TI 2012-058/1
Tinbergen Institute Discussion Paper

Screening for Collusion: A Spatial Statistics Approach

Pim Heijnen[1]
Marco A. Haan[1]
Adriaan R. Soetevent[2]

[1] *Faculty of Economics and Business, University of Groningen;*
[2] *Faculty of Economics and Business, University of Amsterdam, and Tinbergen Institute.*

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at http://www.tinbergen.nl

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: http://www.dsf.nl/

Duisenberg school of finance
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 8579

Screening for collusion: A spatial statistics approach

Pim Heijnen* Marco A. Haan† Adriaan R. Soetevent‡

June 15, 2012

Abstract

We develop a method to screen for local cartels. We first test whether there is statistical evidence of clustering of outlets that score high on some characteristic that is consistent with collusive behavior. If so, we determine in a second step the most suspicious regions where further antitrust investigation would be warranted. We apply our method to build a variance screen for the Dutch gasoline market.

JEL-codes: C11, D40, L12, L41
Keywords: collusion, variance screen, spatial statistics, K-function

1 Introduction

Tracking down and prosecuting cartels are among the most important areas of antitrust enforcement. To track down a cartel, an antitrust authority has numerous instruments at its disposal. For example, it may actively screen markets for price patterns or other markers that suggest collusive behavior. In this paper we develop a method to screen for local cartels. First, we use

*Corresponding author: Faculty of Economics and Business, University of Groningen, P.O. Box 800, 9700AV, Groningen, e-mail: p.heijnen@rug.nl

†Faculty of Economics and Business, University of Groningen, e-mail: m.a.haan@rug.nl

‡University of Amsterdam and Tinbergen Institute, e-mail: a.r.soetevent@uva.nl.
Soetevent's research is supported by the Netherlands Organisation for Scientific Research under grant 451-07-010. The comments of Romy Abrantes-Metz and seminar participants at the IIOC 2010 and the EARIE 2009 meetings are gratefully acknowledged. The usual disclaimer applies.

spatial statistics to test whether outlets that show suspicious behavior are clustered. Second, if so, we provide an algorithm to find the most suspicious cluster of such outlets. We apply our method to data on gasoline prices in the Netherlands.

We thus consider the following problem. Suppose that an antitrust authority has data on all outlets in a particular industry. It suspects that there may be local cartels but it is not sure whether, and if so where, to start an investigation. Yet, it is able to identify a number of outlets that exhibit suspicious behavior. This suspicious behavior can take any form, e.g. particularly high prices or remarkably little variability in prices. The first question then is whether the spatial distribution of suspicious outlets exhibits local clustering, and hence that there is reason to believe that there may be local cartels. If so, the next question is where the most suspicious cluster is located that warrants further investigation. Finally, the question remains whether among remaining outlets there is reason to suspect further local cartels. In this paper, we propose a method that allows the antitrust authority to do exactly this. We thus provide a formal approach to integrate price data and spatial information into a collusion screen that is easily applicable for practitioners.

Abrantes-Metz and Bajari (2009) provide an overview of various screens that are used to detect anticompetitive behavior. Harrington (2008) also surveys methods to screen collusion. He argues (p. 250) that there are at least three requirements for systematic and ubiquitous screening. Evidence of collusion must be discernable by just looking at data that is readily available such as prices or market shares; the procedure should be routinizable so that it can be conducted with minimal human input; and the screen should be costly for the cartel to beat. Our method satisfies these criteria.

Needless to say, a collusion screen like the one we propose can never serve to establish the existence of local cartels. Further research to find evidence for collusion will always be necessary. Yet, a screen can be used to identify areas where the existence of a cartel is most likely.

For the application of our collusion screen, the identification of suspicious

outlets is particularly important. One identification method is the price variance screen, pioneered by Abrantes-Metz, Froeb, Geweke and Taylor (2006) (AFGT henceforth). It is based on the observation that prices of firms that are in a cartel often exhibit less volatility than prices of firms that are not.[1] Yet, this literature does not fully exploit data on the location of outlets.[2] Our contribution to this literature is a formal test for clustering, plus an algorithm to determine where these clusters are located. It is important to stress that the price variance screen is just one possible application of our method. Any other marker for suspicious behavior can also be used as an input, such as high prices, little advertising, or any other behavior or characteristic that an antitrust authority could think of.

As noted, the first step in our method borrows heavily from the literature on spatial statistics and spatial economics. In particular, our test for local clustering largely follows Diggle and Chetwynd (1991). They provide a statistic to test whether type 1 events (in our case: suspicious outlets) are more highly clustered than type 0 events (non-suspicious outlets). In economics, a related approach is that in Duranton and Overman (2005). They do a kernel density estimation of the bilateral distances between all pairs of establishments in an industry, and compare this to a counterfactual in which the establishments in that industry are randomly distributed across all industrial sites. Thus, they also test whether type 1 events (in their case: establishments in a particular industry) and more highly clustered than type 0 events (establishments in all other industries).

The second step in our method, finding the most prominent local cluster, is novel. A large literature in many fields is concerned with testing for local clustering but, to our knowledge, there is no work that addresses the problem of identifying the most prominent clusters. Our method boils down to finding

[1] This is in line with theory. For example, Athey et al. (2004) show that when firms face privately-observed i.i.d. cost shocks, the profit-maximizing cartel agreement often has them setting prices independent of marginal costs.

[2] AFGT (2006) use eyeballing to determine that gasoline stations with low price variability in their data set are not clustered. Jimenez and Perdiguero (2009) look at pre-defined markets.

the local cluster that is least likely to occur by chance.

The main contributions of this paper are thus twofold. To the literature of collusion screening, we add a formal test to identify suspicious clusters. To the literature of agglomeration we add a method to identify the most prominent clusters.

We apply our method to the gasoline market. Price data for this market are abundantly available: for many countries price quotes for most individual outlets can now be obtained on a weekly or even daily basis (e.g. Soetevent *et al.* 2011; Wang, 2009). Moreover, gasoline markets are often suspected to be prone to anti-competitive price manipulation, and in many countries they are subject to close antitrust scrutiny (FTC, 2005). Since 2002, the Federal Trade Commission (FTC) monitors gasoline prices on a daily basis using fleet-card data to detect "anomalous" pricing (Froeb et al., 2005).

We look at a data set of almost daily prices in the Netherlands. We test for local clustering in the period 2005 – 2007. In applying our collusion screen, many choices have to be made. For example, we have to decide on the number of outlets that we qualify as suspicious, and on the distance at which we look for local clustering. Also, we may focus on raw prices, but may also choose to correct prices for station characteristics. Any screen would be of little use if the suspicious clusters that are found would highly depend on these choices. We therefore perform a large number of robustness checks.

Naturally, our results differ somewhat depending on the choice we make, but an area close to Rotterdam persistently pops up as the most suspicious cluster in our data. Hence, if the Dutch antitrust authority would have used this tool in that period, the advise would have been to have a closer look at the gasoline stations in that particular area. If we repeat our analysis for the period 2007 – 2009, however, we find that an area close to Eindhoven is now the most suspicious, although Rotterdam is still among the identified clusters as well. Hence, there may now be a local cartel near Eindhoven and, if so, it is likely to have formed after 2005.

The remainder of this paper is structured as follows. In the next section,

we provide a more detailed overview of our method. In Section 3 we consider the first step of our method: testing for local clustering. We discuss our test statistic and compare it to other methods used in spatial statistics and economics. Section 4 discusses our method to identify the most suspicious region. In Section 5, we apply our method to Dutch gasoline data. We perform a sensitivity analysis in Section 6, and conclude in Section 7.

2 Overview of the method

Our method proceeds in five steps. Before actually doing the analysis, we of course have to collect and prepare the necessary data. This is **step 1**. This can be a nontrivial exercise, as price data are often plagued by missing observations. In our empirical application, we largely follow AFGT by using Markov chain Monte Carlo methods to impute missing data. **Step 2** is to determine which outlets are suspicious and which are not. For simplicity, we will refer to suspicious outlets as type 1, and to non-suspicious outlets as type 0 outlets. In our baseline application, we will consider outlets with a variation coefficient that is among the lowest 5% as suspicious, and the other outlets as non-suspicious.

