On the effects of ignoring parameter uncertainty in antitrust benchmark comparison

Erik Lindén*

Department of Economics and Business Economics & CREATES, Aarhus University

Asger Lunde† ‡

Department of Economics and Business Economics & CREATES, Aarhus University

February 18, 2019

Abstract

In this paper, we discuss the implication of using standard testing procedures, e.g. the t-test, when comparing predicted (forecasted) values to observed values in counterfactual analysis. Standard testing procedures only consider normal sampling uncertainty where the limiting distribution depends on known population parameters, and hence ignore the additional level of uncertainty encompassed in the estimated parameters used for prediction. Any test using standard normal tables, not taking the uncertainty of the estimated parameters into consideration, may lead to a rejection of the null of a zero-mean difference too often. Here we show that it is possible to account for the estimation uncertainty, and thereby to provide reliable test statistics, by adjusting the standard test statistics with a factor that depends on the parameter estimation (i.e. on the explanatory variables in both the estimation- and the prediction period).

Keywords: Forecasting, prediction uncertainty, hypothesis testing, test correction, benchmark analysis, infringement effect estimation

*Corresponding author. E-mail: `e.linden@econ.au.dk` Mailing address: CREATES, School of Business and Social Sciences, Aarhus University, Fuglesangs Allé 4, 8210 Aarhus V. Denmark

†E-mail: `alunde@econ.au.dk`

‡We thank Anders Rahbeck and Giorgio Mirone for useful comments and discussion on early versions on this paper. The usual disclaimers apply.

Contents

1. Introduction

The presumption that competition law violations inflict financial damage on trading partners, has put the quantitative assessment of financial harm in a central position in antitrust litigations'. The objective of the quantitative assessment is to establish the causal link between the law infringement and the financial harm, and hence to evaluate the magnitude of the harm.

Even though financial harm can result from a variety of illegal acts, the calculation of the economic damage typically involves a description of the evolution of the particular economic environment as it would have prevailed in the absence of the law violations (i.e. in a counterfactual world).

To focus the antitrust discussion in this the paper and thereby give it a conceptual context, we consider the antitrust violation at issue to involve price fixing (i.e. a price fixing cartel),[1] where the cartel is considered to have been detected and resolved, and the question is to evaluate its impact (i.e. to estimate the effect the infringement have had on the particular product price in question.)

To isolate the infringement effect from other factors also affecting the evolution of prices, the statistical framework of regression analysis has become an essential tool (Finkelstein & Levenbach (1983)). Using regression analysis, the relation between the product price and its main determinants can be established and used to predict the counterfactual prices that would have prevailed in the infringement period (van Djik & Verboven (2008)).

The exercise of making a counterfactual study and evaluate the infringements effect on prices typically involves a comparison of the product price, either across treated- and untreated economic entities, in a yardstick or cross-section analysis, or for the same economic entities across two distinct time periods with treated- and untreated observations in a benchmark or time-series analysis. The latter is often referred to as a before-during-after analysis made on empirical observations gathered over time, where the price level of the market during the infringement is compared to its price level in a control period before and/or after the infringement. This benchmark approach has become one of the most frequently used methods for assessing the effect of anti-competitive conduct (McCrary & Rubinfeld (2014)) and outlines the framework of this paper.

A standard procedure, advocated and widely used among practitioners, is to evaluate the price impact in a three-step forecasting procedure (White et al. (2006), Godek (2011)). *First*, one estimates a regression model specified to 'explain' the evolution of the product price using data solely from the unharmed benchmark period. In a *second* step, the estimated parameters are used to forecast a counterfactual price that would have prevailed during the infringement in its absence. In the *third* and final step, the predicted price is compared to the actual values observed during the infringement, and the difference, if statistically significant, is considered the effect of the law violation (Davis & Garces (2010)).[2]

[1] A cartel is a competition law violation by object, where colluding parties agree on prices, share market, or share information that illegally strengthen their market power on behalf of their trading partners, an effect typically referred to as increasing prices. As such law infringements do not offer any welfare improving trade-offs against the anti-competitive effects, they are prohibited in most competition regulations around the world (Huschelrath et al. (2013)).

[2] Another widely used approach, often compared to the forecast approach, is to estimate a regression model for the entire period for which data is available, ideally before-during-after the law infringement, using the so-called *dummy variable approach*. The magnitude to which extent the infringement has affected product prices is measured by the coefficient on a dummy variable that differentiate the impact period from benchmark period considered unharmed by the infringement. If statistically significant, the coefficient represents the magnitude of the price effect caused by the infringement. Even though there are quite strong opinions (e.g. White et al. (2006)) in the literature on which of the

To determine whether the predicted prices are statistically different from the observed prices, simple statistical test, e.g. the standard t-test, is typically used (Oxera (2009), p.50). If the test rejects the null of a common mean value, the predicted value can be considered different and it is, in a statistical sense, safe to conclude that the law infringement has had an impact on the product price equal to the difference between the predicted and observed product prices. However, inference using standard test statistics is only valid in the case where population parameters are known.

In the case of comparative benchmark analysis where one of the series (the counterfactual price) is predicted, the population parameters of this series are not known. This follows the fact that the series is predicted using estimated parameters and thereby entail an estimation uncertainty, and this uncertainty is not accounted for in the standard testing procedure (West & McCracken (1998), McCracken (2000), McCracken & West (2001), West (2001*a*,*b*)).

Hence, a comparison between predicted and observed values can be thought of as a comparison between two series with a different relative uncertainty, where only the parameters of the observed values are known. Even though the forecasting approach for assessing anti-competitive effects is widely used, the issue of estimation uncertainty is neglected in the antitrust literature.

McCracken & West (2001) show that the error that is induced when estimating regression parameters affects the asymptotic distribution of the predicted out-of-sample test statistics. This means that forecasts and forecast errors that depend upon estimated parameters do not have the same statistical properties as their population counterparts, and this uncertainty has to be explicitly accounted for to enable valid inference. Testing procedures that rely on Gaussian normal tables with a limiting distribution that depends on the parameters do not, by construction, take this additional level of uncertainty into consideration, and overlooking the issue will lead to an increase of Type I errors, with a too frequent rejection of the null of zero-mean difference. Accordingly, damage calculations based on such results, comparing predicted and actual prices, will erroneously find a significant price difference too often. Given that the estimation uncertainty typically is not accounted for in the competition literature, too much statistical confidence has been addressed the results of the forecasting approach's ability to accurately predict counterfactual prices.

As the estimation error is known, it is possible to incorporate it as a factor in the standard testing procedure. The factor depends directly on the explanatory variable in both the estimation period and the period for prediction, and thereby encompasses the uncertainty inherited from the estimation. By dividing the standard test statistics by this factor, the additional uncertainty that comes with the estimation is controlled for and the testing procedure thereby provides reliable results.

On the terminology used: we frequently refer to the different parts of the sample as the "*estimation period*" and the "*prediction period*", where the former refers to the in-sample estimation, while the latter refers to the out-of-sample estimation. The out-of-sample period is considered for "*prediction*", where prediction refers both to forecasting and backcasting. Further, the term "*estimation uncertainty*" is used when we refer to the uncertainty surrounding the estimated parameters that are used to make the predictions.

approaches yields the most correct approximation of the counterfactual price, it is worth noting that both methods yield numerically identical results when using the same covariates and (i) quantities are constant over the conspiracy period or (ii) the model used is quantity weighted (McCrary & Rubinfeld (2014)).

2. Model specification

2.1. Economic outset

There is a general presumption that illegal price fixing is harmful to society and leads to financial damages for the conspiracy's trading partners. This has made cartels prohibited and punishable in most competition legislation around the world (Huschelrath et al. (2013)). Nevertheless, price fixing schemes are detected, resolved and punished on a rather frequent basis all over the world's jurisdictions (Harrington (2005)).

The aim of price fixing is to maintain or increase prices. Independent of the strategy of the colluding parties to do so, their behavior aims to distort the competition mechanism in the market and thereby decouples the price from its determinants. In turn, this inflicts financial damages on trading partners in form of overcharges, an artificial markup paid by the trading partners, that has noting to do with the development of the price in relation to its determinants.

