MAGKS

Joint Discussion Paper
Series in Economics

by the Universities of

Aachen · Gießen · Göttingen
Kassel · Marburg · Siegen

ISSN 1867-3678

No. 33-2010

Johannes Paha

Empirical Methods in the Analysis of Collusion

Coordination: Bernd Hayo • Philipps-University Marburg
Faculty of Business Administration and Economics • Universitätsstraße 24, D-35032 Marburg
Tel: +49-6421-2823091, Fax: +49-6421-2823088, e-mail: hayo@wiwi.uni-marburg.de

EMPIRICAL METHODS IN THE ANALYSIS OF COLLUSION

Johannes Paha*

ABSTRACT

Regression methods are commonly used in competition lawsuits for, e.g., determining overcharges in price-fixing cases. Technical evaluations of these methods' pros and cons are not necessarily intuitive. Appraisals that are based on case studies are descriptive but need not be universally valid. This paper opens up the black box called econometrics for competition cases. This is done by complementing theoretical arguments with estimation results. These results are obtained for data that is generated by a simulation-model of a collusive industry. Using such data leaves little room for debate about the quality of these methods because estimates of, e.g., overcharges can be compared to their true underlying values. This analysis provides arguments for demonstrating that thoroughly conducted econometric analyses yield better results than simple techniques such as before-and-after comparisons.

Keywords: Collusion, Empirical Methods, Industry Simulation

JEL Codes: C51, D43, K21, L41

---

* Johannes Paha is a research associate at the

Chair for Industrial Organization, Regulation and Antitrust (VWL 1)
Justus-Liebig-University Giessen
Licher Straße 62
D-35394 Giessen

email johannes.paha@wirtschaft.uni-giessen.de
phone +49 – 641 – 99 22052
fax +49 – 641 – 99 22059
web http://wiwi.uni-giessen.de/ma/dat/goetz/Johannes_Paha%2C%20M.A./

I would like to thank Oliver Budzinski, Henning Fischer, Georg Götz, Maarten Janssen, Peter Winker, my colleagues in Giessen, and the participants of the 35th *Hohenheimer Oberseminar* for their valuable comments on this paper.

1 INTRODUCTION

This paper analyzes empirical methods that are increasingly used in the detection and prosecution of cartels (for an overview see, e.g., Harrington 2008). This analysis is important as, first, “[e]conometric modeling has become the world standard for proving cartel damages” (Connor 2008: 54). Regression techniques may, e.g., be used to decide, how much of an observed price change may be attributed to the cartel and how much is caused by other factors such as changes in production costs. Doing such a decomposition manually would require highly subjective and inexact judgment (Finkelstein and Levenbach 1983: 145). Second, analyzing the pros and cons of such methods is important as their application in legal proceedings may have considerable economic effects on the affected parties such as the payment of fines and/or the award of damages. Therefore, it is of crucial interest to the involved parties to have a sound knowledge about the virtues and shortcomings of these methods.

In this paper, I review some of these methods and apply them to the analysis of a particular cartel. Among other things, I show that calculating cartel overcharges by the use of econometric models yields estimates that are better and more reliable than estimates obtained by simpler techniques such as a comparison of average prices in collusive and competitive periods. Moreover, I show that estimates which are derived from a structural model of the industry tend to outperform estimates that are obtained from descriptive econometric analyses. Finding that an econometric analysis, which is firmly rooted in economic theory, outperforms simple comparisons is textbook-knowledge for econometricians (see, e.g., Heij et al. 2004: 274). However, simplistic techniques such as the comparison of average prices are still in the debate for being used in competition cases. This can, e.g., be seen from their discussion in the recent EU-commissioned study on the quantification of antitrust damages (Oxera 2009: 49).

In presenting a simulation-model of collusive industries and econometrically analyzing its outcomes, this paper provides econometricians with a tool for pursuing two purposes. First, econometricians may obtain knowledge on the likely size of estimation errors that result from, e.g., inappropriate estimation-techniques or data that suffers from measurement errors. Second, this paper provides econometricians with a tool for defending the application of correct but sometimes complex econometric methods in an intuitively comprehensible way.

To illustrate the second point, consider that econometricians may explain the characteristics of their methods in three ways that differ in their levels of complexity and generality. First, mathematical proofs of, e.g., estimates' unbiasedness or efficiency provide valuable insights in the quality of econometric methods. Second, case studies of real cartels are valuable in illustrating the points of these proofs but lack generality. Third, case studies of simulated cartels combine the advantages of the two above methods. On the one hand, they are illustrative. On the other hand, the effects of collusion are perfectly observed in simulated industries and, thus, provide an ideal benchmark for a comparison of the estimated coefficients and an assessment of estimation errors, i.e., it addresses the first above purpose. This line of argumentation is detailed in the following paragraphs.

In competition cases, economic experts have a vital interest in explaining the pros and cons of their methods to an audience that has not necessarily been trained in econometrics. Doing this by providing mathematical proofs of estimates' unbiasedness and efficiency cannot be considered the most time-efficient strategy. One alternative is to illustrate the quality of a method by providing a case study or specific cases where a method has proven to be useful (see, e.g., Hausman (1984)). Regarding a particular case, one would like to argue that *a method is good because a cartel caused an increase in prices of 17.72% and the method arrives at an estimated overcharge of 17.09%. On the contrary, another method is less favorable as its estimated overcharge is only 7.85%.* However, with real case-data such a statement is impossible as the true overcharge of 17.72% is unknown. One may only assess the plausibility of estimates by applying statistical tests or economic reasoning. Hence, one may argue that *the 95%-confidence interval around the first estimate is smaller than that around the second estimate. This implies a higher quality of the first method.*

In this paper, I show that the first statement about the comparison of overcharges can be made when the illustrative case study is based on data of a computer-simulated, collusive industry. In this context, my paper bridges a gap in the literature because using simulated data enables us to perfectly generate the counterfactual, competitive situation, i.e. the situation without the infringement of competition laws. This allows for an exact calculation of, for example, the cartel-overcharge, which is unknown in case of real data. Consequently, one may not only make a statement like the second – somewhat vague – one on confidence-bands. One may rather make the first, strong statement on the comparison of overcharges. Additionally, econometricians may learn the likely size of the error that is made by using misspecified variables. In the analyzed example, the overcharge-estimate may well decrease to a value of 4.55% (instead of the above, quite accurate 17.09%) by misspecifying only four out of 56 cartel-periods.

Basing my results on such an analysis of a particular simulated industry, I show that a reduced-form regression, which is derived from the theoretical model of the industry, almost perfectly predicts the true overcharge. This result is not surprising as one would expect a regression, which perfectly depicts the industry, to yield very good results. The point of this exercise is to show that one must formulate a good theoretic model of the industry in order to attain such ideal results.

As an alternative to identifying a model of this type, one might run descriptive regressions or even use simple comparisons of average prices. The below analysis indicates that such simpler methods do not necessarily predict the estimated effects well. One finds that for a specific sample-industry the estimate obtained by a naïve, descriptive econometric model underestimates the true overcharge by almost 10 percentage points while the comparison of collusive and competitive prices even indicates a cartel-related *reduction* of prices.

With this research I contribute to showing that the cogs and wheels of econometric methods can be explained in an intuitive, accessible way. This, ideally, promotes the use of econometric analyses in competition cases. Even today, such sound scientific analyses are sometimes discarded in favor of subjective, personal experience as was the case in the German sour milk cheese decision (Monopolkommission 2010: 267, 269). In the market definition stage of this case, the judge disregarded econometric analyses concerning the substitutability of different types of cheese in favor of his own culinary experience.

In section 2.1, I provide a brief description of the simulation model that is used to generate the data. In section 2.2, I present the dataset that is generated by the simulation model. It includes data on costs, output, revenue, prices, profits, and market shares of the parties involved and of the other participants in the relevant market. Data of this type is considered helpful for competition analyses by the Directorate General Competition (DG Comp) at the European Commission (EU 2010: 13). In section 3.1, I show how a reduced-form pricing equation can be derived for the industry of interest. This reduced-form pricing regression is estimated in section 3.2, where I also analyze the properties of the estimates. In section 4, I evaluate methods for the detection of collusion and the identification of periods in which the cartel was active and effective. Section 5 concludes.

2 SIMULATING COLLUSION

In section 2.1, I present a model of collusive industries with endogenous cartel formation. A more detailed description (of a previous version) of the simulation model is provided by Paha (2010). This model is used to generate data of collusive industries, which is described in section 2.2.

2.1 The Simulation Model

Cartels are formed not only in industries for homogeneous goods such as cement but also in industries for somewhat differentiated products such as beer or bathroom equipment.[1] Therefore,

1 The mentioned cases were investigated by the European Commission under the following case numbers: Bathroom fittings and fixtures 39092, Belgian beer market 37.614

demand is assumed to be given by a representative agent's utility function that allows for modeling different degrees of product differentiation (Shubik and Levitan 1980). Vectors and matrices are denoted in bold.

$$V = q_0 + U(\boldsymbol{q}) = q_0 + \nu \boldsymbol{q}'\boldsymbol{\iota} - \frac{n}{2\cdot(1+\mu)}\left[\boldsymbol{q}'\boldsymbol{q} + \frac{\mu}{n}(\boldsymbol{q}'\boldsymbol{\iota})^2\right] \tag{1}$$

In this function q_0 is the outside option of consumers. $\boldsymbol{q}$ is a ($n \times 1$)-vector whose elements are the quantities q_i of n products. Each product is produced by exactly one firm, so that there are n firms in the industry. The number of firms is modeled as fixed. This may be motivated by sunk costs being sufficiently high such that there are no firms outside the industry for whom it would be profitable to enter. $\boldsymbol{\iota}$ is a ($n \times 1$)-vector where each element takes a value of 1. ν is a positive parameter and $\mu \in [0,\infty)$ represents the degree of substitutability[2] of the n products. According to this utility function the representative agent consumes some quantity of each good.

Each product is produced by a one-product firm at marginal cost c_i. Marginal costs basically have two features: (i) They are firm-specific, which makes firms asymmetric. (ii) Cost-shocks are assumed to occur in every period such that marginal costs follow a random walk (Harrington 2008: 241). This introduces dynamics to the simulation model.