In **step 3**, we establish whether there is statistical evidence for clustering of type 1 outlets. To this end, we use a slight variation of Diggle and Chetwynd's (1991) test statistic. Essentially, this involves testing whether there is random labelling, in the sense that the type 1 'labels' are randomly distributed over all existing outlets. To do so, we compare our population of type 1 outlets with a sample of controls consisting of the same number of outlets that is drawn from the entire population. More precisely, we compare the extent of clustering of type 1 outlets (that is, the average number of type 1 outlets within a fixed distance h of each type 1 outlet), to the extent of clustering in the sample of controls (that is, the average number of sampled outlets within a fixed distance r of each type outlet in the sample). Under random labelling, the difference between these two measures should be 0 on average, where the average is taken over all possible samples of controls. This

step is described more extensively in section 3.

If we find evidence for local clustering, we move to **step 4**. We partition the type 1 outlets into clusters of outlets that are relatively close to each other. For each such cluster, we determine the number of type 1 outlets, and the number of type 0 outlets in the same area. The most suspicious cluster is then the one for which the observed number of type 1 outlets relative to the total number of outlets, is most unlikely to occur under the null hypothesis of random labeling. This step is discussed in more detail in Section 4. **Step 5** then consists of eliminating all outlets within the identified region from the data. After having done so, we move back to step 3 to test whether there is evidence for local clustering in the remaining outlets.

3 Testing for local clustering

In this section, we introduce and motivate our test statistic to determine whether there is evidence for local clustering of type 1 events. Our problem can be stated as follows. We have a set N consisting of n outlets.[3] The location of outlet $i \in N$ is given by $x_i \in \mathbb{R}^2$. On the basis of some observable characteristic, we partition the set N into two subsets; the set N_1 of type 1 outlets (or, more generally, type 1 events) that are "suspicious", and the set N_0 of remaining type 0 outlets. We denote the fraction of outlets that is designated as type 1 as γ: $\gamma \equiv n_1/n$. The main question is whether there is local clustering, in the sense that type 1 outlets are on average more likely to be surrounded by other type 1 outlets.

In economic geography, a number of methods have been developed to test for local clustering or spatial agglomeration. Many of these, including Ellison and Glaeser (1997), and Rysman and Greenstein (2005), look at existing geographic entities (such as states or cities) and then test whether some statistic is significantly different between these entities. Such methods are not suitable for our purpose: when we look for areas where the variability of

[3]Throughout, we use the convention that upper-case letters refer to the set and lower-case letters denote the cardinality of the set.

prices is suspiciously low, these areas do not necessarily coincide with cities, municipalities, or even zip codes. We thus need a distance-based method.

In spatial statistics, the ubiquitous test for spatial dependence is Ripley's planar K-function (Cressie 1991, pp. 615–619). It tests whether events are more clustered than what one would expect purely on the basis of complete spatial randomness (CSR). Under CSR, events occur with the same probability at each location, and event locations are independent from each other.

The planar K-function at radius h counts the average number of other events within h of an event and relates this to the expected number of events under spatial randomness. The canonical K-function has the form:

$$K(h) = \frac{1}{\lambda} E[\text{\# further events within distance } h \text{ of a randomly chosen event}],$$

with λ the intensity of the spatial process, that is, the number of events per unit area. Ripley's planar K-function thus represents the average relative occurrence of events within distance h from any randomly chosen event. With more spatial clustering, events are located close to each other, hence $K(h)$ will be higher. Confidence intervals are determined by Monte Carlo simulation.[4] The null hypothesis of CSR can be tested for a pre-determined distance h, or by using a joint test for a range of values, typically from 0 to some natural upper bound. Applications of Ripley's K include spatial patterns of trees (see e.g. Stoyan and Penttinen, 2000), plant communities (Haase, 1995), and disease cases (Diggle and Chetwynd, 1991), amongst many others (see also Dixon, 2002). Applications in economics include Picone *et al.* (2009) who study spatial clustering of alcohol retailers.

For the problem at hand, this method has one major drawback. It tests whether locations are randomly distributed on a plane. Our problem is slightly different. We have a set of given locations, and are interested in knowing whether type 1 events are randomly distributed over these fixed locations.[5] Thus we are not so much interested in whether type 1 events

[4] Under some additional assumptions on the underlying spatial data generating process, these confidence intervals can also be derived analytically.

[5] As an example, consider an isolated area A in which 4 outlets are located, 2 of which

are clustered *per se*, but rather whether, given the locations of events, the type 1 labels are randomly distributed over all available locations. Diggle and Chetwynd (1991) study a very similar problem in the context of possible clustering of rare diseases. In their approach, a type 1 event is an occurrence of the disease. These occurrences are limited to the places where people live.

Their approach is as follows. Consider Ripley's K for type 1 events. Thus,

$$K_1(h) \equiv \lambda_1^{-1} E[\#\text{ further type 1 events within } h \text{ of random type 1 event}],$$

with λ_1 the intensity of type 1 events. Now take a sample of controls consisting of n_1 events drawn from the entire population. Thus, the number of events in the sample of controls equals the number of type 1 events. We can also calculate Ripley's K for our sample of controls:

$$K_c(h) \equiv \lambda_1^{-1} E[\#\text{ further controls within } h \text{ of random control}].$$

Consider the test statistic $D(h)$, defined as

$$D(h) \equiv K_0(h) - K_c(h).$$

Under random labelling, there is no systematic clustering of type 1 events relative to that in the underlying population. In that case we expect $D(h) = 0$. A value of $D(h) > 0$ indicates that type 1 events are more clustered than what can be expected on the basis of chance. To test whether $D(h)$ significantly differs from 0, Diggle and Chetwynd (1991, pg. 1157-1158) propose to either use the asymptotic distribution of $D(h)$ under the null of random sampling, or to approximate the true distribution by doing a Monte Carlo simulation consisting of a number of random permutations of the type 1 labels over the type 1 events and controls.

---

are type 1. All outlets are located within a distance h of each other. Compare this to area B in which 40 outlets are located, 3 of which are type 1. Arguably, A is more suspicious than B, as the fraction of type 1 outlets is much higher. Still, Ripley's K would flag B as more suspicious, simply because this statistic only looks at the absolute number of type 1 outlets. A test statistic that is appropriate for our purposes should thus correct for the density of stations and look at the relative number of type 1 outlets in an area, rather than at the absolute number.

We closely follow this approach. The only difference between our application and that in Diggle and Chetwynd (1991) is that we have information on the entire population of possible controls (i.e. all actual locations of outlets), while they only have one sample of controls. Therefore, rather than calculating $K_c(h)$ on the basis of one sample of events, we can be somewhat more precise by using $\bar{K}_c(h)$, the average value of $K_c(h)$ over 1000 samples of controls of size n_1. We calculate $\bar{K}_c(h)$ by doing a Monte Carlo simulation.

Summarizing, for a given radius of h we proceed as follows. In **step 3a**, we take a sample of n_1 controls, calculate the corresponding $K_c(h)$, and repeat this procedure 1,000 times to calculate $\bar{K}_c(h)$. In **step 3b**, we take a random sample of n_1 events, assign them a type 1 label and calculate the corresponding $K_1(h)$. On the basis of that, we calculate $D(h) = K_1(h) - \bar{K}_c(h)$. We repeat this procedure 1,000 times to calculate the distribution of $D(h)$ under the null of random labelling. In **Step 3c** we look at the actual incidence of type 1 labels, calculate the corresponding $K_1(h)$ and use the distribution derived in step 3b to test whether the resulting $D(h)$ significantly departs from the null of random labelling. If so, we conclude that there are clusters of low price variation at scale h.

