DISCUSSION PAPER SERIES

DP16900

A Practical Review of Methods to Estimate Overcharges Using Linear Regression

Roman Inderst and Christopher Milde

INDUSTRIAL ORGANIZATION

ISSN 0265-8003

A Practical Review of Methods to Estimate Overcharges Using Linear Regression

Roman Inderst and Christopher Milde

Discussion Paper DP16900
Published 15 January 2022
Submitted 11 January 2022

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

A Practical Review of Methods to Estimate Overcharges Using Linear Regression

Abstract

Arguably the most widely used techniques for estimating price overcharges from competition law infringements are the dummy variable and the forecasting approaches using linear regression analysis. While rarely used in practice, in this note we make use of the fully interacted dummy variable approach to review some basic properties of all three approaches. We show under which conditions and for which estimands of interest these approaches are equivalent and when they differ. We also note some interesting additional choices an interaction approach allows.

JEL Classification: K13, C20

Keywords: overcharge estimation, Forecasting, dummy variable, interactions, interaction model, Treatment effects, damages quantification

Roman Inderst - inderst@finance.uni-frankfurt.de
Goethe University Frankfurt and CEPR

Christopher Milde - christopher.milde@gmail.com
Compass Lexecon

A Practical Review of Methods to Estimate Overcharges using Linear Regression

by

Roman Inderst[*] and Christopher Milde[†]

Abstract

Arguably the most widely used techniques for estimating price overcharges from competition law infringements are the dummy variable and the forecasting approaches using linear regression analysis. While rarely used in practice, in this note we make use of the fully interacted dummy variable approach to review some basic properties of all three approaches. We show under which conditions and for which estimands of interest these approaches are equivalent and when they differ. We also note some interesting additional choices an interaction approach allows.

[*] Johann Wolfgang Goethe University, Frankfurt/Main.

[†] Compass Lexecon, Berlin.

1 Introduction

The first step in the quantification of damages from competition law infringements such as horizontal agreements is the estimation of the price overcharge for the market transactions affected by the infringement. A widely used approach to estimating this price overcharge is the comparison of prices of market transactions affected by the infringement ("affected transactions") with prices of market transactions not affected by the infringement ("comparator transactions"), most often those in the same market before and/or after the infringement. For this comparison to provide a credible estimate for the causal effect of the infringement on prices, differences in price that are not due to the infringement but to other price determining factors must be taken into account. The most widely used technique for doing so is multivariate linear regression analysis.

In this context, the relevant literature and practice typically distinguish two approaches: the "forecasting approach" and the "dummy variable approach".[3] The dummy variable approach is often implemented as a single dummy variable ("single DV") approach, i.e. one that contains a single indicator to denote all transactions potentially affected by the infringement. Variations from this single dummy variable approach play an important role, though. A particularly important variation is a model that allows all other price determining factors to exert a different influence on the price of affected transactions than on the price of comparator transactions. This can be achieved by interacting the dummy variable for the infringement with every other price determining factor (in the following "interactions"). In this note, we focus on the model with interactions for every other price determining factor used, the fully interacted dummy variable model ("fully interacted DV" model), and its relationship to the single DV and forecasting approaches.

Building on general econometric insights regarding linear regression algebra, which to different degrees have been explicitly discussed in a competition damages context by inter alia McCrary and Rubinfeld (2014), White (2006) and Higgins & Johnson (2003), we first show under which assumptions these three approaches, the single DV, the fully interacted DV and the forecasting approach are equivalent and when they differ for the case that the estimand of interest is the average overcharge. Here, our main objective is to make the respective equivalence or difference as transparent as possible.

The fully interacted DV approach helpfully clarifies that at least for the average overcharge the forecasting and the single DV approach are two endpoints of a spectrum of choices, i.e. all interactions versus no interactions at all. A priori there is no reason for choosing either all or nothing though, i.e. it is also possible to follow an intermediate approach. In this respect the fully interacted DV model produces helpful information that allows the analyst to judge whether the presence of an interaction is supported by the data. On the basis of

[3] See e.g. European Commission, Practical Guide, Quantifying Harm in Actions for Damages based on Breaches of Articles 101 or 102 of the Treaty on the Functioning of the European Union; Strasbourg, 11.6.2013, paragraph 72.

this information, the analyst may want to judge whether an observed interaction (that is supported by the data) is caused by the infringement in question or not. This intermediate approach allows calculating an overcharge that is neither possible with the forecasting nor with the single DV approach. This additional flexibility should be handled with care, however.