Marginal costs of firm i, i.e. $c_{i,t}$, are generated according to equation (2) in conjunction with conditions (3) and (4).

$$c_{i,t} = \begin{cases} a_1 \cdot \nu + a_{2i,t} \cdot s_t & \text{if } t=1 \\ c_{i,t-1} + a_{2i,t} \cdot s_t & \text{if } t>1 \end{cases} \tag{2}$$

$$\left.\begin{array}{lcl} a_1 & \in &]0;1[\\ a_{2i,t} & \sim & CN\left(\frac{a_3+1}{2}, \mu^{-2}, a_3, 1\right) \\ a_3 & \in & [0;1] \\ a_4 & \in &]0;1[\end{array}\right\} \tag{3}$$

The base level of marginal costs, i.e. in the initial period $t = 1$, is determined as the percentage a_1 of the variable ν, which is closely related to goods' reservation price. Cost-asymmetry among firms is modeled by adding a firm-specific term $a_{2i,t}s_t$ to the base level of marginal costs. The multiplicative, firm-specific technology-parameter $a_{2i,t}$ is drawn randomly from a censored normal distribution in the interval [a_3;1]. The expected value E($a_{2i,t}$) is the mean of the interval [a_3;1]. The variance μ^{-2} of marginal costs is modeled to decrease in the degree of product homogeneity μ. This is because homogeneous products are produced by similar production technologies and, thus, at the same cost.

2 For $\mu = \infty$ goods are perfect substitutes. For $\mu = 0$ goods are independent. As this paper is interested in analyzing (imperfect) substitutes, μ is set at values greater than 0.

If production costs differed, the less efficient firms would want to exit the industry. This process of convergence is assumed to have happened in the past and is thus beyond the scope of the model. Cost shocks s_t are drawn randomly from a uniform distribution in the interval given by equation (4). $a_4 \in [0;1]$ determines the amplitude of cost-shocks. Setting $a_4 = 0$ yields marginal costs that are symmetric across firms and constant over time. This interval ensures that marginal costs cannot become negative.

$$s_t \in \begin{cases} \left[-a_4 \cdot (a_1 \cdot v); a_4 \cdot (a_1 \cdot v)\right] & \text{if } t=1 \\ \left[-a_4 \cdot \min_i (c_{i,t-1}); a_4 \cdot \min_i (c_{i,t-1})\right] & \text{if } t>1 \end{cases} \tag{4}$$

As in Harrington (2008: 241), marginal costs are assumed to follow a random walk in subsequent periods, i.e. $t > 1$. Thus, in every period t a random, scaled shock term s_t is added to the last period's marginal costs of each firm i. The marginal cost-shock s_t is the same for all firms and can be considered a fluctuation in input prices. It is drawn randomly from a uniform distribution in the interval $\left[-a_4 \cdot min_i(c_{i,t-1}); a_4 \cdot min_i(c_{i,t-1})\right]$, where $min_i(c_{i,t-1})$ is the minimum (over all firms i) of last period's marginal costs. Again, this ensures that marginal costs cannot become negative.

As is well known, a cartelist i may find it profitable to deviate from the collusive agreement and make deviation-profits π_{di} rather than cartel profits π_i. Friedman (1971) shows that cartels can be stabilized in a dynamic game when firms play a grim trigger strategy. Hence, after an observed deviation all cartelists revert to the competitive equilibrium where firm i makes profits π_{ci}. It can be shown that cartel-firm i will not deviate from the collusive agreement when condition (5) applies.

$$\frac{\pi_{di} - \pi_i}{\pi_{di} - \pi_{ci}} \leq \delta \cdot (1-P) \tag{5}$$

In equation (5), $\delta = 1/(1+r)$ denotes firms' common discount factor with discount rate r. P is the probability that the cartel will be discovered by the competition authority.

The evolution of cost-shocks may generate a distribution of costs where some firms in a previously stable cartel find it profitable to deviate from the collusive agreement. Therefore, firms are assumed to lower collusive prices in order to prevent deviations from the cartel. To see how this mechanism works, consider that the marginal costs of all firms are perfectly known to every other firm. Therefore, all firms can perfectly anticipate if (at prevailing prices) it would be profitable for any cartelist to deviate from the collusive agreement. If neither cartelist has an incentive to deviate, i.e. inequality (5) is satisfied for all cartel firms, they jointly set profit maximizing prices (making profit π_{jpi}). If at least one cartel firm finds it profitable to deviate from the collusive agreement, I assume that the cartelists render a deviation unprofitable by setting prices in the current period at the level that would prevail under competition. Hence, I do not assume firms will lower prices until condition (5) is *just* equalized for all firms. Assuming a reversion to competitive prices (Fershtman and Pakes 2000) is convenient as it is difficult to determine the set of prices that satisfies condition (5) in equality for all firms. Therefore, it would be hard to imagine that firms can determine this set of prices. However, it is plausible that they coordinate on the one perfectly known stable equilibrium, i.e. the competitive equilibrium. This strategy is shown in equation (6).

$$\pi_i = \begin{cases} \pi_{jpi} & \text{if } \forall\, j \in [1, m] \;\; \dfrac{\pi_{dj} - \pi_{jpj}}{\pi_{dj} - \pi_{cj}} \leq \delta \cdot (1 - P) \\ \pi_{ci} & \text{if } \exists\, j \in [1, m] \;\; \dfrac{\pi_{dj} - \pi_{jpj}}{\pi_{dj} - \pi_{cj}} > \delta \cdot (1 - P) \end{cases} \tag{6}$$

This modeling assumption as well as the firms' reaction is quite similar to the model by Rotemberg and Saloner (1986) with perfectly observable demand shocks. In contrast to Rotemberg and Saloner (1986), a preventive reduction of prices is not triggered by demand shocks but by cost shocks.

Even if firms know that an existing cartel can be stabilized by the above trigger strategy, they need not have an individual incentive to join or form a cartel. This is because firms will often make higher profits π_{fi} in the competitive fringe than they would in the cartel. The reason for this is that fringe-firms may expand their output-quantity under the cartel's price-umbrella while cartelists must reduce their output in order to raise prices. This freerider-effect is particularly strong for small, inefficient firms (Paha 2010).

Prokop (1999) proposes that cartels can nonetheless be formed if firms play a mixed strategy. Every firm chooses a participation-probability j_i that maximizes its expected payoff. This is equivalent to choosing a participation probability that, in the Nash-equilibrium, makes the competitors of a firm indifferent between joining the cartel or remaining outside. Solving for the Nash-equilibrium of the cartel formation game is difficult as these expected present values of profit depend on the expected size and composition of the cartel which, again, are a function of firms' participation probabilities. Paha (2010) shows that a solution can nonetheless be attained by a Differential Evolution stochastic search heuristic.

2.2 The Simulated Dataset

The above model is used to generate data of a collusive industry as described in Appendix A.[3] From the group of all generated industries, I pick one that gives (i) a stable cartel with price wars and (ii) whose price-overcharge takes a realistic value of 17.72% (see Table 1). The simulated goods are relatively homogeneous with a maximum price-difference between the cheapest and the

3 For readers interested in reproducing the below regressions and calculations, I will be happy to provide the data in EViews-format on request. Please send an email to johannes.paha@wirtschaft.uni-giessen.de.

most expensive good of less than 7% in competition.

One might object by arguing that analyzing just one industry yields results that apply only to this specific industry. This is only true for the absolute value of the calculated effects. Other cartels in other industries may generate overcharges of a size other than the above 17.72%. Therefore, possible misspecifications in the econometric measurement of overcharges cause errors of different absolute values. However, when repeating the below calculations for different cartels one gets the same qualitative interpretation. In this context, a focus on just one industry does not reduce the meaningfulness of the qualitative results but may be considered more illustrative for the non-econometrician.

A comparison of the numeric error may be done by calculating the error relative to the true underlying (e.g., the overcharge). This allows for a comparison of relative errors across industries that may differ in cost-structures, the price elasticity of demand, or the degree of product differentiation. Such a comparison is especially interesting for econometricians themselves as it allows for an empirical determination of the distribution of the relative error that is, for example, caused by a misspecified regression. Exploring such distributions requires extensive research as one can easily imagine at least three different frameworks for comparing relative errors. One might use (i) different cost-evolutions in the same cartel, (ii) different cartel-compositions in the same industry, or (iii) different cartels in different industries. Such extensive research heavily relies on the qualitative insights provided below for a single sample-industry and is beyond the scope of the present paper.

In the selected, infinitely lived industry $T = 100$ periods can be observed. These 100 periods can be thought of as quarterly data for a period of 25 years. The $n = 9$ firms are assumed to compete during the first 20 periods. At the beginning of period 21, they meet and agree on forming a cartel that encompasses all firms with the exception of firm 3. The time of the formation-meeting is exogenously given. Although the cartel is active in periods 21-100, it is only effective in 56 of these periods. These collusive periods are characterized by good economic conditions, i.e. average marginal costs are only 24.34. The remaining 24 (price war) periods are economically stressful for firms because mean marginal costs amount to 36.04. Firms pass on some portion of the increase in production costs to consumers, which leads to higher prices. However, with price-sensitive consumers this pass-on is not perfect causing firms' profits c.p. to decrease. Additionally, in the above model with linear demand curves, higher prices imply a greater price-elasticity of demand, so that a deviator from the collusive agreement may attract much additional business by lowering his price. All three effects, i.e. higher prices and demand elasticities, as well as lower cartel-profits, raise cartelists' critical discount factor and, hence, their desire to deviate from the collusive agreement. This would cause the breakdown of the cartel. The existence of the cartel is preserved by temporarily reverting to the competitive equilibrium, which renders deviations unprofitable.

Table 1 shows the average effects of the cartel on the prices of both cartel- and fringe-firms (i.e. the overcharge), as well as the quantities and profits in the periods during which the cartel is effective. These effects are calculated as the difference of a variable between its collusive and its counterfactual, competitive value divided by the collusive value. These values highlight one of the advantages of using simulated data. With real data, the competitive *but-for* values of prices, quantities, and profits would be unknown. One clearly sees how cartelists reduce quantity in order to raise prices while the fringe firm expands its quantity *and* price under the cartel's price umbrella. Note that the fringe-firm is the least efficient firm which renders its percentage increase in profits lower than that of the cartelists.

