor et Lande (2008) survolent la littérature sur les majorations de prix des cartels et concluent à une augmentation moyenne variant entre 31 % et 49 %. Considérant un échantillon plus grand, Connor (2010) trouve une moyenne de 50,4 % pour les cartels réussis. Cependant, les échantillons utilisés dans ces études sont constitués d'estimations venant de différentes études ou cas et non pas d'observations directes. De ce fait, ces échantillons héritent possiblement d'erreurs de modélisation et d'estimation, ainsi que d'un biais de publication. Une analyse sommaire des surprix dans l'échantillon de Connor révèle une distribution asymétrique, de l'hétérogénéité et la présence d'observations aberrantes. Ainsi, au-delà du fait que les estimations des surprix sont potentiellement biaisées, l'estimation d'un modèle de régression linéaire avec de telles données sans un traitement adéquat des problèmes ci-dessus pourrait produire des distorsions dans les résultats. Nous présentons une méta-analyse dans l'esprit de Connor and Bolotova (2006), mais qui tient compte adéquatement des problèmes mentionnés ci-dessus. Après correction du biais d'estimation, nos résultats suggèrent que la moyenne et la médiane des majorations de prix sont de l'ordre de 13,62 % avec une médiane de 13,63 % pour les cartels dont les majorations de prix se situaient initialement entre 0 % et 50 % et de l'ordre de 17,52 % avec une médiane de 14,05 % pour l'ensemble des cartels. Nos résultats débouchent sur des enjeux importants en politique de la concurrence.

**Mots clés :** Surprix de cartel, Politique de la concurrence, Heckman, Meta-analyse.

*The estimation of cartel overcharges lie at the heart of antitrust policy on cartel prosecution as it constitutes a basic element in the determination of fines. Connor and Lande (2008) conducted a survey of cartels and found a mean overcharge estimates in the range of 31% to 49%. By examining more sources, Connor (2010) finds a mean of 50.4% for successful cartels. However, the data used in those studies are estimates obtained in different ways, sources and contexts rather direct observations. Therefore, these data are subject to model error, estimation error and publication bias. A quick glance at the Connor database reveals that the universe of overcharge estimates is asymmetric, heterogenous and contains a number of influential observations. Beside the fact that overcharge estimates are potentially biased, fitting a linear regression model to the data without providing a carefull treatment of the problems raised above may produce distorted results. We conduct a meta-analysis of cartel overcharge estimates in the spirit of Connor and Bolotova (2006) while providing a sound treatment of those matters. We find typical bias-corrected mean and median overcharge estimates of 13.62% and 13.63% for cartels with initial overcharge estimates lying between 0% and 50% and bias-corrected mean and median overcharges estimates of 17.52% and 14.05% for the whole sample. Clearly, our results have significant antitrust policy implications.*

**Keywords:** Cartel overcharges, Antitrust, Heckman, Meta-analysis.

---

\* We thank Jimmy Royer and René Garcia for their comments on a previous version of this paper, but we remain solely responsible for its content. This paper is a significantly revised version of "The Econometrics of Cartel Overcharges" (CIRANO 2011s-35, march-august 2011).

<sup>†</sup> Emeritus Professor of Economics, Université de Montréal, Fellow of CIRANO and the C.D. Howe Institute.

[www.cirano.qc.ca/~boyerm](http://www.cirano.qc.ca/~boyerm)

<sup>‡</sup> Visiting Scholar in Economics, Université de Montréal.

## 1. Introduction

A cartel is a group of independent firms which collectively agree to coordinate their supply, pricing or other marketing policies in order to make larger profits than they would when "natural competition" prevails. Depending on the market of interest, the natural competition can be pure and perfect, oligopolistic or monopolistic. The price that would prevail absent the cartel conspiracy is called the "but-for price", and the amount charged by the cartel in excess of the but-for price is the "cartel overcharge". For a cartel to be successful, the cumulated overcharges paid by consumers who are able and willing to pay the higher price must be large enough to cover the losses due to the reduction in sales following the price increase. Beside the welfare loss incurred by the latter consumers, the surplus that accrue to consumers who are unable or unwilling to pay the higher price is a deadweight loss for society.

Most advanced economies consider cartels as illegal and are endowed with legislations aimed at deterring their formation. The United States Sentencing Guidelines (USSG) recommends a base fine of 10% of the affected volume of commerce to a firm convicted of cartel collusion, plus another 10% for the harms "*inflicted upon consumers who are unable or for other reasons do not buy the product at the higher price*". This yields a fine of 20% that may undergo some adjustments for aggravating and mitigating factors. The total financial fine ranges from 15% to 80% of affected sales. Moreover, there is a possibility of incarceration for the individuals involved in the collusion. Between 2006 and 2009, the prosecution of cartels resulted in over \$2 billion in criminal fines and more than 162 years in jail time (Kearney and Katz, 2009).

In the European Union, the antitrust policy is implemented by the European Competition Commission. The amount of the fine takes into account the severity of the damages inflicted upon consumers as well as some aggravating and mitigating factors, but the total fine must not exceed 10% of the overall turnover or global sales of the firm. Over the period 2005-2009, 33 cartels were fined for a total of close to 10 billion, of which the highest fine in history, nearly €1.4 billion, was imposed on four glass producers for price fixing and illegal supply strategies aimed at market sharing.

In Canada, penalties for "agreements between competitors to fix prices, restrict production, or allocate sales, customers or territories" (a *per se* offense since March 2010) may reach \$25 million per count and up to 14 years in prison. But actual penalties have not been close to those maxima yet. As in the U.S., the Canadian Competition Bureau uses a proxy for economic harm of 20% of

the volume of affected commerce to set fines, 10% for the basic overcharge and 10% for other harm including the deadweight loss.<sup>1</sup>

Cohen and Scheffman (1989) argued that an increase of 1% of a price above its natural competition level usually results in a reduction of sales of more than 1%. Based on this, they concluded with respect to USSG guidelines that *"at least in price-fixing cases involving a large volume of commerce, ten percent is almost certainly too high"*. However, Connor and Lande (2008) examined a large number of cartel overcharge estimates from the literature and found an average overcharge estimate in the range 31% to 49% and a median in the range 22% to 25%. Based on this, they concluded that *"the current Sentencing Commission presumption that cartels overcharge on average by 10% is much too low, and the current levels of cartel penalties should be increased significantly"*. A similar study conducted by Connor (2010) concludes that *"...penalty guidelines aimed at optimally deterring cartels ought to be increased"*. Combe and Monnier (2009) performed an analysis of 64 cartels prosecuted by the European Commission and arrived at the conclusion that *"fines imposed against cartels by the European Commission are overall sub optimal"*. In criticizing the Canadian Competition Bureau approach, Kearney (2009) writes: *"The assumption of an average overcharge of 10 percent also has been put into question by economic survey evidence which suggests that the median long-run overcharge is much greater than 10 percent. Research conducted by Professor John Connor indicates 'that the median long-run overcharge for all types of cartels over all time periods is 25.0 percent ...' Accordingly, an assessment of economic harm based on an estimated overcharge of 10 percent is not supported by the empirical evidence."* Hence, there is disagreement among economists about the representative magnitude of cartel overcharges, hence of fines.<sup>2</sup>

The sample used for the present study is an extended version of the one used in Connor (2010).<sup>3</sup> The raw database consists of 1178 cartels, from which we exclude 58 cartels with missing information. This leaves us with a sample of 1120 cartels with overcharge estimates ranging from 0% to 1800%. The mean overcharge estimate is 45.5% on the whole sample and 49% for the subsample of strictly positive estimates. The mean is 20.6% for the cartels with overcharge

---

<sup>1</sup> The harm caused to society is equal to the cartel overcharge plus additional social costs, namely the deadweight losses and the resources devoted to antitrust authorities for fighting cartels, plus other effects such as the impact on investment and employment, on entry and exit dynamics, on innovation and learning curve, etc.

<sup>2</sup> There is also disagreement with respect to the fining rule. One such rule is given by the so-called Becker-Landes formula, which prescribes that the optimal fine is equal to the harm caused by cartels to society divided by the probability of detection (See Becker 1968 and Landes 1983). See also Allain et alii (2011a) and Allain et alii (2011b) for discussions of cartel fining in a dynamic context.

<sup>3</sup> We sincerely thank Professor John Connor for generously making his database available to us.

estimates lying strictly between 0% and 50%, representing 70% of the sample. Estimates that are larger than or equal to 50% represent 22.6% of the sample, and the average overcharge estimate for this subsample is 137.3%. A close look at the data shows that the 49% mean overcharge is actually influenced by a small number of outliers. For example, when the 5% largest observations are left out of the sample, the average overcharge estimate drops from 49% to 32%. If an overcharge of 10% is beyond belief as argued by Cohen and Scheffman (1989), it is reasonable to question whether some estimates in the Connor database overstate the actual overcharges. Our paper contributes to the debate on this point.

Indeed, the methodology of converting a Lerner index into an overcharge was used in some instances by Connor (2010) to obtain overcharge estimates, thereby assuming that the but-for situation is perfect competition. This may significantly overstate the cartel overcharge.<sup>4</sup> Based on this result, we suspect other estimation methods to be potentially biased. One way to verify this presumption is to perform a meta-analysis, as in Connor and Bolotova (2006). Two groups of regressors are needed for such a meta-analysis. The first group is comprised of variables that may influence the size of the true overcharge (e.g. the characteristics of the cartel) while the second group gathers subjective factors that capture the estimation bias (e.g. the computation method and the publication source). Connor and Bolotova (2006) provide evidence that part of the variability of overcharge estimates is due to the estimation method and publication source.

We conduct a new meta-analysis in the spirit of Connor and Bolotova (2006) but addressing three important issues raised by the raw overcharge estimates data. The *first* issue is related to the representativity of sample statistics that are distorted by a few highly influential observations. We address this issue by removing the outliers from the sample used for model estimation<sup>5</sup> while including them back in the sample used for prediction. The overall data trimming is controlled in the estimation process by using a Heckman-type correction for sample selection. The *second* issue is the residual heterogeneity in the subsample of “representative cartels” included in the regressions. We address this problem by using a K-means analysis to separate the sample into four “homogenous” clusters and we use an indicator variable in interaction with other variables in the regressions so as to obtain cluster-specific bias correction formulas. The *third* issue is the asymmetry of the distribution of the overcharge estimates. We mitigate this problem by modeling the logarithm of overcharges as a linear function of the explanatory variables. The log-linear specification is more robust to distortions caused by influential observations and it explains the variance of the overcharge estimates to a greater extent than the standard linear specification.