2.2. A back-of-the-envelope approach to damages

A widely used method for calculating a cartel effect, i.e. the overcharge, is to use historical time series on product prices and compare the prices during the infringement to prices from the period just before, and/or after, the infringement in a *before-during-after analysis* (Davis & Garces (2010)). This approach simply considers the price in the benchmark period to be competitive and thereby to constitute a counterfactual price not affected by the law infringement. Even though the actual financial damage falls outside the scope of this paper, it is relevant to know where the quantification of the effect enters the damage equation. The financial damage "FD" can in its simplest form be calculated as the average price difference between the periods, i.e. the infringement period "N" and the non-infringement period "$M-N$", multiplied by the amount traded during the infringement "Q"

$$FD = (P_{t|N} - P_{t|M-N})Q_{t|N}. \tag{1}$$

Even though this model might be a rather drastic simplification of the actual price formation in the market, it may yet provide a relatively good approximation of the damage inflicted on trading partners. At least under the condition that the infringement was stable and that the price determinants did not change too much between the periods. However, the downside of the approach is that these conditions are seldom fulfilled, and thereby model (1) might be completely erroneous (Rubinfeld (2009)).[3] Ignoring the effect of all other price determinants (i.e. any demand, supply and cost factor) may bias the effect in an arbitrary direction where the effect of the infringement is blurred. Hence, a more elaborate modeling of the price formation have been advocated by empirical economists as well as other practitioners. This have given regression analysis a central position in the court room.

[3] Nevertheless, it is worth noting that most European cartel decisions are based on such simple price comparisons (see Jean Francois Laborde, 2017, Cartel damage claims in Europe: How courts have assessed overcharges, Concurrences 4-2017).

2.3. The regression model

Regression analysis has the aim of predicting the prices that would have prevailed in the absence of the infringement while controlling for other factors also causing fluctuations in the product price, and thereby constitutes a systematic framework for isolating the price effect of the infringement (Finkelstein & Levenbach (1983)).

Generally the staring point of the analysis is a reduced form price equation, where the product price is modeled as a function of its main determinants, i.e. its main demand, supply and cost factors. In this study, as well as in many previous studies on price fixing (Boswijk et al. (2018), McCrary & Rubinfeld (2014), Niebering (2006)), a reduced form price model is throughout considered. However, for simplicity, and in contrast to the just mentioned studies, the model we use here ignore dynamic effects and assume that the product price evolves according to the data generating process

$$Y_t = \alpha + \beta' X_t + \theta \delta_t + \varepsilon_t, \tag{2}$$

where Y_t is the actual realization of the product price in period t. X_t constitutes a set of demand, supply, and cost factors explaining the fluctuations in prices. δ_t is a binary shift dummy variable indicating the period of conspiracy and thereby constitutes the time dependent difference between two, or more, consecutive sub-samples of the time series with affected and unaffected prices. The term is defined as $\delta_t = \mathbf{1}(t \in [T_B, T_E]) = \mathbf{1}_{[T_B, T_E]}(t)$ where $\mathbf{1}$ is an indicator function spanning the period N, where T_B and T_E are the first and last dates of the infringement, respectively. In this paper, only one single infringement period is considered. The coefficient θ is the magnitude of the effect the infringement has had on prices in period t. The error term, ε_t, is assumed independent and identically distributed (iid) normal. Finally, the constant α represent the level of measurement of the price series which is, typically, different from zero.

The conspiracy is assumed to not have caused changes in the price determinants. If the infringement did affect the effects of the covariates in the model, as described in (2), the model would be misspecified. It is though likely that the relation between the price and its determinants is different in the period of conspiracy, for example, in markets or time periods where excess profits are being dissipated over time through change in market conditions or market concentration, this would result in a different relation between X_t and Y_t (Rubinfeld (2009)).[4]

In counterfactual terms, the reduced form price model in Equation (2) can be thought of as the two counterfactual outcomes

$$Y_t(1) = \alpha + \theta + \beta' X_t + u_t \tag{3}$$

$$Y_t(0) = \alpha + \beta' X_t + v_t, \tag{4}$$

where $Y_t(1)$ is the price as it evolves under conspiracy conditions in period N, while $Y_t(0)$ is the price formation under normal form of competition in the period $M - N$. Both the error terms u_t

[4] However, it is worth noting that it is possible to allow the relation between Y_t and X_t to be different in the two periods. McCrary & Rubinfeld (2014) show that by including a term $\gamma' \delta_t X_t$ in the model, whereγ takes on the value 0 or 1 in the non-infringement period and the infringement period, the effect of the difference in the relation between the price and its determinants (between the two periods) can be captured without letting the conspiracy to have any causal effect on the level of X_t.

and v_t are zero mean residuals uncorrelated with the covariates in X_t (following the formulation in Rubin (1974), Imbens (2004)). To fulfill the assumption above, $E[\varepsilon_t \delta_t] = 0$, and the same holds for the errors u_t and v_t, which also are uncorrelated with the conspiracy. Under this formulation, the observed market price evolves according $Y_t = \delta_t Y_t(1) + (1-\delta_t) Y_t(0)$. The estimated financial damage in this setting is calculated as an average markup during the infringement times the quantity traded. Hence, if $\hat{Y}_t(1)$ is the estimated average price during the infringement, $\hat{Y}_t(0)$ is the estimated average price in the non-infringement period, and Q_t represents the volumes traded, the estimated financial damage $\hat{FD}$ is given by

$$\widehat{FD} = \frac{1}{M} \sum_{t=1}^{M} \delta Q_t \left\{ \hat{Y}_t(1) - \hat{Y}_t(0) \right\}.$$

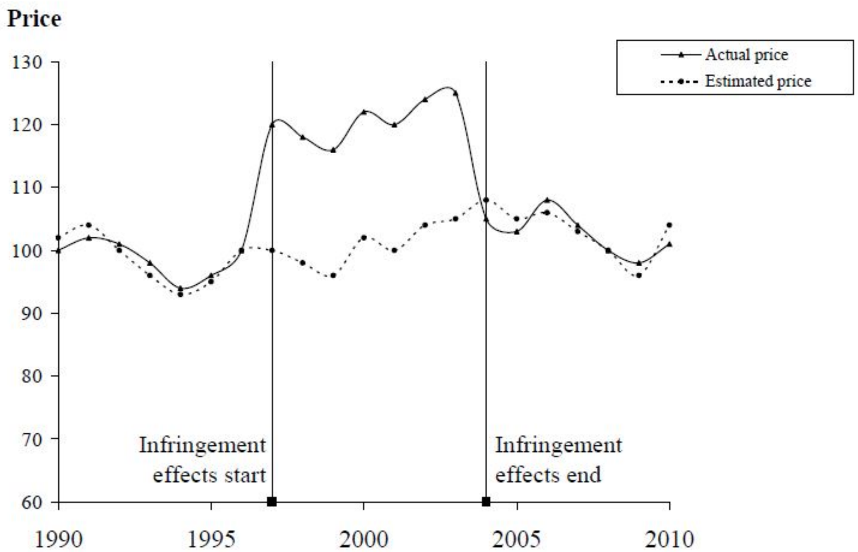
2.4. The forecasting approach to damages

Predicting counterfactual prices for the infringement period using the forecasting approach is a three step procedure. In the *first* step, the competitive price model (4) is estimated for the benchmark period, i.e. in-sample. In the *following* step, the estimated parameters from the first step, representing the relation between the price and its main determinants, are used to predict the counterfactual price that would have prevailed in the infringement period, i.e. out-of-sample, in the absence of price fixing. In the *third* and final step, the predicted counterfactual prices are compared to the observed prices in the infringement period, where the difference corresponds to the overcharge paid by the trading partners. Hence, if Y_t denotes the observed price, Q_t denotes the quantity of sales, and the out-of-sample prediction for period t is $Y_t(0)$, the predicted financial damage (PFD) is given by

$$\widehat{PFD} = \frac{1}{M} \sum_{t=1}^{M} \delta_t Q_t \left\{ Y_t - \hat{Y}_t(0) \right\}.$$

In Figure 1, the second and third steps are displayed. The price model describing the evolution of the product price under normal forms of competition, corresponds to the dotted line in the figure. In the in-sample estimate, i.e. the period unharmed by the infringement, the price is tracing the observed prices (the continuous line in the figure) quite well. In the infringement period, the predicted price and the actual observed price are clearly separated from each other. The distance between the two series is the estimated overcharge, and subtracting the predicted price from the actual price gives the magnitude of the overcharge.

Figure 1: Conceptual illustration of an infringement effect



Source: European Commission (2013) page 23, point 79

Even though there might be, at least visually, a clear distance between the estimated price and the actual price, it is not from a statistical point of view as evident that a difference exists. An essential part of the third step in the forecasting approach is therefore to verify whether the difference between the predicted prices and the actual prices is statistically significant. This is essential because if the prices cannot be concluded to be statistically different, there is no statistical evidence for the infringement to have affected prices during the infringement period. Consequently this means that even though it is proven that a law violation has taken place, and that the claiming party has the right to sue for damages, it cannot be proven that the infringement actually have had an impact on prices, i.e. the financial damage cannot be concluded to be different from zero.