	Cartel firms	**Fringe firm**
Change in price ($\overline{oc}$)	+17.72%	+3.81%
Change in quantity	-56.81%	+36.83%
Change in profit	+59.50%	+47.09%

Table 1: Cartel Effects

The relative change in cartelists' price, i.e. the average overcharge $\overline{oc}$, is computed as the mean (over all cartelists) of their average (over all effective cartel periods, i.e. non-price war periods) price-increase relative to the collusive price. The average overcharge is 17.72%. Following a recent meta-study of overcharge-estimates (Connor and Lande 2008) this is a modest overcharge.

Figure 1 provides an impression of the cartel's effect on prices. This figure displays the average price over all cartelists ± one standard deviation (dark gray time series) and the average marginal costs over all cartelists ± one standard deviation (light gray time series). Moreover, one sees the price (dashed line) and marginal costs (dotted line) of the fringe firm. The shaded areas in the background indicate the periods when the cartel was effective. From the figure, the following conclusions can be drawn.

1. In periods of price war, firms' prices are close to marginal costs. When the cartel is active, the fringe firm undercuts cartelists' prices.

2. The incentive to deviate from the cartel is especially high in stressful economic situations. This is the case when marginal costs are exceptionally high and/or dispersed.

3. With such pronounced cost-shocks it is difficult to detect collusion from analyzing time series of prices alone. When regarding the plot of both prices *and* costs, one might infer

evidence of collusion from the fact that, e.g., in period 40 and 82 prices go up while costs decline. However, this movement of prices is similar for fringe- and cartel-firms, so that the two cannot well be discriminated.

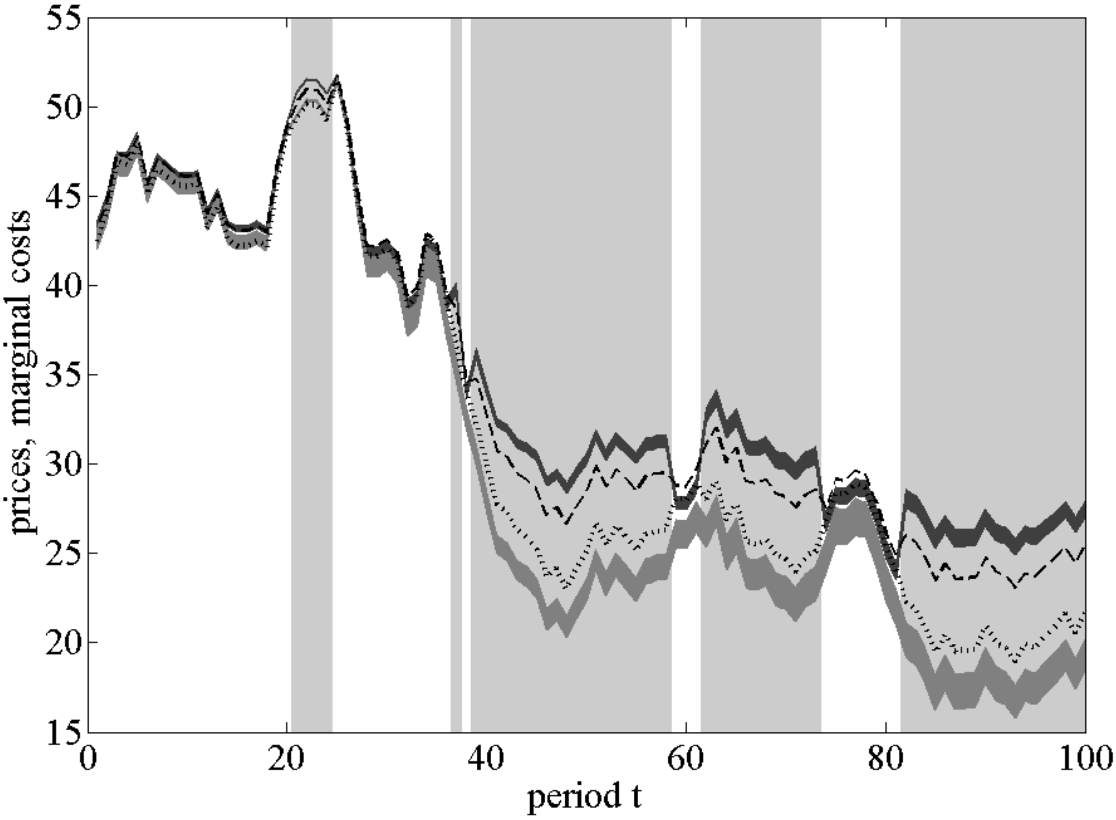

Figure 1: Prices and Costs

3 PRICING REGRESSIONS

Assume the following situation. A competition authority receives information that a collusive agreement might have been established in the above industry. The following analysis is concerned with evaluating how well the competition authority would be able to identify the effects of the cartel by employing econometric techniques. Such techniques have been reviewed by Harrington (2008: 216), who states that "[v]erification of episodes of collusion is a data-intensive and time-intensive task that requires controlling for many determinants of behavior."

In section 3.2, I present a reduced-form regression of prices on supply- and demand-side variables. Regressions of this type are commonly used in competition cases (Clark et al. 2004: 22) and allow for simulating counterfactual, i.e. competitive, prices in the allegedly collusive periods (Baker 1999: 389). This enables researchers to produce estimates of, e.g., the cartel's overcharge. I show that a regression of this type predicts the overcharge quite accurately under two conditions.

First, the fundamental parameters of the model must not change between collusive and competitive periods. Second, the estimated (reduced-form) model must be derived from a structural model that provides a good description of the industry's structure and firms' conduct. I explicate this point more clearly in section 3.1. There, I also show that the use of econometric techniques has decisive advantages over simpler methods such as the comparison of prices in competitive and collusive periods.

Throughout section 3 I assume that (i) the identity of cartelists is known and (ii) the researcher knows in which periods cartelists set collusive prices. This assumption can be motivated by the existence of hard evidence that provides such information. The assumption that D_t is known is relaxed in section 4, which is concerned with detecting the start of the cartel and cartelists' temporary reversions to competitive behavior.

3.1 Deriving a Reduced-Form Pricing Equation

In section 2.1, I present a structural economic model that describes the industry under research in several dimensions (Reiss and Wolak 2007: 4304). (i) The *economic environment* is characterized by n firms who supply a differentiated product to representative agents. (ii) The structural model defines the *economic primitives* such as consumers' utility function (1) and firms' production technology (2)-(4). (iii) The model is characterized by *exogenous variables* such as the degree of product differentiation μ and consumers' willingness to pay, as represented by variable v. (iv) The industry is assumed to be infinitely lived with utility-maximizing consumers and profit-maximizing firms. (v) The *equilibrium solution concept* is a Nash-equilibrium with price as firms' strategic variable, while firms face price-taking consumers.

Structural-form econometric modeling is concerned with identifying the structural parameters of demand and supply in an industry. Such parameters can be, e.g., marginal costs, price-elasticities or modeling parameters such as the parameters μ and v, and/or the cost parameters a_1, $E(a_2)$, a_3, and a_4 as in case of the above model. As such, the structural model consists of a system of several simultaneous equations where a variable may appear as a dependent variable in one equation and as an explanatory variable in another equation. For example, marginal costs are a dependent variable in equation (2) but an explanatory variable in firms' reaction function.[4]

Structure can be imposed on these models in basically two ways, i.e. by making economic and/or statistical assumptions. Economic structural assumptions include assuming functional forms for demand and costs, and supposing that firms compete in, e.g., prices or quantities. These

4 It can be shown that firms' reaction function is $p_i=\frac{c_i}{2}+\frac{n\nu+\mu\sum_{(j=1)\setminus i}^{n} p_j}{2\cdot(n+n\mu-\mu)}$ (see, e.g., Paha (2010)).

assumptions are summarized in the system of equations (7) below. Statistical structural assumptions often relate to the distribution of the error term that is added to the regression equations. These assumptions are made in order to (i) reflect economic realities, (ii) rationalize what is observed in the data, and/or (iii) simplify estimation (Reiss and Wolak 2007: 4285).

After obtaining the structural parameters of an industry, a competition authority can do counterfactual simulations such as determining firms' prices with and without the collusive agreement. However, estimating these parameters of the structural model is rarely a simple task. Often it cannot even be done due to the limited availability of data or due to time-constraints in real-world competition cases. It is typically easier, and thus more common in practice (Clark et al. 2004: footnote 38), to estimate a reduced-form pricing equation.

Reduced-form models may also consist of a system of simultaneous equations. These are obtained by transforming the structural model in such a way such all endogenous variables are treated as dependent variables. These are regressed on exogenous variables only (Reiss and Wolak 2007: 4293). One example for this is a reduced-form pricing regression, in which the price of a good is regressed on “variables related to cost, demand, and market structure, and a series of indicator (dummy) variables that allow the intercept to differ among relevant groups of observations” (Baker 1999: 391).

From the model outlined in section 2.1, one may derive a vector of equilibrium-prices $\boldsymbol{p}_t$ and a vector of marginal costs $\boldsymbol{c}_t$ as presented in the system of equations (7) (for details see Paha 2010: 5). $\boldsymbol{c}_t$ follows from equations (2)-(4). $\boldsymbol{p}_t$ is the vector of profit-maximizing prices when consumers behave according to utility-function (1) and when firms maximize profits à la Bertrand.

$$\begin{array}{rcl} \boldsymbol{p}_t & = & \boldsymbol{X}_t^{-1}\left[\boldsymbol{\iota} n v+\boldsymbol{Y}_t \cdot \boldsymbol{c}_t\right] \\ \boldsymbol{c}_t & = & \boldsymbol{c}_{t-1}+\boldsymbol{a}_{2,t} \cdot s_t \end{array} \tag{7}$$

The matrices $\boldsymbol{X}_t$ and $\boldsymbol{Y}_t$ are defined as follows.

$$\begin{array}{rcl} \boldsymbol{X}_t & = & \boldsymbol{I}(2\mathrm{n}+2\mu n-\mu)-\mu(\boldsymbol{\iota}\boldsymbol{\iota}'+D_t \boldsymbol{A}) \\ \boldsymbol{Y}_t & = & (n+n\mu-\mu)-\mu D_t \boldsymbol{A} \end{array} \tag{8}$$

D_t is a dummy variable that takes a value of 1 when firms have formed a cartel and do not engage in a price war, and a value of 0 otherwise. In this section, D_t is assumed to be perfectly observed by the researcher. This assumption is relaxed in section 4. $\boldsymbol{I}$ is an $(n\times n)$-identity matrix. $\boldsymbol{\iota}$ is an $(n\times 1)$-vector of ones. $\boldsymbol{A}$ is a matrix of dimension $n\times n$. Its elements $a_{i,j}$ take a value of 1 if both firms i and j are cartel firms, and a value of 0 otherwise.

