Collusive **S**cene **I**nvestigation: A Modern Empirical Approach to Uncover Cartels

Pedro Gonzaga*
University of Porto and CEF.UP
eupedrogonzaga@gmail.com

António Brandão
University of Porto and CEF.UP
abrandao@fep.up.pt

Hélder Vasconcelos
University of Porto and CEF.UP
hvasconcelos@fep.up.pt

Natércia Fortuna
University of Porto and CEF.UP
natercia.fortuna@gmail.com

May 11, 2015

*We acknowledge *Fundação para a Ciência e Tecnologia* for the financial support given through the PhD scholarship SFRH/BD/78763/2011 and in the framework of the project Pest-OE/EGE/UI4105/2014.

Abstract

We develop an empirical method to detect price collusion from the analysis of observed data. Our approach describes the supply side of the industry as a switching regression with two regimes, collusion and competition, which is estimated using a modified EM algorithm, based on the two-stages least squares method. We use simulated data to show that our algorithm is able to accurately predict collusion and to obtain unbiased estimates for the parameters of the switching regression. Our conclusions can be extended to other forms of collusion and to other economic problems.

1 Introduction

Collusion is one of the most serious economic crimes committed against society, whose effects contaminate the entire economic system. Indeed, as a result of the actions of cartels, millions of people have access denied everyday to a large variety of goods and services that are too expensive for their own budgets, although the markets could efficiently provide them at low prices. Individuals who afford such expensive goods are forced to pay for them a greater fraction of their income and are able to buy less of alternative goods. Downstream companies that rely on cartelized producers face higher costs and are forced to charge higher prices to their customers as well, having less money available to distribute among their workers and shareholders. By reducing the level of output transacted, cartels are also a source of low employment and even low wages, particularly among highly skilled workers who have invested considerable time and money on formation and training. But worse of all, because collusion is not a zero sum game, the gains earned by cartels are far exceeded by the damage imposed on the consumers, workers, administrators, and even shareholders of competitive firms, causing society to lose as a whole in a manner that can hardly be measured.

Unlike more traditional crimes as murder, robbery, bribery, extortion and fraud, which affect the integrity and wealth of individuals in a very perceivable way, the costs of collusion are dispersed over a large number of victims who rarely realize they have been injured at all, making collusion extremely hard to detect. Even so, every time a cartel fixes prices or restrains competition in the internal market, a trace is left in the pattern of economic data that can be tracked by proper statistical tools. Therefore, in the same way murders and robberies are investigated with advanced techniques of forensic science to analyze DNA, fingerprints, footwear impressions and blood spatter, it is possible to develop analogous advanced econometric methods to screen the data and to constantly seek evidence of collusion.

While the empirical analysis of economic data may be extremely useful to uncover cartels that have successfully remained in secret, it is important to keep in mind that these methods cannot be used as hard evidence to prove guilt in the court of law. Nevertheless, we believe they can still be applied as, using the wording of Harrington (2005), a screening and verification device to identify the industries worthy of further investigation. It should be their main purpose to improve an efficient allocation of antitrust authorities' resources towards the industries with higher likelihood of collusion, contributing to an increasing number of uncovered cartels. Moreover, in case hard evidence is collected and firms are actually condemned, the same empirical methods can be used to estimate the overcharge of the cartel, in order to determine the adequate fee or sentence.

Despite the existence of some empirical models to detect collusion in the economic literature, so far none has been systematically used by competition authorities, whose investigation continues to rely mainly on whistle blowers. In reality, most models are still too complex, hard to implement, require the collection of a lot of data and must be adjusted case-by-case, creating a great problem for competition authorities who lack the time and resources to investigate every industry in such detail. Furthermore, there is still little evidence that those models are able, indeed, to accurately distinguish competition from collusion, once they have not been rigorously tested in the two distinct scenarios.

As a result, it is our purpose in this paper not only to present a parsimonious and computationally efficient method that can be easily applied by antitrust authorities to detect collusion, but also to prove that our approach is accurate and robust. For that, we describe the supply side of the industry as a switching regression with two regimes, collusion and competition, which is estimated by an expectation-maximization algorithm specifically designed to our problem. Using simulated data, we show that our algorithm is able to accurately predict collusion and to obtain unbiased estimates for the parameters of the switching regression. In fact, we explain how to correct any estimation bias that arises from the misidentification of the regimes and how to deal with the endogeneity problem in the supply equation. Resorting to simulated data has the enormous advantage of knowing exactly when firms are colluding and having access to the real parameters of the population, which enable us to determine the success rate of the method and to evaluate how close the estimates obtained are to the true underlying values. A similar procedure was followed by Paha (2011), who simulated an industry in order to evaluate different empirical methods to determine the overcharge of cartels.

In the next section we review some previous models of collusion detection in the literature, in order to sustain the direction of our particular approach. In Section 3 we briefly illustrate switching regressions in a simple application. In Section 4 we explain how to detect collusion with a switching regression, which is estimated using an expectation-maximization algorithm. In Section 5 we identify the estimation bias that results from the misidentification of competitive and collusive periods, and we explain how to correct it. In Section 7 we discuss the identification problem of the supply equation due to the endogeneity of the quantity transacted. In Section 8 we present a new EM algorithm which solves the identification problem and is able to estimate consistently the parameters of the model. Section 9 discusses a statistical test for structural breaks. Lastly, Section 10 concludes.

2 Literature Review

Most empirical models of collusion detection described in the literature can be divided into three main groups: the models based on statistical features of data, the ones based on price-cost margins and those based on structural breaks.

The models in the first group attempt to identify features of data that are consistent with either competition or collusion, by testing, for example, if price levels are correlated after controlling for demand or cost factors. In this respect, Bajari and Ye (2003) test for the presence of collusion in procurement auctions conducted by construction firms of the seal coating industry, in the Midwest, between 1994 and 1998. Their methodology involves checking whether firms' bids are independent and exchangeable (that is, if a permutation of the costs of firms leads to a permutation of their bids), two properties that should be verified in a competitive bidding scenario. Similar tests were conducted earlier by Porter and Zona (1993), who analyzed bidder collusion by highways construction firms, in Long Island, in the 80s; and by Baldwin, Marshall and Richard (1997), who investigated collusion between purchasers of timber of the Forest Service in Pacific Northwest, in the 70s. Unfortunately, these models are usually narrowed to detect collusion in auctions and they cannot be easily extended to the analysis of other industries, where properties like independence and exchangeability may fail to distinguish collusion from competition.

The second group of models measures price-cost margins or other performance indices to access the degree of market power in a particular industry, which is then compared with a competitive benchmark (usually another industry or market).[1] However, attempting to infer collusion from price-cost margins may lead to spurious results. Foremost, there is rarely good cost data available in most databases and one may be forced to use indirect methods based on cost estimations.[2] But even more importantly, since price-cost margins depend on so many economic factors as product differentiation, market size, patents, barriers to entry and regulation, high margins are not exclusively observed in cartels and there are many profitable industries without any evidence of collusion. A more modern approach uses economic theory to decompose the observed price-cost margins into unilateral effects (which may be the outcome of product differentiation or market structure) and coordinated effects, testing then whether the later are statistically significant. Such approach was initially developed by Nevo (2001) to show the absence of coordinated strategies in the breakfast cereal industry in United States and it was followed by Slade (2004) to study the brewing industry in United Kingdom. Unfortunately, this not only requires good cost data, but also the estimation of demand using discrete choice models (see Berry, 1994) and the computation of several price-competition equilibria.