---

<sup>4</sup> Product differentiation and other industry-specific factors may generate market power that allows firms to apply absent collusion significant mark-ups over marginal cost. We show in Appendix A that not accounting for such mark-ups causes overcharge estimates obtained by conversion of Lerner indices to overstate the overcharge by potentially large margins.

<sup>5</sup> The reasons to exclude the 0 raw overcharge observations will be discussed later in subsection 5.1.

Our empirical results show that the bias captured by the estimation method and the publication source is substantial and economically significant. The log-linear meta-analysis model is used to produce two types of predictions. The first prediction is the bias-corrected overcharge estimates for the representative cartels, with raw estimates lying between 0% and 50%. For this subgroup, we find a median of 13.63% and an average of 13.62%. The second prediction is the bias-corrected overcharge estimates for outlying cartels, with raw estimates lying outside the open interval (0%-50%). After merging the predictions for this subsample with those for the representative cartels, we find a median of 14.05% and a mean of 17.52% for all types of cartels.

The rest of the paper is organized as follows. In Section 2, we briefly review the overcharge estimation methods commonly used in the literature. Section 3 presents the raw data and discusses its limitations. Section 4 presents our meta-analysis models. Section 5 presents our treatment of heterogeneity, outliers, and publication bias. Section 6 presents the bias-correction procedure and the main empirical results. Section 7 concludes. Appendix A illustrates the danger of converting Lerner indices into overcharge estimates by ignoring the competitive mark-ups while Appendices B and C present some summary tables.

## 2. The Overcharge Estimation Methods

Let  $\tilde{p}$  be the price imposed by the cartel and  $p$  be the but-for price, i.e. the price that would prevail absent the cartel. The cartel overcharge expressed as a percentage of the but-for price is given by  $\delta = (\tilde{p} - p)/p$ . While the cartel price  $\tilde{p}$  is observed, the but-for price  $p$  needs to be estimated. Many authors acknowledge that overcharges represent the bulk of the damages caused by cartels. This explains why the largest body of the economic literature on cartels is devoted to the price effects. The methods often used by academic researchers and forensic economists to estimate the cartel overcharge can be summarized into five groups:

***The "Before-and-after" Method:*** This method is based on the comparison of the price during the alleged cartel period with the price before and/or after the cartel. It has the main drawback that it tends not to control for external factors. For that reason, Connor (2010) warns that "*Shifts in buyer preferences, appearance or the disappearance of substitutes, or changes in the cost of production of the cartelized product during the affected period can cause overstatement or understatement of the overcharge.*" This lack of robustness is also pointed out by Finkelstein and Levenback (1983): "*[...] This estimate, however, meets the immediate objection that it is likely to be incorrect because changes in factors affecting price other than the conspiracy would have produced changes in competitive prices if there had been competition during the conspiracy period.*"

**The "Price during a price war" Method:** This method uses the price during a price war or lack of collusion to proxy the but-for price. This method is basically an instance of the before-and-after method and thus suffers from the same limitations.

**The Yardstick Method:** This method compares the prices during the conspiracy period with comparable or yardstick, assumed competitive, firms, product or markets. The yardstick method should be used with caution because an increase in price due to domestic market cartelization can cause a partial demand shift toward nearby markets. Similar domestic firms that are not participating in the collusion will tend to follow the cartel price (umbrella effect).

**The " Cost-Based" Method:** This method is based on the observation that changes in price should reflect changes in costs. The direct way to apply this method is to estimate the production costs by using the (accounting) information on firms involved in the cartel. In the lysine cartel case for example, prosecutors introduced confidential production and sales records of ADM as exhibits, now publicly available (See Connor, 2001). But economic experts do not usually have access to such confidential information. Typically, the overcharge is thus approximated by subtracting a "reasonable margin" from the actual cartel profit and dividing by the production volume.

**"Econometric" Methods:** This denomination gathers all methods using more or less sophisticated econometric models to assess the but-for price. Econometric methods can be used to simulate an oligopolistic competition (Cournot, Bertrand), to predict the Lerner index of market power or to estimate a demand and cost functions that account for dynamic market conditions. See for example Appelbaum (1979) and Froeb, Koyak and Werden (1993).

Any reasonable estimation method is expected to produce reasonable estimates when used wisely. In this paper, one of the goals is to assess the bias associated with estimating average overcharges using the raw estimates.

### **3. The Characteristics of the Connor Database**

The Connor sample consists of estimates previously published by different analysts and researchers. Therefore, it is subject to model errors, estimation errors and sample selection. Model errors come from inconsistencies between the nature of the estimation problem and the choice of estimation method; estimation error stems from the randomness of the data used as input while sample selection refers to the extent to which the sample used is representative of the whole population. In the Connor database, the sample selection problem takes the form of the publication bias due to the fact that "...editorial reviewers have a substantial preference for studies with statistically significant results" (Hunter and Schmidt, 2004).

Connor (2010) found that the long run median overcharge for all types of cartels over all periods is 23.3% while the mean overcharge for successful cartels is 50.4%. In the sample used for the

current study, the median and the mean are respectively 23% and 45.5% for all cartels (25.0% and 49.0% for cartels with strictly positive estimates). The difference between means and medians is due to the skewness of the distribution. Figures 1 and 2 confirm that the large magnitude of the empirical mean is due to a few number of outliers: roughly 1% of overcharge estimates are larger than 400% and 22.6% are larger than or equal to 50%. These outliers should be treated carefully when using econometric methods that are sensitive to their presence (OLS regressions).

Figure 1  
 Overcharge estimates: Distribution skewed to the right.  
 Overcharges larger than 400% (1% of the sample) are not shown on this figure.

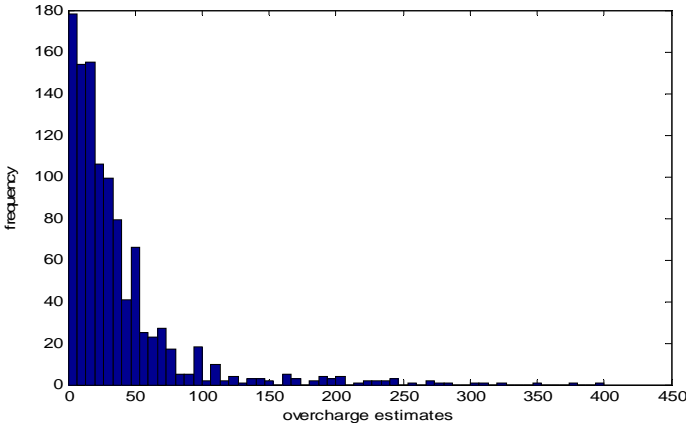
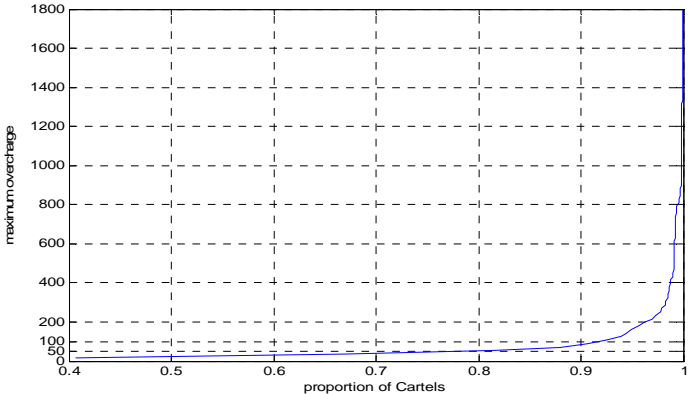


Figure 2  
 Proportion of cartels with overcharge estimates smaller than y,  
 where y is given by the vertical axis.



The raw overcharge data (Table 1) are quite heterogenous across regions, scope (domestic vs international), and time periods.

Table 1  
Mean and median overcharge estimates (OE) per location and types of cartels.  
The prop.% are fractions of the total Connor sample (1120 cartels).

		All Cartels	OE>0%	0%<OE<50%	OE≥50%	Cartels Before 1973	Cartels After 1973
All locations	Mean	45.47	49.01	20.61	137.26	61.98	38.89
	Median	23.00	25.00	18.38	74.00	29.00	21.40
	prop.	100.00	92.77	70.18	22.59	28.50	71.50
US	Mean	38.15	42.03	19.69	126.14	47.79	33.58
	Median	20.50	23.50	17.50	70.20	30.50	16.80
	prop.	30.00	27.23	21.52	5.71	9.64	20.36
EU	Mean	42.65	45.57	19.19	113.65	43.83	41.86
	Median	23.00	25.00	16.10	75.00	24.75	20.40
	prop.	33.48	31.34	22.59	8.75	13.39	20.09
Domestic	Mean	33.60	36.91	18.66	141.31	35.42	32.79
	Median	17.05	19.00	16.45	71.00	20.50	16.45
	prop.	46.79	42.59	36.25	6.34	14.46	32.32
International	Mean	55.90	59.28	22.70	135.68	89.38	43.93
	Median	30.00	31.88	22.00	74.45	37.00	27.50
	prop.	53.21	50.18	33.93	16.25	14.02	39.20

Such heterogeneity raises aggregation problems. Indeed, the average overcharge obtained for the whole sample is meaningful only if the conditions that determine the but-for price are the same across time and markets. Levenstein and Suslow (2003) illustrates this point: *"The reported price increases vary widely by industry and by source. At the low end, for example, we have a reported price increase of ten percent for the thermal fax paper cartel, which was formed as the industry was declining and lasted for less than a year. At the high end there is the stainless steel cartel, which reportedly almost doubled prices. This cartel lasted slightly more than one year (from January 1994 to March 1995) and involved six European steel companies."*

A number of overcharge estimates used in Connor (2010) have been obtained from the conversion of Lerner indices into overcharges. These conversions are usually done assuming that the situation that would prevail absent the cartel conspiracy is pure and perfect competition. However, the natural competition for some markets might be oligopolistic. Morrison (1990, 1993) claims that oligopolistic competition mark-ups typically range between 10% and 30%. Appendix A illustrate the danger of converting a Lerner index into an overcharge by ignoring the presence of such mark-ups.

## 4. A Meta-Analysis to Bias-Correct Overcharge Estimates

Each estimation methodology is susceptible of introducing a particular type of bias. One way to size up the bias contaminating overcharge estimates is to build an econometric meta-analysis model of overcharges. Meta-analyses are used in experimental fields to summarize the findings of studies on a particular topic. They may also be used to verify if the conditions of experiment impact their results. In his preface to the second edition of Hunter and Schmidt (2004), F. Schmidt wrote: “[...] *In our view, the purpose of meta-analysis is to estimate what the results would have been had all the studies been conducted without methodological limitations and flaws*”. The meta-analysis conducted in this section is consistent with this statement.