In the following, the equations in (7) are used to derive the reduced-form pricing equation (9) of the above model. This is done by solving the pricing-equation of $\boldsymbol{p}_{t\text{-}1}$ for $\boldsymbol{c}_{t\text{-}1}$ and plugging the resulting term in the equation of $\boldsymbol{c}_t$. Inserting this term into $\boldsymbol{p}_t$ yields the reduced-form pricing equation (9).

$$\boldsymbol{p}_t \quad = \quad \boldsymbol{X}_{t-1}^{-1}\left(\boldsymbol{I}-\boldsymbol{Y}_t\boldsymbol{Y}_{t-1}^{-1}\right)\boldsymbol{\iota} n v \quad + \quad \boldsymbol{X}_{t-1}^{-1}\boldsymbol{Y}_t\boldsymbol{Y}_{t-1}^{-1}\boldsymbol{X}_t\boldsymbol{p}_{t-1} \quad + \quad \boldsymbol{X}_{t-1}^{-1}\boldsymbol{Y}_t\boldsymbol{a}_{2,t}s_t \tag{9}$$

The decisive advantage of pricing-equation (9) is that firms' prices in period t are expressed as a function of prices in period t-1. However, they do not depend on marginal costs. This is advantageous as prices can easily be observed while marginal costs are often unknown to the econometrician. A further explanatory variable is the common cost-shock s_t that affects the marginal costs of all firms. I assume that this common cost-shock is observed by the researcher who may, e.g., analyze input prices or price-indices that are provided by statistical offices. However, the researcher does not observe the firm- and time-dependent scaling parameter $a_{2i,t}$. This unobserved heterogeneity transforms the deterministic relationship (9) into a stochastic one (10). This is reflected by adding the i.i.d. error term $\varepsilon_{i,t}$ to the regression equation. The α-, β-, and γ-variables are the coefficients of the regression that are estimated in section 3.2.

$$\begin{array}{rclclc} p_{i,t} & = & \alpha_i(D_t-D_{t-1}) & + & \beta_1 p_{i,t-1} & + \\ & & \beta_{2,i} p_{i,t-1}\big((D_t-D_{t-1})>0\big) & + & \beta_{3,i} p_{i,t-1}\big((D_t-D_{t-1})<0\big) & + \\ & & \gamma_{1,i} s_t\big((D_t-D_{t-1})>0\big) & + & \gamma_{2,i} s_t\big((D_t-D_{t-1})<0\big) & + \\ & & \gamma_{3,i} s_t\big(D_t=D_{t-1}=1\big) & + & \gamma_{4,i} s_t\big(D_t=D_{t-1}=0\big) & + \quad \epsilon_{i,t} \end{array} \tag{10}$$

The first summand in (10) is a constant which measures the effect of moving from the competitive equilibrium in period t-1 to the collusive equilibrium in period t, or the other way around. Only in these switching periods does the coefficient α_i take a value greater than zero. This value provides a measure of the average absolute change in prices that firm i brings about as a consequence of the cartel being active. Equation (9) implies that α_i only varies with the affiliation of firm i to the group of cartel-firms or the group of fringe-firms. Hence, α_i must take one value for all cartelists and another value for all firms in the fringe. All other firm-specific effects are accounted for by the remaining coefficients. Economic theory predicts that fringe firms charge a lower overcharge than cartelists and, thus, profit from increased prices *and* an increased quantity sold. Therefore, the value of α_i should be smaller for fringe-firms than for cartelists. Equation (9) implies that the entire summand should be zero when firms either continue behaving competitively or collusively from period t-1 to period t. To see this, consider that in this case, the condition $\boldsymbol{Y}_t = \boldsymbol{Y}_{t-1}$ applies, such that $(\boldsymbol{I}-\boldsymbol{Y}_t\boldsymbol{Y}^{1}_{t-1})$ turns out to be a matrix of zeros.

In times when firms do not change their competitive conduct from period t-1 to period t, firms take their price $p_{i,t-1}$ as a base value (see the second summand in (9)) and only adjust it by the factors (here in particular production costs) that have changed between these two periods (see the third summand in (9)). Equation (9) simplifies to (11), because in these periods the conditions $\boldsymbol{Y}_t = \boldsymbol{Y}_{t-1}$ and $\boldsymbol{X}_t = \boldsymbol{X}_{t-1}$ apply.

$$\boldsymbol{p}_t = \boldsymbol{p}_{t-1} + \boldsymbol{X}_{t-1}^{-1} \boldsymbol{Y}_t \boldsymbol{a}_t s_t \tag{11}$$

In this case, the estimated coefficient β_1 in (10) should take a value of one. Finding a statistically significant value of $\beta_1 \neq 1$ would indicate misspecification of the model due to, e.g., omitted relevant variables. When moving from collusion to competition, the value of $p_{i,t-1}$ is elevated by the overcharge. Therefore, it is incorrectly high when being used as a base value for determining $p_{i,t}$. This effect is considered by including the variable $((D_t\text{-}D_{t-1})>0)$ that takes a value of one when moving from collusion to competition, and a value of zero otherwise. For the coefficient $\beta_{2,i}$ of this variable, one expects (i) a negative value that is (ii) the same for all cartel-firms and a (iii) somewhat higher but, again, identical value for fringe-firms. When moving from competition to collusion $((D_t\text{-}D_{t-1})<0)$, this effect is reversed. This can be seen from coefficient $\beta_{3,i}$, which is expected to be greater than zero. The equality of these effects for cartelists on the one hand and fringe firms on the other follows directly from equation (9).[5]

The third term in equation (9) requires the inclusion of four dummy variables in regression (10). These dummies capture the effect of the cost-shock s_t on prices in the following four types of periods: (i) Transition from collusion to competition ($\gamma_{1,i}$), (ii) transition from competition to collusion ($\gamma_{2,i}$), (iii) collusion in both periods ($\gamma_{3,i}$), and (iv) competition in both periods ($\gamma_{4,i}$). Again, the term $\boldsymbol{X}^{-1}_{t-1}\boldsymbol{Y}_t$ implies that the values of γ_1-γ_4 differ for the group of cartelists and the group of firms in the fringe. Multiplying this product with the vector of firm-specific technology-parameters $\boldsymbol{a}_{2,t}$ would imply a firm-specificity of γ_1-γ_4 only if the values of a_2 were persistent over time. However, as the values of a_2 evolve according to a random walk, the condition $E(a_{2,i}) = E(a_{2,j})$ applies for all firms i and j. Consequently, γ_1-γ_4 only depend on the affiliation of a firm to the cartel or the fringe but do not differ across firms in either of these two groups.

Estimating a reduced-form equation is typically easier than estimating the structural model. Therefore, reduced-form estimations are frequently used in competition cases (Clark et al. 2004: footnote 38). An estimation of equation (10) is provided in section 3.2. If the parameters of the structural model – such as the number of firms n or the degree of product homogeneity μ – remain constant over the period of observation, reduced-form estimations are a valuable tool for determining counterfactual values of, e.g., prices. However, if these parameters change, the matrices

5 To see this even more clearly, one may choose arbitrary values for n and μ, and consider an arbitrary composition of the cartel. Then, writing equation (9) in non-matrix notation for collusive periods, competitive periods, and transition-periods supports the above statements.

$\boldsymbol{X}_t$ and $\boldsymbol{Y}_t$ change over time. In this case, the parameters of the reduced-form model cannot be estimated consistently. Therefore, one would want to estimate the structural model and account for the changes in its parameters. This allows for simulating the effect of the structural change on market outcomes. Some refer to this approach as simulation method or market-structure-based method (Oxera 2009: v, 83). The step of estimating the coefficients of the structural model is needed to numerically calibrate the model to the features of the industry under examination. In the above case, no structural changes occur. Hence, I present the reduced-form estimation only.

3.2 Estimation and Comparison of a Reduced-Form Pricing Equation

Table 2 presents the results of the reduced-form regression (10) for the industry described in section 2.2. The equation is estimated by a pooled ordinary least squares (OLS) panel-regression. The second column presents the regression-coefficients for all firms. The third column indicates by how much the coefficients of the fringe-firm deviate from the common coefficient. As an example, for a cartelist, one finds α_{cartel}=11.070, while for a fringe firm α_{fringe}=11.070-5.712=5.358 applies.

dependent variable $p_{i,t}$		
	all firms	Δ fringe
(D_t-D_{t-1})	11.070 (0.056)***	-5.712 (0.164)***
$p_{i,t-1}$	1.000 (0.000)***	
$p_{i,t-1}((D_t-D_{t-1})>0)$	-0.215 (0.001)***	0.107 (0.004)***
$p_{i,t-1}((D_t-D_{t-1})<0)$	0.194 (0.001)***	-0.095 (0.004)***
$s_t((D_t-D_{t-1})>0)$	0.612 (0.008)***	-0.004 (0.024)
$s_t((D_t-D_{t-1})<0)$	0.717 (0.006)***	-0.005 (0.019)
$s_t(D_t=D_{t-1}=1)$	0.597 (0.003)***	0.046 (0.010)***
$s_t(D_t=D_{t-1}=0)$	0.721 (0.002)***	-0.005 (0.006)
R^2	99.99%	
$\bar{R}^2$	99.99%	

Table 2: Reduced-Form Pricing Regression (10)

One finds that this regression is of a very high quality because all coefficients take the expected values and signs as described in section 3.1. The value of the coefficient of determination R^2 is higher as one would expect for real data. This is because simulated data is neither affected by measurement errors nor by firms who, e.g., set sub-optimal prices due to information asymmetries. Hence, there is little white noise in the data in the sense of firms not behaving perfectly according to the above structural model.