An additional contribution of this article is a comparison of group-specific or even transaction-specific overcharges. By definition, such specific overcharges are ruled out in the single DV approach, which assumes constant overcharges for all affected transactions. Notably, also the fully interacted DV and the forecasting approach differ in this respect, at least without an explicit adjustment. The choice of calculation of individual overcharges has implications for the calculation of interest.

This note ends with a discussion of potential additional differences between the methods, e.g. regarding estimation efficiency as well as the preferred choice in case of few observations.

2 The counterfactual framework and the estimand of interest

The by now widely accepted guiding principle for the quantification of damages from competition law infringements is the counterfactual framework: in order to obtain the overcharge for an individual transaction affected by an infringement, oc_i, the actually observed price of this transaction, p_{infr_i}, must be compared to the price that would have been observed for this transaction had there been no infringement (the "counterfactual price"), p_{cf_i}:[4]

$$oc_i = p_{infr_i} - p_{cf_i}.$$

The counterfactual price is of course not observable, a conundrum also known as the fundamental problem of causal inference.[5]

The above difference between the actual and counterfactual price for an individual observation may in itself be of interest (in the following "individual overcharge"). Notably, when i refers to time or a specific group and when, for economic reasons or as indicated by the nature of the infringement, it is likely that the overcharge varies over time or between certain groups, the individual overcharge may need to be estimated. If the overcharge changes over time, one may also wish to calculate interest using such estimated individual, i.e. time-specific, overcharges.

---

[4] In the policy evaluation or treatment effects literature this counterfactual framework is commonly referred to as the potential outcome model of causality or the Rubin causal model. See Cameron and Trivedi (2005), chapter 2.7 for a short introduction, or Imbens and Rubin (2015) for a longer one.

[5] See Holland (1986) for a lucid discussion of the counterfactual approach to causal inference.

Often, however, the magnitude of interest is the average overcharge for all transactions affected by the infringement, i.e.

$$\overline{oc_{infr}} = \frac{1}{N_{infr}} \sum_{i=1}^{N_{infr}} \left(p_{infr_i} - p_{cf_i}\right) = \overline{p_{infr}} - \overline{p_{cf}}.$$

The above formula for the average overcharge assumes that quantity is constant over transactions i. If quantity is not constant across transactions i the average overcharge for the affected transactions is given by

$$\overline{oc_{infr}} = \frac{1}{Q_{infr}} \sum_{i=1}^{N_{infr}} q_i * \left(p_{infr_i} - p_{cf_i}\right),$$

where $Q_{infr} = \sum_{i=1}^{N_{infr}} q_i$.

As the counterfactual prices are not observable, they have to be estimated. With market comparator based approaches it is assumed that prices in a suitable comparator market provide a measure of the competitive price given the values of the other price determining factors in the comparator market. In order to estimate the counterfactual price it is acknowledged that the observed prices in the comparator market must be adjusted if the price determining factors (other than the infringement) differ between the comparator market and the affected market. In this note we focus on results that follow from using linear regression to make these adjustments, which arguably is the most widely used technique in current practice. [6]

3 Approaches towards overcharge estimation

3.1 The forecasting approach

The desire to extrapolate prices from a comparator market to a market affected by an infringement naturally gives rise to the forecasting approach: First, for any given set of price determining factors, X, a linear regression of price on a constant and X is run on the comparator data only. This gives rise to the least squares decomposition of the price of comparator transactions (p_{comp}):[7]

$$p_{comp_i} = a_{comp} + b_{comp} * X_{comp_i} + e_{comp_i},$$

---

[6] Multivariate linear regression as a method to make such adjustments has been subject to criticism in the econometrics literature. As a result of this criticism many refinements and alternatives have been developed and discussed in this literature. Some of these refinements and alternatives may also be relevant in the context of damage quantification. The discussion of these is not the subject of this note though.

[7] For notational convenience, we disregard the conventional notation for matrix multiplication in this note, as there is no scope for misunderstandings. I.e. the term b*X denotes the product of the relevant coefficients in vector b with the corresponding variables in X, or b1*X1 + b2*X2 + … or in conventional matrix notation Xb, where X is an nxk-matrix and b is a kx1-vector.

where a_{comp} and b_{comp} are the estimated coefficients and e_{comp_i} are the fitted residuals from the linear regression. The coefficients a_{comp} and b_{comp} are then used to forecast the counterfactual price in the infringement period by applying them to the values of the price determining factors in the infringement period, X_{infr_i} , i.e.

$$\widehat{p_{cf_i}} = a_{comp} + b_{comp} * X_{infr_i}.$$

Thus the estimated overcharge for an individual transaction in the infringement period using the forecasting approach is

$$\widehat{oc_{fc_i}} = p_{infr_i} - \widehat{p_{cf_i}} = p_{infr_i} - a_{comp} - b_{comp} * X_{infr_i},$$

and the simple average overcharge for all affected transactions is

$$\overline{\widehat{oc_{fc}}} = \overline{p_{infr}} - a_{comp} - b_{comp} * \overline{X_{infr}}.$$