The statistical uncertainty surrounding the price difference

A widely used procedure to verify the validity of an estimated damage is to perform statistical testing. Typically this is done using a standard t-test (see for example Oxera (2009) page 50), i.e. any statistical hypothesis test where the test statistic follows a student's t-distribution under the null hypothesis, where the relative difference between the observed and predicted price is compared to the underlying variability in the data.

The procedure is essential since in a statistical sense the observed price and the predicted price can only be considered different if one with some certainty can reject the null hypothesis that the mean, μ, of the two series are the same, i.e. reject $H_0: \mu_{Y_t} = \mu_{\hat{Y}_t(0)}$ in favor of the alternative $H_a: \mu_{Y_t} \neq \mu_{\hat{Y}_t(0)}$. In a standard setting, comparing the two samples, the test statistic can be formulated as

$$t = \frac{\frac{1}{N}\left(Y_t - \hat{Y}_t(0)\right)}{\sqrt{\frac{1}{N}\left(s^2_{Y_t} + s^2_{\hat{Y}_t(0)}\right)}}, \tag{5}$$

where s_{Y_t} and $s_{\hat{Y}_t(0)}$ are respectively samples deviation from the mean (see for example Wooldridge (2006)). However, it is straightforward from the formula in (5) that such test statistics rely on the assumption that both series are sample observations. In a forecasting context, where one of the series are predicted based on estimated parameters, this prerequisite can never be fulfilled(!) As the forecasted values are predicted using estimated parameters they also encompass an estimation uncertainty, which in contrast to the ordinary sampling variability, is not taken in consideration in the standard formulation of the test. The consequence of not accounting for this additional level of uncertainty underlying the predicated series is that the test will have a too high rejection frequency. This invalidates the use of the standard test as it too often will find the predicted and the actual series to be statistically different, and hence, too often conclude that an eventual infringement effect is statistically significant, even though it in fact is not.

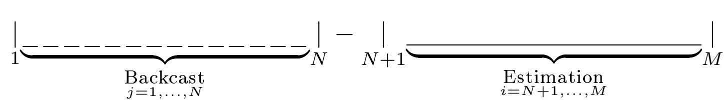
3. Controlling for estimation uncertainty in predicted samples

When the population parameters are known, then the t-statistics is, typically, asymptotically normal with a limiting variance equal to 1. When the population parameters are not known, the limiting variance may be greater than 1. Hence, any test using standard normal tables, without accounting for the additional uncertainty induced by the parameter estimation, will reject the null of a zero-mean difference too often. Here we illustrate this, first by presenting the outline of the test and then by deriving its correction.

3.1. The outline of the test

For simplicity and illustrative purpose, we consider only two time periods with affected and unaffected price observations, respectively. Consequently this means a non-infringement period covering the observations $M - N$, which is considered the estimation period, and an infringement period N considered for prediction (i.e. for forecasting, or in the example below, for backcasting), that is the following situation.[5]

Figure 2: Schematic description of the sample division



To derive the asymptotic distribution of the t-test, comparing the predicted values to the actual

[5]It is worth noting that in the antitrust context this is, in fact, a realistic assumption about the accessibility of data. Often the before-during-after analysis boils down to a during-after analysis where the after period solely constitutes the competitive benchmark.

values, the starting point is the ratio between the length of the prediction period N in relation to the period $M-N$ considered for estimation

$$N,\ M-N \to \infty \quad \text{while} \quad \tfrac{N}{M-N} \to \pi < \infty, \tag{6}$$

where π is a constant in the range $0 < \pi < \infty$.

Taking into consideration the fact that the parameters of the predicted sample are estimated, the test has to be corrected for their influence. To adjust the test for the estimation uncertainty, the usual test statistics has to be divided by a factor that depends on the model in both the estimation sample and the predicted sample, i.e. a factor that depends on the explanatory variables in both periods and in relation to π.

The procedure presented here is a straightforward adaptation of the procedure presented by McCracken & West (2001). We start in the relation in Figure 2 and rewrite it in matrix notation as

$$\begin{aligned} y_j &= \beta' X_j + \varepsilon_j, \qquad & j &= 1, \ldots, N \\ y_i &= \beta' X_i + \varepsilon_i \qquad & i &= N+1, \ldots, M, \end{aligned}$$

where β is a vector with model parameters and X is a vector with covariate(s), in the receptively period, i.e. in the infringement period (j) and the post-infringement period (i). Under the assumption that the variables are stationary and well behaved in line with Diebold & Mariano (1995), West & McCracken (1998), West (2001*b*), we claim:

Claim 1 *The correct t-statistics is $t/\sqrt{\kappa}$ where t is the usual t-statistics for the difference between the actual and predicted value*

$$\frac{1}{N}\sum_{j=1}^{N}(y_j - \hat{\beta}^{\star} X_j)$$

with $\hat{\beta}^{\star}$ estimated on the $i = N+1, \ldots M$ sample, and κ is the influence from the prediction

$$\kappa = 1 + \frac{N}{M-N}\left(\frac{1}{N}\sum_{j=1}^{N} X_j\right)'\left(\frac{1}{M-N}\sum_{i=N+1}^{M} X_i X_i'\right)^{-1}\left(\frac{1}{N}\sum_{j=1}^{N} X_j\right).$$

If the prediction uncertainty equals zero, i.e. the whole second term on the right hand side equal zero, κ equals one, meaning that the asymptotic distribution of the test statistic would be standard normal.

3.2. Derivation of the correction of the test

To prove that Claim 1 is an appropriate way to account for the estimation uncertainty, it is first shown that the prediction uncertainty affects inference, and later it is proven that the asymptotic distribution of the test statistics can only be standard normal if the statistics is corrected for the prediction uncertainty.

Remark 1 *Throughout, the asymptotic arguments are based on*

$$N,\ M-N \to \infty \quad \textit{with} \quad \tfrac{N}{M-N} \to \pi$$

to ensure that the parameters' uncertainty is accounted for in the limiting distribution.

STEP 1: The prediction is based on the estimated relations between the price and its determinants in the benchmark period. The considered model to be estimated is

$$y_i = \beta' X_i + \varepsilon_i, \qquad E(\varepsilon_i) = 0, \quad Var(\varepsilon_i) = \sigma_\varepsilon^2, \quad \text{for} \quad i = N, \dots M \ ,$$

where β and X_i are vectors with the model parameters and covariates, respectively. We consider the ordinary least squares (OLS) estimator which is given by

$$\hat{\beta}^\star = S_{xx}^{\star-1} S_{yx}^\star,$$

where

$$S_{yx}^\star = \frac{1}{M-N} \sum_{i=N+1}^{M} y_i X_i \quad \text{and} \quad S_{xx}^\star = \frac{1}{M-N} \sum_{i=N+1}^{M} X_i X_i^{'}.$$

The superscript * is used to denote quantities that are based on data from the estimation period. Following from conventional OLS results, the usual limiting distribution of the estimator is

$$\sqrt{M-N}(\hat{\beta}^\star - \beta) = \sqrt{M-N}(S_{xx}^{\star-1} S_{\varepsilon x}^\star) \xrightarrow{D} N(0, \sigma_\varepsilon^2 E[XX']^{-1}),$$

where $M-N$ is the observations in the estimation period, and $S_{\varepsilon x}^\star$ is defined analogously to $S_{yx}^\star$.

Now turning to the test statistics.

STEP 2: We consider the usual two sample t-test formulated for testing a zero mean difference between the observed price and predicted price by imposing $H_0: \ \mu = 0$ in the regression

$$\omega_j = \mu + e_j, \quad \text{where } \omega_j = y_j - \hat{\beta}^{\star'} X_j, \text{ for} \quad j = 1, \dots N \ .$$

The t-statistic is

$$t = \frac{\frac{1}{\sqrt{N}} \sum_{j=1}^{N} \omega_j}{\sqrt{\frac{1}{N} \sum_{j=1}^{N} (\omega_j - \bar{\omega})^2}}, \quad \text{with} \quad \bar{\omega} = \tfrac{1}{N} \sum_{j=1}^{N} \omega_j \ .$$

If there was no estimation uncertainty, the asymptotic distribution of this statistic would be standard normal. However, in the present setting one has the following result

RESULT:

$$t \xrightarrow{D} N\left(0, 1 + \pi(E[X]')(E[XX'])^{-1}(E[X])\right). \tag{7}$$

The proof of the result in (7) follows a standard three step procedure where in the initial two steps the limit under H_0 is found separately for the numerator and denominator, respectively. Hence, it is shown that the numerator of the t-statistics is asymptotically normal with limiting variance Ω, while the denominator is shown to converge in probability to $\Omega^{1/2}$. In the final third step, the result follows applying the Slutsky's Theorem[6] and the t-statistic can be concluded to be asymptotically standard normal.