I use this regression to determine estimates of firms' counterfactual, competitive prices and their overcharges.[6] One obtains an average overcharge for cartel firms amounting to 17.09% as compared to the true overcharge of 17.72% (see Table 1). Therefore, the estimated overcharge is quite close to its true

6 The calculation is done as follows: One observes that firm 1 sets a collusive price $p_{1,42}$=31.994 in period 42 and a collusive price of $p_{1,43}$=31.173 in period 43. The common cost-shock in period 23 is s_{43}=-1.354. Applying the above

underlying. With a value of 8.92%, the model overestimates the overcharge of the fringe firm, whose true value is 3.81%. This is mainly due to the fact that there is just one firm in the fringe so that only little variation in the data can be used to infer the behavior of fringe-firms.

The high predictive value of the regression has several reasons. These may be formulated as requirements for good reduced-form pricing regressions.

1. Ideally, pricing regressions should be based on reduced-form equations instead of having a descriptive nature. Hence, the regression-specification should be derived from a structural model of the industry under research.
2. The assumed structural model must be a good description of the industry's characteristics and its firms' conduct.
3. A reduced-form regression can only be applied when there is no structural change in industry-characteristics and firm-behavior.

To see the importance of the first requirement – deriving the regression-specification from a structural model – consider a linear pricing regression as shown by equation (12) which is sometimes referred to as *the dummy variable model* (Clark et al. 2004: 22).

$$p_{i,t}=\alpha+\boldsymbol{w}_{i,t}\boldsymbol{\beta}+\boldsymbol{y}_{i,t}\boldsymbol{\gamma}+\boldsymbol{h}_{i,t}\boldsymbol{\delta}+\boldsymbol{D}_{i,t}\boldsymbol{\lambda}+\epsilon_{i,t} \tag{12}$$

Regressions of this type are among the earliest econometric analyses in competition cases (see, e.g., Finkelstein and Levenbach 1983: 153ff.) that have been frequently applied since then. Baker (1999: 393) even considers a regression of this type to be a typical example of a reduced-form pricing equation.[7]

In this notation, α is a constant. $\boldsymbol{w}_{i,,t}$ is a vector of variables that affect cost (e.g., the above cost-shock s_t). $\boldsymbol{y}_{i,,t}$ is a vector of variables affecting demand (e.g. the price of substitute products). $\boldsymbol{h}_{i,,t}$ is a vector of variables related to market structure (e.g. seller concentration or measures of entry conditions). The variable $\boldsymbol{D}_{i,t}$ is a vector of dummy variables that allow the intercept α to vary among relevant groups of observations (e.g. the dummy variable D_t that indicates whether a cartel was effective in period t).

When specifying such a dummy variable model, one must be extremely careful as to which

regression, one predicts a hypothetical competitive price of $\hat{p}_{c,1,43}=-11.070+(1+0.194)\cdot p_{1,42}+0.717\cdot s_{43}=26.16$. From the industry-simulation, one can infer the true competitive price $p_{c,1,43}$=25.956 that would be unknown in a real case. Therefore, the predicted price is found to be a good predictor for its true underlying.

7 Please note that the variables used for the coefficients of the regression have a different meaning in equation (12) than in equation (10) above. On the one hand, this choice is made for reasons of parameter parsimony and, on the other hand, for being more in line with Baker's (1999) notation.

variables to include. When estimating the model for competitive periods only, including competitors' prices in $\boldsymbol{y}_{i,,t}$ can be helpful. When estimating the model for collusive periods, including these prices may cause the estimated coefficients to be biased. This is especially the case, when the effect of collusion on prices is not (appropriately) considered in the regressors and, thus, shows up in the error term. Then, the prices of other firms will be correlated with the error term (Clark et al. 2004: 23). A similar argument can be made for the vector $\boldsymbol{h}_t$. It contains variables related to market structure such as concentration measures and market shares. The values of both measures are affected by collusion. As cartelists restrict their output quantity, their market shares decline while fringe firms' quantity and market shares increase. This raises the asymmetry across firms and drives the Herfindahl-Hirschmann-index (HHI) up. The resulting correlation of concentration measures and the error term in collusive periods causes the coefficient δ to be biased.

dependent variable $p_{i,t}$		
	all firms	Δ fringe
$p_{i,t-1}$	1.001 (0.001)***	
$p_{i,t-1}D_{t-1}$	-0.082 (0.002)***	0.046 (0.007)***
s_t	0.619 (0.013)***	
D_t	2.601 (0.078)***	-1.528 (0.222)**
R^2	99.21%	
$\bar{R}^2$	99.21%	

Table 3: Naïve Pricing Regression

Starting with equation (12) and after a few steps of model-selection, a researcher may end up with a regression such as the one presented in Table 3.[8] One would economically interpret this regression as follows: Firm i takes its price in period t-1 as a base value for its price in period t (Finkelstein and Levenbach 1983: 159). This base value is biased upwards if the market in t-1 is characterized by collusion. This is econometrically reflected in the interaction term $p_{i,t-1}D_{t-1}$. This base price is adjusted by the changes that occurred between period t-1 and t, i.e. cost shocks s_t, and collusion D_t.

The pros of this naïve approach are twofold. First, one arrives at an economically sensible regression-specification by employing purely statistical techniques in addition to some economic thinking. Second, this regression provides estimates of mean overcharges at a reasonable order of

8 In particular, I started with a specification that closely resembled the final specification in Table 3 but included firm-specific interaction effects for all firms. By successively applying F-tests it was possible to show that the fringe firm is the only one for whom the coefficients of $p_{i,t-1}D_{t-1}$ and D_t differ from those of the other firms in a statistically significant sense. Including the prices of competitors and the Herfindahl-Hirschman-Index supports the above proposition of biased estimates as these variables are correlated with the error term. The consequence can be seen particularly well in the coefficient of $p_{i,t-1}$ that is biased away from its economically plausible value of one.

magnitude, i.e. 7.85% for cartelists and 3.71% for fringe firm 3.

The cons of this naïve approach are the following: First, its overcharge-estimates do not match the true overcharges as well as those obtained from the reduced-form regression (10). In fact, the estimated overcharge is about 56% (9.87 percentage points) lower than the true overcharge. This is because the naïve model does not appropriately reflect changes in firms' behavior. To name one example, the naïve model does not reflect the fact that in the case of effectively colluding firms, cost-changes are passed on to consumers to a lesser degree than in the case of competition. It is needless to say that with sufficient efforts even this naïve approach might finally yield the correct regression specification (10). However, the correct specification can be much more easily derived from an adequate theoretical model of the industry. A second disadvantage of the naïve regression-approach is that the researcher can hardly form expectations about the sign and absolute value of the estimated coefficients. These expectations are necessary for testing the plausibility of the estimates.

One may summarize that a more descriptive econometric approach as shown in Table 3 is not necessarily doomed to fail. In contrast, it may well deliver reasonable estimates of cartel-overcharges. However, if the researcher has a good theoretical understanding of the industry's structure, he does well in obtaining a truly reduced-form model from the structural model of the industry. In the above example, using a structural model as a basis for the pricing regression makes the difference between a 0.63- and a 9.87-discrepancy (in percentage points) between estimated and true overcharges.

Additionally, I argue that an econometric estimation of overcharges is generally preferable to conceptually simpler methods. To see this, a "period of collusive activity can be compared with (1) a competitive period prior to the beginning of such activity; (2) a "white sale" period within the collusive period, or (3) a period after the termination of the conspiracy" (Finkelstein and Levenbach 1983: 161). In competitive periods (i.e. periods 1-20 plus the price war periods), a cartelist charges an average price of 40.01. In collusive periods, the average price of cartelists is at a level of 31.18. This implies an *undercharge* of 28.32%. Based on these "estimates" the defendant's lawyer might argue that the cartel was not effective at all. However, we as omniscient observers know that the cartel was effective, charging an average overcharge of 17.72%.

How can this puzzle be resolved? In periods with high marginal costs of production, margins are low and cartelists have a high incentive to deviate from the cartel (see section 2.2). Firms end up in a price war, setting prices competitively. However, these competitive prices are nonetheless set at high levels because of the high production costs. When production costs decrease, it becomes profitable for firms to restore collusion. The low level of production costs allows firms to set collusive prices which are even below their previously competitive levels. As all these effects occur in an artificial, simulated industry, it is easy to determine the true competitive price level in periods where collusive behavior is observed. This price level is 25.99. Comparing this unobserved, competitive price level to the observed cartel-price of 31.18, one gets an overcharge-estimate of 16.65%. There is no chance to obtain this value employing the above simple comparison of prices. Not surprisingly, even the naïve regression in Table 3 outperforms the simple comparison of prices.

4 DETECTING COLLUSION

An important input parameter to pricing-equation (10) is dummy-variable D_t that measures which firms participated in the cartel and in which periods the cartel was active and effective. In the previous section, I assume that the researcher has perfect knowledge about D_t. To see the importance of defining D_t correctly, consider that misspecifying D_t does result in biased regression-coefficients and, thus, biased overcharge estimates.[9] Ideally, the researcher can infer D_t from hard evidence such as protocols of cartelists' meetings that were seized by the competition authority. However, such evidence need not necessarily be available, complete, or reliable. In these cases, one might want to employ statistical tests that, e.g., are designed to detect breakpoints in time-series or identify changes in the behavior of firms, which is described by the coefficients of pricing regression (10).

Section 4 presents and evaluates methods for specifying the dummy-variable D_t. In doing so, one must answer two questions. First, did firms in the industry under investigation behave collusively? If yes, which firms participated in the cartel? Second, in which periods did cartelists set jointly profit-maximizing prices? Methods for answering the first question are presented and evaluated in section 4.1. Section 4.2 is concerned with the second question.

4.1 Detecting Cartelists

One possibility to detect if a firm behaved collusively is to compare its behavior to some competitive benchmark. (i) Benchmark values can be provided by a theoretical model of the industry as is shown below. Other benchmarks can be provided by competitive markets of (ii) similar goods or in (iii) other geographic areas. Further benchmarks are provided by (iv) fringe-firms in the same market or by (v) the behavior of cartelists in indisputably competitive periods.