Finally, the third group of models searches for structural breaks in time series, in order to identify the moments in time when cartels are created or closed. As long as there is some a priori information that may suggest breakpoints in data, structural changes can easily be checked using a Chow test (1960). If the econometrician does not have

[1]This analysis is related to the *structure, conduct and performance paradigm* attributed to Bain (1951), which dominated empirical industrial economics in the second half of the twentieth century.

[2]In this respect, see Bresnahan (1989).

any clues about possible breakpoints, a structural change in the whole time series can still be sought by following Quandt (1960). In any case, one must be very cautious with this approach, since the structural break may be triggered by other reasons than collusion. A much more refined technique was proposed by Porter (1983) to identify the time periods when the Joint Executive Committee was operational, a famous railroad cartel in the 1880s that faced alternative periods of collusion and reversion to Nash equilibrium. His method consists in the joint estimation of two simple linear equations, a homogeneous market demand and a supply relationship described by a switching regression with two regimes, collusion and competition. Using a version of the EM algorithm for the estimation of switching regressions, Porter was able to identify the regime observed in each period. Since then some authors have extended Porter's model to other industries, as Almoguera, Douglas and Herrera (2007), who studied the Organization of Petroleum Exporting Countries (OPEC) between 1974 and 2004 and tested for switches between collusive and non-cooperative periods.

The model proposed in this paper belongs to the last group and is closely related to Porter (1983), which among the methods previously described, has the advantage of requiring few data, having good precision and being universally applied to most industries with little modifications. Indeed, as a long as a cartel is able to successfully affect prices in any degree, some breaks in data must be observed at some point. Still, there are two important features that distinguish our model from Porter's. Firstly, we do not estimate the demand equation, which is particularly useful when we do not know its functional form or we do not observe many demand side variables. Secondly, our switching regression is estimated with a modified EM algorithm that uses only analytical estimators at each iterative step (like TSLS). This makes our method faster, easier to compute and less sensitive to initial points.

Before actually presenting our own methodology to detect collusion, we will now illustrate the role and predictive power of switching regressions in a more simple example.

3 An illustratitive example

Switching regression models have been increasingly applied not only in economics, but also in natural sciences, to study the behavior of populations whose structural parameters vary across two or more regimes. In what follows, we briefly illustrate one of many possible applications of switching regressions.

Imagine a large population composed of adult men and women, not necessarily in the same proportion, and suppose it is our task to estimate the average height per gender from a random sample of observations. However, as a major drawback, suppose that the gender of each individual is unknown and there is no information about the fraction of men and women in the population. In fact, there is absolutely no data available but a finite sample of randomly extracted heights. In such circumstances, how can we predict the average height of males and females?

At first sight, this problem appears to be unsolvable. Yet, due to the fact that men are, in average, taller than women, the dataset of heights of the entire population can be actually approximated to a mixture of two normal distributions, as illustrated in Figure 1. Then, all we need is to identify the two uniform normal distributions for men and women that originated the envelop distribution, as depicted in Figure 2. Naturally, one must account that some tall women are higher than short or average men, just as the opposite is also true, reason why the two density curves in Figure 2 intersect each other.

Fortunately, that is precisely what switching regressions do. By describing the height of individuals as a switching regression with two unobserved regimes, male and female, we are able to estimate with good precision men's average height, women's average height, the fraction of men and women in the total population and, most impressive, the probability of each individual being a man or a woman given the height observed.

The estimation of the illustrative problem previously described is certainly as impressive as it is entirely useless, once there are hardly any databases where the gender of individuals is unknown. Still, if we apply the same principles to a dataset of prices and other control variables, we may have come up with a good mechanism to distinguish competitive from collusive regimes, as we will now show.

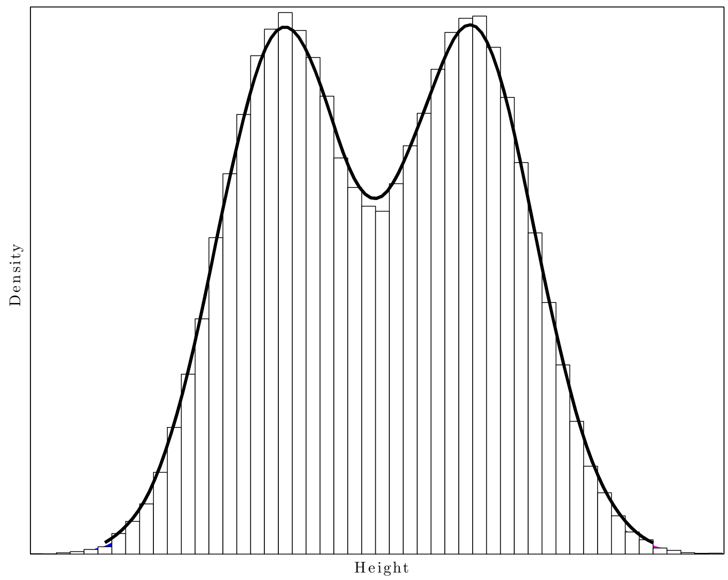

Figure 1: Distribution of heights in the population

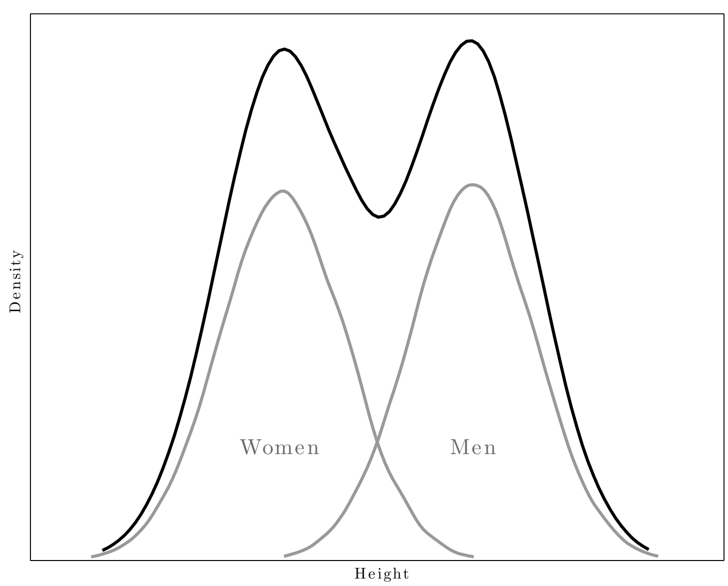

Figure 2: Distribution of heights by gender

4 Detecting Collusion

It is well established in the economic literature that cartels alternate consecutive periods of cooperation with temporary periods of punishment, or reversion to Nash, whenever a deviation occurs.[3] This typical behavior imprints a peculiar pattern in observed prices or other relevant variables, whose data generating process becomes characterized by switches between collusive and competitive regimes. Even a cartel that is successfully sustained without any deviations from the collusive path must, at some point in time, be either created or terminated, leading to a structural break in data that separates the two opposed regimes.