It is reasonable to expect that the true overcharge depends on the conspiracy period, the duration of the cartel, the characteristics of the firm involved in the collusion and other similar factors. However, we do not observe the true overcharge. Instead, we observe an estimate which is equal to the actual overcharge plus a bias, positive or negative. Hence in addition to factors that affect the true overcharge, we can expect the estimate to be sensitive to subjective factors that cause the bias, namely the estimation method, the source of publication and other factors “posterior” to the occurrence of the conspiracy period. Formally, the bias is defined as the influence of factors that affect the overcharge estimates, but not the true overcharge values themselves. To bias-correct the estimates, we build a meta-analysis model that relates overcharge estimates to both types of factors.

### 4.1. A Linear Meta-Regression Analysis of Cartel Overcharges

Assume that the true overcharge  $\theta_i$  is linearly related to its determinants, gathered in a matrix of regressors  $Y$ :

$$\theta_i = \alpha + Y_i\phi + u_i \tag{1}$$

where  $u_i$  is an error term such that  $E(u_i) = E(u_i Y_i) = 0$ . Examples of variables to include in  $Y$  are duration of the cartel, its organizational characteristics, its scope (domestic vs. international), the conspiracy period, the industry characteristics (e.g, concentration, the elasticity of demand, etc).

Appendix A shows that the conversion of a Lerner index into an overcharge generates a multiplicative bias. Hence in general, we can expect the estimated overcharge to take the form:

$$X_i = (1 + Z_i\lambda)\theta_i \tag{2}$$

where  $X_i$  is the overcharge estimate from study  $i$ ,  $Z_i$  is the set of variables that explain the size of the bias. This amounts to say that the overcharge is inflated or deflated by a factor  $Z_i\lambda$ , i.e. a linear

combination of the  $Z$  variables. Example of variables to include in  $Z$  are the estimation method, the publication source, the presumption of collusion when computing the estimator, etc. By substituting for  $\theta_i$  into the expression of  $X_i$ , we obtain:

$$X_i = \alpha + Y_i\beta + Z_i\gamma + Y_iZ_i\delta + \varepsilon_i, \quad (3)$$

where  $Y_iZ_i$  is the set of interaction variables,  $\beta = \alpha\phi$ ,  $\gamma = \alpha\lambda$ ,  $\delta = \lambda\phi$  and  $\varepsilon_i = (Z_i\lambda + 1)u_i$ . If the number of  $Y$  and  $Z$  variables are large, this may result in a huge number of interaction variables that would cause the model to lack parsimony. To mitigate this drawback, we perform a prior cluster analysis from which we infer four homogenous clusters. We use the indicators of the cluster as summary of the  $Y$  variables and use them in interaction with the  $Z$  variables. Thus, the model that we estimate is:

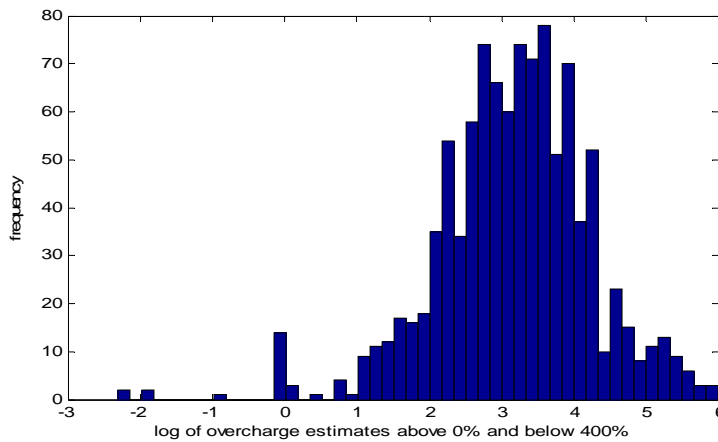
$$X_i = \alpha + Y_i\beta + \sum_{k=1}^4 Z_{\cdot,i}^{(k)} \gamma_k + \varepsilon_i \quad (4)$$

where  $Z_{\cdot,i}^{(k)} = (Z_{1,i}^{(k)}, \dots, Z_{K,i}^{(k)})$ ,  $Z_{j,i}^{(k)} = Z_{j,i}$  if the cartel  $i$  belongs to cluster  $k$ , and  $Z_{j,i}^{(k)} = 0$  otherwise. This model presumes that the parameters of the bias-correction factors vary across clusters; as we will see, they do.

#### 4.2. A Log-linear Meta-Analysis of Cartel Overcharges

The distribution of the logarithm of estimated overcharges is close to symmetry and less subject to distortion caused by extreme values, as shown in Figure 3.

Figure 3  
Log of overcharge estimates that are above zero and below 400% (91.6% of the sample).



Indeed, the log-transformation shifts part of the right skewness into left skewness. Thus, as an alternative to the linear model, we consider the log-linear specification:

$$\log X_i = \alpha + Y_i\beta + \sum_{k=1}^4 Z_{\cdot,i}^{(k)} \gamma_k + \varepsilon_i \quad (5)$$

This model implies that  $X_i = \exp(\alpha + Y_i\beta + \varepsilon_i)\exp\left(\sum_{k=1}^4 Z_i^{(k)}\gamma_k\right)$ . If the true overcharge is given by  $\theta_i = \exp(\alpha + Y_i\beta + \varepsilon_i)$ , then the bias is given by:

$$\text{Bias}(\theta_i) \equiv X_i - \theta_i = \exp(\alpha + Y_i\beta + \varepsilon_i) \left[ \exp\left(\sum_{k=1}^4 Z_i^{(k)}\gamma_k\right) - 1 \right] \quad (6)$$

This formula allows the bias to be positive, null, or negative depending on whether  $\exp\left(\sum_{k=1}^4 Z_i^{(k)}\gamma_k\right)$  is larger than, equal to, or smaller than 1. In the sequel, the estimation results of the two model specifications are compared.

#### 4.3. An Example of Meta-Analysis: Connor and Bolotova (2006)

Connor and Bolotova (2006) performed a meta-analysis of cartel overcharge estimates, in which they used the following variables:

- $Y_1$  – Duration: equals 1 if duration is less than 5 years, 2 if duration is from 6 to 10 years, 3 if duration is from 11 to 15 years, and 4 if duration is 16 years or more.
- $Y_2$ - Scope: equals 1 if domestic and 0 if international.
- $Y_3$ - Bid rigging: equals 1 if Yes and 0 if No.
- $Y_4$ - Found guilty or pleads guilty: equals 1 if Yes and 0 if No.
- $Y_5$ - Geographic market: dummy variables for US, EU, ASIA, ROW including Latin America, and WORLD cartels that cannot be associated to a head region.
- $Y_6$ - Antitrust law regime: dummy variables for P1 (1770-1890), P2 (1891-1919), P3 (1920-1945), P4 (1945-1973), P5 (1974-1990), and P6 (1991-2004).
- $Z_1$ - Overcharge estimation method: dummy variables for Price before conspiracy, Price war, Price after conspiracy, Yardstick, Cost based or normal profit, Econometric modeling, Historical case study with no method specified, and Other.
- $Z_2$ - Type of publication: dummy variables for Peer reviewed journal, Chapters in a book, Monograph or book, Government report, Court or antitrust authority source, Working paper, and Speech or conference presentation.

With this notation, the model estimated by Connor and Bolotova (2006) is:

$$X_i = \alpha + \sum_i Y_i\beta_i + \sum_i Z_i\gamma + \varepsilon_i \quad (7)$$

They estimated different restrictions of the full model. Here we consider the estimation results for the full model (column [7] of Table 6 in their paper). They found that the estimated overcharge is positively related to the duration, but does not depend on whether the firm is "guilty" or not; it is lower for domestic cartels and for cartels that have operated in the EU; it is neither higher nor

lower for bid-rigging cases;<sup>6</sup> and that the size of overcharges has declined over time. The authors attribute the later to the increased severity of antitrust regulation. However, increased competition and free trade may produce similar effects.

Interestingly, they found that the  $Z$  variables also have significant impacts on overcharge estimates. For example, they found “yardstick” estimation method to produce estimates that are at least 10% higher than the “after the conspiracy” method. For the publication sources, they found that “government reports” and “court reports” produce estimates that are respectively 22% lower and 15% higher than “monograph or book”. The fact that the  $Z$  variables show significant effects in the regression makes it difficult to interpret the 23.3% median overcharge and the 50.4% average overcharge of Connor (2010) as meaningful indicators of the representative behavior of “*all types of cartels over all periods of time*”.

## 5. Treatment of Heterogeneity, Outliers and Sample Selection

By definition, outliers represent a relatively small proportion of highly influential observations and should be treated with caution in linear regressions. Hunter and Schmidt (2004, page 196) warns: “*The use of least squares statistical methods [...] is based on the assumption that the data contain no aberrant values (i.e. outliers). When this assumption does not hold, the statistically optimal properties (efficiency and unbiasedness) of least squares estimates disappear. Under these circumstances, least squares estimates become very inaccurate because of their extreme sensitivity to outliers.*” Hidden forms of heterogeneity affecting a relatively large proportion of observations also need a careful treatment. Finally, when outliers are removed from the regressions, a Heckman-type correction for the resulting sample selection bias is mandatory.

### 5.1 Avoiding the Distorsion Caused by Influential Observations

The presence of influential observations in a sample can be due to the fat-tailedness of the distribution that generates the data. For example, stable distributions display frequent extreme values compared to normal distributions. Influential observations can also be due to heterogeneity in the sample. For illustration, consider observations that are generated by two normal distributions, one with mean  $\mu_1$  and variance  $\sigma^2$ , and the other with mean  $\mu_2$  and variance  $\sigma^2$ . Let  $\pi_1 = 1 - \varepsilon$  be the probability that an arbitrary observation comes from the first distribution and  $\pi_2 = \varepsilon$  denote the complementary probability. For quite small  $\varepsilon$ , the number of observations coming from the second distribution can be as small as one in a large sample. Without loss of generality, assume that we have a sample with only the first observation coming from the second distribution, and that an econometrician who ignores the heterogeneity runs the following regression:

---

<sup>6</sup> Contrary to what is claimed by Cohen and Scheffman (1989).

$$X_i = \alpha + \varepsilon_i \quad (8)$$

Then the OLS estimator of  $\alpha$  is the sample mean  $\hat{\alpha} = \frac{1}{N} \sum_{i=1}^N X_i$ .