Such benchmark-comparisons are conducted by estimating a pricing-regression under the hypothesis of competitive behavior and test whether the estimated coefficients match a competitive

9 Suppose that the above cartel was found to have been active in periods 21-24, 39-73, and 82-100. This specification only neglects one cartel-period and three price war periods in times, where short price wars interrupt longer cartel-periods. Estimating equation (10) with this faint misspecification results in a downwards biased overcharge estimate of 4.55%.

benchmark (Harrington 2008: 222). Rejecting this hypothesis with statistical tests provides evidence that firms did not behave competitively. In the above model, equation (9) indicates that firms' pricing behavior in clearly competitive periods can be mathematically described by reduced-form pricing regression (13).

$$p_{i,t} \quad = \quad \alpha_i + \beta_i p_{i,t-1} + \gamma_i s_t + \epsilon_{i,t} \tag{13}$$

After estimating equation (13) for the period under research, the estimated coefficients α_i, β_i, and γ_i may be compared to the above competitive benchmarks. The comparison to benchmarks (i), (iv) and (v) is illustrated below. In addition to detecting non-competitive behavior, equation (13) may be used to detect structural breaks in the pricing behavior of firms. This helps to identify in which periods cartelists behaved non-competitively, which is done in section 4.2.

I start with benchmark (i) for detecting behavior that is inconsistent with competition, i.e. I compare the estimated coefficients to their theoretical competitive values. The deterministic reduced-form pricing-equation (9) implies that in competition, the constant α_i should take a value of zero. β_i should take a value of one. γ_i measures to which extent cost shocks are passed on to consumers. Equation (2) shows that the marginal costs of firm i rise on average by $E(a_2)s_t$ in response to some cost-shock s_t. In perfect competition, firms set prices at the level of marginal costs as long as these are at or above average costs. Therefore, in perfect competition cost-shocks would be fully passed on to consumers, so that one would expect a value of $\gamma_i = E(a_2)$ (i.e. 0.76 in the above example). As the differentiation of products makes competition imperfect, cost-shocks are under-proportionally passed on to consumers so that one expects a value of γ_i in the interval [0; 0.76]. In the following, I estimate equation (13) for periods t=1-100 and test whether the estimated coefficients match the predicted competitive values. The results of pricing regression (13) are presented in Table 4.[10] These results are obtained by a stacked OLS panel-regression.

This regression does not clearly indicate a deviation from competitive behavior by firms 1, 2, and 4-9. The estimated constants are not significantly different from zero and the β-coefficients are close to one. Only the γ-coefficients are perceptibly lower than the upper bound of the above interval [0; 0.76]. It is also not possible to reject the hypothesis that firms behave competitively. The p-values of two Wald-tests are relatively high, which leads to the inability to reject the hypotheses that (i) the constants α_i jointly take a value of zero and that (ii) the β-coefficients jointly take a value of one.

10 Please note that the last column presents the regression-results of equation (13) for firm 2 in periods t=1-20. The last four rows summarize the results of tests for structural breaks. These results are not interpreted here, as they are used further below.

dependent variable $p_{i,t}$	t = 1-100									t = 1-20
	firm 1	firm 2	firm 3	firm 4	firm 5	firm 6	firm 7	firm 8	firm 9	firm 2
α_i	0.416 (0.500)	0.445 (0.528)	0.090 (0.465)	0.423 (0.489)	0.428 (0.493)	0.409 (0.515)	0.468 (0.516)	0.431 (0.520)	0.472 (0.527)	-0.321 (0.670)
$p_{i,t-1}$ β_i	0.988 (0.014)***	0.988 (0.015)***	0.998 (0.013)***	0.988 (0.014)***	0.988 (0.014)***	0.989 (0.014)***	0.987 (0.014)***	0.988 (0.014)***	0.987 (0.014)***	1.008 (0.015)***
s_t γ_i	0.531 (0.060)***	0.537 (0.060)***	0.626 (0.060)***	0.544 (0.060)***	0.546 (0.060)***	0.538 (0.060)***	0.536 (0.060)***	0.537 (0.060)***	0.549 (0.060)***	0.730 (0.011)***
R^2	98.12%									99.75%
$\bar{R}^2$	98.06%									99.72%
Information Criteria			Wald Test		p-value					
Akaike	3.173		$\alpha_i = 0 \; \forall \, i$		75.09%					-1.67
Schwarz	3.318		$\beta_i = 1 \; \forall \, i$		72.30%					-1.52
H-Q	3.229		$(\alpha_i = 0) \wedge (\beta_i = 1) \; \forall \, i$		99.52%					-1.65
recursive coefficients	mostly correct									
recursive residuals	correct									
CUSUM	incorrect (does not indicate structural break at all)									
CUSUMSQ	correct (no information about timing)									

Table 4: Competitive Reduced-Form Pricing Regression (13)

In the following, I assume that the competition authority knows that firm 3 had acted as a competitive fringe firm. This knowledge is supported by the above regression, as the estimation-results for fringe firm 3 are closer to the above theoretical benchmarks than the estimated coefficients of the other firms. Therefore, I examine if firm 3 can serve as a competitive benchmark (i.e. the above benchmark (iv)) so that the other firms can be shown to have behaved in a significantly different, i.e. non-competitive, manner. An F-test indicates that the coefficients of cartelist 1 and cartelist 2 are identical with a probability of 99.96%. However, the probability that the coefficients of cartelist 2 and fringe-firm 3 are identical is only 72.27%. This is not low enough to reject the hypothesis (at standard confidence-levels) that cartelist 2 and fringe-firm 3 behaved in the same way. This finding highlights one drawback of this approach. When the cartel is active, fringe firms do not exactly behave as they would in competition. They rather expand their price under the price-umbrella of the cartel. Therefore, cartelists and fringe firms cannot be expected to behave very dissimilar.

Therefore, I resort to benchmark (v), i.e. comparing supposedly collusive behavior to clearly competitive behavior. The last column in Table 4 provides the results of the reduced-form pricing regression (13) for cartel-firm 2 during the clearly competitive periods 1-20. The estimated coefficients β_2 and γ_2 provide good estimates of the theoretically true, competitive coefficient-values. However, as the time series is relatively short, α takes a value below zero that is not statistically significant. Therefore, the quality of this regression must be considered too low to be useful for further statistical tests.

To summarize the above discussion, interpreting the values of the estimated coefficients supports the assumption that firm 3 behaved competitively while all other firms behaved collusively. However, the above statistical tests cannot reject the hypothesis that all firms behaved competitively. Nonetheless, such a finding must not be considered a proof of competition. It should be considered inconclusive evidence that requires further investigation.

4.2 Detecting Collusive Periods

In the following, assume that the identity of cartelists and fringe-firms is known but the start of the cartel as well as the timing of price war periods is unknown. Therefore, I now address the second question, i.e. identifying periods where firms probably behaved non-competitively. This is done by performing four popular tests for detecting structural breaks in the data. These tests are (i) analyzing plots of recursive coefficients, (ii) analyzing plots of recursive residuals, (iii) the CUSUM-test, and (iv) the CUSUMSQ-test.

Before addressing these tests, I argue that the simplest test for detecting collusive periods is often unlikely to yield meaningful results. This test simply consists of looking at the time series of prices and identifying possible breakpoints by visual observation (Harrington 2008: 221). The inadequacy of this test can be seen by visually examining the prices of cartelists (dark gray line) in Figure 1 above. One would probably identify the price wars that start in periods 25 and 59. However, I doubt that one would correctly identify the start of the cartel-period in t=39, as the price-trend keeps falling, being driven by negative cost-shocks. One might object and argue that additional switches between competitive and collusive behavior can be found by identifying periods where costs decline but prices rise. This is clearly inconsistent with competitive behavior. Even for the above simple industry, this is a tedious exercise and is even more difficult in real cases in which goods are not produced by a single but by several input factors, where each has a unique evolution of prices. As a consequence, Harrington (2008: 219) suggests applying a statistical test for breakpoints in firms' pricing behavior such as the ones described below.

To perform these tests, regression (13) must be run for each firm separately. This yields the same results as those presented in Table 4 for the stacked panel-regression. For ease of notation, equation (13) is rewritten as $p_{i,t} = \boldsymbol{Z}_{i,t}\boldsymbol{\theta}_{i,t} + \varepsilon_{i,t}$, where $\boldsymbol{\theta}_{i,t}$ is the vector of regression-coefficients and $\boldsymbol{Z}_{i,t}$ is the vector of regressors. Here, the vector includes a constant, lagged prices, and the contemporaneous common cost-shock. All of the below tests for structural breaks rely on recursive regressions of equation (13). Therefore, a sequence of coefficient-vectors, $\hat{\boldsymbol{\theta}}_{i,k+1}, \hat{\boldsymbol{\theta}}_{i,k+2}, \hat{\boldsymbol{\theta}}_{i,T}$, is generated by running separate regressions for the k+1, k+2, ..., T first observations, where k is the number of explanatory variables.

The first test for structural breaks relies on plots of these recursive coefficients and their standard errors. A structural break may be supposed if some coefficient $\hat{\theta}_{\iota,k*}$ moves out of the bound provided by $\hat{\theta}_{\iota,k*-1}$ plus or minus two times the standard error of $\hat{\theta}_{\iota,k*-1}$. This is approximately equivalent to a 5%-level of significance. The plots of firm 1's recursive estimates are provided in Figure 2 in Appendix B. It is sufficient to provide the plots for one firm only because the figures of the other firms look very similar. This test of recursive coefficients correctly detects cartel-periods around the start of the cartel where many competitive observations and only few collusive observations are included in the regression. Therefore, the inclusion of collusive periods causes perceptible changes in the values of the coefficients. In particular, the test detects the start of collusion in periods 21, 37, and 39. As more and more collusive observations enter the estimation, the test indicates *that* a structural break has occurred but not necessarily *when*. The evolution of the coefficient γ_i provides some evidence of a structural change, as it moves from a level of about 0.73 in competitive periods to the level shown in Table 4 when all periods are considered in the regression. However, it provides exact information about the timing of neither collusion nor price wars.

Second, the results of this recursive estimation can be used for computing recursive residuals, i.e. one-step ahead prediction-errors.

$$\omega_{i,t} = p_{i,t} - \boldsymbol{Z}_{\boldsymbol{i},\boldsymbol{t}} \hat{\boldsymbol{\theta}}_{i,t-1} \quad . \tag{14}$$

Information about a structural break in period $k*$ can be derived from a plot of recursive residuals. Hence, a structural break is assumed when $\omega_{i,k*}$ moves out of the bound provided by $\omega_{i,k*}$ plus or minus two times its standard error for $k*$-1. Based on this test, the start and end of all collusive periods is determined accurately from the regressions of all firms. This can be seen from Figure 3. The effect of collusion shows up in the error term quite perceptibly. However, the establishment of the cartel in period 21 only shows up as a very faint structural break. This is because marginal costs and, thus, prices had already been on the rise in the previous periods. In the same context, a structural break shows up in period 28 where prices dropped quite strongly because of a decrease in marginal costs. However, in this period cartel-activity did not change.