To see how this information can be used to detect collusion, consider a simple industry composed of a small number of firms selling a homogeneous product and repeatedly interacting along a horizontal time period. Then, if firms coordinate prices in a formal cartel, the supply relationship of the industry will be described by a regression of the type:

$$P_t = \begin{cases} MC_t + \beta_c Q_t + u_t^c, & \text{if } S_t = 1 \\ MC_t + \beta_n Q_t + u_t^n, & \text{if } S_t = 0 \end{cases} \tag{1}$$

where P_t is the price, MC_t are marginal cost related variables, Q_t is the quantity produced, u_t^c and u_t^n are normal unobserved errors with constant variance σ^2, and S_t is a state variable that takes the value 1 in collusive periods and 0 in competitive periods. Introducing S_t in the supply relationship, equation (1) can be also expressed as:

$$P_t = MC_t + \beta_c Q_t S_t + \beta_n Q_t (1 - S_t) + u_t. \tag{2}$$

Naturally, if the state variable St was observable, equation (2) could be easily estimated by least squares. Notwithstanding, just as in the illustrative example in the previous section we did not know which observations corresponded to men or women, we also ignore here in which periods firms compete or collude. As a result, we face the more complex challenge of estimating not only the structural parameters β_c and β_n, but also the state variable S_t. This problem can be solved by estimating equation (2) as a switching regression model following the methodology in Quandt (1972),[4] as we will now illustrate.

[3]See Green and Porter (1984).

[4]For an alternative methodology to estimate switching regressions, see Goldfeld and Quandt (1972).

Suppose that, in the industry under analysis, the regime operational at time t is independently set as collusion or competition, with a constant but unknown probability:

$$S_t = \begin{cases} 1, & \text{with probability } \lambda \\ 0, & \text{with probability } 1-\lambda \end{cases} . \tag{3}$$

This is equivalent to say that S_t follows a Bernoulli distribution with parameter λ, where the value of λ can be interpreted as the fraction of collusive periods in the whole time series or, similarly, as the unconditional probability of collusion, while $1-\lambda$ has the analogous interpretation for competition. We could also assume S_t to be generated by other stochastic processes such as a Markov chain, with few implications on the estimation results.

Under this simplifying assumption, Kiefer (1978) proved that there is an unique, consistent and asymptotically efficient estimator for the coefficients of the switching regression in equation (2), which corresponds to a local maximum of the likelihood function

$$L = \prod_{t=1}^{T} \Big\{ \lambda f(P_t|S_t=1) + (1-\lambda) f(P_t|S_t=0) \Big\}, \tag{4}$$

where $f(.)$ is the probability density function of P_t conditional on that a specific regime was observed.

Unfortunately, because the maximizing conditions of the likelihood function above are non-linear and have several roots, it is not easy to detect which root corresponds to our consistent estimator. One common solution is to use an EM algorithm, a computational method that iteratively repeats an expectation step and a maximization step, until converging to the consistent root.

In order to illustrate the mechanisms behind the EM algorithm, consider again the switching regression

$$P_t = MC_t + \beta_c Q_t S_t + \beta_n Q_t (1-S_t) + u_t, \quad S_t = \begin{cases} 1, & \text{prob. } \lambda \\ 0, & \text{prob. } 1-\lambda \end{cases} \tag{5}$$

and define W_t as the probability that the regime at time t is collusion conditional on the price observed:

$$W_t = f(S_t = 1|P_t). \tag{6}$$

At the maximization step of the algorithm, given an initial guess for the unobserved series of W_t, we must obtain initial estimates for the parameters β_c, β_n and σ in equation (5). This can easily be done by maximizing the following likelihood function:

$$L = \prod_{t=1}^{T} f(P_t|W_t), \quad P_t \sim N(MC_t + \beta_c Q_t W_t + \beta_n Q_t (1 - W_t), \sigma). \tag{7}$$

Alternatively, one can follow Kiefer (1980) and simply estimate the next regression by ordinary least squares:

$$P_t = MC_t + \beta_c Q_t W_t + \beta_n Q_t (1 - W_t) + u_t. \tag{8}$$

In addition, the initial guesses for W_t can be used further to compute the unconditional probability of collusion λ as the mean of all conditional probabilities:

$$\lambda = \sum_{t=1}^{T} W_t. \tag{9}$$

Next, at the expectation step of the algorithm, we must revise our expectations of the conditional probabilities of collusion W_t, taking into account the estimates of β_c, β_n, σ and λ obtained during the previous step. Using Bayes rule,

$$W_t = f(S_t = 1|P_t) = $$
$$= \frac{f(S_t = 1) f(P_t|S_t = 1)}{f(S_t = 1) f(P_t|S_t = 1) + f(S_t = 0) f(P_t|S_t = 0)}. \tag{10}$$

Replacing the unconditional probabilities with λ:

$$W_t = \frac{\lambda f(P_t|S_t = 1)}{\lambda f(P_t|S_t = 1) + (1 - \lambda) f(P_t|S_t = 0)}. \tag{11}$$

Given the new values of W_t computed from equation (11), the maximization and expectation steps can be then iteratively repeated until convergence is reached. The algorithm stops when the coefficient of correlation between two successive estimates of W_t series is near one.

Kiefer (1980) proves that the estimates obtained by the EM algorithm correspond, in fact, to a local maximum of the likelihood function in (4). In the next section we will actually implement this algorithm to simulated data, to check whether it converges, indeed, to the consistent estimator of the switching regression.

5 Simulation

We have presented a powerful algorithm to identify structural breaks, which can be used to detect switches between collusion and competition along a time series. But before this algorithm can be actually applied to real data and start detecting cartels, it must be tested in a controlled environment, so that its predictive power can be properly evaluated. The best way to do this is to create a simulated industry, for which we define the true values of the structural parameters and determine when firms compete or collude. Then, without using such information, we should be able to consistently estimate the structural parameters and to identify the switches between regimes.

We simulate seven different populations or industries, whose true underlying parameters are listed in Table 1. For each population, we extract a sample of 100 000 observations as follows: the variables marginal cost MC_t and quantity transacted Q_t are randomly withdrawn from two bounded normal distributions; the state variable S_t is randomly drawn from a Bernoulli distribution that takes value 1 (collusion) with success probability λ; and the dependent variable price P_t is generated according to equation (5) and using the beta coefficients in Table 1.

Note also that the first six populations in study have the same underlying parameters, except for an increasing standard deviation of the error term, as it is our desire to assess how the algorithm behaves in increasingly worse conditions. The fraction of collusive periods, λ, is set only to 0.2, because estimation is harder when one of the regimes dominates the time series. Then, in population 7 we define λ equal to 0.8 to see what happens when the other regime dominates.

Population	β_c	β_n	σ	λ
1	0.5	0.25	1	0.2
2	0.5	0.25	2	0.2
3	0.5	0.25	3	0.2
4	0.5	0.25	4	0.2
5	0.5	0.25	5	0.2
6	0.5	0.25	6	0.2
7	0.5	0.25	5	0.8

Table 1: Parameters of the population

Population	$\hat{\beta}_c$	$\hat{\beta}_n$	$\hat{\sigma}$	$\hat{\lambda}$	R^2
1	0.5007	0.2508	0.9992	0.2000	0.9984
2	0.5023	0.2516	1.9769	0.2001	0.8766
3	0.5107	0.2463	2.7441	0.1990	0.7937
4	0.5319	0.2393	3.3878	0.1921	0.7368
5	0.5821	0.2452	4.1345	0.1675	0.6731
6	Na	Na	Na	Na	Na
7	0.5142	0.1772	4.1430	0.8300	0.6861

Table 2: Estimation output of the OLS EM algorithm

Using solely the simulated data on prices, quantities and marginal costs (and not the serie S_t, which is unobservable), we run the EM algorithm described in the last section, whose Matlab code is available in Appendix 1. Table 2 displays the estimation results for the seven populations.