To measure the impact of the unique observation generated by the second distribution, we write:

$$\hat{\alpha} = \frac{1}{N} X_1 + \frac{N-1}{N} \left( \frac{1}{N-1} \sum_{i=2}^N X_i \right) \quad (9)$$

The expectation of  $\hat{\alpha}$  is:

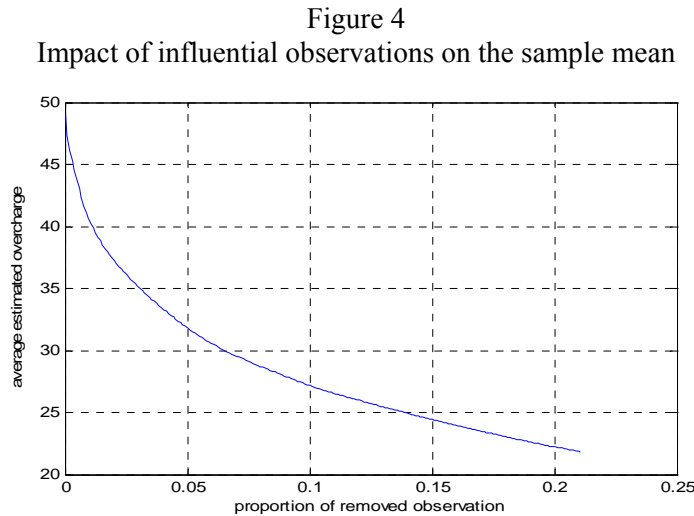
$$E(\hat{\alpha}) = \frac{1}{N} \mu_2 + \frac{N-1}{N} \mu_1 \quad (10)$$

First, we see that  $E(\hat{\alpha})$  lies strictly between  $\mu_1$  and  $\mu_2$  so that  $\hat{\alpha}$  is biased for both  $\mu_1$  and  $\mu_2$ . Second, assuming that  $\mu_2 = K\mu_1$  for quite large  $K$ , we have  $E(\hat{\alpha}) = \frac{K+N-1}{N} \mu_1$ . If the aim of the study is to draw conclusions that are applicable to the majority of the population, then the econometrician should design an estimator that targets  $\mu_1$  as precisely as possible. But the bias of  $\hat{\alpha}$  for  $\mu_1$  is given by:

$$E(\hat{\alpha}) - \mu_1 = \frac{K-1}{N} \mu_1 \quad (11)$$

This bias is increasing in both  $K$  and  $\mu_1$  and can be substantial in many cases of practical interest.

Similar results hold for estimated coefficients of a regression. As an empirical illustration, we computed the empirical mean of the strictly positive overcharge estimates in the Connor sample by sequentially leaving *the largest observation* out: as shown on Figure 4, the empirical mean drops from 50% to approximately 32% when the 5% most influential observations are removed from the sample, and to approximately 22% when the 20% most influential observations are left out.



To address the issue of outliers, we remove in this study overcharge estimates that are larger than or equal to 50% from the sample used for the regression (22.6% of the sample). Zero overcharge

estimates are also removed from the sample for four reasons. First, these estimates represent 7.23% of the sample, suggesting that the probability of an overcharge *estimate* taking exactly the value zero is not equal to zero.<sup>7</sup> Second, our preferred empirical framework consists of regressing log-overcharge estimates on a number of explanatory variables. Hence zero overcharge estimates must be discarded for technical reasons (log-of-zero problem). Third, there are reasons to expect that the true underlying overcharge estimates are not all exactly equal to zero.<sup>8</sup> And fourth, even if the zero overcharge estimates that are present in the sample are unbiased, their empirical proportion is highly susceptible to be distorted by the publication bias. Note however that zeros and outliers are not excluded from our analysis, as they are included in the sample used for predicting bias-corrected overcharges.

## 5.2. Treatment of the residual heterogeneity in the sample

Removing 0% and  $\geq 50\%$  overcharge estimates does not provide a complete solution to the heterogeneity problem. We control for the residual heterogeneity through a K-means analysis, a nonparametric cluster analysis aimed at partitioning a sample of observations into  $K$  groups minimizing the within groups heterogeneity while maximizing the between groups one. A typical K-means algorithm starts with a random draw of  $K$  arbitrary points  $(c_1, \dots, c_K)$  from the sample, called the centroids. First, each observation is allocated to the centroid most similar or closest to it according to the Euclidean distance:

$$x_i \in G_k \Leftrightarrow k = \underset{1 \leq i \leq K}{\operatorname{argmin}} \{ \|x_i - c_i\|^2 \}, i = 1, \dots, n \quad (12)$$

This step partitions the sample  $(x_1, \dots, x_n)$  into  $K$  initial groups  $G_1, \dots, G_K$ . Second, one replaces the previous centroids by the average of the groups, that is:

$$c_k \leftarrow \frac{1}{n_k} \sum_{x_i \in G_k} x_i, k = 1, 2, \dots, K \quad (13)$$

where  $n_k$  is the number of observations in the group  $G_k$ . Third, one iterates the previous two steps until the groups are stable or converge. When this is achieved, one obtains a decomposition of the

---

<sup>7</sup> A truncated distribution would be needed to fit the portion of the sample consisting of strictly positive overcharges, given that the true distribution admits a mass of probability at zero.

<sup>8</sup> In practice, zero overcharge estimates may arise either from censoring (unreported negative estimates that are set to zero) or from a mass of probability (a strictly positive proportion of unsuccessful cartels). A zero overcharge estimate belonging to the first category is likely to be biased. An example of a negative estimate is discussed by Finkelstein and Levenbach (1983), citing an unpublished 1979 study “Economic Analysis of Potential Damages” by E.W. Erickson & W. Henry on the overcharge in an alleged broiler chicken conspiracy: “Unfortunately for the plaintiffs, the study showed a statistically significant *negative* coefficient of \$1.36 per pound for the dummy variable, indicating that prices for broilers in the price-fixing period were about 5% *lower* than would have been expected on the basis of the supply and demand factors included in the regression equation model.”

original sample into  $K$  groups such that the within group variance is minimized and the between group variance is maximized, as the sum of the two gives the total variance.

We segregate the cartels into four clusters from overcharge estimates and the  $Y$  variables of the meta-analysis. It turns out that the resulting ranges of overcharge estimates across clusters do not overlap, which reflects the fact that heterogeneity in the sample mainly comes from the distribution of overcharge estimates. The clusters are made respectively of 215 cartels with overcharge estimates between 0.1% and 11.1%, 234 cartels with overcharge estimates between 11.4% and 21.6%, 197 cartels with overcharge estimates between 21.7% and 33.7%, and finally 140 cartels whose overcharge estimates between 33.9% and 49.9%.

### 5.3. Controlling for the Sample Selection Bias

Leaving out roughly 30% of cartels<sup>9</sup> for the model estimation may cause some loss of generality in the results. Indeed, the removal of those cartels raises a sample selection problem. Heckman (1979) shows that the bias induced by sample selection on linear regression coefficients is similar to a missing regressor problem. Indeed, let us assume there exists a latent index  $X_i^*$  indicating the quality of the data. In the present context, the quality of an observation is defined in relation with its representativity and its contribution to the quality of the estimation results. For example, very large positive overcharge estimates are not representative of the average cartel. Suppose that the quality indicator  $X_i^*$  takes the following value for cartel  $i$ :

$$X_i^* = A + Y_i B + Z_i C + u_i^* \quad (14)$$

where  $u_i^*$  follows a standard normal distribution. Assume that this latent variable is such that the cartel  $i$  is included in the meta-analysis if and only if  $X_i^* > 0$ , while the cartel is excluded otherwise.<sup>10</sup> By definition, Equation (7) is estimated only on the portion of the sample where  $X_i^* > 0$ . The expectation of the overcharge estimates on this subsample is:

$$E(X_i | Y, Z, X_i^* > 0) = \alpha + Y_i \beta + Z_i \gamma + E(\varepsilon_i | X_i^* > 0) \quad (15)$$

Hence, the sample selection bias may be controlled by estimating the following equation, which is Equation (7) augmented with the term  $E(\varepsilon_i | X_i^* > 0)$ . We have:

$$X_i = \alpha + Y_i \beta + Z_i \gamma + E(\varepsilon_i | X_i^* > 0) + e_i, \quad (16)$$

---

<sup>9</sup> In total, 22.59% of the observations have overcharge estimates  $\geq 50\%$  and 7.23% have overcharge estimates of 0%.

<sup>10</sup> This model is consistent with a situation where the relationship between  $X_i$  and  $X_i^*$  is nonlinear and U-shaped:  $X_i^* \geq b$  for zero overcharge estimates,  $X_i^* < b$  for estimates lying strictly between 0% and 50% and  $X_i^* \geq b$  the for observations above 50%.

where  $e_i \equiv \varepsilon_i - E(\varepsilon_i|X_i^* > 0)$  has zero expectation by construction. It can be shown that:<sup>11</sup>

$$E(\varepsilon_i|X_i^* > 0) = \theta \frac{\varphi(A+Y_iB+Z_iC)}{\Phi(A+Y_iB+Z_iC)} \quad (17)$$

where the theoretical value of  $\theta$  is  $\text{Corr}(\varepsilon_i, u_i)\widehat{\sigma}_\varepsilon^2$  and  $\varphi$  and  $\Phi$  are the standard normal density and cumulative distributions respectively. Similarly, we have:

$$E(\varepsilon_i|X_i^* < 0) = -\theta \frac{\varphi(A+Y_iB+Z_iC)}{1-\Phi(A+Y_iB+Z_iC)} \quad (18)$$

Let us consider the inverse Mills ratio (IMR) defined as:

$$\text{imr}_i = \frac{\varphi(A+Y_iB+Z_iC)}{\Phi(A+Y_iB+Z_iC)} \quad \text{if } X_i^* > 0 \quad (19)$$

$$\text{imr}_i = \frac{-\varphi(A+Y_iB+Z_iC)}{1-\Phi(A+Y_iB+Z_iC)} \quad \text{if } X_i^* < 0 \quad (20)$$

When restricted to the subsample of cartels defined by  $X_i^* > 0$ , Equation (16) is equivalent to:

$$X_i = \alpha + Y_i\beta + \sum_{k=1}^4 Z_i^{(k)} \gamma_k + \text{imr}_i\theta + e_i, \quad (21)$$

A similar methodology applies to the log-linear model.