3.2 The single DV model

In the single DV model not only comparator transactions are used to estimate the relationship between price and other price determining factors (as is the case for the forecasting model) but also affected transactions. In the single DV model it is assumed that the relationship between the other factors X and price does not change between the affected and comparator transactions. This gives the following least squares decomposition of the price:

$$p_i = g + f * D_i + k * X_i + u_i,$$

where p_i is actual price, D is a dummy variable that indicates transactions affected by the infringement, and g, f, and k are estimated coefficients, and u_i are the fitted residuals from linear least squares regression. The estimate for the average overcharge from this model is the coefficient f. Note that estimates for individual overcharges are not possible with the single DV model, or more precisely it is assumed that the overcharge is the same for every individual transaction.

3.3 The fully interacted DV model

In the fully interacted variant of the DV model the estimated relationship between price and each other price determining factor in X can differ for the affected and comparator transactions. This is achieved by including interactions between the dummy variable for the infringement and each factor in X.

The fully interacted DV model results in the following decomposition of price from a linear regression:

$$p_i = a + d * D_i + b * X_i + c * D_i * X_i + e_i,$$

where D is a dummy variable that indicates transactions affected by the infringement, and a, d, b, and c are the estimated coefficients from a linear least squares regression and e_i are

the fitted residuals from this regression. In this model the effect of the infringement is the change in price that can be attributed to D, i.e. the difference between the fitted price for D=1:

$$\widehat{p_{infr_i}} = a + d + b * X_{infr_i} + c * X_{infr_i},$$

minus the fitted price (of the affected transactions) for D=0:

$$\widehat{p_{cf_i}} = a + b * X_{infr_i}.$$

This results in individual (transaction) overcharges of

$$\widehat{oc_{inter_i}} = d + c * X_{infr_i}.$$

Averaging this over all transactions affected by the infringement gives:[8]

$$\overline{\widehat{oc_{inter}}} = d + c * \overline{X_{infr}}.$$

4 Equivalence and Non-Equivalence

4.1 Forecasting and fully interacted DV model

4.1.1 Equivalence for average overcharge

Recall that the average overcharge estimate from the forecasting approach is given by

$$\overline{\widehat{oc_{fc}}} = \overline{p_{infr}} - a_{comp} - b_{comp} * \overline{X_{infr}},$$

i.e. the difference between the actual average price of the affected transactions minus the estimated counterfactual price on the basis of the least squares coefficients estimated over the comparator transactions.

The average of the actual price over the affected transactions, however, is the same as the average of the fitted price from a linear regression of price on X (only over these affected transactions). That is, running a regression of p_{infr_i} on a constant and X_{infr_i} allows decomposing the price affected by the infringement into

$$p_{infr_i} = a_{infr} + b_{infr} * X_{infr_i} + e_{infr_i},$$

where a_{infr} and b_{infr} are the estimated coefficients and e_{infr_i} are the fitted residuals from linear least squares regression. As the sum of e_{infr_i} over the infringement transactions is by construction zero, the average of the actual price in the infringement period is

[8] Standard errors for this expression can be obtained using standard methods, see e.g. Wooldridge (2002), chapter 18.

$$\overline{p_{infr}} = a_{infr} + b_{infr} * \overline{X_{infr}}.$$

Plugging the latter result into the overcharge formula for the forecasting approach gives:

$$\overline{p_{infr}} - a_{comp} - b_{comp} * \overline{X_{infr}}$$

$$= a_{infr} + b_{infr} * \overline{X_{infr}} - a_{comp} - b_{comp} * \overline{X_{infr}}$$

$$= (a_{infr} - a_{comp}) + (b_{infr} - b_{comp}) * \overline{X_{infr}}.$$

Compare this to the decomposition of price from running a linear regression with a dummy and a full set of interactions:

$$p_i = a + d * D_i + b * X_i + c * D_i * X_i + e_i.$$

The interpretation of the coefficients of the above decomposition is:[9]

- a is the constant when D=0, i.e. for the comparator transactions, or a = a_{comp};
- b is the coefficient for the covariates for the comparator transaction, or b = b_{comp};
- a+d is the constant for the infringement transactions, or a + d = a_{infr}, which implies that d = $a_{infr} - a = a_{infr} - a_{comp}$;
- b+c is the coefficient on the covariates in the comparator transactions or b + c = b_{infr} , which implies that c = $b_{infr} - b_{comp}$.