The estimates obtained for populations 1 and 2 are quite accurate, as they are close to the true values in Table 1. Nonetheless, in populations 3 to 5, where the volatility of the error is successively higher, there is a clearly growing estimation bias. In fact, the parameters σ and λ are systematically underestimated, whereas the absolute difference between the parameters β_c and β_n is overestimated. A similar pattern is found in population 7, with the exception that λ is overestimated, due to the fact that collusive periods dominate the sample. Finally, the algorithm fails to converge in population 6, where the standard deviation of the error is too high.

The results in Table 2 suggest that the EM algorithm, as currently described in the literature, is unable to obtain consistent estimates for the coefficients of the switching regression, at least when the error is sufficiently volatile. The dimension of the estimation bias is particularly serious if taken into account that we used a very large sample, with 100 000 observations. Much greater bias are expected to occur for samples with a more realistic size.

After a deep analysis of the code and running several tests to isolated parts of the algorithm, it came to our conclusion that the bias had to be originated by the introduction of the conditional probability W_t in regression (8), during the maximization step. In fact, by substituting the state variable S_t, which only takes the extreme values 0 or 1, with the conditional probability W_t, which ranges between 0 and 1, we allow the switching regression to fit better the data and thus to underestimate the standard deviation of the error. Intuitively, when an arbitrary price at time t is observed somewhere between the collusive and competitive expected prices, the algorithm estimates a conditional probability of collusion W_t around 50% and the fitted price becomes the average value between the collusive and competitive levels, leading to a very small residual. The estimation of residuals below the true errors for all intermediate prices leads to the underestimation of σ in Table 2.

Hereafter, there is a contamination effect on the remaining estimated values. Because the value of σ affects the density function of P_t, the conditional probabilities of collusion W_t are miscalculated in equation (11), being all high prices excessively attributed to collusion and all low prices associated to competition, when this is not necessarily the case. As a result, when competitive regimes dominate the sample, W_t's are in average under evaluated and, through equation (9), λ is underestimated. When collusive regimes are more frequent, W_t's are in average over evaluated and λ is overestimated.

In addition, the excessive association of high prices to collusion and low prices to competition leads to the overestimation of β_c and underestimation of β_n. Intuitively, when a price observed under collusion is very high, we are able to correctly identify the operative regime, but when a collusive price is particularly low, we misidentify the regime as competition. As a result, average collusive prices appear to be higher than what they really are and the parameters associated with collusion are overestimated. The opposite occurs at the competitive regime, whose average prices and parameters tend to be underestimated.

This phenomenon is graphically illustrated below. In Figure 3, where the standard deviation of the error is very small (like in populations 1 and 2), all observations are classified at the correct regime and the betas are consistently estimated as the slope of the two lines that intersect the midpoint between each pair of observations. However, in Figure 4, where the volatility of the error is higher, the two intermediate observations are misclassified at the wrong regime, causing betas to be wrongly estimated as the slope of the two dashed lines.

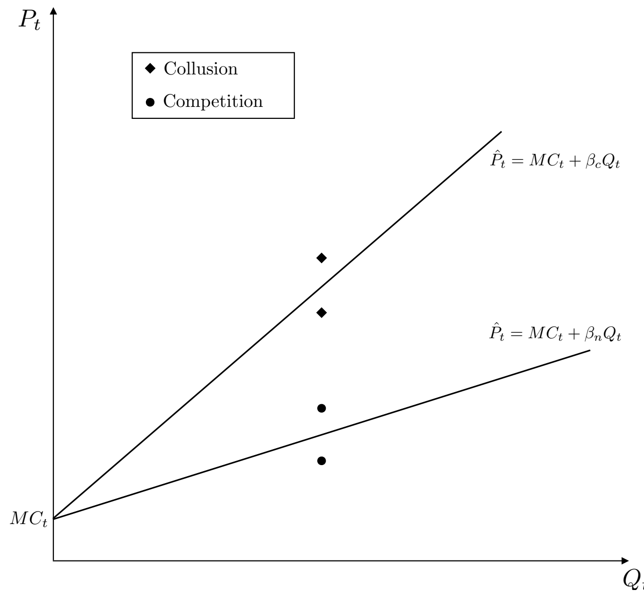

Figure 3: Consistent Estimation of Betas

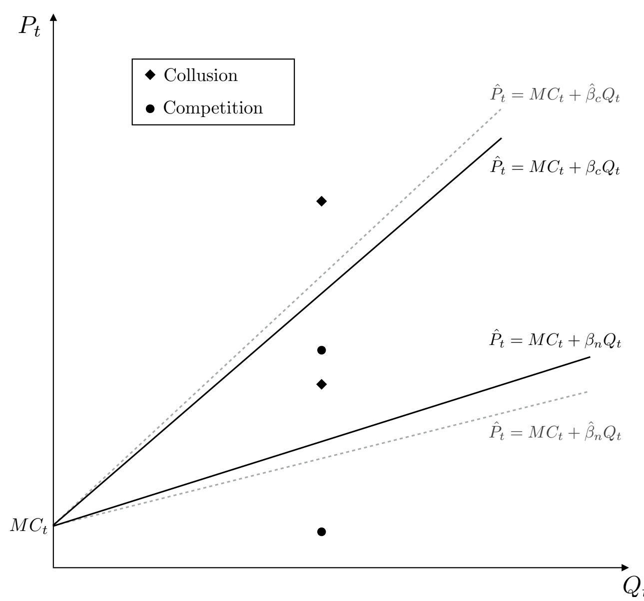

Figure 4: Biased Estimation of Betas

6 Correcting the Estimation Bias

The inability of the EM algorithm in Kiefer (1980) to converge, in some conditions, to the consistent root of the switching regression raises a serious concern to our empirical approach. Indeed, if it is our desire to detect collusion by identifying switching regimes in data, we must correct somehow the estimation bias detected in the last section. Clearly, any solution proposed should involve removing the conditional probability W_t from the maximization step and finding an alternative mechanism to estimate the beta coefficients and the standard deviation of the error.

A possible solution is to replace W_t in equation (8) with some estimated series of the regime, $\hat{S}_t$, which only takes values 0 or 1. This way, by considering that each observation is either explained by competition or collusion, and not by a linear combination of the two regimes, we should be able to measure the true size of the errors. The central question is, therefore, how to get an estimated series $\hat{S}_t$.

The most obvious hypothesis is to consider that collusion is observed at periods with high probability of collusion, according to the following rule:

$$\hat{S}_t = \begin{cases} 1, & \text{if } W_t \geq 0.5 \\ 0, & \text{if } W_t < 0.5 \end{cases} . \tag{12}$$

Nevertheless, equation (12) has the disadvantage of considering that all higher prices are observed under collusion and all lower prices correspond to competition, which seems a very unlikely distribution. Instead, we propose a more subtle and sophisticated rule, which consists in randomly attributing a regime to each time period according to the probability of observing that regime, that is:

$$\hat{S}_t = \begin{cases} 1, & \text{with probability } W_t \\ 0, & \text{with probability } 1 - W_t \end{cases} . \tag{13}$$

In other words, each observation of $\hat{S}_t$ is withdrawn from a Bernoulli distribution with probability of success W_t. To understand the reasoning behind equation (13), suppose that ten observations of the sample have 90% probability of belonging to the collusive regime. Then, nine of those observations will be classified as collusion and one as competition, to reflect the fact that sometimes competitive prices are higher due to unobserved factors, just as some women are taller than men.

This way, the rule in equation (13) not only guarantees that the estimated series $\hat{S}_t$ is a dummy variable composed of zeros and ones, but also that it follows a similar distribution to that of the unobserved S_t.