This method is relatively easy to implement and interpret. As the density of type 1 events is given by λ_1, we have that $\lambda_1 D(h)$ represents the average number of extra type 1 events within distance h of a typical type 1 event over and above the number expected by random labeling.[6]

The null hypothesis of random labeling can either be tested for a predetermined distance h, or by using a joint test for a range of values, see e.g. the discussion in Diggle and Chetwynd (1991) for such tests. We have chosen to look at a fixed h. One natural interpretation is that an antitrust authority

[6] An alternative could have been to use the approach used by Marcon and Puech (2010). In the context of our application they essentially look for all type 1 events at the fraction of all events within a distance r of that event that is also of type 1, take the average of that number over all type 1 events and compare that average to a Monte Carlo simulation. We prefer to use Diggle and Chetwynd (1991), as that method has clear theoretical properties. One advantage of Marcon and Puech (2010) in other applications is that is easy to allow for different weights of events. When studying clustering of industries, for example, one can weigh different outlets with their level of employment.

first determines at which distance h firms still (should) effectively compete with each other. Alternatively, different distances of h could be used as a robustness check. That is also what we will do in our application.

Our approach is close in spirit to Duranton and Overman (2005). They do a kernel density estimation of the bilateral distances between all pairs of establishments in an industry, and compare this to a counterfactual in which the establishments in that industry are randomly distributed across all industrial sites. The main difference with our approach is that where Duranton and Overman (2005) look at the density of establishments *at* a distance h, we look at the density *within* a distance h. For our application, this makes more sense. After all, the natural way to define a market is to look at competitors within h kilometer, rather than the competitors at a distance h.[7].

4 Identifying the location of clusters

Suppose that, using the method described in the previous section, we have found evidence for local clustering at a distance of h kilometer. To judge which cluster is the most suspicious one, we determine the likelihood of the number of type 1 outlets in that cluster, taking into account the number of type 0 outlets in the same area. More formally, this step is split into three. In **step 4a**, we determine clusters of type 1 outlets. In **step 4b**, we determine the geographical areas where these clusters are located. In **step 4c**, we determine which of these areas is the most suspicious.

For **step 4a**, we first have to decide which type 1 outlets are part of a cluster. We will consider two type 1 outlets to be part of a cluster if they are within a distance of h kilometer from each other. If there exists another type 1 outlet that is also within h kilometers of any of the outlets in our tentative cluster, then that outlet is also considered to be part of the cluster. Repeating

[7]Note that, as the number of establishments within a given distance is a much smoother function than the number of establishments at a given distance, we can refrain from doing kernel density estimates.

this procedure leads to partitioning of all type 1 outlets into clusters. By construction, any type 1 outlet is less than h kilometer removed from at least one other outlet in that cluster, and more than h kilometer removed from an outlet in any other cluster.

Consider the set of type 1 outlets N_1. We consider two type 1 outlets as being *adjacent* if they are located less than h kilometers from each other. Linking adjacent outlets yields an undirected graph of type 1 outlets. We define a cluster as any connected component of that graph, that is, any subset of N_1 in which any two outlets are connected to each other by paths, and which is connected to no additional outlet. This yields a set of clusters that is a partition of N_1.

Suppose that this procedure yields s clusters of more than 1 outlet; $S_1, S_2, \ldots S_s$. The cardinality of cluster S_i is denoted s_i. Without loss of generality, we order clusters from largest to smallest, so $s_i \geq s_{i+1}, \forall i < s_s$. Although S_1 is the cluster with the largest number of type 1 outlets, it is not necessarily the most suspicious cluster. For example, it may well be the case that, say, S_1 has 10 type 1 outlets but is located in an area where also 20 type 0 outlets are active, while S_2 has 8 outlets, but is located in an area where only 1 type 0 outlet is active. Then S_2 is arguably more suspicious than S_1.

To formalize this, we define the area where cluster S_i is located as the convex hull of the locations of all outlets in S_i: $A_i = \text{Conv}(S_i)$.[8] This is **step 4b** of our procedure. The number of type 1 outlets in A_i obviously is s_i, while we denote the number of type 0 outlets in A_i as s_i^0. Note that a fraction γ of all outlets is of type 1. Under the null hypothesis of random labelling, we can calculate the probability that, given that there are a total of $s_i + s_i^0$ outlets in A_i, at least s_i are of type 1. We will denote this as the p-value of cluster

[8]Of course, it would also be possible to take into account type 0 outlets in the close proximity but outside the convex hull, as arguably these stations also compete with our type 1 stations. We have chosen not to do so, as that would imply that type 0 stations can be part of more than 1 cluster. Anyhow, we do not believe that this would greatly affect our analysis: an alternative clustering method that we will consider in the next section is close in spirit to this approach but yields a similar outcome.

S_i:

$$p(S_i) = \sum_{j=s_i}^{s_i+s_i^0} \binom{s_i + s_i^0}{j} (\gamma)^j (1-\gamma)^{s_i+s_i^0-j}. \quad (1)$$

It is important to note that these p-values should not be compared to traditional significance levels: as we deliberately look for clusters of type 1 outlets, the p-values that we find are necessarily low. There are infinitely many possible clusters that can be defined. Hence, necessarily, there always are some clusters that are very unlikely to occur when looked at in isolation.

For ease of exposition, we will report the negative of the log of p: this is a positive number that is usually between 0 and 10. In the example above, it turns out that $-\log p(S_1) = 5.9$, while $-\log p(S_2) = 9.5$. Hence, S_2 is indeed identified as the more suspicious cluster. In **step 4c** of our procedure we identify the most suspicious cluster, which is the cluster with the lowest p-value or, alternatively, the largest value of $-\log p(S)$:

$$S^M = \arg \max_{S \in \{S_1, \ldots S_S\}} (-\log p(S)).$$

After having identified this cluster, we remove it from our data. We then move back to step 3 as described in the previous section to test whether among the remaining outlets, there is still evidence for clustering of type 1 outlets. If that is the case, we again perform the procedure described above to find the now most suspicious cluster.

5 Empirical application

5.1 Introduction

In this section, we apply our method to data on the Dutch gasoline market. Following AFGT, our measure of price variability of station i is the variation coefficient v_i. This variation coefficient is defined as the standard deviation σ_i of i's retail price, divided by its mean price μ_i. Members of a cartel often exhibit low price variability and charge high prices. Both adversely affect

the variation coefficient, thus making it a useful instrument to screen for collusion. We denote as type 1 stations those that have a particularly low v_i.

In the remainder of this section we go through the five steps of our procedure. We first describe our data, and then discuss how we impute missing data. Second, we identify the type 1 stations. Third, we determine whether there is statistical evidence for local clustering. That turns out to be the case. In step 4 we determine the most suspicious cluster. After removing that cluster we find no further evidence for clustering in the remaining data.

Before being able to apply our method, we have to make a number of choices. First, we have to decide on the time period to consider. On the one hand, we want a period that is long enough for the presence of possible local cartels to be fully captured by the price variability of those stations relative to others. On the other hand, we do not want a period that is too long: local cartels may be temporary, so if we look at a period that is too long we may not be able to catch them. In our application, we look at a period of almost 2 years, between May 2005 and March 2007.

Second, we have to decide on the distance h at which we test local clustering. In our baseline, we will use $h = 5$ km. Third, we have to decide on the fraction γ of stations that we flag as suspicious. We will use $\gamma = 0.05$. Fourth, we have to decide whether we look at the raw price data, or whether we correct these for e.g. local circumstances. Initially we go for raw prices. In the next section, we will do a sensitivity analysis to check how sensitive the results are for all of these choices.

5.2 Step 1: Data

5.2.1 Step 1a: data collection

We use a fleet card data set which contains regular price quotes for 3,259 gasoline retail outlets in the Netherlands[9]. Price data were downloaded on a

[9]For comparison, the Dutch competition authority NMa (2006a, p. 8) cites a total number of 3,625 outlets in the Netherlands in 2004. An estimate of Bovag (the Dutch industry association for the automotive sector) mentions 4,319 outlets in 2005.

daily basis from the website of Athlon, the largest independent car leasing company in the Netherlands with a fleet of over 125,000 cars. For now, we limit attention to the period October 1, 2005 - June 30, 2007.

The price at a particular gasoline station on a given day is observed only if at least one fleet card owner bought gasoline there. On each day we observe a price quote for on average 37.5% of all stations. We restrict attention to prices of regular unleaded gasoline, the most common type and hence the one for which the most data is available. Using point of interest-data and Google Earth, we append our station data with geographic coordinates. It is important to note that we have the *exact* location of each station, rather than merely an approximation of that location based on e.g. the zip code, a method that is often used in other applications.