We thus have as well:

$$\overline{\widehat{oc_{inter}}} = d + c * \overline{X_{infr}} = (a_{infr} - a_{comp}) + (b_{infr} - b_{comp}) * \overline{X_{infr}} = \overline{\widehat{oc_{fc}}}.$$

Hence the average estimated overcharge for the affected transactions from the standard forecasting approach is numerically equivalent to that from the fully interacted DV approach.[10] This result holds for any choice of price determining factors, X, as long as the same factors X are used for the forecasting and the fully interacted DV approach and as long as each factor exhibits variation within the comparator data. This equivalence result is therefore entirely independent of the question whether the specific choice of X constitutes the "correct" model specification.

---

9 See e.g. Greene (2002), section 8.2 or any other standard econometrics textbook.

10 It is noteworthy that there is another dummy variable model that gives numerically equivalent results to those from the forecasting model, namely a model in which each transaction in the infringement period obtains a separate dummy variable (Salkever (1986)). This procedure is a convenient way to obtain standard errors when the forecasting approach is used. A more practical implementation could be an approximation to the full set of dummies by including separate dummies for several subgroups of interest, such as years, customers, products etc.

However, the equivalence result hinges on the assumption that quantity is constant across transactions. McCrary and Rubinfeld (2014) have shown that when quantity varies over transactions the results from both approaches differ. For completeness, within our simple framework we derive and restate this result formally in the Appendix.

4.1.2 Non-equivalence for individual overcharges

The overcharge for an individual transaction from the forecasting approach (see section 3.1) is

$$\widehat{oc_{fc_i}} = p_{infr_i} - \widehat{p_{cf_i}} = p_{infr_i} - a_{comp} - b_{comp} * X_{infr_i}.$$

As discussed in section 4.1.1 the infringement price can be decomposed by linear regression into

$$p_{infr_i} = a_{infr} + b_{infr} * X_{infr_i} + e_{infr_i},$$

so that the individual overcharge from the forecasting approach can be written as

$$\widehat{oc_{fc_i}} = a_{infr} + b_{infr} * X_{infr_i} + e_{infr_i} - a_{comp} - b_{comp} * X_{infr_i},$$

or

$$\widehat{oc_{fc_i}} = (a_{infr} - a_{comp}) + (b_{infr} - b_{comp}) * X_{infr_i} + e_{infr_i}.$$

The individual overcharge from the fully interacted DV model is

$$\widehat{oc_{inter_i}} = d + c * X_{infr_i},$$

and as discussed in section 4.1.1 $d = (a_{infr} - a_{comp})$ and $c = (b_{infr} - b_{comp})$.

Thus the individual overcharges from the forecasting and the fully interacted DV approach do not give an equivalent result. They differ by the component of price that is not explained by the assumed price model (but which is assumed to be uncorrelated with the explained part), the residual e_{infr_i}:[11]

$$\widehat{oc_{fc_i}} - \widehat{oc_{inter_i}} = e_{infr_i}.$$

While the forecasting approach does include this unexplained price component in the individual overcharges, the fully interacted DV model does not. However, only the fully interacted DV model allows separately identifying the unexplained price component

[11] Incidentally, the formulations in the Commission's Practical Guide (cf. footnote 1 above) themselves do not provide guidance as to whether the individual overcharge should be calculated by using observed or by using estimated (infringement) prices. At various accounts, the Practical Guide explicitly refers to a comparison to the "price actually paid" or the "observed price" (e.g. §79 or §101), but more generally the treatment of the before-during-and-after method with a constant overcharge is more in line with the latter interpretation.

whereas the forecasting approach always exhibits the entire individual overcharge $p_{infr_i} - \widehat{p_{cf_i}}$, which includes the unexplained price component. Thus the fully interacted DV approach allows distinguishing between variation in individual overcharges that is due to the other factors X and variation that is due to other, unknown factors, and it allows analysing this unexplained price variation.

Depending on the nature of the infringement, it may be appropriate or desirable to take such unexplained price variation into account in calculating individual overcharges. Suppose, for instance, that as part of a cartel infringement, cartelists agreed to pass on increases in the price of a particular commodity to a larger extent than what would have been the case under competition, or that they use such increases in costs as triggers for price increases. Following an increase in the respective component of X_{infr_i}, say W, the individual overcharge should thus increase as well. Both, the forecasting and the fully interacted DV model would be able to capture this specific structure by allowing a different coefficient on W for the affected transactions than the comparator transactions. In addition, the fully interacted DV approach would allow analysing the evolution of the unexplained price variation, the residual. If, for example, one observed that whenever W increased also the residual increased, this may indicate that the effects of the cartel were not fully captured by the different coefficient on W for the affected transactions. This may thus provide a rationale for including these residuals in the individual overcharges. The then economically justified estimation of an individual overcharge would in turn provide a more adequate and accurate basis for the calculation of interest.