After computing $\hat{S}_t$, the beta parameters and standard deviation of the error can be estimated by running the next regression by least squares:

$$P_t = MC_t + \beta_c Q_t \hat{S}_t + \beta_n Q_t (1 - \hat{S}_t) + u_t. \tag{14}$$

We introduce equations (13) and (14) in the maximization step of the EM algorithm, whose new code is available in Appendix 2. In order to evaluate the quality of our proposed corrections, we use the modified algorithm to reestimate the coefficients of the seven populations with the same data samples than before. The new estimation results are displayed in Table 3.

Population	$\hat{\beta}_b$	$\hat{\beta}_n$	$\hat{\sigma}$	$\hat{\lambda}$	R^2
1	0.5007	0.2508	0.9992	0.2000	0.9652
2	0.5014	0.2516	2.0014	0.2001	0.8735
3	0.4991	0.2492	3.0035	0.1998	0.7528
4	0.4979	0.2477	3.9951	0.1994	0.6339
5	0.5011	0.2512	4.9856	0.1995	0.5247
6	0.4976	0.2465	5.9591	0.2007	0.4393
7	0.5081	0.2591	5.0196	0.7974	0.5393

Table 3: Estimation output of the modified OLS EM algorithm

Few comments need to be made to the new estimation output, as the results speak for themselves. In fact, all estimates obtained for the seven populations are very close to the true underlying parameters, even for population 6 for which the previous algorithm failed to converge to any solution. Besides, there appears to be no bias in any particular direction, pointing to consistency of results.

Given the considerable improvement of our estimation method, does this mean we are finally ready to implement the algorithm to actual data and start detecting real cartels? Not yet. But we are getting there.

7 Endogeneity Problem

So far, for exposition purposes, we have discussed the estimation of a switching supply function in a very simplified scenario, where the variable price is endogenously set by firms and the quantity transacted is exogenously determined by a random distribution. Nonetheless, the later assumption does not usually hold in reality, since the quantity purchased by consumers is also affected by the market price, creating an endogeneity problem. For that reason, we will now focus our analysis on the estimation of a more realistic model of the industry, where prices and quantities observed at time t are the simultaneous solution of the following system of demand and supply:

$$\begin{cases} Q_t = \beta_0 + \beta_1 Y_t + \beta_2 P_t + v_t \\ P_t = MC_t + \beta_c Q_t S_t + \beta_n Q_t (1 - S_t) + u_t \end{cases}, \tag{15}$$

where Y_t is the income, v_t and u_t are the error terms of the demand and supply functions (with zero mean and constant standard deviation), and all the remaining variables have the same meaning as in Section 4.

In order to assess the predictive power of our empirical methodology under this more realistic setting, we simulate six new populations whose parameters are reported in Table 4. As before, a random sample of 100 000 observations is extracted for each population: the exogenous series Y_t and MC_t are withdrawn from bounded normal distributions, S_t is withdrawn from a Bernoulli distribution with probability of success λ and the endogenous series P_t and Q_t are calculated as the solution of system (15).

Pop	β_0	β_1	β_2	β_c	β_n	σ_v	σ_u	λ
8	50	5	-2	0.5	0.25	1	1	0.2
9	50	5	-2	0.5	0.25	2	2	0.2
10	50	5	-2	0.5	0.25	3	3	0.2
11	50	5	-2	0.5	0.25	4	4	0.2
12	50	5	-2	0.5	0.25	5	5	0.2
13	50	5	-2	0.5	0.25	6	6	0.2

Table 4: Parameters of the population

After running the modified EM algorithm proposed in the last section, we obtain the estimation output in Table 5.

Population	$\hat{\beta}_c$	$\hat{\beta}_n$	$\hat{\sigma}_u$	$\hat{\lambda}$	R^2
8	0.2153	0.2057	4.2756	0.3568	0.3828
9	0.4740	0.2292	1.9898	0.1903	0.8705
10	0.4417	0.2017	3.0286	0.1639	0.7249
11	0.4035	0.1671	4.0849	0.1290	0.5438
12	0.3641	0.1340	5.0733	0.0931	0.3580
13	0.3391	0.1070	5.9015	0.0694	0.2308

Table 5: Estimation output of the OLS EM Algorithm

As we can see, the estimates obtained for the parameters of the supply equation are erratic and far way from the true underlying values. With the exception of population 8, where the algorithm appears to diverge to a meaningless solution, we can actually find a pattern for the bias observed: the two beta coefficients and the probability of collusion λ are systematically underestimated. Furthermore the overall estimation bias grows exponentially with the standard deviation of the errors.

Unlike before, the inconsistent results are a direct consequence of endogeneity. In fact, when an unobserved shock in supply (u_t) affects the market price, there is a feedback effect on the quantity transacted through the demand equation, leading to correlation between Q_t and u_t. For that reason, any attempt during the maximization step to estimate the supply equation by least squares is inconsistent and leads to the underestimation of betas, due to the negative slope of the demand equation.

The correlation between Q_t and u_t also distorts the expectation step, leading to the under evaluation of the probability of collusion at low price observations and the over evaluation of the probability of collusion at high price observations (note that when a price is low, demand raises the quantity purchased and the relation between the two variables is better explained by a competitive coefficient). Once in populations 9 to 13 most observations correspond to competition, the probability of collusion is more often under evaluated and λ is underestimated.

It is therefore imperative to introduce further modifications in the maximization and expectation steps of the algorithm.

8 The 2SLS EM Algorithm

It has become clear that when prices and quantities observed are the outcome of the supply and demand relations described in system (15), the supply equation of the industry cannot be estimated by traditional methods based on least squares. In these circumstances, the econometrician typically observes a map of dispersed points (see Figure 5), from which he is not able to identify a supply relationship between prices and quantities. In order to solve the identification problem, he must at least observe one exogenous variable for each equation, like the income and marginal cost, so that each price-quantity combination can be explained as a particular equilibrium resulting from shifts in demand and supply (see Figure 6).

If the series S_t was observed, there would be several methods available to estimate the coefficients of system (15). Using an instrumental variable approach, we could either estimate the supply equation by two-stages least squares (2SLS) or the whole system by three-stage least squares (3SLS). Alternatively one could use limited information maximum likelihood (LIML) or full information maximum likelihood (FIML) as the analogous maximum likelihood estimators. Since in our model S_t is unknown, one of these methods must be combined with the switching regression techniques exposed in the earlier sections, turning the analysis much more complex.

Such procedure was conducted by Porter (1983), who successfully introduced FIML in all iterative steps of the EM algorithm, in order to identify switches between collusive and competitive regimes at the Joint Executive Committee railroad cartel. However, following his approach comes with a high cost. Firstly, FIML estimation is computationally heavy, time consuming and extremely sensitive to initial points, making it particularly hard to converge to the solution when the sample is large. Secondly, the good properties of FIML, like consistency and asymptotic efficiency, depend on the assumption that the distribution of the errors is well specified. Thirdly, FIML requires the estimation of the complete system of equations, even if there is few information available about demand.