These results indicate a conceptual difficulty in the use of structural break tests. They show all structural breaks and not only those related to the cartel. Therefore, the researcher needs to be able to identify those structural breaks that can be attributed to phenomena other than collusion. This is particularly important when structural breaks occur as a consequence of structural changes in the industry such as, e.g., the entry/exit of firms, mergers, technology shocks, or demand shocks. Therefore, the researcher must have good knowledge of the industry to determine whether an identified structural break can be attributed to such alternative explanations rather than to collusion (Harrington 2008: 220).

Third, the CUSUM-test provides a rather general test for a structural break, i.e. one which gives no indication of the time-period when the structural break may have occurred. A structural break is assumed if the cumulative sum of the above recursive residuals – each scaled (i.e. divided) by its standard error – differs significantly from zero in any time period t. Let the scaled recursive residuals be denoted by $w_{i,t}$. The CUSUM-test does not detect a structural break for any of the above firms as can be seen from Figure 4. This failure of the CUSUM-test can easily be explained. When the cartel is established, the increase in prices shows up as a positive spike in the error term. When the cartel enters a price-war period, the decrease in price shows up as a negative spike in the error term. By summing up these errors, the effect of switching from competition to collusion and the effect of switching from collusion to competition cancel each other out.

Fourth, the CUSUMSQ-test statistic

$$S_{i,t} = \frac{\sum_{j=k+1}^{t} w_{i,j}^2}{\sum_{j=k+1}^{T} w_{i,j}^2} \quad , \tag{15}$$

which takes account of this effect by taking the square of the scaled recursive residuals, is therefore calculated. Again, one must reject the null-hypothesis of “no structural break” if the test-statistic $S_{i,t}$ exceeds some critical value. Figure 5 shows that this test indicates a structural break. This finding applies for all firms. However, due to the construction of the test, one gets no information about the timing of collusion.

In the case of cartels that become active and remain effective, one might also apply a Chow breakpoint-test to verify the supposed and unique structural break. This test compares the sum of squared residuals obtained by fitting a single equation to the entire sample with the sum of squared residuals when separate equations are fit to each sub-sample of the data (Chow 1960). However, this test is of little use for the above industry that is characterized by firms which switch frequently between competition and collusion.

In summary, one finds that (i) tests for structural breaks may be used to identify periods of competition or collusion in order to specify the cartel-dummy D_t. (ii) Some of these tests, such as the CUSUM-test, are practically never applicable for this use. (iii) Other tests can be applied in principle, but may – by construction – have drawbacks. For example, the CUSUMSQ-test is not suitable to identify the time of the structural break. The Chow-test is only a good choice when firms rarely switch between competitive and collusive behavior. (iv) When these behavioral changes occur frequently, tests that are based on recursive coefficients or recursive residuals may be valuable tools. (v) Generally, one should use several tests, which ideally provide complementary evidence. (vi) Before attributing a structural break to the formation or demise of a cartel, one should look out for other events that might possibly cause a structural break.

5 CONCLUSION

Reduced-form pricing regressions are commonly used in the detection and prosecution of cartels. Based on the analysis of a simulated sample-industry, I provide evidence for the proposition that a pricing regression should be preferably derived from a theoretical model of the industry under research. In contrast, descriptive pricing-regressions may, at first sight, look like reduced-form regressions because they contain the same or quite similar variables. However, the functional form of a naïve descriptive regression may be inappropriate and/or miss the complex interplay of the included variables. Nonetheless, my analysis illustrates that even such a naïve approach outperforms non-econometric approaches such as the comparison of prices in collusive and competitive periods.

Based on the above reduced-form regressions I perform tests of coefficients' statistical significance and tests for structural breaks in the data. These tests aim to identify cartelists and collusive periods. I show that such tests can suffer from two types of errors. On the one hand, there may be technical obstacles such as time series that are too short, or functional forms that inaccurately describe the underlying economic relationship. On the other hand, the testable hypotheses might be economically flawed. Sometimes it is argued that the behavior of fringe-firms provides a benchmark for the behavior of cartelists in the same market. I show that this benchmark may be used but is certainly imperfect. This is because fringe-firms adjust their behavior to the existence of their competitors' cartel and, thus, only behave quasi-competitively.

I arrive at these findings with a new and innovative method. I use a theoretically supported and detailed simulation model to generate panel-data of a collusive industry, and apply the above methods to this data. This approach is advantageous as it allows econometricians to use the generated data to explain and illustrate the pros and cons of their methods to non-econometricians. In doing so, the recipients need not go through the technical details of a regression-approach or resort to a formal proof that both are not necessarily easy to understand. Using examples based on simulated data is convenient, as the accuracy of estimates can be determined by a direct comparison of the estimated coefficient and its true underlying. This allows for an intuitive assessment of the results and is more meaningful than a case study where the empirical methods are applied to real case data. Such studies are less illustrative because the true value of, e.g., an overcharge is unknown.

Analyzing the properties of different empirical methods on the basis of simulated data has a further advantage other than simply being illustrative. One may also obtain quantitative estimates of both correctly specified and misspecified models and, thus, get an idea of the size of misspecified estimates' distortion. This can be seen from the above example. Therefore, future research might be directed at deriving the distribution of possible errors. Such distributions might be determined for (i) different cost-evolutions in the same cartel, (ii) different cartel-compositions in the same industry, or (iii) different cartels in different industries.

The superiority of the theory-based reduced-form pricing regression rests on two requirements. First, there must not be an (undetected) structural change in the data such as innovations, entry/exit of firms, changes in demand, or changes in the substitutability of products. Such changes would have to be accounted for in the specification of the pricing regression. Therefore, future research should be devoted to including such effects in the simulation model and deriving the pricing-regression accordingly. A second requirement for the superiority of the theory-based estimation is its characteristic of being a good description of the underlying (simulated) industry. An important question for future research is how one would arrive at this model of the industry in reality.

This article constitutes a first step in the simulation-based analysis of econometric analyses in competition policy. Its contribution lies in providing a framework for such analyses. Moreover, it gives a first indication of the size of estimation errors that are caused by inappropriate specifications, techniques, or data. Analyzing these effects in greater detail constitutes an interesting avenue for further research.

REFERENCES

[1] Baker, J.B. (1999). "Empirical Methods in Antitrust Litigation: Review and Critique." American Law and Economics Review. Vol. 1 No. 1/2, pp. 386-435

[2] Bajari, P. and Ye, L. (2003). "Deciding Between Competition and Collusion." The Review of Economics and Statistics. Vol. 85 No. 4, pp. 971-989

[3] Bos, I. and Harrington, J.E. (2010). "Endogenous cartel formation with heterogeneous firms." The RAND Journal of Economics. Vol. 41 No. 1, pp. 92-117

[4] Chow, G.C. (1960). "Tests of Equality Between Sets of Coefficients in Two Linear Regressions." Econometrica. Vol. 28 No. 3, pp. 591-605

[5] Clark, E. and Hughes, M. and Wirth, D. (2004). "Study on the Conditions of Claims for Damages in Case of Infringement of EC Competition Rules." http://ec.europa.eu/competition/antitrust/actionsdamages/comparative_report_clean_en.pdf

[6] Connor, J.M. (2008). "Forensic Economics: An Introduction with Special Emphasis on Price Fixing." Journal of Competition Law and Economics. Vol. 4 No. 1, pp. 31-59

[7] Connor, J.M. and Lande, R.H. (2008). "Cartel Overcharges and Optimal Cartel Fines." in Issues in Competition Law and Policy. ABA Section of Antitrust Law. Chicago: ABA Book Publishing, 2008

[8] Directorate General Competition (2010). "Best Practices for the Submission of Economic Evidence and Data Collection in Cases Concerning the Application of Articles 101 and 102 TFEU and in Merger Cases." http://ec.europa.eu/competition/consultations/2010_best_practices/best_practice_submissions.pdf

[9] Donsimoni, M.P. (1985). "Stable Heterogeneous Cartels." International Journal of Industrial Organization. Vol. 3 No. 4, pp. 451-467

[10] Fershtman, C. and Pakes, A. (2000). "A dynamic oligopoly with collusion and price wars." The RAND Journal of Economics. Vol. 31 No. 2, pp. 207-236

[11] Finkelstein, M.O. and Levenbach, H. (1983). "Regression Estimates of Damages in Price-Fixing Cases." Law and Contemporary Problems. Vol. 46 No. 4, pp. 145-169

[12] Hausmann, W.J. (1984). "Cheap Coals or Limitation of the Vend? The London Coal Trade, 1770-1845." Journal of Economic History. Vol. 44 No. 2, pp. 321-328

[13] Harrington, J.E. (2008). “Detecting Cartels.” in Handbook of Antitrust Economics. Edited by P. Buccirossi. Cambridge, MA: The MIT Press, 2008

[14] Heij, C., de Boer, P., Franses, P.H., Kloek, T. and van Dijk, H.K. (2004). “Econometric Methods with Applications in Business and Economics.” Oxford: Oxford University Press, 2004

[15] Monopolkommission (2010). “Mehr Wettbewerb, wenig Ausnahmen.” Achtzehntes Hauptgutachten der Monopolkommission gemäß § 44 Abs. 1 Satz 1 GWB. http://www.monopolkommission.de/haupt_18/mopoko_volltext_h18.pdf

[16] Oxera (2009). “Quantifying Antitrust Damages. Towards Non-Binding Guidance for Courts.” Study prepared for the European Commission. http://ec.europa.eu/competition/antitrust/actionsdamages/quantification_study.pdf

[17] Paha, J. (2010). “Simulation and Prosecution of a Cartel with Endogenous Cartel Formation.” MAGKS Discussion Paper No. 07-2010. http://www.uni-marburg.de/fb02/makro/forschung/magkspapers/07-2010_paha.pdf

[18] Prokop, J. (1999). “Process of dominant-cartel formation.” International Journal of Industrial Organization. Vol. 17 No. 2, pp. 241-257