Proof of Result:

First, consider the numerator

$$\begin{aligned}
\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\omega_j &= \frac{1}{\sqrt{N}}\sum_{j=1}^{N}(y_j - \widehat{\beta^\star}' X_j) \\
&= \frac{1}{\sqrt{N}}\sum_{j=1}^{N}(\beta' X_j + \varepsilon_j - \widehat{\beta^\star}' X_j) \\
&= \frac{1}{\sqrt{N}}\sum_{j=1}^{N}(\varepsilon_j - (\hat{\beta}^\star - \beta)' X_j) \\
&= \frac{1}{\sqrt{N}}\sum_{j=1}^{N}\varepsilon_j - (\hat{\beta}^\star - \beta)'\frac{1}{\sqrt{N}}\sum_{j=1}^{N}X_j \\
&= \frac{1}{\sqrt{N}}\sum_{j=1}^{N}\varepsilon_j - \sqrt{\frac{N}{N-M}}\sqrt{M-N}S_{xx}^{\star-1}S_{\varepsilon x}^{\star}\frac{1}{N}\sum_{j=1}^{N}X_j \\
&\xrightarrow{D} N\left(0, \sigma_\varepsilon^2 + \sigma_\varepsilon^2 \pi (E[X])'(E[XX'])^{-1}(E[X])\right),
\end{aligned} \tag{8}$$

where $E[X] \cong \frac{1}{N}\sum_{j=1}^{N} X_j$ which relates to the prediction period, while $E[XX'] \cong \frac{1}{M-N}\sum_{j=N+1}^{M} X_i X_i'$ relates to the estimation period.

Now turning to the denominator.

Remark 2 *First note that* $\frac{1}{N}\sum_{j=1}^{N}(\omega_j - \bar{\omega})^2 = \frac{1}{N}\sum_{j=1}^{N}\omega_j^2 - (\bar{\omega})^2$, *and then that* $\bar{\omega}$ *converges in probability to zero* ($\bar{\omega} \xrightarrow{P} 0$) *and hence* $(\bar{\omega})^2 \xrightarrow{P} 0$.

Hence, it can be shown that $\frac{1}{N}\sum_{j=1}^{N}\omega_j^2$ can be expanded as

$$\frac{1}{N}\sum_{j=1}^{N}\omega_j^2 = \underbrace{\frac{1}{N}\sum_{j=1}^{N}\varepsilon_j}_{\xrightarrow{P}\sigma_\varepsilon^2} + \underbrace{(\hat{\beta}^\star - \beta)'\frac{1}{N}\sum_{j=1}^{N}X_j X_j'(\hat{\beta}^\star - \beta) + 2\frac{1}{N}\sum_{j=1}^{N}X_j'(\hat{\beta}^\star - \beta)}_{\xrightarrow{P}0} \xrightarrow{P} \sigma_\varepsilon^2. \tag{9}$$

[6]**Slutsky's Theorem** Let X_n and $\{Y_n\}$ be sequences with random elements. If X_n converges in distribution to a random element in X; and Y_n converges in probability to a constant c, then $X_n + Y_n \xrightarrow{D} X + c$; $X_n Y_n \xrightarrow{D} cX$; $X_n/Y_n \xrightarrow{D}$ X/c given that c is invertible, and $\xrightarrow{D}$denotes convergence in distribution.

Finally, by applying the Slutsky's Theorem

$$t = \frac{\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\omega_j \xrightarrow{D} N\left(0, \sigma_\varepsilon^2 + \sigma_\varepsilon^2 \pi (E[X])'(E[XX'])^{-1}(E[X])\right)}{\frac{1}{N}\sum_{j=1}^{N}\omega_j^2 \xrightarrow{D} \sigma_\varepsilon^2} \xrightarrow{D} N\left(0, 1+\pi(E[X])'(E[XX'])^{-1}(E[X])\right). \tag{10}$$

Hence, to correct the test statistic for the estimation uncertainty, and thereby yield a standard normal distribution, one has to divide the statistic by the square root of the variance in (10)

$$\frac{t}{\sqrt{\kappa}},$$

where

$$\kappa = 1 + \pi\left((E[X])'(E[XX'])^{-1}(E[X])\right) \tag{11}$$

which can be estimated by

$$\hat{\kappa} = 1 + \frac{N}{M-N}\left(\frac{1}{N}\sum_{j=1}^{N}X_j\right)'\left(\frac{1}{M-N}\sum_{i=N+1}^{M}X_iX_i'\right)^{-1}\left(\frac{1}{N}\sum_{j=1}^{N}X_j\right). \tag{12}$$

4. Numerical results

Given that $\kappa \neq 0$, the uncertainty encompassed in the estimated population parameters is by definition existing, but its effect on the statistical inference using the standard formulation of the test is less obvious. Hence, in this part of the paper, we present results of a small simulation study, investigating the accuracy of the asymptotic approximation of the test correction, and thereby also the correction's practical relevance. In Subsection 4.1, the data generating mechanism used in the analysis is described. This is followed by Subsection 4.2, where results related to the definition of κ are presented. From (12) it can be seen that estimation uncertainty is directly related to the sample size, and the relative difference between the estimation- and prediction period. It can also be seen that the estimation uncertainty is directly related to the covariates. Hence, the starting point of the simulation study is therefore to alter over these factors, and investigate how the estimation uncertainty varies over different simple sizes and the dimension of the sets of covariates. In Subsection 4.3, the basic results are extend by altering the data generating process to show how specific characteristics of the data affect the estimation uncertainty.

4.1. The data generating mechanism

The price, Y_t, formation mechanism is initially designed around a set X_t of covariates including just one covariate, defined as

$$x_t = e_t, \tag{13}$$

where the initial value of the series is set to zero, $X_0 = 0$, and the error term, e_t, is identically distributed standard normal $N(0,1)$. The specification of X_t is trend and autoregressively neutral, meaning that the set of covariates is well behaved. The evolution of the simulated series in the infringement period and the competitive period follows the definition in equations (3) and (4), specifically as

$$Y_t(1) = \alpha + \theta + \beta' X_t + u_t$$

and

$$Y_t(0) = \alpha + \beta' X_t + v_t,$$

where the error terms u_t and v_t are *iid* $N(0,1)$. Hence, combining these two equations the model forming the product price can be summarized as

$$Y_t = \alpha + \beta X_t + \theta \delta_t + \varepsilon_t, \tag{14}$$

where $\varepsilon_t = \delta_t u_t + (1-\delta_t)v_t$, δ_t indicate the infringement period (which obviously is analog to the prediction period, N). The parameters in model (14) are from the outset selected as $\alpha = 1$, $\beta = 1$, and $\theta = 0$, i.e. a framework for testing the empirical size of the test correction.

To mimic empirical sample sizes, both in terms of M and N, typically relevant in the considered context, data series in the range of $25 - 1,000$ observations are simulated. These sample sizes can be thought of as data, in different frequencies, i.e. as yearly, quarterly, monthly, or daily, which are about equally likely to be relevant in different cases. Table 1 also includes a sample size of $10,000$, which is of minor practical interest but is of theoretical interest for the asymptotic validity. This is, however, left out in the other tables as there are simply no significant relevant differences between the results of $1,000$ and $10,000$ observations. Finally, to make the results comparable, the seed of the random number generator is fixed such that the observations are generated equally across the sets of simulations and are thereby replicable.[7]

4.2. Benchmark analysis - altering M, π, and X_t

In the first analysis, the interest lies in the performance of the simple test in relation to its κ-corrected formulation. From Equation (12) in Section 3, the intuition behind the results follows from the relative difference between the estimation period, $M - N$, and the period for prediction, N. If the estimation period is relatively small compared to the prediction period, i.e. for larger π, the uncertainty in the estimated parameters is larger than if the opposite relation is true. This implies that the correction will be of larger importance in small samples where a large share of the sample is predicted, i.e. the importance of the correction is decreasing in relation to the sample size but increasing in relation to π. In addition, larger models with a larger number of covariates also bring larger uncertainty from the estimation.

To investigate the small sample effect of M separated from π, results are initially presented for a fixed π and varying sample size M. In the next step, the size effect is put in relation to the effect of varying π, where the ratio between the prediction period relative the estimation period is altered over the set of sample sizes.

[7]The simulations are performed using Stata 15, with a fixed seed of 2019. The results and thereby the conclusions are not dependent on the seed used.

The last aspect determining the size of κ is its relation to the estimator and the estimated model. As the OLS estimator is considered, the size of κ is investigated in relation to the model estimated, i.e. in relation to the dimension of the set of covariates, X_t. A larger model implies a larger number of estimated parameters and hence a larger κ.