While the estimation of a switching regression by FIML is achievable for a professional researcher focusing entirely on the analysis of a single industry, it does not seem so attractive for competition authorities who have limited time and resources. For that reason, we propose instead a new EM algorithm that solves the identification problem

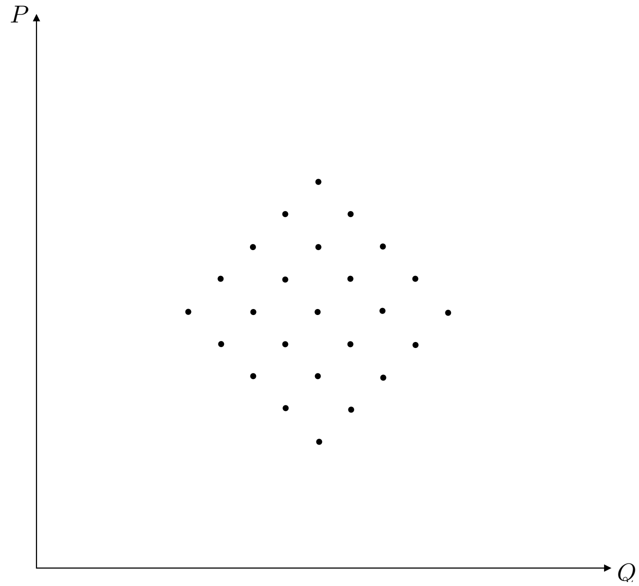

Figure 5: Map of dispersed observations

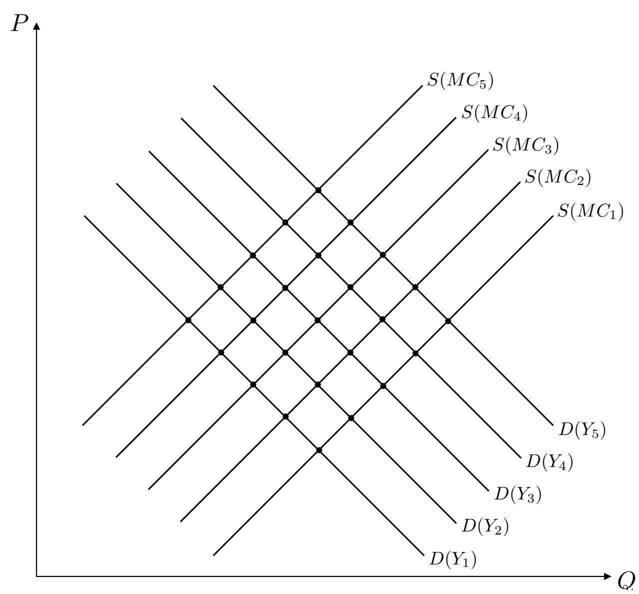

Figure 6: Identification of demand and supply

using two-stage least squares, a simple and parsimonious analytical method that is easy to compute, has good properties and does not rely on much information a priori. In what follows, we describe in detail how to introduce 2SLS in the maximization and expectation steps of the EM algorithm.

First of all, we start by solving system (15) with respect to the endogenous variables P_t and Q_t, in order to obtain the reduced form equations for prices and quantities:

$$\begin{aligned} Q_t =& \alpha_c S_t + \gamma_c S_t MC_t + \delta_c S_t Y_t + v_t^c + \\ & \alpha_n (1 - S_t) + \gamma_n (1 - S_t) MC_t + \delta_n (1 - S_t) Y_t + v_t^n. \end{aligned} \tag{16}$$

$$\begin{aligned} P_t =& \rho_c S_t + \theta_c S_t MC_t + \phi_c S_t Y_t + u_t^c + \\ & \rho_n (1 - S_t) + \theta_n (1 - S_t) MC_t + \phi_n (1 - S_t) Y_t + u_t^n. \end{aligned} \tag{17}$$

During the maximization step, the first stage of 2SLS consists in estimating equation (16) by least squares, using updated expectations of the regimes ($\hat{S}_t$). Then, the estimation output can be used to compute fitted values for the quantity transacted $\hat{Q}_t$:

$$\begin{aligned} \hat{Q}_t =& \hat{\alpha}_1 \hat{S}_t + \hat{\alpha}_2 \hat{S}_t MC_t + \hat{\alpha}_3 \hat{S}_t Y_t + \\ & \hat{\alpha}_4 (1 - \hat{S}_t) + \hat{\alpha}_5 MC_t (1 - \hat{S}_t) + \hat{\alpha}_6 Y_t (1 - \hat{S}_t). \end{aligned} \tag{18}$$

At the second stage of TSLS, Q_t is replaced with $\hat{Q}_t$ in the original supply function:

$$P_t = MC_t + \beta_c \hat{Q}_t S_t + \beta_n \hat{Q}_t (1 - S_t) + \epsilon_t. \tag{19}$$

Because the fitted quantity $\hat{Q}_t$ is a function of exogenous variables only and is not correlated with ϵ_t, equation (19) can now be run by OLS to obtain consistent estimates for the beta coefficients.

With regards to the expectation step of the algorithm, we must revise the conditional probabilities of collusion W_t, given the estimates obtained for the coefficients. Once the new model is composed of two endogenous variables driven by demand and supply, W_t must now be defined as the probability of collusion conditional on the observations of P_t and Q_t, in order to incorporate the fact that the collusion not only affects prices but also quantities. Using Bayes rule:

$$\begin{aligned} W_t = f(S_t = 1 | P_t \cap Q_t) = \\ = \frac{\lambda f(P_t \cap Q_t | S_t = 1)}{\lambda f(P_t \cap Q_t | S_t = 1) + (1 - \lambda) f(P_t \cap Q_t | S_t = 0)}. \end{aligned} \tag{20}$$

As before, λ is calculated by equation (9) as the average value of all W_t's obtained in the previous iteration of the algorithm. The probabilities $f(P_t \cap Q_t | S_t = 1)$ and $f(P_t \cap Q_t | S_t = 0)$ are computed from the normal multivariate distribution that results from equations (16) and (17):

$$\begin{aligned}
&[Q_t, P_t | S_t] \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \\
&\qquad\qquad \text{where} \quad \boldsymbol{\mu} = \begin{bmatrix} E(Q_t) = \alpha_i + \gamma_i MC_t + \delta_i Y_t \\ E(P_t) = \rho_i + \theta_i MC_t + \phi_i Y_t \end{bmatrix} \\
&\qquad\qquad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2(v_t^i) & \sigma(v_t^i, u_t^i) \\ \sigma(v_t^i, u_t^i) & \sigma^2(u_t^i) \end{bmatrix} \\
&\qquad\qquad i = \begin{cases} c, & \text{if } S_t = 1 \\ n, & \text{if} S_t = 0 \end{cases} .
\end{aligned} \tag{21}$$

Finally, after revising the conditional probabilities W_t, the predicted regimes $\hat{S}_t$ can be determined according to the rule in equation (13). The maximization and expectation steps are then iteratively repeated until convergence is reached.

All computations previously described were introduced in the TSLS EM algorithm, which can be used to estimate once more the parameters of populations 8 to 13 (code available in Appendix 3). After running the TSLS EM algorithm to the 100 000 observation samples previously collected, we obtained the estimation output in Table 6.

Population	$\hat{\beta}_c$	$\hat{\beta}_n$	$\hat{\sigma}_u$	$\hat{\lambda}$	R^2
8	0.4996	0.2497	1.0010	0.2000	0.9662
9	0.5001	0.2498	1.9930	0.2004	0.8701
10	0.5003	0.2505	3.0020	0.1997	0.7297
11	0.5011	0.2505	4.0195	0.2007	0.5583
12	0.4963	0.2488	5.0048	0.2004	0.3752
13	0.4994	0.2503	5.9700	0.2001	0.2128

Table 6: Estimation output of the TSLS EM Algorithm

The new algorithm was clearly able to correct endogeneity and converge to the consistent root for all populations, once the estimates obtained in Table 6 are very close to the true values in Table 4. Most impressive, even in population 13, which is poorly described by a regression with a 0.2128 R^2, the results obtained have great precision.

We are finally getting closer to an accurate method of collusion detection that can be actually applied to real data.