A probit can be estimated over the whole sample to infer fitted values of  $X_i^*$  and the IMR. We do not observe the latent variable, but we do know if an observation is excluded ( $I_i=0$ ) or not ( $I_i=1$ ). If observation  $i$  is included, then it must be the case that  $X_i^* > 0$ , which in turn implies that  $A + Y_iB + Z_iC + u_i^* > 0$  or equivalently  $-u_i^* < A + Y_iB + Z_iC$ . Because  $u_i^*$  is standard normal, the likelihood of this observation is:

$$L_i = \Phi(A + Y_iB + Z_iC) \quad (22)$$

Likewise, if observation  $i$  is excluded, then it must be the case that  $X_i^* < 0$ , which in turn implies that  $-u_i^* > A + Y_iB + Z_iC$ . Hence the likelihood of an excluded observation is given by  $1 - L_i$ . The sample log-likelihood of the probit is thus given by:

$$\mathcal{L}(A, B, C) = \sum_{i=1}^T \{I_i * \log L_i + (1 - I_i) * \log(1 - L_i)\} \quad (23)$$

Maximizing this log-likelihood with respect to parameters  $A$ ,  $B$  and  $C$  gives the probit estimators. Table 2 presents the probit estimates  $\widehat{A}$ ,  $\widehat{B}$  and  $\widehat{C}$ . Such a probit analysis is interesting per se because it permits to see characteristics of excluded cartels since variables with positive [negative] coefficients are correlated with inclusion [exclusion].

---

<sup>11</sup> See Heckman (1979) for details.

Table 2  
A probit model for the inclusion of cartels into the meta-analysis

	Coefficients	Student-t <sup>12</sup>
Constant	0.03	0.10
Duration	-0.01	-0.22
Domestic	0.61	5.29
BidRig	0.49	3.65
Guilty	0.15	1.29
US	-0.41	-2.83
EU	-0.15	-1.19
ASIA	-0.56	-2.86
ROW	0.29	1.02
P1	0.44	0.83
P3	0.14	0.69
P4	0.22	1.04
P5	0.02	0.10
P6	0.28	1.37
OTHER	-0.77	-3.75
HISTOR	-1.40	-4.01
PBEFOR	-0.24	-1.55
PWAR	-0.63	-2.02
PAFTER	-0.29	-1.66
COST	-0.48	-2.14
YARDST	-0.44	-2.54
LEGAL	0.09	0.46
JOURNAL	0.25	1.31
MONOGR	0.37	2.08
EDBOOK	0.15	0.61
COURT	0.35	1.73
WORKP	0.81	3.76
SPEECH	1.16	1.71
Sample size <sup>13</sup>	1119	

We see that international cartels have been excluded more than domestic cartels as the coefficient the “Domestic” dummy variable is positive and significant.<sup>14</sup> Cartels involved in bid-rigging have been excluded less than those involved in other price fixing collusions. Cartels that pleaded guilty have been excluded less than those that did not. Cartels of the ROW have been excluded less than those of the benchmark category WORLD, but the difference is not statistically significant. The WORLD cartels have been excluded less than those of the US, EU and ASIA groups. Finally,

<sup>12</sup> A coefficient is significant in a one-sided 10% level test if the corresponding student-t is larger than 1.28 in absolute value.

<sup>13</sup> In the original database, there is one overcharge estimate with publication source labelled as “OTHER”. This observation has been excluded from the meta-analysis.

<sup>14</sup> We consider a one-sided test and 10% significance level.

cartels of the period P6 have been excluded less than those of periods P1 to P5. Estimates obtained by the econometric method (used as benchmark) have been excluded less than those computed by other methods as the coefficients of all the methods are negative and significant. Finally, estimates published in government report (used as benchmark) have been discarded more than those published in other publication sources.

Let  $\hat{X}_i^* = \hat{A} + Y_i\hat{B} + Z_i\hat{C}$  be the fitted values of  $X_i^*$ . Empirically, the sample selection bias is controlled in the meta-analysis by including the estimated inverse Mills ratio as an additional regressor. Thus, the estimating equation is given by (21), rewritten as:

$$X_i = \alpha + Y_i\beta + Z_i\gamma + \widehat{\text{imr}}_i\theta + e_i. \quad (24)$$

where  $\widehat{\text{imr}}_i = \frac{\varphi(\hat{X}_i^*)}{\Phi(\hat{X}_i^*)}$  for included cartels. The coefficients estimated from (24) using the trimmed sample can be used to bias-correct *all* overcharge estimates.

## 6. Bias-correcting the Raw Overcharge Estimates

We estimate our linear and log-linear model using the same Y and Z variables as in Connor and Bolotova (2006), with and without controlling for sample selection bias.<sup>15</sup> The estimated coefficients for the Y variables are shown in Table 3 while those of the Z variables are shown in Appendix C. The coefficients of the Z variables for clusters 1 and 2 are generally negative and significant while those of clusters 3 and 4 are generally positive and significant. These discrepancies play an important role in explaining the large magnitude of the bias.

Our results differ in two main ways from those of Connor and Bolotova (2006). First, the coefficient of the duration variable is not statistically significant in a conventional 5% level two-sided test. In the log-linear model, this coefficient is significant in a one-sided 10% level test, but is relatively small (0,03%). This suggests that the ability of a cartel of five to ten year duration to raise its price is not too different from that of a cartel of zero to five year duration. Second, the results from the log-linear model suggest that cartels resolved with guilty pleas display significantly higher overcharges than other cartels. Otherwise, our results are qualitatively similar to those of Connor and Bolotova (2006) in terms of the signs of the estimated coefficients. The percentage of explained variance (variance ratio) is 11.16% for the linear model and 13.40% for the log-linear model. Hence, the log-linear model explains the variance of the original estimates to a greater extent and is therefore the better model.

---

<sup>15</sup> Even under no interaction, the coefficients of the Z variables are not directly comparable with the results of Connor and Bolotova (2006) because we have changed the reference variables. Connor and Bolotova (2006) used PAFTER and MONOGR as reference variables for the method of calculation and source of publication, while we use ECON and GOVREP respectively. Also, the modality P6 of the antitrust law regime now covers periods from 1991 to 2010.

Table 3

## Estimation Coefficients for the Y variables

*Mean overcharge 1* is the estimated average bias-corrected (bc) overcharge for the subsample of positive initial estimates below 50%. *Mean overcharge 2* is the same for the whole sample.

		Linear regression		Log-linear regression		
		Coefficients	Student-t	Coefficients	Student-t	
No selection bias control	Constant	17.26	13,80	2,35	16,32	
	Duration	0.06	0,34	0,03	1,49	
	Domestic	-0.51	-1,17	-0,03	-0,67	
	BidRig	-0.37	-0,86	-0,03	-0,58	
	Guilty	0.68	1,53	0,20	3,88	
	US	0.38	0,65	0,02	0,32	
	EU	0.54	1,06	0,01	0,21	
	ASIA	1.17	1,59	0,05	0,64	
	ROW	0.34	0,38	0,10	1,00	
	P1	-1.31	-0,66	-0,01	-0,03	
	P3	-1.19	-1,30	-0,15	-1,45	
	P4	-1.39	-1,54	-0,18	-1,75	
	P5	-0.92	-1,00	-0,11	-1,00	
	P6	-0.56	-0,65	-0,14	-1,45	
		Var(epsilon)	18.15		0,24	
		Mean bc overcharge 1	17.18		12.09	
	Variance ratio	11.16		13,40		
	Sample size		786			
Controlling Selection bias	Constant	16.72	4,60	2,85	6,80	
	Duration	0.05	0,33	0,03	1,56	
	Domestic	-0.32	-0,25	-0,21	-1,41	
	BidRig	-0.25	-0,28	-0,14	-1,38	
	Guilty	0.74	1,24	0,14	2,04	
	US	0.25	0,25	0,14	1,21	
	EU	0.49	0,81	0,06	0,84	
	ASIA	1.00	0,78	0,21	1,40	
	ROW	0.40	0,41	0,05	0,45	
	P1	-1.16	-0,53	-0,14	-0,57	
	P3	-1.16	-1,22	-0,19	-1,71	
	P4	-1.33	-1,35	-0,24	-2,11	
	P5	-0.90	-0,96	-0,13	-1,20	
	P6	-0.46	-0,44	-0,23	-1,91	
		IMR	0.66	0,16	-0,61	-1,25
		Var(epsilon)	18.18		0,24	
	(Mean, median) bc overchg 1	(17.11, 17.19)		(13.62, 13.63)		
	(Mean, median) bc overchg 2	(15.64, 16.87)		(17.52, 14.05)		
	Variance ratio	11.15		15.56		
	Sample size		786			

In a standard Heckman framework, the portion of the sample that is excluded is not observed. In our case, the regressors are observed for the all observations. We trimmed the sample only because of the representativity issue raised by the dependent variable. If the quality of the dependent variable observations were the same for the whole sample, an approach inspired from the treatment effect literature could have been used. In the latter approach, all observations are included in the regression of  $X$  or  $\log X$  onto  $Y$ ,  $Z$  and the IMR, with the formula of the IMR depending on whether the realization of  $X$  lies above 50% or not. The approach that we advocated is halfway. First, the parameters of the models are estimated using a Heckman regression as if the outlying observations were not available. Second, the estimated coefficients are used together with the values of the regressors to predict the dependent variables for the outlying observations.

Bias-corrected estimates are obtained by eliminating the impact of the  $Z$  variables from an initial estimator  $X_i$ . In the linear model, the bias-corrected overcharge estimates are given by:

$$\hat{X}_i = \hat{\alpha} + Y_i \hat{\beta} + \widehat{\text{IMR}}_i \hat{\theta} \quad (25)$$

where  $\widehat{\text{IMR}}_i$  is given by Equation (19) for included cartels and by Equation (20) for excluded cartels.<sup>16</sup> In the log-linear model, the bias-corrected overcharge estimate is given by:

$$\hat{X}_i = \exp(\hat{\alpha} + Y_i \hat{\beta} + \widehat{\text{IMR}}_i \hat{\theta} + \hat{\sigma}_e^2/2), \quad (26)$$

assuming that  $e_i \sim N(0, \sigma_e^2)$ . The contribution of the IMR reflects the amount by which the average over a subsample of interest deviates from the average over the whole sample.