The size of κ decreases in M

To benchmark the simulation study statistical size of the simple test is compared to the corrected test for different sample sizes with a fixed π. The sample sizes are arbitrarily chosen, but as typically relevant in the antitrust context in question. The aim of the analysis is to give an indication of the relation between the size of the sample and how serious it is to ignore the estimation uncertainty encompassed in the predicted values. Table 1 reports the actual size of the simple test and the corrected test, when the critical values are ± 1.96 meaning that the nominal size of the test is equal to 5 %. As a benchmark, the ratio $\pi = 1$, i.e. the estimation and prediction sample have the same size. This is used in all subsequent tables below if not stated otherwise. Table 1 presents the results and also displays the actual difference between the predicted and observed values in the infringement period, i.e. $Y_t - \hat{Y}_t(0)$, to show that the difference between the samples is in fact zero. This zero difference holds in all tables below, and is for simplicity left out from the rest of the tables.

The results show $\kappa > 1$, a result that is relatively stable over the sample sizes. It also shows that the correction of the test significantly improves its performance, where a rejection frequency close to 0.05 is considered accurate. However, this conclusion seems not to be as clear in very small samples, where it can be seen that the corrected test performs slightly worse than the standard test.

Table 1: Actual size of the Simple and Corrected tests

		Rejection frequency (5%)		
M	$Y_t - \hat{Y}_t(0)$	Standard test	Corrected test	κ
25	-0.002	0.189	0.105	1.446
50	-0.008	0.178	0.082	1.415
75	0.004	0.185	0.076	1.424
100	0.010	0.181	0.075	1.415
250	-0.003	0.170	0.060	1.414
500	0.000	0.175	0.060	1.414
1000	0.001	0.175	0.056	1.414
10000	0.000	0.168	0.057	1.414

Note: $\pi = 1$ i.e. equally sized predicted period and estimation period. The results are based on the tests relative rejection frequency of the null of a zero mean difference, even though it is true, when the critical value is ± 1.96. The results are based on 5,000 iterations. The data generating process is defined as (14) with a fixed seed.

The size of κ increases in π

Turning to the relative size of the estimation and prediction period, π, it can be seen from the definition of κ in Equation (11) that its size is dependent on π. As $\pi \to 0$ there is no uncontrolled estimation uncertainty in the simple test, while arbitrarily large π causes an arbitrary large uncontrolled estimation uncertainty, which in turn implies a too large rejection frequency. Hence, in conjunction with π it is assumed that κ is relatively smaller when the π is small relatively to when π is arbitrarily larger.

Dependent on the area of application, π varies for natural reasons (West (2001*b*)). Within the area of antitrust, π is typically large due to regulation. Under European Legislation,[8] the period within which it is possible to bring an action for damages is limited to at least 5 years after it can be regarded as known that a law infringement have occurred (c.f. Directive 2014/104/EU, Article 10, point 2.).[9] This means that in law infringements detected by authorities, potential claimant(s) has to evaluate the price effect of the infringement and bring it to court within 5 years otherwise the limitation period expires. Hellwig & Huschelrath (2016) show that a typical law violation regarding price fixing has on average been active for 8 years (but spans from as little as a couple of months to 35 years).[10] This means that the period after the infringement considered for estimation, is relatively small in comparison to the period for prediction, which implies a typically large π.

Investigating the effect of different size of π shows that the performance of the standard test underestimates the underlying estimation uncertainty and leads the test to reject the null of a zero effect significantly more often than the corrected test (c.f. Table (2)). The bias is less pronounced when the prediction period is relatively small compared to the estimation period.

[8] The Directive was implement in all member states' legal systems by 27 December 2016.

[9] The minimum rule of five years limitation has become the norm in most of the member states' legislation, with some exceptions as, e.g. Scotland, where a six years limitation applies.

[10] Over the observation period from 2001 to 2015, Hellwig & Huschelrath (2016), find that an average infringement (cartel) duration is 97 month, based on 90 individual cartel cases decided by the European Commission.

Table 2: The implication of π variation

Size of κ					
M	$\pi = 9$	$\pi = 3$	$\pi = 1$	$\pi = 0.3$	$\pi = 0.1$
10	13.66	2.466	1.434	1.244	1.249
25	4.258	1.917	1.440	1.297	1.055
50	2.354	1.880	1.422	1.346	1.051
75	2.436	1.895	1.425	1.156	1.057
100	2.606	2.050	1.415	1.156	1.053
250	2.541	2.008	1.414	1.156	1.054
500	3.163	2.000	1.414	1.155	1.054
1000	2.335	2.000	1.414	1.155	1.054

Rejection frequency (5%)										
	Simple test					Corrected test				
M	$\pi = 9$	$\pi = 3$	$\pi = 1$	$\pi = 0.3$	$\pi = 0.1$	$\pi = 9$	$\pi = 3$	$\pi = 1$	$\pi = 0.3$	$\pi = 0.1$
10	0.233	0.262	0.149	0.096	0.089	0.033	0.060	0.122	0.238	0.031
25	0.372	0.250	0.182	0.137	0.045	0.026	0.067	0.094	0.097	0.274
50	0.275	0.256	0.171	0.135	0.055	0.085	0.101	0.096	0.120	0.157
75	0.325	0.279	0.197	0.081	0.060	0.070	0.078	0.086	0.070	0.097
100	0.325	0.329	0.176	0.105	0.070	0.060	0.073	0.072	0.105	0.105
250	0.383	0.323	0.267	0.111	0.067	0.061	0.063	0.174	0.067	0.068
500	0.488	0.328	0.175	0.095	0.056	0.063	0.059	0.060	0.059	0.052
1000	0.356	0.333	0.140	0.099	0.074	0.056	0.052	0.035	0.059	0.060

Note: The results are based on the tests relative rejection frequency of the null of a zero mean difference, even though it is true, when the critical value is ± 1.96. The results are based on 5,000 iterations. The data generating process is defined as (14) with a fixed seed.

The size of κ increases in dimension, i, of X_t

In the results above the evolution of Y_t is designed around just one single covariate, X_t, defined as $X_t = e_t$, where $e_t \sim i.i.d. - N(0,1)$. The fact that just one covariate is included in the mechanism forming the series also means that all the uncertainty in the estimated parameters stems from this single covariate. Hence, extending the set of covariates also increases the estimation uncertainty, i.e. the size of κ, and the need for correction increases. This regardless of the behavior of the covariates. Here the covariates have zero correlation, are not trending, and lack serial correlation.

The set of covariates is extended with samples drawn from a multivariate normal distribution of orthogonal data with mean 0 and variance 1.[11] The inclusion of more covariates increases the size of κ and so the need for correction, c.f. Table 3. In line with the results presented above, the estimation uncertainty is prone to be more severe in small samples.

[11] It is worth noting that this result is unaffected by the correlation between the covariates in X_t, when the number of covariates is larger than three. For a lower number of covariates, a lower degree of correlation implies a larger κ. For full results on multicolinearity see Appendix I and Tables 7, 8, and 9 within.

Table 3: The implication of variation in dimension of X_t

		Size of κ		
M	$i=1$	$i=3$	$i=5$	$i=10$
10	1.460	1.826	14.963	∞
25	1.449	1.518	1.845	3.764
50	1.416	1.480	1.552	1.798
75	1.425	1.440	1.511	1.648
100	1.414	1.371	1.477	1.568
250	1.414	1.378	1.438	1.469
500	1.414	1.353	1.426	1.441
1000	1.414	1.417	1.420	1.426

	Rejection frequency (5 %)							
	Simple test				Corrected test			
M	$i=1$	$i=3$	$i=5$	$i=10$	$i=1$	$i=3$	$i=5$	$i=10$
10	0.108	0.130	0.074	0.089	0.140	0.055	0.008	0.000
25	0.189	0.159	0.146	0.140	0.101	0.055	0.025	0.002
50	0.181	0.161	0.163	0.145	0.085	0.060	0.046	0.016
75	0.185	0.169	0.179	0.174	0.075	0.062	0.057	0.033
100	0.180	0.171	0.179	0.161	0.070	0.086	0.065	0.038
250	0.170	0.149	0.172	0.176	0.060	0.096	0.055	0.049
500	0.175	0.157	0.174	0.151	0.060	0.082	0.056	0.070
1000	0.177	0.153	0.163	0.172	0.055	0.042	0.048	0.053

Note: $\pi = 1$ i.e. equally sized predicted period and the estimation period. The results are based on the tests' relative rejection frequency of the null of a zero mean difference, even though it is true, when the critical value is ± 1.96. The results are based on 5,000 iterations. The data generating process is defined as (14) with a fixed seed.