9 Testing for Structural Breaks

We have developed a fast and parsimonious algorithm to consistently estimate the supply side of an industry as a regression that switches between two regimes, competition and collusion. All the previous analysis was developed under the assumption that structural breaks do exist, that is, it was assumed that each regime is observed at least in some periods. A relevant question that must now be addressed is how the algorithm behaves when no structural break occurs (λ is equal to zero or one).

For that, consider the switching supply regression rewritten in matrix notation:

$$P_t = MC_t + \beta_c Q_t S_t + \beta_n Q_t (1 - S_t) + u_t \iff \mathbf{P} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U},$$

$$\text{where } X = \begin{bmatrix} MC_1 & Q_1 S_1 & Q_1(1-S_1) \\ ... & ... & ... \\ MC_T & Q_T S_T & Q_T(1-S_T) \end{bmatrix} \quad \text{and} \quad \boldsymbol{\beta} = \begin{bmatrix} 1 \\ \beta_c \\ \beta_n \end{bmatrix}. \tag{22}$$

When firms either compete or collude along the whole time series, the variable S_t or $1 - S_t$ is a row of zeros and the matrix of regressors X is singular. Therefore, if the EM algorithm is able to converge to the true values of the state variable, it will attempt at some point to invert a singular matrix during the maximization step, generating an error.

Interestingly, the algorithm often converges instead to a meaningless random solution, estimating a value for lambda between zero and one and identifying both regimes in the dataset. Indeed, despite the absence of structural breaks, the EM algorithm allows the regime to switch along time in order to improve the fitting of the data, attributing observations with highly positive errors to collusion and observations with highly negative errors to competition. Unfortunately this means we cannot actually rely on the estimation output of the switching regression, unless we know for sure that the data was generated by a mixture of the two regimes. To overcome this problem we must implement some statistical test to check whether there is evidence of structural breaks.

We may be initially tempted to use a Wald test or a likelihood ratio test to verify whether the fitted state variable $\hat{S}_t$ is statistically significant, case in which we conclude the time series is a mixture of two distinct regimes. However, the test statistics of the switching regression model do not have the traditional distributions under the null and cannot be used to conduct such analysis. For that reason, several authors have proposed a modified likelihood ratio test to check the hypothesis of a homogeneous model against a mixture of two or more regimes,[5] as Chen and Kalbfleisch (2004) and Zhu and Zhang (2003). Still, these methods are quite hard to implement, as they involve finding lower and upper bounds, as well as the asymptotic behavior of the distribution of the test statistic.

In this paper we propose a much more simple and convenient approach. Instead of studying the complex distribution of the switching regression estimator and respective test statistics, we focus our analysis on the original TSLS estimator. To illustrate this, consider again a dataset generated by the system in (15) and suppose we estimate the following simple regression by TSLS, using Y_t as an instrumental variable:

$$P_t = MC_t + \beta Q_t + u_t. \tag{23}$$

On the one hand, if the regime remains unchanged along time, the TSLS estimator is consistent and it follows directly that the residuals have a normal distribution, as observed in Figure 7. On the other hand, if the time series has structural breaks separating the two regimes, the TSLS estimator is no longer consistent and the residuals are given by:

$$e_t = \begin{cases} (\beta_n - \hat{\beta}_{TSLS})Q_t + u_t, & \text{if } S_t = 0 \\ (\beta_c - \hat{\beta}_{TSLS})Q_t + u_t, & \text{if } S_t = 1 \end{cases}. \tag{24}$$

where $\hat{\beta}_{TSLS}$ is the TSLS estimator of regression (23). In other words, in the presence of structural breaks, the residuals are generated by a mixture of two normal distributions with different mean, as observed in Figure 8.

The clear distinction between the two distributions in the homogeneous and mixture models can be used to obtain evidence for structural breaks. Indeed, we can test for normality of the residuals following Jarque and Bera (1987), whose null hypothesis is that the residuals

[5] For hypotheses tests of two against more regimes see Dannemann and Holzmann (2010).

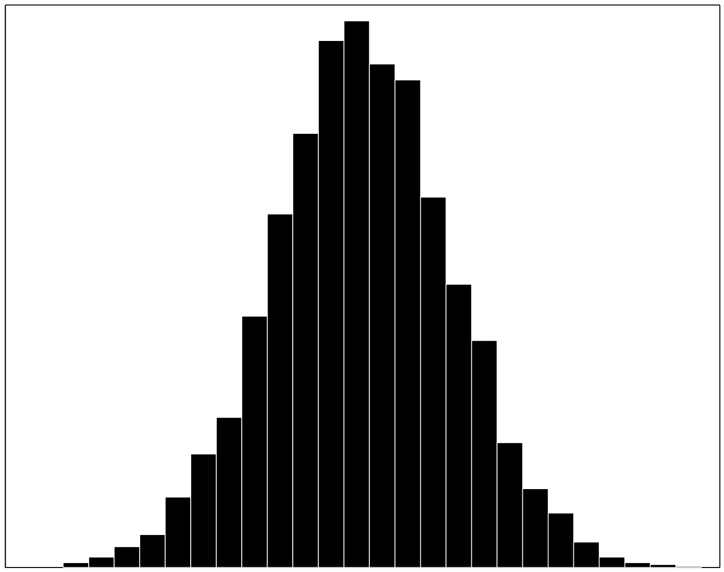

Figure 7: Histogram of the residuals of a homogeneous model

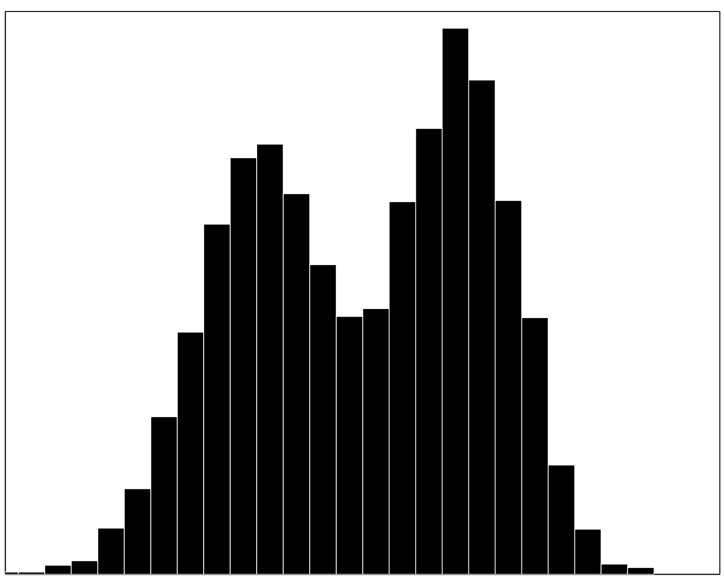

Figure 8: Histogram of the residuals of a mixture model

follow a normal distribution with unknown mean and variance. If the null is rejected, it raises suspicions that there are structural breaks and we can then run the EM algorithm to verify whether the industry is well explained by a switching regression model.

We conduct several simulation experiments to evaluate the Jarque-Bera test as a tool to detect the presence of structural breaks in the time series. As before, we simulate an industry with random levels of income and marginal costs and we generate the prices and quantities transacted as a solution of the system (15). The true values of the parameters are as follows:

β_0	β_1	β_2	β_c	β_n	σ_v
50	5	-2	0.5	0.25	1

Table 7: Parameters of the population

For each simulation experiment we collect a random sample of observations from the population, estimate equation (23) by TSLS to obtain the residuals and run the Jarque-Bera test for a 5% significance level. Then we replicate the previous steps several times and calculate the type I and type II errors, whose results are displayed in Table 8.