We applied the above bias-correction formulas only to cartels with strictly positive overcharge estimates. Two reasons support this choice. First, the empirical distribution of overcharge estimates exhibits a mass at zero. This means that neither the linear model nor the log-linear model can fit this distribution at zero.<sup>17</sup> Second, the predictions of the log-linear model are necessarily positive, which indicates that this model always overestimates 0% overcharges. Thus, we will use a 0% bias-corrected overcharge for a cartel with an initial overcharge estimate of 0%.<sup>18</sup>

Table 4 replicates Table 1 with bias-corrected overcharge estimates as input. For the subsample with initial estimates lying in the range 0% to 50%, the log-linear model predicts a mean overcharge of 13.62% with a median of 13.63%. For the whole sample, the predicted mean is 17.52% while the median is 14.05%. Note that the median for the whole sample is quite close to the mean and median of the trimmed sample. Mean and median bias-corrected overcharge

---

<sup>16</sup> In a previous version of this paper, we did not use the IMR in the bias-correction formula of excluded cartels. It appears that this is an important omission since excluded cartels are also subject to a bias proportional to (31).

<sup>17</sup> In fact, these models are not rich enough to bias-correct zero overcharges.

<sup>18</sup> This may affect the average bias-corrected overcharge over the whole sample (the ‘‘All cartels’’ column of Table 4), but it does not affect the average of positive estimates.

estimates suggest a more homogenous behaviour of cartels across different types, geographical locations and periods than suggested by the raw data of Table 1, hence that *a cartel is a cartel is a cartel*.

The log-linear model predicts for US cartels a mean overcharge estimate of 13.73% (with a median of 13.75%) for the subsample and 16.45% (with a median of 14.01%) for the whole sample. For EU cartels, the corresponding figures are 13.08% (13.32%) and 17.75% (13.78%). The model predicts moreover that the mean overcharge estimate of international cartels are larger than that of domestic cartels by 0.73 (or 5.5% more profitable) and 4.20 percentage points (or 27.5% more profitable) for the subsample and the whole sample respectively, not considering the potentially different costs of forming and maintaining international versus domestic cartels. Finally, it predicts that post-1973 cartels achieved slightly lower mean but slightly larger median overcharges than pre-1973 cartels.

Table 4  
Bias-corrected mean and median overcharge estimates (OE)  
The prop.% are fractions of the total Connor sample (1120 cartels)

		All Cartels	OE > 0%	0% < OE < 50%	OE ≥ 50%	Cartels Before 1973	Cartels After 1973
All locations	Mean	17,52	18,89	13,62	35,28	17,87	17,39
	Median	14,05	14,35	13,63	34,23	13,45	14,22
	prop.%	100,00	92,77	70,18	22,59	28,50	71,50
US	Mean	16,45	18,13	13,73	34,70	17,92	15,76
	Median	14,01	14,27	13,75	32,52	13,94	14,14
	prop.%	30,00	27,23	21,52	5,71	9,64	20,36
EU	Mean	17,75	18,97	13,08	34,16	17,70	17,79
	Median	13,78	13,90	13,32	32,55	13,40	14,03
	prop.%	33,48	31,34	22,59	8,75	13,39	20,09
Domestic	Mean	15,29	16,80	13,26	37,00	16,86	14,59
	Median	13,40	13,55	13,32	35,80	13,40	13,39
	prop.%	46,79	42,59	36,25	6,34	14,46	32,32
International	Mean	19,49	20,67	13,99	34,61	18,90	19,70
	Median	15,44	15,64	14,03	33,91	13,67	15,52
	prop.%	53,21	50,18	33,93	16,25	14,02	39,20

Table 5 presents *mean* bias-corrected overcharge estimates for different categories of cartels according to whether they are domestic or international, in bid-rigging cases or not, and/or were found or pleaded guilty or not. Table 6 presents *median* bias-corrected overcharge estimates for the same subgroups. The differences between the raw overcharge estimates (Table 1 and the left-hand side of Tables 5 and 6) and the bias-corrected ones (Table 4 and the right-hand side of Tables 5 and

6) are striking, both in terms of levels and in terms of orderings. The medians of the subgroups do not change too much as one moves from the whole sample to the trimmed sample. This suggests that the subsample is quite representative of the whole universe of cartels as long as the median bias-corrected overcharge estimate is given more consideration than the mean. Finally, note that the mean and median overcharge estimates in Tables 5 and 6 are sometimes higher in the subsample than in the whole sample. This stems from the fact that zero estimates are part of the whole sample while they are excluded from the subsample.

Table 5

Raw versus Bias-corrected *mean* cartel overcharge estimates.

The values reported in the last group of columns are the empirical averages of the bias-corrected overcharges predicted by the log-linear model, *controlling for sample selection*.

“n.a.” means that the corresponding category of cartels is not represented in the sample or subsample.

Cartel characteristics			Raw Average Estimates					Bias-Corrected Average Estimates				
domestic	bidrig	guilty	US	EU	ASIA	ROW	WORLD	US	EU	ASIA	ROW	WORLD
<b>Subsample with initial overcharge estimates lying between 0% and 50%</b>												
Yes	Yes	Yes	19,69	15,56	25,65	17,40	18,70	13,61	13,06	13,47	12,77	12,62
No	Yes	No	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Yes	No	No	18,88	17,60	20,49	18,67	n.a.	12,05	11,8	11,46	12,21	n.a.
No	No	Yes	19,27	21,25	17,25	21,21	23,85	15,50	14,42	15,43	15,39	15,03
Yes	Yes	No	11,57	16,67	4,88	14,50	n.a.	11,44	10,68	12,32	11,73	n.a.
No	Yes	Yes	22,20	18,20	25,53	28,73	10,23	14,60	14,07	14,78	16,50	14,41
Yes	No	Yes	18,70	17,77	22,39	11,41	n.a.	15,05	13,95	14,11	13,82	n.a.
No	No	No	29,50	23,29	29,00	n.a.	23,97	11,12	10,94	10,92	n.a.	11,35
<b>Whole sample</b>												
Yes	Yes	Yes	26,73	17,38	29,47	17,40	18,70	15,93	13,06	17,61	12,77	12,62
No	Yes	No	n.a.	430,00	n.a.	n.a.	n.a.	n.a.	28,55	n.a.	n.a.	n.a.
Yes	No	No	46,72	20,15	77,15	17,12	n.a.	16,31	12,21	15,23	11,19	n.a.
No	No	Yes	27,67	52,78	37,60	24,49	71,79	19,28	21,41	21,35	18,32	21,11
Yes	Yes	No	21,18	16,67	4,88	14,50	n.a.	15,68	10,68	12,32	11,73	n.a.
No	Yes	Yes	38,05	32,89	35,40	28,73	20,18	20,44	21,19	19,97	16,50	19,26
Yes	No	Yes	44,51	20,52	51,24	11,41	n.a.	15,63	16,39	18,87	13,82	n.a.
No	No	No	46,56	72,20	29,00	50,00	34,29	16,47	19,80	10,92	23,24	12,56

Table 6  
 Raw versus Bias-corrected *median* cartel overcharge estimates.  
 The values reported in the last group of columns are the empirical medians of the bias-corrected overcharges predicted by the log-linear model, *controlling for sample selection*.  
 “n.a.” means that the corresponding category of cartels is not represented in the sample or subsample.

Cartel characteristics			Raw Median Estimates					Bias-Corrected Median Estimates				
domestic	bidrig	Guilty	US	EU	ASIA	ROW	WORLD	US	EU	ASIA	ROW	WORLD
Subsample with initial overcharge estimates lying between 0% and 50%												
Yes	Yes	Yes	17,40	12,30	28,80	17,40	18,70	13,75	12,59	13,47	12,77	12,62
No	Yes	No	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Yes	No	No	18,25	15,65	19,00	17,50	n.a.	12,03	11,85	11,17	11,95	n.a.
No	No	Yes	10,40	19,23	17,50	20,30	22,30	15,57	14,12	15,64	15,62	15,44
Yes	Yes	No	8,00	17,00	4,83	14,50	n.a.	11,34	10,65	12,90	11,73	n.a.
No	Yes	Yes	20,50	14,00	18,10	30,55	11,90	14,45	14,01	13,93	16,58	13,67
Yes	No	Yes	15,50	15,10	20,00	10,00	n.a.	15,19	13,55	13,90	13,46	n.a.
No	No	No	30,50	23,50	29,00	n.a.	26,25	11,39	10,98	10,92	n.a.	11,66
Whole sample												
Yes	Yes	Yes	18,05	12,15	29,00	17,40	18,70	13,76	12,59	13,47	12,77	12,62
No	Yes	No	n.a.	430,00	n.a.	n.a.	n.a.	n.a.	28,55	n.a.	n.a.	n.a.
Yes	No	No	24,70	13,50	24,50	13,60	n.a.	12,30	11,48	11,47	11,95	n.a.
No	No	Yes	27,80	29,00	17,50	21,65	29,60	16,25	15,40	15,64	15,68	15,86
Yes	Yes	No	14,50	17,00	4,83	14,50	n.a.	11,50	10,65	12,90	11,73	n.a.
No	Yes	Yes	25,75	17,00	29,80	30,55	13,45	15,18	14,65	15,48	16,58	15,00
Yes	No	Yes	14,90	15,95	23,25	10,00	n.a.	15,01	13,76	14,46	13,46	n.a.
No	No	No	36,50	50,00	29,00	50,00	25,00	11,59	25,39	10,92	23,24	11,66

## 7. Conclusion

Our study identified the mean and median overcharges of representative cartels by performing a meta-analysis on an extended version of the database used in Connor (2010). Each observation in the sample is subject to potential model and estimation errors and publication bias. A quick examination of the sample reveals a highly positively skewed distribution, the presence of outliers and a significant amount of heterogeneity. Although few in number, the outliers are quite influential. In order to improve the representativity of the meta-analysis results, we removed outliers for model estimation but included them back for predictions. We controlled for the resulting sample selection bias by augmenting the estimating equations with an inverse Mills ratio inferred from a prior probit estimation (Heckman 1979). We mitigated the asymmetry problem by taking the logarithm of overcharge estimates. Using the trimmed subsample, we regressed the log of overcharge estimates on a number of explanatory variables. A first group of regressors captures the true overcharge while a second group captures potential biases. Moreover, to mitigate the impact of heterogeneity on the results, we used a cluster analysis to segregate the cartels into four homogenous clusters and we interacted the cluster indicators with the variables capturing the biases, yielding a log-linear model in which the coefficients of the bias-correction formulas differ across clusters. The log-linear model produced a better fit compared to the linear model.<sup>19</sup>

Our meta-analysis derives a mean and median bias-corrected overcharge estimate of 13.62% and 13.63% for the subsample of cartels with raw overcharge estimates in the interval (0%, 50%)<sup>20</sup> and a mean and median of 17.52% and 14.05% for the whole sample with raw overcharge estimates in the interval [0%, 1800%]. These results differ significantly from the raw mean and median figures. The comparison of mean and median bias-corrected overcharge estimates reveals also a fairly homogenous behavior of cartels accross different types, geographical locations and periods.