5.2.2 Step 1b: data imputation

There are 3,259 outlets that we follow for 637 days. We should thus have $3,259 \times 637 = 2,075,983$ price quotes. Yet, we only observe $669,000$ quotes. If we ignore the missing data and compute the variation coefficient on the basis of observed prices, a number of problems arise. First, this may bias our estimates of the station-specific variation coefficient. Second, additional uncertainty as a consequence of missing data is ignored. To confront these problems, we follow the method proposed AFGT to impute the missing data. We only give a sketch of this approach here.[10]

The essence of the approach is to draw multiple imputations from a Bayesian predictive distribution. A Markov chain Monte Carlo method is then used to draw from this distribution, using Gibbs sampling that incorporates the Metropolis-Hastings algorithm. To find the predictive distribution. first decompose the price p_{it} of firm i at date t into the average price on that day $\bar{p}_{\cdot t}$ and the deviation of firm i from this average z_{it}:

$$p_{it} = \bar{p}_{\cdot t} + z_{it}.$$

[10]Full details can be found in AFGT, pg. 475-478.

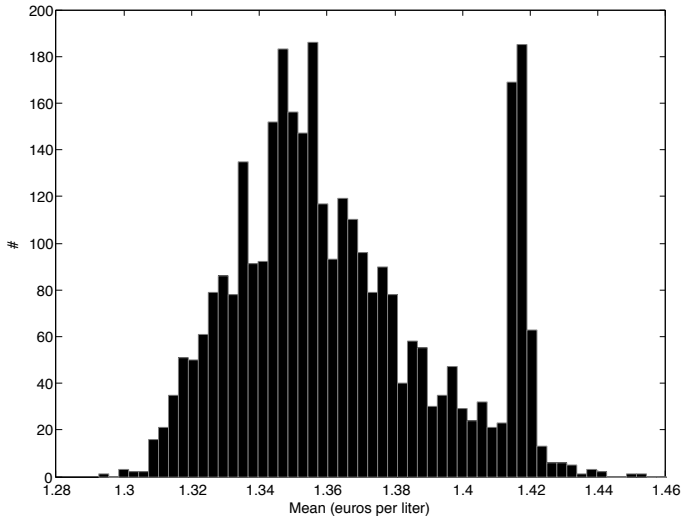
Assume that these deviations z_{it} follow a stationary AR(1)-process with mean μ_i and correlation ρ_i:

$$z_{it} - \mu_i = \rho_i(z_{it-1} - \mu_i) + \epsilon_{it}, \tag{2}$$

where $\epsilon_{it} \sim \mathcal{N}(0, \sigma_i^2)$ and the error terms are independent.

Take a flat prior over the parameters $(\mu_i, \rho_i, \sigma_i^2)$ and the missing prices. Given observed prices, the likelihood function of the data-generating process as given in (2) is a posterior distribution of the parameters and the missing prices. Missing prices can now be replaced by a draw from the posterior distribution. We then proceed with the analysis using the imputed data.

Figure 1: Histogram of average price, station level



Sample period May/05-March/07 (in €/liter)

Figure 1 shows the average price per site for the time period considered. The distribution is clearly bimodal, with the second peak caused by stations located close to or along the highway. These stations systematically charge higher prices. In two competition cases, the European Commission has also judged that highway stations constitute a separate product market.[11] In our

[11]See e.g. European Union, 1999, where it is argued in the Exxon/Mobil case that "in

analysis we therefore exclude the 224 highway stations and limit attention to the 3,035 remaining non-highway stations.

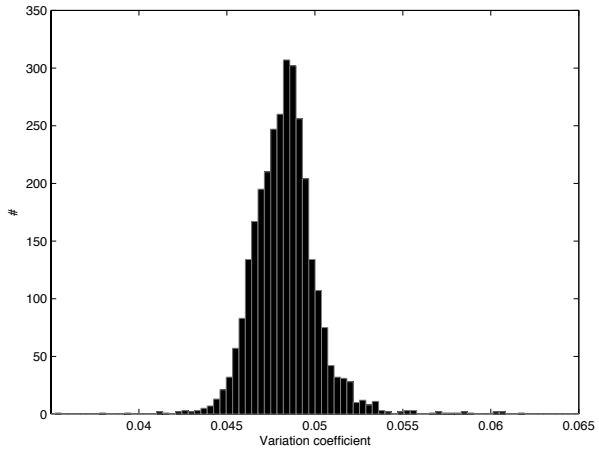
5.3 Step 2: identifying type 1 stations

For each station we calculate the variation coefficient v_i on the basis of the observed and imputed data. For each value of γ, we classify as type 1 stations those that have a v_i that is among the lowest γn. Formally,

$$N_1(\gamma) \equiv \{i \in N | v_i \leq \bar{v}(\gamma)\},$$

where $\bar{v}(\gamma)$ is defined such that $n_1 = \gamma n$, whereas $N_0(\gamma) \equiv N \setminus N_1(\gamma)$. As noted, we will use $\gamma = 0.05$.

Figure 2: Histogram of the variation coefficient

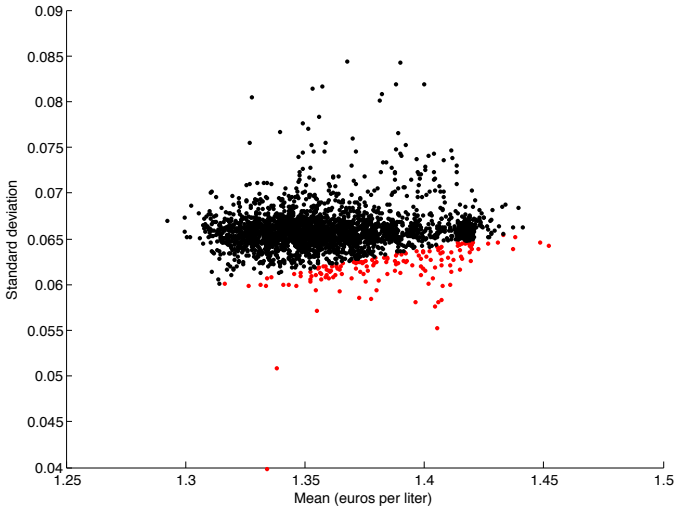


Sample period May/05-March/07

A histogram of the variation coefficients for all stations is given in Figure 2. The distribution is unimodal and roughly symmetric. Figure 3 gives a scatter plot of the standard deviation against the mean for all stations in the data. Type 1 stations are depicted as red dots. As in AFGT, stations with

some countries, it is possible to consider fuel retailing on motorways as a separate product market" (point 436).

Figure 3: Relation between mean price and standard deviation



Non-highway stations, October/2005-June/2007). Mean price (in € per liter) on the horizontal axis, standard deviation on the vertical axis. Stations with a variation coefficient in the lowest 5% in red.

higher means tend to have (slightly) higher variance, and there are no clear outliers in terms of stations with a high mean and low standard deviation.

5.4 Step 3: testing for local clustering

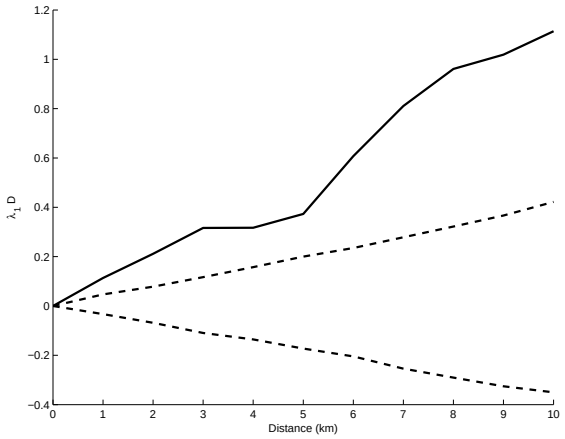
In Figure 4, we plot our D-function for different values of h. As noted, we focus on clusters at a distance of 5 kilometers. At $h = 5$, the D-function shows clear evidence for clustering of type 1 stations. The average type 1 station has 0.4 more type 1 neighbors than expected. This is a substantial excess, as the average circle with a radius of 5 km only has 0.3 type 1 stations[12].