4.2 Single DV and fully interacted DV models

Recall that the estimate for the average overcharge in the single DV model is given by the coefficient f in the model:

$$p_i = g + f * D_i + k * X_i + u_i,$$

where g, f, and k are estimated coefficients and u_i are the fitted residuals from linear least squares regression. How does this estimate, f, relate to the estimate from the fully interacted DV approach? To answer this question, compare the above model with the regression decomposition from the fully interacted DV approach:

$$p_i = a + d * D_i + b * X_i + c * D_i * X_i + e_i.$$

Obviously the difference between both models is the term $c * D_i * X_i$. The coefficient c measures to what extent the relationship between price and the other price determining factors, X, is different for the affected transactions compared to the comparator transactions. If c=0, so that this relationship is not different, then both models give equivalent results for the average overcharge.[12]

[12] Note that in any real world dataset, even if in the true data generating process there are no interactions, it is unlikely that c, the estimated coefficients on the interaction terms, would be exactly 0.

The infringement may however have directly impacted on the effects of the covariates X_i on transaction prices. Suppose X_i comprises costs. If the main effect of the infringement is to increase the cartelists' joint market power, so that their behaviour should resemble more closely that of a textbook monopolist, then economic theory would predict that typically the pass-on rate should decrease. However, also the opposite is conceivable, that is a higher pass-on rate, notably if through the infringement cartelists reduce strategic uncertainty as to their individual reactions to such cost increases. In addition, cartelists may use such cost increases also as trigger points for price hikes.

If therefore c cannot be expected to be equal to zero, what does the coefficient f measure when in truth such interactions are present? If the true model is the one with interactions, but the model without interactions is estimated, the single DV model is misspecified. To understand the effects of this misspecification it is helpful to consider the missing interaction terms as omitted variables. Then, the estimated overcharge from a dummy variable model without interaction terms is unbiased only if

$$\overline{X_{infr}} = \overline{X_{comp}},$$

i.e. only if the average of every price determining factor for the affected transactions is the same as the average for the transactions not affected by the infringement.[13] Note however that the fact that these averages differ between the infringement and comparator transactions is precisely the reason for why we wish to take them into account in estimating the overcharge. In fact if this condition were true, there would be no need to take these price determining factors into account at all as the estimate of the average overcharge is not affected by them (in either the model with or without interactions or the forecasting approach).[14]

If the condition does not hold, however, the estimate from the single DV approach is biased, and the direction of this bias is generally (that is, when c is unknown) ambiguous.

Thus c=0 should more precisely be referred to as the corresponding true coefficient being 0, or the estimated coefficient c being not statistically significant.

[13] This has been derived in detail by Higgins and Johnson (2003). In a linear regression setting the omitted variable bias ("OVB) in the single DV model due to leaving out interactions is equal to

$$OVB = \left[\overline{X_{infr}} - \Delta\left(\overline{X_{infr}} - \overline{X_{comp}}\right)\right] * c,$$

where Δ is the matrix of coefficients on X from a regression of the interaction terms D*X on a constant, the infringement dummy and X. Accordingly the estimate from the single DV model is "d + OVB". Comparing this to the unbiased overcharge estimate from the fully interacted DV model:

$$\widehat{\overline{oc_{inter}}} = d + c * \overline{X_{infr}}$$

shows that in the presence of interactions the estimated overcharge from a dummy variable model without interaction terms is unbiased only if $\overline{X_{infr}} = \overline{X_{comp}}$.

[14] The estimate will be more precise (i.e. the standard error more narrow) if these factors are taken into account in the case that in truth they are related to price though.

5 Broader discussion of approaches

To summarize the preceding, the forecasting and the fully interacted DV approach give numerically equivalent results for the average overcharge if quantity is constant and each factor in X exhibits variation in the comparator transactions. Thus in terms of the pure result for the average overcharge it does not matter which approach is used.

As discussed below, however, the fully interacted DV approach provides a richer set of results in addition to the average overcharge for the affected transactions, namely the interactions themselves, which may be desirable to assess the credibility of the model overall and/or to choose and justify an intermediate approach. The two approaches also differ when it comes to estimating individual overcharges.

The fully interacted DV approach and the single DV approach are equivalent only under fairly strong assumptions. If they are, the preceding obviously implies that the single DV approach estimates the same average overcharge as the forecasting approach as well. If these assumptions are not true, however, the single DV approach will generally give different results than the forecasting and interaction approaches and, a priori, there are good reasons to assume that it will be biased for the average overcharge.