Our results have significant implications for antitrust policy. Indeed, a major element in the prosecution of cartel is their capacity to exert upward pressures on prices. Allain et alii (2011a) have used our bias-corrected overcharge estimates to conduct a study on the optimal fining of cartels by the European Court and found that there is no need to increase the level of fines as the current level are on average optimal.<sup>21</sup> The US and Canadian guidelines are based on the presumption of an average overcharge of 10%, which is close to the mean bias-corrected overcharge of the representative cartel.<sup>22</sup>

---

<sup>19</sup> The variance ratio of explained variance to total variance is 40% higher under the log-linear model.

<sup>20</sup> For the interval [0%, 50%), the mean would fall to 12.35%.

<sup>21</sup> See also Allain et alii (2011b).

<sup>22</sup> Our meta-analysis speaks of mean and median representative overcharges, but of course nothing can replace when available specific data in specific cartel cases.

## Appendix A: Misuses of the Lerner Index

An overcharge calculation approach based on the Lerner index can fall within the family of econometric methods or cost-based methods depending on how this index is estimated. The Lerner index of market power is defined as:

$$L = \frac{p-c}{p} \quad (i)$$

where  $p$  is the market price and  $c$  is the marginal cost (MC). In a perfectly competitive market,  $p = c$  so that  $L = 0$ . If the condition that would prevail in the absence of cartels is perfect competition, then the but-for price is given by  $p = c$ . The Lerner index for the cartelized market is then given by  $L = \frac{\tilde{p}-c}{\tilde{p}}$  and the true overcharge is:

$$\delta = \frac{\tilde{p}-p}{p} = \frac{\tilde{p}-c}{c} \Leftrightarrow \delta = \frac{L}{1-L} \quad (ii)$$

where  $p = c$  is the but-for price in this case. Hence the formula above permits to retrieve the overcharge from the Lerner index if the but-for world is assumed to be characterized by pure and perfect competition.

In a natural competition (real life) context, the price is in general equal to the MC ( $c$ ) plus a margin over MC ( $m$ ). The but-for price is then given by  $p = c + m$ . The Lerner index in the cartelized market is:

$$L \equiv \frac{\tilde{p}-c}{\tilde{p}} = \frac{\tilde{m}}{c+\tilde{m}} \quad (iii)$$

where  $\tilde{p} = c + \tilde{m}$  is the price in the cartelized market. Hence the (true) overcharge by the cartel over the natural competition price becomes:

$$\delta \equiv \frac{\tilde{p}-p}{p} = \frac{\tilde{m}-m}{c+m} \quad (iv)$$

However, the overcharge that would be inferred from the Lerner index (wrongly) assuming perfect competition as benchmark is:

$$\tilde{\delta} \equiv \frac{L}{1-L} = \frac{\tilde{m}}{c} = \delta + \frac{m}{c}(\delta + 1) \quad (v)$$

Typically, oligopolistic competition mark-ups range between 10% and 30%, and sometimes more (Morrison 1990, 1993). If the true overcharge is  $\delta = 10\%$  and  $\frac{m}{c} = 20\%$ , then the Lerner index delivers the biased estimate  $\tilde{\delta} = 32\%$ , i.e. more than three times the true value. Note that the bias is increasing in both the true  $\delta$  and the natural competition mark-up to marginal cost ratio  $\frac{m}{c}$ . If the

unit cost  $c$  is small, the bias  $\frac{m}{c}(\delta + 1)$  will tend to be large, no matter how small the true  $\delta$  is. This is illustrated by the following table.

Table A1  
Pitfall in the Conversion a Lerner Index Into an Overcharge Estimate: Constant overcharge, constant margin, and bias increasing as the marginal cost decreases.

Parameters	Values				
$\delta$	10%	10%	10%	10%	10%
$m$	0.02	0.02	0.02	0.02	0.02
$c$	0.20	0.1625	0.1250	0.0875	0.05
$\frac{m}{c}$	10%	12.3 %	16%	22.9%	40%
Bias = $\frac{m}{c}(\delta + 1)$	11%	14%	18%	25%	44%
$\tilde{\delta} = \delta + Bias$	21%	24%	28%	35%	54%
$\frac{Bias}{\delta}$	52.4%	58.3%	64.3%	71.4%	81.5%

## Appendix B: Summary of the Variables

Table B1. Whole Sample

	mean	std	min	max
Overcharge Estimates	45.47	102.86	0,00	1800,00
Duration	9.25	11.86	1	109
duration (discrete)	1.86	1.06	1	4
Domestic	0.47	0.50	0	1
BidRig	0.20	0.40	0	1
Guilty	0.66	0.48	0	1
US	0.30	0.46	0	1
EU	0.33	0.47	0	1
ASIA	0.09	0.28	0	1
ROW	0.04	0.19	0	1
WORLD	0.24	0.43	0	1
P1	0.01	0.08	0	1
P2	0.10	0.30	0	1
P3	0.11	0.31	0	1
P4	0.15	0.35	0	1
P5	0.57	0.50	0	1
P6	0.07	0.25	0	1
OTHER	0.06	0.24	0	1
HISTOR	0.02	0.13	0	1
PBEFOR	0.27	0.44	0	1
PWAR	0.02	0.14	0	1
PAFTER	0.13	0.34	0	1
COST	0.05	0.21	0	1
YARDST	0.14	0.35	0	1
ECON	0.14	0.34	0	1
LEGAL	0.18	0.38	0	1
JOURNAL	0.22	0.41	0	1
MONOGR	0.23	0.42	0	1
EDBOOK	0.06	0.24	0	1
GOVREP	0.23	0.42	0	1
COURT	0.18	0.38	0	1
WORKP	0.01	0.07	0	1
SPEECH	0.07	0.25	0	1
Sample size		1120		

Table B2. Cartels with Raw Overcharge Estimate in the open interval (0%, 50%)

	mean	std	min	max
Overcharge Estimates	20.61	12.34	0.10	49.90
Duration	9.26	11.95	1	98
duration (discrete)	1.84	1.05	1	4
Domestic	0.52	0.50	0	1
BidRig	0.24	0.43	0	1
Guilty	0.72	0.45	0	1
US	0.31	0.46	0	1
EU	0.32	0.47	0	1
ASIA	0.09	0.29	0	1
ROW	0.04	0.20	0	1
WORLD	0.24	0.43	0	1
P1	0.01	0.09	0	1
P2	0.09	0.28	0	1
P3	0.10	0.30	0	1
P4	0.13	0.34	0	1
P5	0.62	0.49	0	1
P6	0.05	0.23	0	1
OTHER	0.05	0.21	0	1
HISTOR	0.01	0.08	0	1
PBEFOR	0.27	0.44	0	1
PWAR	0.02	0.13	0	1
PAFTER	0.13	0.34	0	1
COST	0.04	0.20	0	1
YARDST	0.13	0.34	0	1
ECON	0.16	0.37	0	1
LEGAL	0.19	0.39	0	1
JOURNAL	0.22	0.41	0	1
MONOGR	0.21	0.41	0	1
EDBOOK	0.05	0.21	0	1
GOVREP	0.25	0.43	0	1
COURT	0.22	0.41	0	1
WORKP	0.01	0.08	0	1
SPEECH	0.05	0.22	0	1
Sample size		786		

Table B3. Cartels with Raw Overcharge Estimate of 0%

	mean	std	min	max
Overcharge Estimates	0	0	0	0
Duration	9.38	12.27	1	72
duration (discrete)	1.86	1.10	1	4
Domestic	0.58	0.50	0	1
BidRig	0.07	0.26	0	1
Guilty	0.37	0.49	0	1
US	0.38	0.49	0	1
EU	0.30	0.46	0	1
ASIA	0.09	0.28	0	1
ROW	0.04	0.19	0	1
WORLD	0.20	0.40	0	1
P1	0.00	0.00	0	0
P2	0.19	0.39	0	1
P3	0.05	0.22	0	1
P4	0.25	0.43	0	1
P5	0.47	0.50	0	1
P6	0.05	0.22	0	1
OTHER	0.16	0.37	0	1
HISTOR	0.14	0.34	0	1
PBEFOR	0.20	0.40	0	1
PWAR	0.01	0.11	0	1
PAFTER	0.11	0.32	0	1
COST	0.04	0.19	0	1
YARDST	0.09	0.28	0	1
ECON	0.02	0.16	0	1
LEGAL	0.23	0.43	0	1
JOURNAL	0.44	0.50	0	1
MONOGR	0.21	0.41	0	1
EDBOOK	0.10	0.30	0	1
GOVREP	0.09	0.28	0	1
COURT	0.05	0.22	0	1
WORKP	0.00	0.00	0	0
SPEECH	0.11	0.32	0	1
Sample size		81		

Table B4. Cartels with Raw Overcharge Estimate  $\geq 50\%$ 

	mean	std	min	max
Overcharge Estimates	137.26	188.29	50.00	1800.00
duration	9.18	11.50	1	109
duration (discrete)	1.93	1.08	1	4
domestic	0.28	0.45	0	1
BidRig	0.10	0.30	0	1
Guilty	0.57	0.50	0	1
US	0.25	0.44	0	1
EU	0.39	0.49	0	1
ASIA	0.08	0.28	0	1
ROW	0.02	0.12	0	1
WORLD	0.26	0.44	0	1
P1	0.01	0.09	0	1
P2	0.13	0.34	0	1
P3	0.15	0.35	0	1
P4	0.16	0.37	0	1
P5	0.45	0.50	0	1
P6	0.11	0.31	0	1
OTHER	0.08	0.28	0	1
HISTOR	0.01	0.09	0	1
PBEFOR	0.28	0.45	0	1
PWAR	0.03	0.18	0	1
PAFTER	0.14	0.35	0	1
COST	0.06	0.24	0	1
YARDST	0.20	0.40	0	1
ECON	0.08	0.28	0	1
LEGAL	0.11	0.31	0	1
JOURNAL	0.16	0.37	0	1
MONOGR	0.30	0.46	0	1
EDBOOK	0.10	0.30	0	1
GOVREP	0.22	0.42	0	1
COURT	0.10	0.30	0	1
WORKP	0.00	0.06	0	1
SPEECH	0.11	0.31	0	1
Sample size		253		

## Appendix C: Estimated coefficients for the bias factors by cluster.

In the tables below,  $Xk$  denotes the interaction variable between  $X$  and cluster indicator  $k$ . HISTOR2 and HISTOR3 are missing because they are empty. We see that the coefficients vary across clusters. For example, the coefficient of JOURNAL1 (in cluster 1) is  $-9.51$  while that of JOURNAL2 is  $-0.65$ , that of JOURNAL3 is  $9.88$  and that of JOURNAL4 is  $15.54$ .<sup>23</sup> These large variations in the coefficients of the  $Z$  variables reflect the large magnitude of the bias of the raw overcharge estimates.