5.5 Step 4: identifying the location of clusters

In this step we determine the most suspicious cluster. Table 1 gives all clusters with more than 2 type 1 stations, listing the coordinates of the midpoint of

[12]We have 153 type 1 stations, the Netherlands is roughly 40,000 km^2. That yields one type 1 station per 261 km^2; a circle with radius 5 km has an area of 78 km^2.

Figure 4: D-function



Non-highway stations, October/2005-June/2007. The solid line is the D-function, 95% confidence interval is indicated by the dashed lines, events are gasoline stations whose variation coefficient is among the lowest 5%

the cluster, the number of type 1 stations it contains, the number of type 0 stations enclosed by the cluster, and the resulting p-value. For ease of reference (and for the benefit of readers with an intimate knowledge of Dutch geography), we have also included for each cluster the city closest to it.

A cluster of suspicious stations may be due to a local cartel, but it may also simply reflect the presence of a local monopoly. For that reason we have calculated the Hirschman Herfindahl Index for the entire cluster (HHIT) and for the subset of suspicious stations within a cluster (HHIS). That information is also included in Table 1. For reasons of data availability, we calculated HHIs on the basis of brand share (i.e. the relative number of station within an area that carries a certain brand) rather than market share. For reference, note that the HHI on a nationwide level is 0.161.

The most suspicious cluster turns out to be an area slightly to the north of the city of Rotterdam, that includes 11 type 1 and 13 type 0 stations, yielding a $-\log(p)$-value of 5.1. The values of the HHI that we find for this cluster do not indicate that this is due to high market concentration.

Table 1: Clusters in the data

cluster	midpoint	s_i	s_i^0	$-\log p$	HHIT	HHIS	nearest city
S_1	(95,443)	11	13	8.2	0.163	0.174	Rotterdam
S_2	(141,473)	5	10	3.2	0.138	0.280	Hilversum
S_3	(110,452)	4	2	4.1	0.278	0.375	Bodegraven
S_4	(137,449)	4	0	5.2	0.375	0.375	Nieuwegein
S_5	(159,402)	4	3	3.7	0.184	0.375	Veghel
S_6	(77,393)	3	0	3.9	0.556	0.556	Bergen op Zoom
S_7	(131,480)	3	0	3.9	0.556	0.556	Weesp
S_8	(134,520)	3	0	3.9	0.556	0.556	Hoorn
S_9	(229,527)	3	0	3.9	0.333	0.333	Hoogeveen

Only clusters consisting of more than two stations. Sample: October 2005 - June 2007. 5% of stations classified as type 1, cluster size 5 km

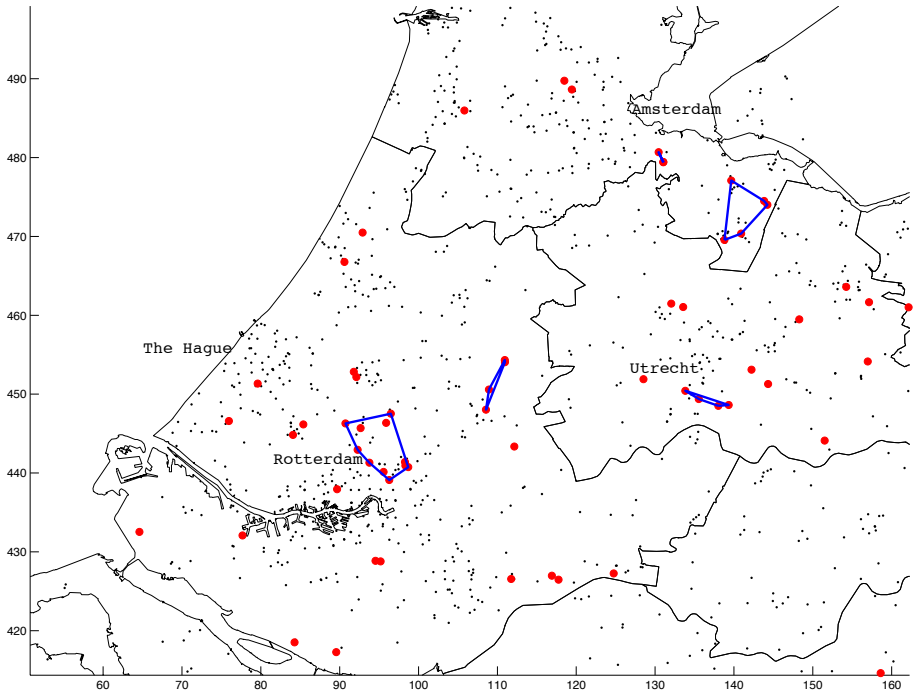
The most suspicious clusters (including our prime suspect) are located in the south-west corner of the country. Figure 5 zooms in on this area, depicting both type 1 (red) and type 0 (black) stations. The most suspicious area is the one slightly north of Rotterdam, S_1 in the table. Some other suspicious clusters are depicted as well: S_2 (Hilversum) is the area to the slight south-east of Amsterdam, S_3 roughly halfway between Rotterdam and Utrecht, and S_4 is close to Utrecht. The other clusters are outside this map.

5.6 Step 5: iterative elimination of clusters

After eliminating this most suspicious cluster, we move back to step 3 to test whether there is evidence for local clustering in the remaining data. Note that the number of type 1 stations has now decreased by 11, while the number of type 0 stations has decreased by 13. This implies that the value of γ, which is the fraction of all stations that is of type 1, has also changed. Naturally this change has to be taken into account when deriving the new D-function.

Figure 6 shows the resulting D-function after the elimination of the most suspicious cluster. The function is now no longer significant at $h = 5$ kilometer. It is at almost all other values of h, but that is largely by construction: our precise aim was to reduce clustering at 5 km, and we achieved that by

Figure 5: Type 1 stations: Rotterdam-The Hague-Amsterdam-Utrecht region.



Map of the western part of the Netherlands. Grey lines indicate province boundaries. Red circles are type 1 stations, black dots type 0 stations. Blue lines reflect convex hulls of clusters of suspicious stations. Sample October 2005 - June 2007. 5% of stations classified as type 1, cluster size 5 km

removing the most suspicious clusters at that distance.

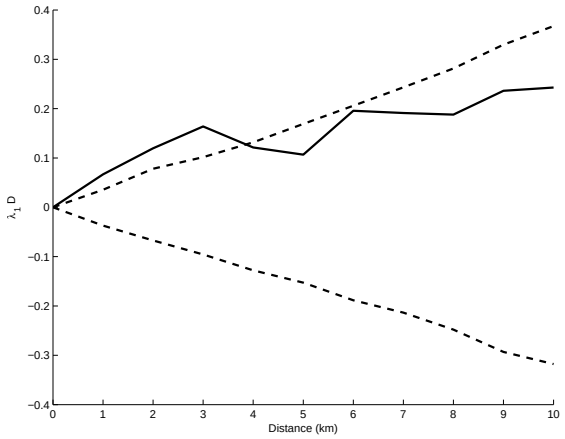
Table 2: Identified suspicious clusters

#	midpoint	type 1	type 0	$-\log p$	HHIT	HHIS	nearest city
1	(95,443)	11	13	8.2	0.163	0.174	Rotterdam

Sample October 2005 - June 2007. 5% of stations classified as type 1, cluster size 5 km.

For future reference, table 2 gives the output of our collusion screen in terms of suspicious clusters that are identified. Based on this, the advice to an antitrust authority would be to have a close look at the area around Rotterdam, where a local cartel may be in place. Of course, this in no way

Figure 6: D-function after removal of first suspicious cluster



The solid line is the D-function, 95% confidence interval is indicated by the dashed lines. Sample October 2005 - June 2007. 5% of stations classified as type 1, cluster size 5 km

provides evidence for collusion. Still, there is an unusually large concentration of stations that exhibit behavior consistent with collusive practices.