To calculate the type I error we set λ equal to 0, so that firms always compete and there are no structural breaks in the data. Next we compute the fraction of replications where the Jarque-Bera test rejects the null hypothesis of normality, suggesting structural breaks. To calculate the type II error we set λ equal to 0.2 to obtain a sample with a mixture of collusive and competitive periods. Then we compute the fraction of replications where the Jarque-Bera test fails to reject the null hypothesis, suggesting a homogeneous model.

From the observation of Table 8 we conclude that, for all experiments, the type I error is very close to the 5% significance level, meaning that the test is well elaborated and can be used to check for structural breaks. In addition, from the analysis of the type II error, we conclude that the power of the test is maximum (equal to one) when the number of observations is very large. Howsoever, when we collect a smaller sample of 100 or 200 observations, the test may lose some power if the error term becomes too volatile. For instance, in experiment 13, for a sample of 100 observation and a standard deviation of the error equal to 3, the Jarque-Bera test fails to reject the null hypothesis of normality in 12,4% of the cases where the null is false. In order words, the power of the test is only 0.876.

Experim.	Observations	Replications	σ_u	Type I Er.	Type II Er.
1	10 000	100	2	0.05	0
2	10 000	100	3	0.04	0
3	10 000	100	4	0.09	0
5	1000	500	2	0.06	0
6	1000	500	3	0.032	0
7	1000	500	4	0.044	0
9	200	500	2	0.05	0
10	200	500	3	0.05	0
11	200	500	4	0.048	0.236
12	100	500	2	0.054	0
13	100	500	3	0.05	0.124
14	100	500	4	0.058	0.666

Table 8: Jarque-Bera test

The ability of the Jarque-Bera test to detect structural breaks depends, of course, on the assumption that the error term of the supply equation follows a normal distribution (actually almost all statistical inference depends on that assumption). While this should not represent a major problem for large samples in which the central limit theorem can be applied, for small samples it may be useful to run alternative tests to check whether the results remain valid using different distributions, as the logistic.

In conclusion, the Jarque-Bera test appears to be an easy, fast and functional method to test for structural breaks. When it rejects the hypothesis of a homogeneous model, it can be complemented with the EM algorithm to verify if the industry can be accurately described by a switching regression of collusive and competitive regimes.

10 Conclusions

The pattern of economic data is the result of multiple complex interactions in the economy, some of which correspond to socially valuable transactions and others to criminal activities with high social costs. Modern econometric tools can be used to dissect the available data and detect some of those criminal behaviors that would otherwise remain unknown. In this paper, we discussed the estimation of switching regressions as a particularly useful method to detect collusion.

Although switching regression models have been broadly applied in various fields of social and natural sciences, the difficult estimation techniques behind them remain a black box for many researchers. We attempted here to get inside the black box in a controlled simulation environment in order to identify some common estimation problems, provide feasible solutions and present a better intuitive understanding of the results obtained.

Firstly, we have discussed some conditions under which the traditional EM algorithm fails to converge to the consistent root of the switching regression, while providing the computational procedures required to correct the estimation bias. Secondly, we have addressed the endogeneity problem in the estimation of supply, when the data observed is the result of a system of supply and demand equations. We overcame this problem by extending the rationality of the TSLS estimator to the maximization and expectation steps of the EM algorithm. Thirdly, we have shown that estimating a switching regression alone does not provide absolute evidence that the time series is composed by a mixture of collusive and competitive regimes. To test whether there are, indeed, structural breaks in data, we proposed the implementation of the Jarque-Bera normality test for the residuals.

We believe the empirical methods presented in this paper are easy to implement, computationally efficient and do not rely on much information a priori. And so, they can be actually implemented by competition authorities who face time and resource constrains. There is, of course, room for improving our empirical analysis of collusion. Future research should focus, for instance, on how to deal with industries with differentiated products, firms with multiple products and more complex supply functions that account for other forms of collusion. Perhaps the continuous development of advanced econometric tools to detect collusion will turn competition authorities, over time, into real *collusive scene investigators*.

References

Abrantes-Metz, R., Froeb, L., Geweke, J., and Taylor, C. T. (2006). A variance screen for collusion. *International Journal of Industrial Organization*, 24(3):467–486.

Almoguera, P. A., Douglas, C. C., and Herrera, A. M. (2011). Testing for the cartel in opec: non-cooperative collusion or just non-cooperative? *Oxford Review of Economic Policy*, 27(1):144–168.

Bain, J. S. (1951). Relation of profit rate to industry concentration: American manufacturing, 1936-1940. *The Quarterly Journal of Economics*, 65(3):293–324.

Bajari, P. and Ye, L. (2003). Deciding between competition and collusion. *Review of Economics and Statistics*, 85(4):971–989.

Baldwin, L. H., Marshall, R. C., and Richard, J. (1997). Bidder collusion at forest service timber sales. *Journal of Political Economy*, 105(4):657–699.

Berry, S. T. (1994). Estimating Discrete-Choice Models of Product Differentiation. *RAND Journal of Economics*, 25(2):242–262.

Bresnahan, T. F. (1989). Empirical studies of industries with market power. In Schmalensee, R. and Willig, R., editors, *Handbook of Industrial Organization*, volume 2 of *Handbook of Industrial Organization*, chapter 17, pages 1011–1057. Elsevier.

Chen, J. and Kalbfleisch, J. D. (2005). Modified likelihood ratio test in finite mixture models with a structural parameter. *Journal of Statistical Planning and Inference*, 129(1–2):93 – 107. {IISA} 2002 DeKalb Conference.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, 28(3):591–605.

Dannemann, J. and Holzmann, H. (2010). Testing for two components in a switching regression model. *Computational Statistics & Data Analysis*, 54(6):1592–1604.

Goldfeld, S. and Quandt, R. (1972). *Nonlinear methods in econometrics*. Contributions to economic analysis. North-Holland Pub. Co.

Green, E. J. and Porter, R. H. (1984). Noncooperative collusion under imperfect price information. *Econometrica*, 52(1):pp. 87–100.

Joseph E. Harrington, J. (2005). Detecting Cartels. Economics Working Paper Archive 526, The Johns Hopkins University,Department of Economics.

Kiefer, N. (1978). Discrete parameter variation: Efficient estimation of a switching regression model. *Econometrica*, 46(2):427–34.

Kiefer, N. M. (1980). A note on switching regressions and logistic discrimination. *Econometrica*, 48(4):1065–1069.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica*, 69(2):307–342.

Paha, J. (2010). Empirical methods in the analysis of collusion. MAGKS Papers on Economics 201033, Philipps-Universit´at Marburg, Faculty of Business Administration and Economics, Department of Economics.

Porter, R. H. (1983). A study of cartel stability: The joint executive committee, 1880-1886. *The Bell Journal of Economics*, 14(2):301–314.

Porter, R. H. and Zona, J. D. (1993). Detection of bid rigging in procurement auctions. *Journal of Political Economy*, 101(3):518–538.

Quandt, R. E. (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes. *Journal of the American Statistical Association*, 55(290):324–330.

Quandt, R. E. (1972). A new approach to estimating switching regressions. *Journal of the American Statistical Association*, 67(338):306–310.

Slade, M. E. (2004). Market power and joint dominance in u.k. brewing. *The Journal of Industrial Economics*, 52(1):133–163.

Zhu, H.-T. and Zhang, H. (2004). Hypothesis testing in mixture regression models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(1):3–16.