4.3. Extended analysis - altering the data generating process

The first part of the simulation study investigates the relation between the amount of estimation uncertainty, κ, and the sample size, M, the ratio between the prediction and the estimation period, π, and the dimension of the set of covariates, X_t. In turn, this part of the simulation study relates to the data generating process and investigates how the statistical features of the time series affect the estimation uncertainty. Here, the interest lies in the performance of the simple test in relation to its corrected formulation, when the underlying data possess, for empirical data, common statistical features.

From the definition of κ, in Equation (12), it is straightforward to see that the variance of the error term in the data generating process (14) does not affect the estimation uncertainty.[12] Nevertheless, it is not straightforward to see how the estimation uncertainty is affected by a trending or serially correlated behavior among the covariate(s) in the model.

The size of κ increases both if the covariate(s) is trending or if it is serially correlated

Empirical data typically possess the properties of being trending and/or serial correlated. Both these features bias the estimation and may cause spurious results, and are therefore typically controlled for in applied work. However, sometimes these issues are ignored or failed to be controlled for, and

[12]See Appendix I for simulation results related to the error variance.

in this context it is not obvious how they affect the estimation uncertainty. Hence, this subsection further investigates these features.

Here, the definition of the covariate(s) in (13) is expanded, which are now considered to evolve according to

$$X_t = \tau t + \rho X_{t-1} + e_t, \tag{15}$$

where e_t is an *iid* error with mean zero and variance one, τ is the slope of the deterministic linear time trend around which the covariate is randomly generated, while ρ is the autoregressive term indicating to what extent the previous value adds to the current realization of the time series. The results are presented in Table 4, where each feature is dealt with separately, i.e. for the left panels, related to τ, the serial correlation is set to zero, and vise-versa for the panels on the right, related to ρ.

Starting with the former feature, the effect of τ on κ, a trending behavior violates the constant mean assumption which implies that independent drawing from the series at two different points in time will have a different mean. For κ, a trending behavior will be more pronounced in larger samples as, in direct relation to the magnitude of τ, the absolute difference between the mean value in the estimation period $(M-N)$, and prediction period (N) will be larger in larger samples. On the other hand, a trending behavior may cause two unrelated time series, but with the same or similar trend, to appear related to each other even though they in fact are not.

The serial correlation, the effect of ρ on κ, implies that each realization of the series consists of a fraction ρ of previous value. A shock to such a series will, as long as $|\rho| < 1$, vanish in time. In the present context, the serial correlation will therefore only affect the size of κ, in smaller samples. This follows from the fact that information is carried from the in-sample period $(M-N)$ of estimation into the out-of-sample period (N) used for prediction. For this reason, serial correlation will have an effect on κ in small samples, but not in larger samples.[13]

In terms of rejection frequency, the simple test is more sensitive to a trending behavior of the covariate than its degree of serial correlation. With a moderate trend, i.e. $\tau = 0.05$ or $\tau = 0.1$ the simple test falsely rejects the null of a zero mean difference in about 50 % of the cases compared to the true 5 %. Compared to a moderate, or even high, serial correlation of $\rho = 0.25 - 0.75$, where the rejection frequency is around 16 %. This indicates that the trending behavior of the covariate is a larger problem for the test performance. Nevertheless, for both cases, where τ and ρ are different from zero, the corrected test performs fairly accurate with values close to 0.05.

[13]Note that the Unit Root environment, i.e. $\rho = 1$, is here excluded and only a stationary environment where $|\rho| < 1$ is considered.

Table 4: The implication of trend or serial correlation in the covariate

Size of κ								
	Trend				Serial correlation			
M	$\tau = 0$	$\tau = 0.015$	$\tau = 0.05$	$\tau = 0.10$	$\rho = 0$	$\rho = 0.25$	$\rho = 0.50$	$\rho = 0.75$
10	1.430	1.443	1.469	1.541	1.439	1.507	1.582	1.709
25	1.431	1.459	1.578	1.889	1.449	1.325	1.558	1.603
50	1.426	1.460	1.831	2.434	1.413	1.325	1.337	1.512
75	1.428	1.492	2.192	2.910	1.419	1.326	1.333	1.489
100	1.415	1.595	2.390	3.118	1.415	1.334	1.251	1.472
250	1.419	1.799	3.241	3.582	1.629	1.330	1.301	1.447
500	1.414	2.416	3.603	3.685	1.492	1.666	1.418	1.427
1000	1.414	2.237	3.437	3.721	1.436	1.479	1.666	1.421
Rejection frequency (5 %)								
Simple test								
	Trend				Serial correlation			
M	$\tau = 0$	$\tau = 0.015$	$\tau = 0.05$	$\tau = 0.10$	$\rho = 0$	$\rho = 0.25$	$\rho = 0.50$	$\rho = 0.75$
10	0.152	0.124	0.129	0.140	0.126	0.093	0.152	0.175
25	0.184	0.195	0.221	0.290	0.190	0.156	0.216	0.221
50	0.166	0.163	0.289	0.400	0.188	0.153	0.150	0.201
75	0.185	0.201	0.368	0.474	0.177	0.149	0.152	0.195
100	0.175	0.230	0.396	0.512	0.180	0.151	0.128	0.195
250	0.177	0.270	0.554	0.582	0.148	0.147	0.138	0.188
500	0.171	0.397	0.590	0.616	0.171	0.151	0.176	0.177
1000	0.173	0.332	0.557	0.580	0.179	0.168	0.146	0.176
Corrected test								
	Trend				Serial correlation			
M	$\tau = 0$	$\tau = 0.015$	$\tau = 0.05$	$\tau = 0.10$	$\rho = 0$	$\rho = 0.25$	$\rho = 0.50$	$\rho = 0.75$
10	0.114	0.155	0.153	0.155	0.155	0.093	0.168	0.160
25	0.086	0.107	0.109	0.118	0.103	0.090	0.110	0.094
50	0.095	0.123	0.090	0.085	0.088	0.073	0.079	0.073
75	0.079	0.091	0.085	0.087	0.065	0.068	0.063	0.068
100	0.073	0.082	0.072	0.086	0.068	0.064	0.065	0.066
250	0.066	0.066	0.064	0.079	0.057	0.059	0.058	0.063
500	0.062	0.070	0.061	0.047	0.059	0.061	0.058	0.057
1000	0.058	0.055	0.051	0.055	0.063	0.057	0.058	0.055

Note: The results are based on the tests' relative rejection frequency of the null of a zero mean difference even though it is true, when the critical value is ± 1.96. The results are based on 5,000 iterations, of the DGP (14) with a fixed seed. In the panels to the left, related to τ, the serial correlation is set to zero, and vice-versa for the panels to the right, related to ρ.

4.4. Mimic empirical data

To mimic the behavior of κ in an empirical environment, the price formation mechanism is designed around a set of five covariates, i.e. $X_t' = [x_1, \cdots, x_3]'$, with individual trending and/or autoregressive behavior. Each x_t in the set of covariates is generated according to:

$$x_{it} = \alpha_i + \tau_i t + \rho_i x_{i,t-1} + e_{it},$$

where, in contrast to the benchmark analysis above, the individual trends of the covariates are arbitrarily set to $\tau' = [0, -0.015, 0.02]'$ and the autoregressive coefficients to $\rho' = [0.25, 0, 0.35]'$, respectively. The initial values of the series are all set to zero, and the error term, e_t, is distributed identically standard normal with mean zero and variance of 1.[14]

The evolution of the price series in the infringement period and the competitive period again follows equations (3) and (4), specifically as

$$Y_t(1) = \alpha + \theta + \beta' X_t + u_t$$
$$\text{and}$$
$$Y_t(0) = \alpha + \beta' X_t + v_t.$$

However, now u_t and v_t are heteroscedastic error terms generated as $u_t = Z_t \tilde{u}_t$ and $v_t = Z_t \tilde{v}_t$ respectively. Both $\tilde{u}_t$ and $\tilde{v}_t$ are distributed *iid* bivariate normal with mean 0 and variance 10. The Z_t is distributed *iid* standard normal with no correlation to $\tilde{u}_t$, $\tilde{v}_t$, and e_t.

Hence, the price series follows the data generating process

$$Y_t = \alpha + \beta' X_t + \theta \delta_t + \varepsilon_t, \tag{16}$$

where $\varepsilon_t = \delta_t u_t + (1 - \delta_t) v_t$, the intercept α is arbitrarily set to 10, and the vector of parameters $\beta' = [1.5, 2, 2.5]'$. Regressing Y_t on the set of covariates X_t, yields a coefficient of determination, R^2, of about one third, which is typical for the context in question (McCrary & Rubinfeld (2014)). It is worth noting that for large values of π and in large samples, R^2 is generally larger.