6 Sensitivity analysis

In applying our collusion screen, we had to make many choices. For example, we fixed the number of type 1 stations at 5%, which is a rather arbitrary choice. Also, we focused on local clustering at 5 kilometer, and choose one particular method for identifying the most suspicious cluster. We used data from 2005-2007, rather than focusing on a smaller, larger, or different time period. Finally, we chose to focus on listed prices, rather than to correct these prices for station characteristics. In this section, we test the sensitivity of the method in our empirical application with respect to these choices. Any screen would be of little use if the suspicious clusters that are found would highly depend on these choices. Moreover, in any practical application of our screen, it is wise not to fully rely on one particular set of choices, but to consider some other choices as well.

6.1 An alternative cluster size

In our baseline, we looked for evidence for local clustering at a distance of 5 kilometers. In this section we vary this distance by looking at distances of 3 and 7 kilometers, respectively. Note that this will affect both step 3 and step 4 of our method. In step 3, we will now look at whether there is statistical evidence for clustering at 3 (7) kilometers, while in step 4 we will look at clusters of stations that are at least 3 (7) kilometers from each other.

Table 3: Identified suspicious clusters at $h = 3$

#	midpoint	type 1	type 0	$-\log p$	HHIT	HHIS	nearest city
1	(95,442)	9	11	6.7	0.155	0.210	Rotterdam
2	(137,449)	4	0	5.2	0.375	0.375	Nieuwegein
3	(131,480)	3	0	3.9	0.556	0.556	Weesp
4	(134,520)	3	0	3.9	0.556	0.556	Hoorn

Sample October 2005 - June 2007. 5% of stations classified as type 1, cluster size 3 km.

Figure 4 shows that there is statistical evidence for local clustering at both 3 and 7 kilometers. Table 3 gives the list of suspicious clusters that are generated at a distance of 3 kilometers. A number of observations stand out. First, the most suspicious cluster is the same as that in our baseline case: close to Rotterdam. Second, the second most suspicious cluster (Nieuwegein) was also the second most suspicious in our baseline, as can be seen from Table 1. Yet, in our baseline this cluster was not flagged, as the elimination of the first cluster already yielded lack of statistical evidence for further clustering. That is no longer the case here. Third, our procedure now also generates 2 suspicious clusters of only 3 type 1 stations. Herfindahl indices indicate that these clusters each contain 2 stations that carry the same brand.

By construction, using a distance of 3 km generates smaller clusters. This suggests that in the implementation of our screen, it is important not to look at distances that are too small.

Table 4 lists the clusters that are found when looking at a distance of 7 km. Again, the area close to Rotterdam yields the most suspicious cluster,

Table 4: Identified suspicious clusters at $h = 7$

#	midpoint	type 1	type 0	$-\log p$	HHIT	HHIS	nearest city
1	(93,444)	16	31	9.3	0.112	0.180	Rotterdam
2	(163,406)	12	17	8.3	0.119	0.190	Veghel

Sample October 2005 - June 2007. 5% of stations classified as type 1, cluster size 7 km.

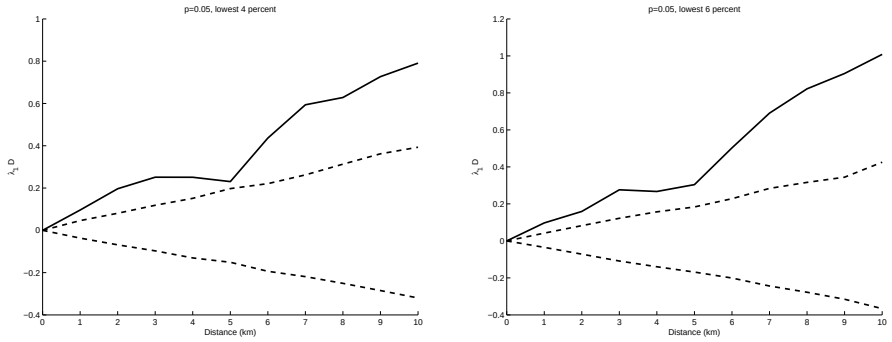
although it is now larger than in the baseline. The second most suspicious cluster is now an area close to Veghel. Comparing Table 4 to Table 1, clusters are obviously (much) larger now, but also have a higher share of type 0 stations. Choosing the right value of h thus implies a tradeoff between finding many clusters that are too small in the sense that they include only a few type 1 stations, and finding a few clusters that are too large in the sense that they include many type 0 stations.

6.2 An alternative fraction of type 1 stations

In our baseline, we classified stations with a variation coefficient among the 5% lowest as type 1. In this section, we consider different definitions. We will look at the lowest 4% and the lowest 6% respectively. This may seem a slight change in the number of type 1 stations, but it does imply a decrease, resp. in an increase of 20 % in the number of type 1 stations that we consider.

In both cases, we again find statistical evidence for local clustering at 5 km. Table 5 shows that the most suspicious cluster is now Nieuwegein, with the same midpoint and number of stations as in the baseline. Apparently, classifying only 4% of stations as suspicious implies that the area near Rotterdam is broken up in a number of smaller clusters. In Table 6 the most suspicious cluster is again near Rotterdam, and has the same midpoint and number of stations as in the baseline. In both cases, removing 1 cluster is sufficient for concluding that there is no statistical evidence for local clustering in the remaining data.

Figure 7: D-function



The solid line is the D-function, 95% confidence interval is indicated by the dashed lines. Sample October 2005 - June 2007. Cluster size 5 km. 4% (left panel) and 6% (right panel) of stations classified as type 1.

Table 5: Identified suspicious clusters, 4% classified as type 1

#	midpoint	type 1	type 0	$-\log p$	HHIT	HHIS	nearest city
1	(137,449)	4	0	5.2	0.375	0.375	Nieuwegein

Sample October 2005 - June 2007. Cluster size 5 km.

Table 6: Identified suspicious clusters, 6% classified as type 1

#	midpoint	type 1	type 0	$-\log p$	HHI	HHI[1]	nearest city
1	(95,443)	11	13	8.2	0.163	0.174	Rotterdam

Sample October 2005 - June 2007. Cluster size 5 km.

6.3 An alternative method to identify clusters

In Section 4, we proposed a method to identify local clusters, and a way to find the most suspicious of those clusters. Yet, it is conceivable that there may be circumstances in which that method does not perform optimally. For example, if an area has a high density of stations, it is also more likely that that area includes more type 1 stations that are thus misidentified as a suspicious cluster.[13] Also, the identity of a cluster may be sensitive to whether one station in the middle is qualified as type 1. We therefore also consider

[13]At the same time, in such a case, it is unlikely that that cluster has the lowest p-value

an alternative method to identify the most suspicious cluster.

Consider a particular type 1 station, which we denote i. If there is local clustering, then this station is particularly suspicious if there are more type 1 stations in its neighborhood. Consider a circle of k kilometer around i, and denote that area as $A(i;k)$. If we count the number of type 1 and type 0 stations in $A(i;k)$, we can calculate the p-value of that area, just as we did in (1). We denote this p-value as $p(A(i;k))$. We repeat this analysis for all other type 1 stations, and also allow for different cluster sizes by varying h between 3 and 7. With some abuse of notation, our must suspicious cluster is then given by

$$A^M = \arg \min_{i \in N_1, k \in \{3,4,5,6,7\}} p\left(A\left(i;k\right)\right)$$

Admittedly, this method is much cruder than the one proposed in section 4, but at the same time it may be more robust to e.g. the classification of one single station.[14]

Table 7: Identified suspicious clusters with circle-based method

#	midpoint	type 1	type 0	$\log p$	HHI	HHI^1	nearest city	radius
1	(92,443)	12	43	4.9	0.146	0.194	Rotterdam	7 km
2	(165,402)	6	11	3.9	0.185	0.222	Veghel	6 km

Sample October 2005 - June 2007. Last column gives radius of circle as found by algorithm. 5% of stations classified as type 1.

Table 7 shows that, also when using this method, Rotterdam and Veghel are identified. The cluster near Rotterdam is now larger than in the baseline.