The higher complexity and likely data needs of the fully interacted DV approach as compared to the single DV approach may deter many practitioners from using it. In this section we therefore discuss some actual or perceived impediments to the use of the fully interacted DV approach. We briefly also discuss other considerations that may prove relevant in practice.

5.1 Efficiency

A popular motivation for using the single DV approach is the notion that the estimates from this model are more precise (in the sense of lower variance of the estimate) compared to the model with interactions if the restrictions associated with the single DV approach are true. McCrary and Rubinfeld (2014) clarify, however, that this statement is not necessarily true. While we do not wish to repeat their argument, it is probably helpful to point out one likely point of confusion regarding this issue.

Compare again the single DV approach:

$$p_i = g + f * D_i + k * X_i + u_i$$

with the fully interacted DV approach:

$$p_i = a + d * D_i + b * X_i + c * D_i * X_i + e_i.$$

Note that the coefficient on the infringement dummy in the former model is f and in the latter model it is d. If in truth there are no interactions, it is generally true that the estimate

of f is more precise (has lower variance) than the estimate of d. This follows from the textbook discussion of the inclusion of irrelevant variables.[15] This observation may lead practitioners to conclude that the single DV approach may be preferable even when there may be some, albeit not too large interactions, as then some bias may be tolerable if at the same time the variance is lower.

A focus on the coefficients on the dummy, i.e. respectively f and d alone, however, is not the relevant comparison in the present case. Instead the estimate for the average overcharge from both approaches must be compared. Therefore the precision of coefficient f, the average overcharge estimate from the single DV model has to be compared to the precision of the average overcharge estimate from the fully interacted DV model, i.e. the term "d + c*X". The variance on the latter not only depends on the variance of d, but also on that of c and on the covariance between d and c. As this covariance can be negative the total variance of "d + c*X" can be smaller than the variance of f, even if the variance of d alone is higher than the variance of f. Therefore it is impossible to say a priori whether the single DV approach will yield an estimate with lower variance.[16]

It is also worthwhile pointing out that in the case that interaction effects are present, which as outlined above is prima facie plausible, not only is f a biased estimate of the true overcharge, but also the estimate of the variance is biased upward, i.e. the estimated variance is larger than the true variance.[17] Thus using the single DV approach when in truth interactions are present provides a biased point estimate of the average overcharge and a biased, too large, variance of this point estimate.

5.2 Model specification in case of few observations

Another popular motivation for using the single DV approach is a lack of sufficient data. Indeed, when only few observations are available, it may be impossible to estimate a dummy variable model with a full set of interactions. For example, there may be 3 cost and 2 demand factors and 10 product characteristics dummies, resulting in 15 other price determining factors X. Adding a constant and the dummy for the infringement results in 32 variables to be included in a dummy variable model with a full set of interaction terms. With a dataset of limited size, it may be impossible to estimate such a model.

This does not imply, however, that the single DV model is automatically the model of choice. Instead of automatically dropping all interactions, one could drop likely less important factors entirely (e.g. those with little average difference between affected and comparator transactions) but keep interactions which are a priori expected to be important (e.g. those on an important cost factor). Overall, it may cause less bias when dropping some factors from the model that can be considered to have only a small effect on price and/or a

---

15 See e.g. Greene (2003), section 8.4.3.

16 Higgins and Johnson (2003) provide theoretical results for when this is the case, namely when there are no differences in the means of the other price determining factors. This, of course, is very unlikely.

17 See e.g. discussion on omitted variable bias in Greene (2003), section 8.4.2.

small correlation with the infringement and instead allow a likely important factor to have an interaction.

Another situation occurring in practice may be that there are either very few observations for the infringement or for the comparator transactions. This may be the case, for example, when there is one price observation per year and only few years have passed since the end of the infringement. Then it may be impossible to test whether interactions are present or to measure their magnitude precisely. It should be noted, however, that not being able to test or measure an interaction effect does not allow the conclusion that there is none. As is generally the case when data are not sufficiently informative, prior knowledge should be used to determine whether an interaction effect is likely or not.

5.3 Model specification with interaction vs forecasting approach

It is also worthwhile to point out that the fully interacted DV approach offers greater flexibility in terms of model specification than the forecasting approach. By construction, the forecasting approach only allows taking into account price determining factors that affect both, the affected transactions and the comparator transactions. However, there often are price determining factors that only affect the affected transactions but not the comparator transactions, e.g. dummies for product characteristics that only existed during the infringement period. Such factors can only be taken into account with the interaction approach (or, for that matter, with the single DV approach) but not with the forecasting approach.[18] If such factors exist and are taken into account, the fully interacted DV and the forecasting approach do not give equivalent results, of course.