In this more exotic environment, the same conclusions from above seem to hold straight through, c.f. Table 5, but the set of exotic covariates inflates the size of κ quite significantly. Especially for larger values of π, where the simple test now approaches a 100% rejection frequency, even though there is no difference between the series. The corrected test still performs well, with a rejection frequency close to 5%.

In conclusion, this implies that the estimation uncertainty plays a significant role in the performance of the tests in empirical multivariate regression models. The simple test rejects the null of a zero mean different in the absolute majority of cases, and only approaches the correct 5% rejection frequency in cases where both M and π are very small.

[14] It is worth noting that in empirical data it can often appear that price determinants evolve differently in the period of infringement compared to the benchmark period(s). This behavior can only be modeled in a complex system describing the underlying time series process of the covariates.

Table 5: Mimic an empirical environment

			Size of κ		
M	$\pi = 9$	$\pi = 3$	$\pi = 1$	$\pi = 0.3$	$\pi = 0.1$
25	∞	3.255	1.737	1.301	1.089
50	5.858	2.468	1.538	1.224	1.082
75	4.793	2.312	1.504	1.197	1.083
100	4.050	2.179	1.472	1.185	1.073
250	3.444	2.073	1.437	1.169	1.063
500	3.292	2.032	1.425	1.161	1.059
1000	3.225	2.016	1.420	1.158	1.056

	Rejection frequency (5%)									
	Simple test					Corrected test				
M	$\pi = 9$	$\pi = 3$	$\pi = 1$	$\pi = 0.3$	$\pi = 0.1$	$\pi = 9$	$\pi = 3$	$\pi = 1$	$\pi = 0.3$	$\pi = 0.1$
25	0.423	0.292	0.187	0.104	0.042	0	0.015	0.061	0.104	0.261
50	0.453	0.326	0.197	0.115	0.057	0.006	0.035	0.075	0.090	0.144
75	0.475	0.326	0.194	0.110	0.063	0.009	0.047	0.064	0.079	0.097
100	0.483	0.324	0.184	0.102	0.058	0.021	0.047	0.061	0.066	0.083
250	0.506	0.333	0.187	0.104	0.057	0.046	0.060	0.058	0.063	0.056
500	0.529	0.337	0.172	0.099	0.064	0.050	0.056	0.056	0.055	0.056
1000	0.531	0.329	0.174	0.093	0.066	0.057	0.051	0.056	0.050	0.054

Note: The results are based on the tests' relative rejection frequency of the null of a zero mean difference, even though it is true, when the critical value is ± 1.96. The results are based on 5,000 iterations. The data generating process is defined as (14) with a fixed seed.

5. Conclusions

In this paper we make four contributions. *First*, by giving a brief introduction to the comparative forecasting approach used to assess the effects of a competition law infringement. *Second*, by displaying the issue of estimation uncertainty, and how the issue feeds into the forecasting approach through the estimated parameters used for prediction. It is argued that if this is not controlled for, it may invalidate the confidence on which conclusions are drawn upon. Even though the issue of estimation uncertainty is known in the econometric literature, it is neglected in the competition literature, at least to the knowledge of the author of this paper, which may have caused an over confidence in results produced using the forecasting approach. *Third*, by presenting the theory for how the estimation uncertainty feeds into the testing procedure and how to correctly account for it in standard testing procedures. Hence, also how to provide valid statistical inference. It is shown that there is a relation between size of the uncertainty and the sample size and the relative size between the prediction and estimation period, and finally that the estimation uncertainty depends on the dimension of the set of covariates. *Fourth*, by presenting numerical results, showing the magnitude of the estimation uncertainty in relation to typical prerequisites in the antitrust field, namely the small finite samples, relatively large prediction period relative to the estimation period, multivariate set of covariates, multicollinearity among covariates, and finally trending and serially correlated covariates. The results show that the bias is more prone to affect the test statistic when the prediction period is relatively large compared to the estimation period, but also generally in smaller samples. Especially, when the number of covariates is large and or when the covariate(s) are trending.

By displaying the issue of estimation uncertainty and by showing a relatively easy correction of the issue for standard testing procedures, we hope that we have contributed with a tool that will be used to improve statistical results of the forecasting approach in future applied work.

References

Boswijk, H. P., Bun, M. J. & Schinkel, M. P. (2018), Cartel dating, Working Paper 2016-05, University of Amsterdam - Amsterdam Center for Law & Econmics.

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

Diebold, F. X. & Mariano, R. S. (1995), 'Comparing predictive accuracy', *Journal of Business & Economic Statistics* **13**(3), 253–263.

Directive 2014/104/EU (2014), Directive 2014/104/eu on certain rules governing actions for damages under national law for infringements of the competition law provisions of the member states and of the european union, Legislation, The European Parlament and the Council of the European Union.

European Commission (2013), Practical guide: Quantifying harm in actions for damages based on breaches of article 101 or 102 of the treaty on the functioning of the european union, techreport, European Commission.

Finkelstein, M. O. & Levenbach, H. (1983), 'Regression estimates of damages in price-fixing cases', *Law and Contemprary Problems* .

Godek, P. E. (2011), 'Time-series models for estimating economic damages in antitrust (and other) litigation: the relative merits of predictive versus dummy-variable appraoches', *Competition Policy International - Antitrust chronicle* **1**.

Harrington, J. J. (2005), 'Optimal cartel pricing in the presence of an antitrust authority', *International Economic Review* **46**(1), 145–169.

Hellwig, M. & Huschelrath, K. (2016), Cartel cases and the cartel enforcement process in the european union 2001-2015: a quantitative assessment, Discussion Paper 16-063, Center for European Economic Research.

Huschelrath, K., Muller, K. & Veith, T. (2013), 'Concrete shoes for competition: The effect of the german cement cartel on market price', *Journal of Competition Law & Economics* **9**(1), 97–123.

Imbens, G. W. (2004), 'Nonparametric estimation of average treatment effects under exogeneity: a review', *Review of Economics and Statistics* .

McCracken, M. W. (2000), 'Robust out-of-sample inference', *Journal of Econometrics* **99**, 195–223.

McCracken, M. W. & West, K. D. (2001), Inference about predictive ability, *in* M. P. Clements & D. F. Hendry, eds, 'A Companion to Economic Forecasting', Oxford, U.K.: Blackwell Publishers, chapter 14, pp. 299 – 321.

McCrary, J. & Rubinfeld, D. L. (2014), 'Measuring benchmark damages in antitrust litigation', *Journal of Econometric Methods* **3**(1), 63–74.

Niebering, J. (2006), 'Estimating overcharges in antitrust cases using a reduced-form apparoach: methods and issues', *Journal of Applied Econometrics* **9**(2), 361–380.

Oxera (2009), Quantifying antitrust damages: Towards non-binding guidance for courts, Technical report, European Commission.

Rubin, D. B. (1974), 'Estimating causal effects of treatments in randomized and nonrandomized studies', *Journal of Educational Psychology* **66**(5), 688–701.

Rubinfeld, D. L. (2009), *Research Handbook on the Economics of Antitrust Law*, chapter Antitrust Damages.

van Djik, T. & Verboven, F. (2008), Quantification of damages, *in* 'Competition law and Policy', Vol. 3, ABA Section of Antitrust Law, pp. 2331–2348.

West, K. D. (2001*a*), 'Encompassing tests when no model is encompassing', *Journal of Econometrics* **105**, 287–308.

West, K. D. (2001*b*), 'Test for forecast encompassing when forecasts depend on estimated regression', *Journal of Business & Economic Statistics* **19**(1), 29–33.

West, K. D. & McCracken, M. W. (1998), 'Regression-based test of predictive ability', *International Economics Review* **39**(4), 817–840.

White, H., Marshall, R. & Kennedy, P. (2006), 'The measurement of economic damages in antitrust civil litigation', *Section of Antitrust Law - Economics Committee Newsletter* .

Wooldridge, J. (2006), *Introductory Econometrics - A modern apparoch*, third edition edn, Thomson Higer Education.

Appendix I

The size of κ is unaffected by the error variance

In the price forming mechanism in (14), the price is initially formed around one normally distributed covariate and an error term $\varepsilon_t = \delta_t u_t + (1-\delta_t)v_t$, where u_t and v_t are independent normal components which both have a mean 0 and variance 1, representing the error in the infringement period and the competitive period, respectively. Increasing the variance in the error term ε_t relative to the variation in X_t, downplays the role of the covariates in the prediction of Y_t as less variation will be inherited form the covariate(s). From the definition of κ in (8)-(10), the variance in ε_t does not affect the size of κ, a result confirmed in Table 6. Not even a significant increase in the variance of the error term affects the size of κ.