Table C1. No control for sample selection bias

	Linear regression		Log-linear regression	
	Coefficients	Student-t	Coefficients	Student-t
OTHER1	-6.19	-4.58	-0.77	-4.94
HISTOR1	-4.26	-1.62	-0.70	-2.32
PBEFOR1	-2.87	-3.03	-0.15	-1.33
PWAR1	2.50	0.57	0.65	1.28
PAFTER1	-3.18	-2.65	-0.55	-3.99
COST1	-5.33	-2.72	-0.97	-4.28
YARDST1	-3.23	-3.04	-0.21	-1.69
LEGAL1	-3.12	-3.15	-0.23	-1.98
JOURNAL1	-9.51	-9.30	-0.74	-6.31
MONOGR1	-6.36	-5.17	-0.21	-1.46
EDBOOK1	-9.50	-5.08	-0.57	-2.65
COURT1	-7.37	-6.52	-0.43	-3.32
WORKP1	-6.82	-5.82	-0.34	-2.52
OTHER2	-0.12	-0.06	0.05	0.19
PBEFOR2	-0.16	-0.17	0.12	1.15
PWAR2	-0.27	-0.14	0.10	0.42
PAFTER2	0.26	0.26	0.10	0.86
COST2	-0.22	-0.17	0.08	0.53
YARDST2	0.34	0.34	0.11	0.91
LEGAL2	0.54	0.52	0.13	1.12
JOURNAL2	-0.65	-0.58	0.30	2.36
MONOGR2	-0.82	-0.71	0.30	2.26
EDBOOK2	0.58	0.39	0.50	2.90
COURT2	-1.74	-1.48	0.19	1.41
WORKP2	-1.77	-1.61	0.24	1.91
SPEECH2	-4.99	-1.04	0.06	0.10

<sup>23</sup> A parameter is significant at 10% [5%] in a two-sided test if its Student-t is larger than 1.64 [1.96] in absolute value. In a one sided test, the threshold is 1.28 [1.64].

Table C1. No control for sample selection bias (Continued)

	Linear regression		Log-linear regression	
	Coefficients	Student-t	Coefficients	Student-t
OTHER3	4.71	2.45	0.37	1.68
PBEFOR3	1.20	1.33	0.18	1.70
PWAR3	1.60	0.61	0.21	0.70
PAFTER3	0.80	0.65	0.18	1.30
COST3	2.70	1.62	0.29	1.50
YARDST3	1.22	1.04	0.14	1.04
LEGAL3	0.10	0.08	0.12	0.80
JOURNAL3	9.88	8.63	0.81	6.16
MONOGR3	9.18	7.90	0.76	5.64
EDBOOK3	9.05	5.45	0.84	4.41
COURT3	9.09	6.86	0.70	4.57
WORKP3	8.73	7.75	0.72	5.57
SPEECH3	5.68	1.74	0.64	1.71
<hr/>				
OTHER4	12.74	7.93	0.69	3.73
HISTOR4	10.01	3.09	0.55	1.48
PBEFOR4	10.13	8.56	0.57	4.21
PWAR4	6.83	2.71	0.44	1.51
PAFTER4	9.81	7.02	0.57	3.55
COST4	11.38	4.20	0.56	1.79
YARDST4	10.01	7.29	0.57	3.60
LEGAL4	8.62	5.70	0.51	2.92
JOURNAL4	15.54	12.79	0.85	6.06
MONOGR4	13.29	9.79	0.75	4.82
EDBOOK4	21.13	13.22	1.19	6.45
COURT4	13.06	9.44	0.68	4.23
WORKP4	14.60	10.54	0.73	4.59
SPEECH4	10.42	3.01	0.69	1.73

Table C2. Controlling for sample selection bias

	Linear regression		Log-linear regression	
	Coefficients	Student-t	Coefficients	Student-t
OTHER1	-6.43	-3.13	-0.55	-2.31
HISTOR1	-4.77	-1.14	-0.24	-0.49
PBEFOR1	-2.93	-2.89	-0.09	-0.80
PWAR1	2.31	0.50	0.83	1.57
PAFTER1	-3.26	-2.50	-0.48	-3.20
COST1	-5.46	-2.55	-0.84	-3.42
YARDST1	-3.35	-2.56	-0.10	-0.64
LEGAL1	-3.10	-3.09	-0.25	-2.15
JOURNAL1	-9.41	-7.77	-0.84	-6.01
MONOGR1	-6.22	-4.04	-0.34	-1.92
EDBOOK1	-9.45	-4.99	-0.62	-2.82
COURT1	-7.26	-5.44	-0.54	-3.48
WORKP1	-6.57	-3.30	-0.57	-2.50
OTHER2	-0.31	-0.13	0.22	0.79
PBEFOR2	-0.21	-0.22	0.17	1.49
PWAR2	-0.44	-0.19	0.25	0.95
PAFTER2	0.20	0.18	0.16	1.27
COST2	-0.35	-0.23	0.20	1.14
YARDST2	0.23	0.19	0.20	1.45
LEGAL2	0.57	0.54	0.11	0.89
JOURNAL2	-0.56	-0.44	0.22	1.52
MONOGR2	-0.68	-0.47	0.18	1.06
EDBOOK2	0.64	0.41	0.45	2.52
COURT2	-1.63	-1.18	0.09	0.57
WORKP2	-1.53	-0.82	0.02	0.12
SPEECH2	-4.64	-0.88	-0.26	-0.42

Table C2. Controlling for sample selection bias (Continued)

	Linear regression		Log-linear regression	
	Coefficients	Student-t	Coefficients	Student-t
OTHER3	4.48	1.85	0.59	2.10
PBEFOR3	1.14	1.17	0.23	2.05
PWAR3	1.40	0.48	0.39	1.18
PAFTER3	0.73	0.56	0.25	1.65
COST3	2.57	1.38	0.41	1.90
YARDST3	1.09	0.76	0.26	1.57
LEGAL3	0.11	0.09	0.10	0.72
JOURNAL3	9.97	7.74	0.73	4.90
MONOGR3	9.32	6.43	0.63	3.78
EDBOOK3	9.11	5.36	0.79	4.06
COURT3	9.22	6.02	0.59	3.33
WORKP3	8.98	4.63	0.50	2.21
SPEECH3	6.03	1.53	0.32	0.71
OTHER4	12.49	5.53	0.92	3.53
HISTOR4	9.48	2.03	1.04	1.93
PBEFOR4	10.05	7.89	0.64	4.38
PWAR4	6.65	2.38	0.61	1.90
PAFTER4	9.73	6.50	0.65	3.76
COST4	11.23	3.91	0.70	2.10
YARDST4	9.89	6.31	0.68	3.76
LEGAL4	8.63	5.70	0.50	2.87
JOURNAL4	15.64	11.33	0.75	4.74
MONOGR4	13.44	8.09	0.62	3.22
EDBOOK4	21.18	12.95	1.14	6.04
COURT4	13.20	8.14	0.55	2.96
WORKP4	14.86	6.90	0.50	2.00
SPEECH4	10.73	2.70	0.41	0.89

## References

- Appelbaum, Elie (1979). Testing Price Taking Behavior. *Journal of Econometrics* 9 283-294.
- Allain, Marie-Laure, Marcel Boyer, Rachidi Kotchoni and Jean-Pierre Ponssard (2011a). The Determination of Fines in Cartel Cases: The Myth of Underdeterrence. CIRANO WP (March 2011, revised September 2011)
- Allain, Marie-Laure, Marcel Boyer and Jean-Pierre Ponssard (2011b). The determination of optimal fines in cartel cases: Theory and practice. *Concurrences* 4-2011, 32-40 (Selected as *Best Academic Economics Article - 2012 Antitrust Writing Awards*, Institute of Competition Law and George Washington University Law School).
- Becker, Gary (1968). Crime and Punishment: An Economic Approach. *Journal of Political Economy* 76: 169-217.
- Cohen, Mark A. and David T. Scheffman (1989). The Antitrust Sentencing Guideline: Is the Punishment Worth the Costs? *Journal of Criminal Law* 27: 331-366.
- Connor, John M. (2010). *Price-fixing overcharges, 2<sup>nd</sup> Edition*. (First edition available online at [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=787924](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=787924))
- Connor, John M. and Yuliya Bolotova (2006). A Meta-Analysis of Cartel Overcharges. *International Journal of Industrial Organization* 24: 1109-1137
- Connor, John M. and Robert H. Lande (2008). Cartel Overcharges and Optimal Cartel fines. *Issues in Competition Law and Policy*, S.W. Waller (ed.), Volume 3, AMA Section of Antitrust Law, chapter 88, pp. 2203-2218.
- Finkelstein, M. and H. Levenback. Regression Estimates of Damages in Price-Fixing Cases (1983). *Law and Contemporary Problems* 46: 145-169.
- Froeb, Luke M., Robert A. Koyak, and Gregory J. Werden (1993). What Is the Effect of Bid-Rigging on Prices? *Economics Letters* 42: 419-423
- Heckman James (1979). Sample selection bias as a specification error. *Econometrica*.47, 153–161.
- Hunter, J. and F. Schmidt (2004). *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*, Second Edition, SAGE Publications.
- Kearney, Elisa (2009). The Competition Bureau “20 Percent Solution”: Does the Punishment Fit the Crime? *The Marker*, CBA National Competition Law Section, March.
- Kearney, Elisa and Mark Katz (2009). Anti-Cartel Enforcement in Canada: Still More Bark Than Bite. Cartel and Criminal Practice Committee, American Bar Association, December.
- Landes, William H. (1983). Optimal Sanctions for Antitrust Violations. *University of Chicago Law Review* 50: 652-678.
- Levenstein, Margaret C. and Valerie Y. Suslow (2003). Contemporary International Cartels and Developing Countries: Economic Effects and Implications for Competition Policy. *Antitrust Law Journal* 71: 801-852.
- Morrison, Catherine J. (1990). Market Power, Economic Profitability, and Productivity Growth Measurement: An Integrated Structural Approach, *NBER Working Paper* No. 3355.
- Morrison, Catherine J. (1993). Productive and Financial Performance in U.S. Manufacturing Industries: An Integrated Structural Approach. *Southern Economic Journal* 69: 376-392.