That said, it should be noted that this particular specification freedom also carries a risk of overfitting that accordingly is not present in the forecasting approach. Overfitting denotes the practice of adding more explanatory variables in order to obtain a better in-sample fit without adding or even decreasing out of sample explanatory power. The freedom to add explanatory variables that only affect the affected transactions could thus be abused to spuriously increase or decrease an overcharge estimate. Only using variables with explanatory power also for the comparator transactions may force some discipline against this kind of overfitting, so that the use of variables only relevant for the affected transactions should occur with caution.[19]

5.4 Attribution of interaction effects

A final issue concerns the possibility that some or all interaction effects, i.e. the observed changes in the relation between price and other price determining factors between affected and comparator transactions, are not caused by the infringement. For instance, when a before-during-after comparator analysis is used, the transition between e.g. a cartel

---

[18] The interaction approach in this case will only produce an unbiased estimate of the average overcharge if it is assumed that the relationship between price and this factor is not affected by the infringement.

[19] A more general discussion of model specification is beyond the scope of this note.

and the competition period may be marked by a change of competition at the downstream level, which is not causally related to the infringement (albeit the start or the end of the infringement may be due to these changes). Differences in, for instance, the cost pass-through rates of cartelists between the two regimes may then derive from causes that are not related to the infringement. If this is true these estimated interaction effects should not contribute to the average overcharge.

As the fully interacted DV model makes these interaction effects explicit it is simple to ignore them for the purpose of overcharge calculation; it is also possible to select only those interactions that are deemed to be caused by the cartel. In other words, given the individual overcharge estimated with the fully interacted DV model

$$\widehat{oc_{inter_i}} = d + c * X_i \ ,$$

if all interaction effects are deemed not to be caused by the infringement, the correct estimate for the average overcharge is simply the coefficient d. It is important to note, that this coefficient d is not equivalent to the coefficient on the dummy in the single DV model, denoted with f above. The coefficient f from the single DV model will implicitly take into account all interactions, albeit in a wrong way via the omitted variable effect, whether the interactions are caused by the infringement or not.

If the interactions with a subset, Z, of the factors in X, are deemed to be caused by the infringement, the overcharge is simply calculated by only considering these factors Z:

$$\widehat{oc_{inter_i}} = d + c_Z * Z_i .$$

These choices and calculations are not possible with the forecasting approach as it automatically takes into account all interaction effects. Needless to say, the single DV model does not allow these choices either, as it does not exhibit the interaction effects (but, as noted above, takes them into account in some way).

These considerations naturally give rise to an intermediate approach, i.e. an interacted DV model but not a fully interacted DV model, that may be more tractable given data availability than the fully interacted DV model but at the same time more in line with plausible assumptions about the market in question than the single DV approach. Still, once again such flexibility should be used with caution, notably as the assumption as to what differences are attributable to the infringement may not be testable and must then rely on economic considerations alone. That said, the fully interacted DV model may still provide a good starting point also for this intermediate approach as it encourages making assumptions about missing interactions explicit.

6 Conclusion

One conclusion from the clarifications in this note is that, where possible, the interaction approach may provide a natural starting point for overcharge estimation when using linear regression. It encompasses both the forecasting model and the dummy variable model without interactions as special cases if certain testable conditions hold. If some or all of these conditions are not rejected, a more parsimonious model without (any or some of the) interactions may be estimated. If these conditions are rejected, however, where feasible the interaction approach may be more appropriate to obtain unbiased estimates of the average overcharge for the affected transactions.

The interaction approach also allows the choice of attributing some interactions but not others to the infringement and thus to take them into account for overcharge calculation accordingly. This is not possible either with the forecasting or the single DV approach. Such flexibility needs to be handled with care, however, and the respective choices should obtain an economic underpinning.

A potentially interesting difference between the fully interacted DV model and the forecasting approach concerns the use of observed (in the forecasting method) vs. estimated (in the fully interacted DV model) infringement prices to calculate overcharges. While this is of no concern when the average or total overcharge is the estimand of interest, when instead there is a particular focus on accurately estimating individual overcharges, e.g. time-specific overcharges, notably so as to calculate interest more precisely, also this difference between methods should be taken into account. In this regard it is noteworthy that the fully interacted DV model (or for that matter also an intermediate DV model) allows the choice of including unexplained price components into the overcharge, whereas the forecasting approach does so automatically.

A References

Cameron, A.C. and P.K. Trivedi, “Microeconometrics: Methods and Applications”, Cambridge University Press, New York, USA, 2005.

Greene, W., “Econometric Analysis, Fourth Edition”, Prentice Hall, New Jersey, USA, 2000.