Table 6: The implication of an increased error variance

	Size if κ		
M	$Var = 1$	$Var = 10$	$Var = 25$
10	1.439	1.440	1.429
25	1.449	1.449	1.434
50	1.413	1.415	1.413
75	1.422	1.422	1.435
100	1.415	1.415	1.415
250	1.414	1.429	1.429
500	1.414	1.431	1.415
1000	1.414	1.415	1.414

	Rejection frequency (5%)					
	Simple test			Corrected test		
M	$Var = 1$	$Var = 10$	$Var = 25$	$Var = 1$	$Var = 10$	$Var = 25$
10	0.126	0.127	0.151	0.155	0.155	0.121
25	0.190	0.190	0.183	0.103	0.103	0.086
50	0.188	0.183	0.188	0.088	0.084	0.088
75	0.178	0.178	0.191	0.070	0.070	0.096
100	0.182	0.182	0.179	0.072	0.072	0.077
250	0.183	0.157	0.159	0.064	0.093	0.085
500	0.172	0.183	0.181	0.064	0.083	0.073
1000	0.170	0.174	0.168	0.063	0.065	0.058

Note: $\pi = 1$, i.e. equally sized predicted period and the estimation period. The results are based on the tests' relative rejection frequency of the null of a zero mean difference, even though it is true, when the critical value is ± 1.96. The results are based on 5,000 iterations. The data generating process is defined as (14) with a fixed seed.

The size of κ is unaffected by multicollinearity

Table 7: The implication of multicolinearety on κ - the *three* variables case

3-dimensional X_t						
			Size of κ			
M	$r = 0$	$r = 0.10$	$r = 0.50$	$r = 0.75$	$r = 0.90$	$r = 0.99$
10	2.433	2.398	1.826	1.744	1.818	1.996
25	2.205	2.281	1.518	1.463	1.459	1.533
50	2.540	2.574	1.480	1.344	1.464	1.481
75	2.521	2.257	1.440	1.332	1.563	1.466
100	2.473	2.563	1.371	1.359	1.445	1.445
250	1.854	2.554	1.378	1.355	1.426	1.426
500	2.550	1.422	1.353	1.242	1.458	1.420
1000	1.659	1.429	1.417	1.430	1.432	1.417

	Rejection frequency (5 %)											
	Simple test						Corrected test					
M	$r = 0$	$r = 0.10$	$r = 0.50$	$r = 0.75$	$r = 0.90$	$r = 0.99$	$r = 0$	$r = 0.10$	$r = 0.50$	$r = 0.75$	$r = 0.90$	$r = 0.99$
10	0.257	0.281	0.130	0.108	0.078	0.104	0.058	0.064	0.055	0.050	0.054	0.053
25	0.308	0.333	0.159	0.139	0.141	0.168	0.062	0.058	0.055	0.057	0.054	0.059
50	0.396	0.384	0.161	0.117	0.177	0.176	0.061	0.064	0.060	0.046	0.068	0.069
75	0.385	0.343	0.169	0.140	0.175	0.182	0.062	0.058	0.062	0.142	0.100	0.065
100	0.385	0.382	0.171	0.133	0.183	0.195	0.061	0.053	0.086	0.128	0.060	0.068
250	0.261	0.382	0.149	0.150	0.177	0.174	0.058	0.061	0.096	0.087	0.070	0.062
500	0.380	0.167	0.157	0.113	0.187	0.175	0.057	0.094	0.082	0.068	0.067	0.062
1000	0.200	0.178	0.153	0.168	0.168	0.170	0.092	0.074	0.042	0.056	0.052	0.049

Note: $\pi = 1$, i.e. equally sized predicted period and the estimation period. The results are based on the tests' relative rejection frequency of the null of a zero mean difference, even though it is true, when the critical value is ± 1.96. The results are based on 5,000 iterations, The data generating process is defined as (14) with a fixed seed.

Table 8: The implication of multicollinearity on κ - the *five* variables case

5-dimensional X_t						
Size of κ						
M	$r = 0$	$r = 0.10$	$r = 0.50$	$r = 0.75$	$r = 0.90$	$r = 0.99$
10	14.963	14.963	14.963	13.732	14.963	14.964
25	1.845	1.845	1.845	1.845	1.845	1.845
50	1.552	1.552	1.552	1.552	1.552	1.552
75	1.511	1.511	1.511	1.511	1.511	1.511
100	1.477	1.477	1.477	1.488	1.477	1.477
250	1.438	1.438	1.439	1.435	1.438	1.438
500	1.426	1.426	1.425	1.426	1.426	1.426
1000	1.42	1.42	1.42	1.42	1.42	1.42

Rejection frequency (5 %)												
	Simple test						Corrected test					
M	$r = 0$	$r = 0.10$	$r = 0.50$	$r = 0.75$	$r = 0.90$	$r = 0.99$	$r = 0$	$r = 0.10$	$r = 0.50$	$r = 0.75$	$r = 0.90$	$r = 0.99$
10	0.074	0.074	0.074	0.084	0.074	0.074	0.008	0.008	0.008	0.013	0.008	0.008
25	0.146	0.146	0.146	0.146	0.146	0.146	0.025	0.025	0.025	0.025	0.025	0.025
50	0.163	0.163	0.163	0.163	0.163	0.163	0.046	0.046	0.046	0.046	0.046	0.046
75	0.179	0.179	0.179	0.179	0.179	0.179	0.057	0.057	0.057	0.057	0.057	0.057
100	0.179	0.179	0.179	0.163	0.179	0.179	0.065	0.065	0.065	0.093	0.065	0.065
250	0.172	0.172	0.198	0.171	0.172	0.172	0.055	0.055	0.023	0.055	0.055	0.055
500	0.174	0.174	0.166	0.174	0.174	0.174	0.056	0.056	0.057	0.056	0.056	0.056
1000	0.163	0.163	0.163	0.163	0.163	0.163	0.048	0.048	0.048	0.048	0.048	0.048

NoNote: $\pi = 1$, i.e. equally sized predicted period and the estimation period. The results are based on the tests' relative rejection frequency of the null of a zero mean difference, even though it is true, when the critical value is ± 1.96. The results are based on 5,000 iterations. The data generating process is defined as (14) with a fixed seed.

Table 9: The implication of multicollinearity on κ - the *ten* variables case

10-dimensional X_t						
Size of κ						
M	$r=0$	$r=0.10$	$r=0.50$	$r=0.75$	$r=0.90$	$r=0.99$
10	∞	∞	∞	∞	∞	∞
25	3.764	3.764	3.764	3.764	3.764	3.764
50	1.798	1.798	1.798	1.798	1.798	1.798
75	1.648	1.648	1.648	1.648	1.648	1.648
100	1.568	1.568	1.568	1.568	1.568	1.568
250	1.469	1.469	1.469	1.469	1.469	1.469
500	1.441	1.440	1.440	1.440	1.440	1.440
1000	1.426	1.427	1.427	1.427	1.427	1.427

Rejection frequency (5 %)												
	Simple test						Corrected test					
M	$r=0$	$r=0.10$	$r=0.50$	$r=0.75$	$r=0.90$	$r=0.99$	$r=0$	$r=0.10$	$r=0.50$	$r=0.75$	$r=0.90$	$r=0.99$
10	0.089	0.123	0.091	0.090	0.086	0.079	0.000	0.006	0.000	0.000	0.000	0.000
25	0.140	0.140	0.140	0.140	0.140	0.140	0.002	0.002	0.002	0.002	0.002	0.002
50	0.145	0.145	0.145	0.145	0.145	0.145	0.016	0.016	0.016	0.016	0.016	0.016
75	0.174	0.174	0.174	0.174	0.174	0.174	0.033	0.033	0.033	0.033	0.033	0.033
100	0.161	0.161	0.161	0.161	0.161	0.161	0.038	0.038	0.038	0.038	0.038	0.038
250	0.176	0.176	0.176	0.176	0.176	0.176	0.049	0.049	0.049	0.049	0.049	0.049
500	0.151	0.167	0.167	0.167	0.167	0.167	0.070	0.047	0.047	0.047	0.047	0.047
1000	0.172	0.168	0.168	0.168	0.168	0.168	0.053	0.051	0.051	0.051	0.051	0.051

Note: $\pi = 1$, i.e. equally sized predicted period and the estimation period. The results are based on the tests' relative rejection frequency of the null of a zero mean difference, even though it is true, when the critical value is ± 1.96. The results are based on 5,000 iterations. The data generating process is defined as (14) with a fixed seed.