[19] Rotemberg, J.J. and Saloner, G. (1986). “A Supergame-Theoretic Model of Price Wars during Booms.” The American Economic Review. Vol. 76 No. 3, pp. 390-407

[20] Shubik, M. and Levitan, R. (1980). “Market Structure and Behavior.” Cambridge, MA: Harvard University Press, 1980

[21] Reiss, P.C. and Wolak, F.A. (2007). “Structural Econometric Modeling: Rationales and Examples from Industrial Organization.” in Handbook of Econometrics, Vol. 6A. Edited by J.J. Heckman and E.E. Leamer. Amsterdam: Elsevier North-Holland, 2007

Appendix A Generating a Collusive Industry

The dataset that is used in the above regressions is generated by running the simulation model with the parameter values given in Table 5. Except for the number of cartelists *m*, these parameter values are randomly determined from uniform distributions within the bounds provided by Table 5.

	n	m	v	μ	a_1	a_3	a_4	P	r
lower boundary	7		50	1	0.05	0	0.05	0.05	0.05
upper boundary	15		150	100	1	1	0.15	0.4	0.25
selected value	9	8	56	18	0.85	0.52	0.12	0.16	0.05

Table 5: Parameter Values

Choosing $n \in [7,15]$ is reasonable because, first, all firms would join the cartel with probability 1 if the number of firms was too small. Second, the time for calculating the participation-equilibrium rises exponentially in the number of firms and, thus, would be undesirably long for $n>15$. From the viewpoint of economic theory, the size of v is irrelevant as it only affects the size of prices and quantities but does not have an impact on the ratio of profit-measures. Using $\mu_{upper}=100$ as an upper bound is reasonable as it suffices to produce rather homogeneous goods. Choosing $a_1 \in [0.05,1.0]$ is reasonable because values below 0.05 would indicate that marginal costs are quite negligible. The production of such firms may be supposed to rather generate fixed costs that, however, are beyond the scope of this model. Choosing $a_4 \in [0.05,0.15]$ gives economically meaningful yet not unrealistically large cost shocks. Drawing P from the wide interval [0.05,0.4] reflects our lack of knowledge about the effectiveness of competition authorities. This is because one knows the number of discovered cartels but can hardly determine the number of undiscovered ones. The interval encloses the 15-20% detection probability that some studies suggest. Choosing $r \in [0.05,0.25]$ suggests that firms' discount rate is somewhere between the return of government bonds and some (ambitious) firms' target value of their return on equity. The number of cartelists m is determined endogenously as the mixed-strategy Nash-equilibrium of the cartel-formation game.

Appendix B Testing for Structural Breaks – Figures

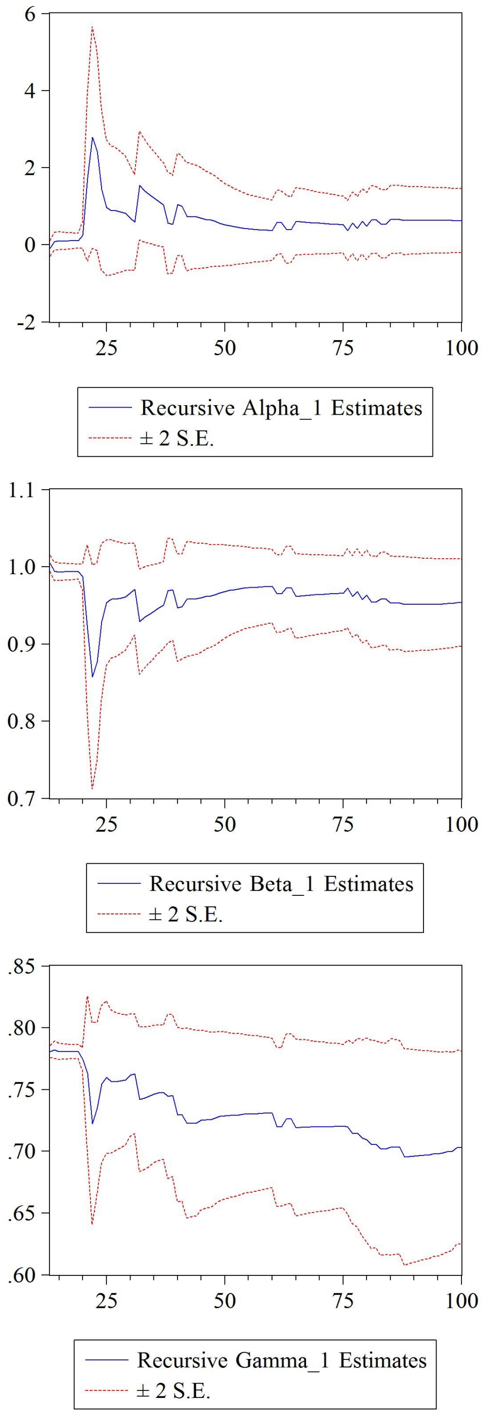

Figure 2: Recursive Estimates

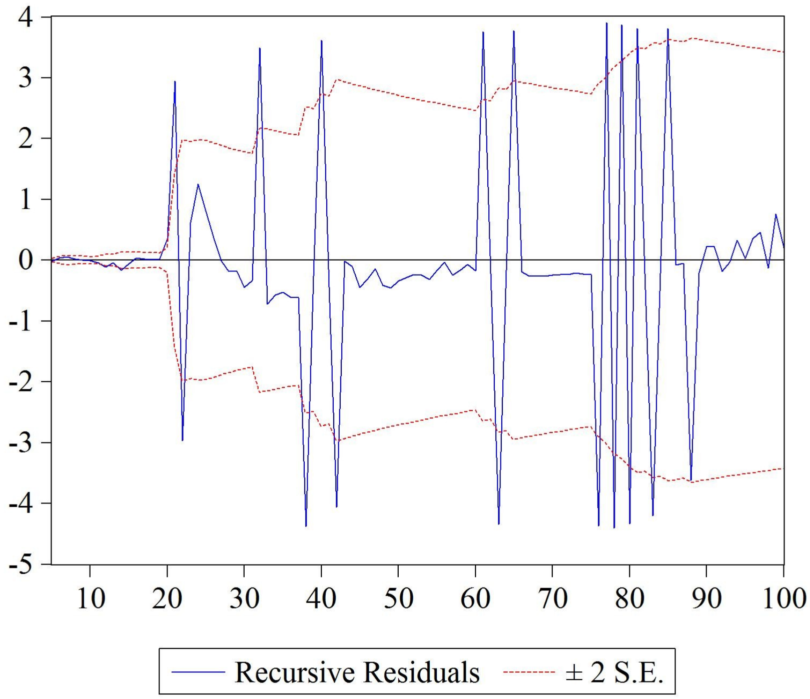

Figure 3: Recursive Residuals

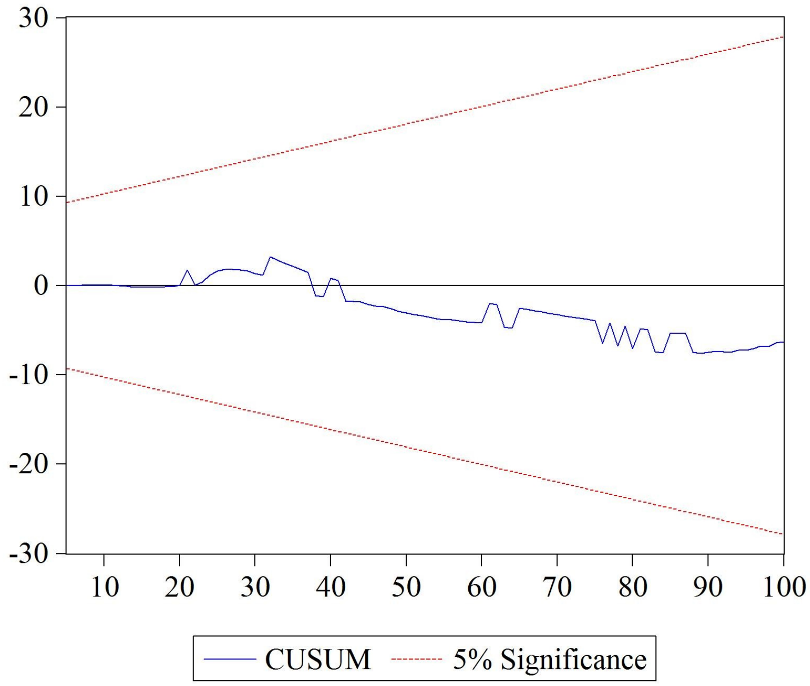

Figure 4: CUSUM-test

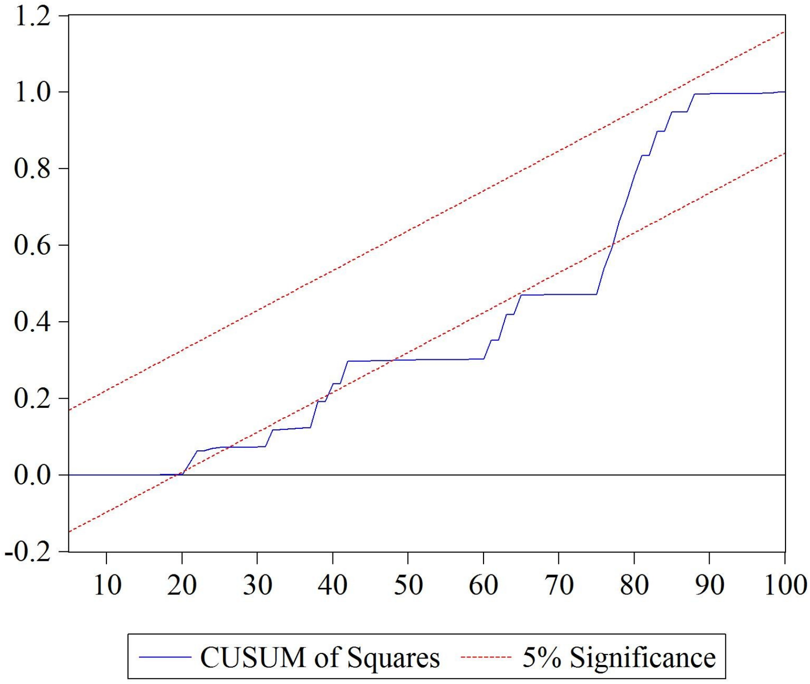

Figure 5: CUSUMSQ-test