6.4 Accounting for site characteristics

The variance screen that we use looks at the variance of prices relative to the mean of prices at a particular station. Yet, there may be reasons for a high price other than a lack of competitive pressure. For example, stations

[14]Note that it is even possible to combine the two methods by comparing the p-value of the most suspicious cluster found in section 4 to the p-value of the most suspicious area from the method above, and favoring whichever has the overall lowest p-value.

may offer a better service, they may be located close to the border, close to a highway, they may have higher demand, face higher marginal costs, etc. If there are such perfectly valid reasons for a station to charge higher prices, then we may underestimate its variation coefficient if we do not adjust for these factors.[15] In turn, that may affect the classification of stations that are of type 1, and also the clusters of type 1 stations that our method identifies.

In this section, we therefore look at prices that are adjusted for such factors. For each station, we determine the average price it would charge if it would have the average characteristics of a station in our data set. Using these adjusted prices, we calculate the adjusted variation coefficients which we then use as input for our collusion screen.

Formally, denote the average price at station i as μ_i. Let x_i denote a vector of station characteristics given as a deviation from the sample mean. To estimate the effect of the different site characteristics on the average level of a particular station, we perform the regression

$$\mu_i = \mu + x_i\beta + \varepsilon_i, \tag{3}$$

The adjusted average price at station i is its counterfactual average price if it would have the characteristics of the average station in our sample, so if $x_i = 0$. Hence $\mu_i^a = \mu + \varepsilon_i$: the adjusted average price at station i is the sample average plus the unexplained part of its true price.

As station characteristics, we include[16] number of pumps; plot size; size of shop area, and dummies for being close to the German or Belgian border, being company owned, carrying one of the four major brands[17], serving hot drinks, having a car wash and being fully automated ('unmanned'). We also include the log of the numbers of cars owned by private households within 20 kilometer of the station as a measure of local demand.[18] Inclusion of these

[15] Of course, factors that increase prices by a fixed amount do not have an effect on the variance of prices.

[16] Data on the characteristics of each gasoline station were obtained from Experian Catalist Ltd.

[17] Esso, Shell, Texaco and BP.

[18] These data are available for all but 34 stations.

variables is motivated by Soetevent, Haan and Heijnen (2011), where we find that these indeed affect gasoline prices.

Table 8: Regression of average price on explanatory variables (Sample: non-highway stations; Period: October/2005-June/2007)

	(1)		(2)	
Local competition measures:	Excluded		Included	
	coefficient	s.e.	coefficient	s.e.
sample mean	1.3641		1.3641	
Geographical characteristics				
German border	-0.0042	(0.0029)	-0.0064*	(0.0029)
Belgian border	0.0118**	(0.0033)	0.0102**	(0.0033)
Site characteristics				
Company owned	-0.0145**	(0.0010)	-0.0126**	(0.0011)
Major brand	0.0110**	(0.0010)	0.1142**	(0.0010)
# pumps	-0.0003	(0.0005)	-0.0001	(0.0005)
Unmanned	-0.0031*	(0.0015)	-0.0037**	(0.0014)
Hot drinks	0.0037*	(0.0015)	0.0036*	(0.0015)
Carwash	0.0020†	(0.0011)	0.0021*	(0.0011)
Plot size (area)	-2.73e-07	(3.73e-07)	-3.54e-07	(3.69e-07)
shop area	3.69e-05	(2.32e-05)	3.91e-05†	(2.29e-05)
Local demand				
# priv. owned cars $\leq$ 20km	0.0056**	(0.0007)	0.0079**	(0.0012)
Local market concentration				
ln(# non-highway stations+1) at...				
$\leq$ *1 km*			-0.0032**	(0.0009)
$1-2$ *km*			-0.0048**	(0.0076)
$2-5$ *km*			-0.0008	(0.0007)
$5-10$ *km*			-0.0008	(0.0012)
ln(# highway stations+1) at...				
$\leq$ *1 km*			0.0094*	(0.0052)
$1-2$ *km*			0.0018	(0.0023)
$2-5$ *km*			0.0018†	(0.0010)
$5-10$ *km*			0.0004	(0.0008)
R^2	0.1308		0.1557	
obs.	3035		3035	

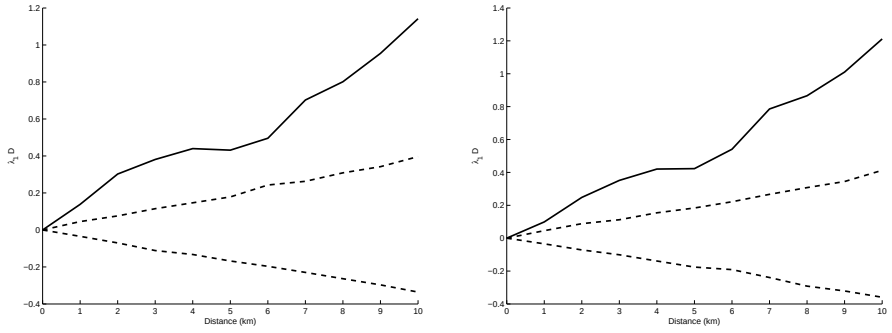
Plot size area and shop area in sq. m; privately owned cars in '000.000.
†: Significant at the 10% level; * : Significant at the 5% level; ** : Significant at the 1% level.

We could also add concentration measures to our list. Yet, this is tricky. If local cartels are more likely to exist in areas with low market concentration, then we are effectively destroying possible evidence for collusion if we use local market concentration as a basis to adjust prices for stations that are active

in such areas. With this caveat in mind, we calculate two sets of adjusted prices: one in which concentration measures are also taken into account, an one in which they are not. As concentration measures we use the logs of the number of highway stations and the number of other non-highway stations within distances of 1, 2, 5 and 10 kilometer. Regression results are given in Table 8.

The estimates in column (1) of Table 8 show that gasoline prices at outlets close to the Belgian border are some 1.2% higher. *Ceteris paribus*, outlets of one of the major brands charge prices that are on average 1% higher, whereas company owned outlets charge prices that are 1.5% lower. Prices at fully automated stations are 0.3% lower on average. Column (2) also includes local concentration measures. The estimates show that the presence of other non-highway stations within 2 kilometer distance puts a downward pressure on prices, while having highway stations nearby increases prices. Most probably this picks up the positive demand effect of being close to a highway exit.[19]

Figure 8: D-function, adjusted prices



The solid line is the D-function, 95% confidence interval is indicated by the dashed lines. Left panel: local concentration variables excluded. Right panel: local concentration variables included. (Sample: non-highway stations; Period: October/2005-June/2007). 5% of station classified as type 1, cluster size 5 km.

Figure 8 shows the D-function for the adjusted prices. In the left-hand panel, prices have been adjusted for differences in site characteristics; in the

[19]The estimates are very similar to those in Soetevent *et al.* (2011, Table 3).

right-hand panel, they have also been adjusted for regional differences in market concentration.

Table 9: Identified suspicious clusters, adjusted for station characteristics

#	midpoint	type 1	type 0	$-\log p$	HHIT	HHIS	nearest city
1	(91,443)	15	25	9.4	0.135	0.173	Rotterdam

Sample October 2005 - June 2007. 5% of stations classified as type 1, cluster size 5 km

Table 9 lists clusters generated by applying our method to prices that are only adjusted for station characteristics, not for concentration measures. Again, Rotterdam is the only suspicious cluster that our method generates, although the identified area is now somewhat larger than in the baseline. The same is true if we also adjust for concentration measures, in Table 10.