Higgins, R.S. and P.A. Johnson, “The mean effect of structural change on the dependent variable is accurately measured by the intercept change alone“, Economics Letters 80, 2003; pp. 255-259.

Holland, P.W., “Statistics and Causal Inference”, Journal of the American Statistical Association, Vol. 81, No. 396, Dec. 1986, pp. 945-960.

Imbens, G.W. and D.B. Rubin, “Causal inference for statistics, social, and biomedical sciences, An introduction”, Cambridge University Press, New York, USA, 2015.

McCrary, J. and D.L. Rubinfeld, „Measuring benchmark damages in antitrust litigation", Journal of Econometric Methods 2014, (3) 1; pp. 63-74.

Salkever, D., "The use of dummy variables to compute predictions, prediction errors, and confidence intervals.", Journal of Econometrics 4 (4), 1976; pp. 393-397.

White, H., "Time Series Estimation of the Effects of Natural Experiments." Journal of Econometrics, Volume 135, Issue 1-2, November-December 2006; pp. 527 – 566.

Wooldridge, J.W., "Econometric analysis of cross section and panel data", The MIT Press, Cambridge, Massachusetts, 2002.

B Non-equivalence of the fully interacted DV and the forecasting method in case of varying quantity

If quantity is not constant across transactions the formula for the average overcharge from the forecasting approach is

$$\overline{\widehat{oc_{fc}}} = \frac{1}{Q_{infr}} \sum_{i=1}^{N_{infr}} q_i * \left(p_{infr_i} - \widehat{p_{cf_i}}\right)$$
$$= \frac{1}{Q_{infr}} \sum_{i=1}^{N_{infr}} q_i * \left(p_{infr_i} - a_{comp} - b_{comp} * X_{infr_i}\right).$$

Plugging in the least squares decomposition for p_{infr_i} gives

$$\overline{\widehat{oc_{fc}}} = \frac{1}{Q_{infr}} \sum_{i=1}^{N_{infr}} q_i * \left(a_{infr} + b_{infr} * X_{infr_i} + e_{infr_i} - a_{comp} - b_{comp} * X_{infr_i}\right)$$
$$= \frac{1}{Q_{infr}} \sum_{i=1}^{N_{infr}} q_i$$
$$* \left((a_{infr} - a_{comp}) + (b_{infr} - b_{comp}) * X_{infr_i} + e_{infr_i}\right)$$
$$= (a_{infr} - a_{comp}) + (b_{infr} - b_{comp}) * \frac{1}{Q_{infr}} \sum_{i=1}^{N_{infr}} q_i * X_{infr_i}$$
$$+ \frac{1}{Q_{infr}} \sum_{i=1}^{N_{infr}} q_i * e_{infr_i}.$$

With varying quantity, the overcharge from the interaction approach is

$$\overline{\widehat{oc_{inter}}} = \frac{1}{Q_{infr}} \sum_{i=1}^{N_{infr}} q_i * \widehat{oc_{inter_i}} = \frac{1}{Q_{infr}} \sum_{i=1}^{N_{infr}} q_i * (d + c * X_{infr_i})$$
$$= d + c * \frac{1}{Q_{infr}} \sum_{i=1}^{N_{infr}} q_i * X_{infr_i},$$

which after substituting d and c, as shown in the main text, is equivalent to

$$\overline{\widehat{oc_{inter}}} = (a_{infr} - a_{comp}) + (b_{infr} - b_{comp}) * \frac{1}{Q_{infr}} \sum_{i=1}^{N_{infr}} q_i * X_{infr_i}.$$

Comparing the overcharge expressions from the forecasting and interaction approaches shows that

$$\overline{\widehat{oc_{fc}}} = \overline{\widehat{oc_{inter}}} + \frac{1}{Q_{infr}} \sum_{i=1}^{N_{infr}} q_i * e_{infr_i}$$

Thus the average overcharges from both approaches differ if quantity is not constant and the correlation between quantity and the fitted residual from a linear regression is not zero. This result is due to McCrary and Rubinfeld (2014).

Such correlation could arise when price is affected by unobserved variables and when these unobserved variables also affect demand directly. This observation still leaves the question unresolved which of the two estimators are preferable in the case of varying quantity. McCrary and Rubinfeld (2014) provide a result for a particular, albeit often plausible case. They show that when quantity is correlated with unobservable price determining factors but this correlation is the same for the actual and counterfactual price, the estimated overcharge from the interaction approach, $\overline{\widehat{oc_{inter}}}$, is consistent for the average overcharge (ore equivalently, total overcharge) whereas the estimate from the forecasting approach, $\overline{\widehat{oc_{fc}}}$, is not.