Table 10: Identified suspicious clusters, prices adjusted for station characteristics and concentration measures

#	midpoint	type 1	type 0	$-\log p$	HHIT	HHIS	nearest city
1	(91,443)	15	25	9.4	0.135	0.173	Rotterdam

Sample October 2005 - June 2007. 5% of stations classified as type 1, cluster size 5 km

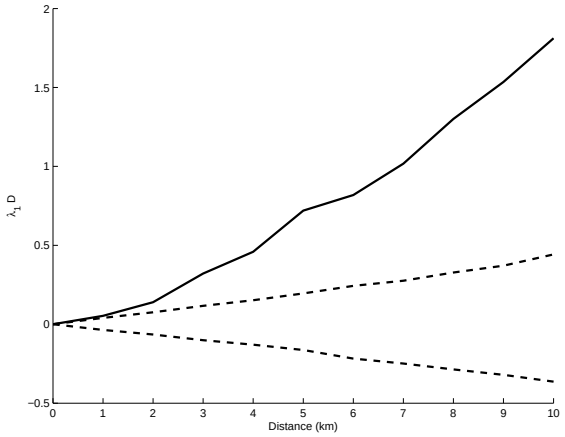
6.5 An alternative time period

Next, we investigate how our method is affected when we consider a different time period. In Figure 9, the D-function is plotted based on imputed price data for the period July 2007 to April 2009. For this period too, we find clustering of suspicious stations for all possible choices of h.

When looking at the most suspicious clusters for $h = 5$ kilometer, the picture looks different. Our method now generates 5 clusters before there is lack of evidence for further clustering, see Table 11. Eindhoven is flagged as the most suspicious cluster, although Rotterdam still makes the list as well.

One observation that stands out in Table 11 is the extremely high value of the Herfindahl index among type 1 stations in cluster 3. Out of 10 type

Figure 9: D-function, July/2007-April/2009)



The solid line is the D-function, 95% confidence interval is indicated by the dashed lines. 5% of stations classified as type 1, cluster size 5 km.

1 stations in this cluster, 9 carry the Texaco brand. Among the 25 type 0 stations, there is not a single Texaco station. Hence, rather than a local cartel, this cluster is more likely to reflect market dominance of Texaco in this particular area.

Table 11: Identified suspicious clusters, sample period July 2007 - April 2009

cluster	midpoint	type 1	type 0	$-\log p$	HHI	HHI[1]	nearest city
S_1	(160,377)	9	12	6.5	0.152	0.259	Eindhoven
S_2	(103,490)	5	1	5.8	0.222	0.280	Haarlem
S_3	(80,453)	10	25	5.3	0.171	0.820	Den Haag
S_4	(94,440)	10	29	4.8	0.120	0.180	Rotterdam
S_5	(92,464)	4	1	4.5	0.440	0.625	Leiden

Sample July 2007 - April 2009. 5% of stations classified as type 1, cluster size 5 km

One explanation for the different picture that we see now is that the market environment may have changed substantially; areas that were a cartel in 2005-2007 may not be so anymore in 2007-2009. That is confirmed if we look at the robustness checks that we also did for the period 2005-2007.

Changing the cartel distance or the fraction of type 1 stations consistently yields 5 or 6 clusters are flagged as suspicious, with Eindhoven and Haarlem being the most suspicious cluster equally often, and with substantial overlap among the other clusters that are generated as well. Yet, when correcting for site characteristics, the most suspicious cluster is close to Arnhem, an area that was not flagged before. That is true when we correct only for site characteristics, but also when we correct for both site characteristics as well as concentration measures. Different from the situation in 2005-2007, correcting for site characteristics now makes a substantial difference in the outcome of the collusion screen. When using the alternative method to determine clusters, we again find Eindhoven as the prime suspect.

7 Conclusion

In this paper, we developed a method to screen for local cartels. Our method takes as an input information on which outlets score high on some characteristic that is consistent with collusive behavior. It then tests whether there is statistical evidence that these suspicious outlets are clustered and, if so, provides an algorithm to find which clusters are the most suspicious. Our method can readily be used in applications outside the realm of competition policy or economics.

Our approach has a number of advantages. It uses data that are readily available, is easy to implement and hard for a cartel to beat. It only identifies suspicious clusters if there is statistical evidence for such clustering. It continues to identify suspicious clusters as long as there still is evidence for clustering in the remaining data.

We applied our method to the Dutch gasoline market. Using daily price data on virtually all gasoline stations in the Netherlands, we classified as suspicious those stations with a particularly low variation coefficient, following the literature on variance screens initiated by Abrantes-Metz et al (2006). For the period 2005-2007 we find clustering in an area close to Rotterdam. This areas is persistent, in the sense that it is robust to variations in our method.

Naturally, this can never be construed as evidence for collusion, but it does suggest that an antitrust authority may have a closer look at the stations active in that area. For the period 2007-2009, the picture is less clear, and areas around Eindhoven, Haarlem and Arnhem turn up as most suspect, depending on the exact method that is used. our method returns many small clusters, plus some larger clusters that also contain many non-suspicious stations.

Needless to say, any method that screens for collusion can only be as good as the data that are used as its input. In the end, it is up to antitrust practitioners to come up with criteria to determine whether a station is suspicious or not. The variance screen is one such criterion, but without doubt, many others can be thought of. Other inputs of the variance screen, such as cluster size or the fraction of outlets that are classified as suspicious, may also influence its output, although as we saw in Section 6 that is reasonably robust for such choices. Just like any other tool, our collusion screen should be applied with care. Its output can only serve as a starting point for further investigation.

References

Abrantes-Metz, R., and P. Bajari (2009): "Screens for Conspiracies and their Multiple Applications," *The Antitrust Magazine*, 24(1), 66–71.

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

Athey, S., K. Bagwell, and C. Sanchirico (2004): "Collusion and Price Rigidity," *Review of Economic Studies*, 71, 317–349.

Cressie, N. (1991): *Statistics for Spatial Data*. Wiley and Sons, New York.

Diggle, P., and A. Chetwynd (1991): "Second-Order Analysis of Spatial Clustering for Inhomogeneous Populations," *Biometrics*, 47, 1155–1163.

Dixon, P. (2002): "Ripleys K function," in *Encyclopedia of Environmetrics*, ed. by A. H. El-Shaarawi, and W. W. Piegorsch, vol. 3, pp. 1796–1803. John Wiley & Sons, Chichester.

Duranton, G., and H. Overman (2005): "Testing for localisation using micro-geographic data," *Review of Economic Studies*, 72, 1077–1106.

Ellison, G., and E. L. Glaeser (1997): "Geographic concentration in US manufacturing industries: a dartboard approach," *Journal of Political Economy*, 105(5), 889–927.

European Union (1999): "Merger Procedure Case No IV/M.1383 - Exxon/Mobil," Regulation (EEC) No 4064/89L.

Froeb, L., J. Cooper, M. Frankena, P. Pautler, and L. Silvia (2005): "Economics at the FTC: Cases and Research, with a Focus on Petroleum," *Review of Industrial Organization*, 27, 223–252.

Haase, P. (1995): "Spatial pattern analysis in ecology based on Ripleys K-function: Introduction and methods of edge correction," *Journal of Vegetation Science*, 6, 575–582.

Harrington, J. E. (2008): "Detecting Cartels," in *Handbook in Antitrust Economics*, ed. by P. Buccirossi, chap. 6, pp. 213–258. MIT Press.

Jimenez, J. L., and J. Perdiguero (2009): "(No) competition in the spanish retailing gasoline market: a variance filter approach," mimeo.

Marcon, E., and F. Puech (2010): "Measures of the geographic concentration of industries: improving distance-based methods," *Journal of Economic Geography*, 10, 745–762.

NMa (2006): "Benzinescan 2005/2006," Discussion paper.

Picone, G. A., D. B. Ridley, and P. A. Zandbergen (2009): "Distance Decreases with Differentiation: Strategic Agglomeration by Retailers," *International Journal of Industrial Organization*, 27, 463–473.

Rysman, M., and S. Greenstein (2005): "Testing for agglomeration and dispersion," *Economics Letters*, 86, 405–411.

Soetevent, A., M. Haan, and P. Heijnen (2011): "Do Auctions and Forced Divestitures increase Competition? Evidence for Retail Gasoline Markets," Tinbergen Institute Discussion Paper 2008-117/1.

Stoyan, D., and A. Penttinen (2000): "Recent applications of point process methods in forestry statistics," *Statistical Science*, 51, 61–78.

Wang, Z. (2009): "(Mixed) Strategy in Oligopoly Pricing: Evidence from Gasoline Price Cycles Before and under a Timing Regulation," *Journal of Political Economy*, 117(6), 987–1030.