

University of Groningen

## Screening for collusion

Heijnen, Pim; Haan, Marco A.; Soetevent, Adriaan R.

*Published in:*  
Journal of Economic Geography

*DOI:*  
[10.1093/jeg/lbu012](https://doi.org/10.1093/jeg/lbu012)

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2015

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Heijnen, P., Haan, M. A., & Soetevent, A. R. (2015). Screening for collusion: A spatial statistics approach. *Journal of Economic Geography*, 15(2), 417-448. <https://doi.org/10.1093/jeg/lbu012>

### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

# Screening for collusion: a spatial statistics approach

Pim Heijnen<sup>\*†</sup>, Marco A. Haan<sup>\*\*</sup> and Adriaan R. Soetevent<sup>\*\*\*</sup>

<sup>\*</sup>Faculty of Economics and Business, University of Groningen, PO Box 800, 9700AV Groningen, The Netherlands

<sup>\*\*</sup>Faculty of Economics and Business, University of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands

<sup>\*\*\*</sup>Faculty of Economics and Business, University of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands, and Tinbergen Institute

<sup>†</sup>Corresponding author: Faculty of Economics and Business, University of Groningen, PO Box 800, 9700AV Groningen, The Netherlands. *email* <p.heijnen@rug.nl>

## 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.

**Keywords:** Collusion, variance screen, spatial statistics

**JEL classifications:** C11, D40, L12, L41

**Date submitted:** 26 September 2012 **Date accepted:** 2 April 2014

## 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 various instruments at its disposal. One is that it may actively screen markets for price patterns or other markers that suggest collusive behavior. In this article, we develop a method to screen for local cartels. First, we use 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.

Prominent cartel cases often entail large or even international conspiracies to raise prices. Yet many cartels, and especially those that directly affect consumers, are of a more local nature and only involve a limited number of local suppliers making agreements in one particular region.<sup>1</sup> Indeed, in industries with many independent local suppliers, it would be very hard to organize and police a nationwide cartel.

The fact that many cartels are of a local nature provides antitrust authorities with a potentially powerful instrument to track them down. Suppose that firms in a cartel

1 Naturally, it is hard to find prominent example of such local cartels. Recent example in the Netherlands include e.g. a cartel of window cleaners in The Hague (ACM, Besluit in zaak 6425 [decision in case 6425], 22 December 2011, <https://www.acm.nl/nl/download/bijlage/?id=7554>), and one of taxi services in the greater Rotterdam area (ACM, Besluit in zaak 7131 [decision in case 7131], 5 March 2013, <https://www.acm.nl/nl/download/publicatie/?id=11222>).

exhibit behavior that systematically differs from firms that are not part of a cartel. For example, they may charge higher prices. An antitrust authority concerned about the existence of local cartels is then well-advised to monitor whether there is indeed a geographical cluster of firms that exhibit such behavior. Finding local cartels then implies screening for clusters where the suspicious behavior is particularly prevalent. In this article, we develop a method for such screening. Of course, the identification of such a cluster can never be construed as legal evidence in a cartel case. But it would surely justify further scrutiny using more conventional methods, such as dawn raids or forensic accounting.

High prices may not be the best candidate for such a test, however. These may be caused by collusion, but can also simply be due to e.g. high local demand, or high land prices that makes entry of additional competitors prohibitively expensive. Abrantes-Metz, Froeb, Geweke and Taylor (2006) (AFGT henceforth) suggest a better marker for collusive behavior. Firms that are part of a cartel often charge prices that are less volatile than firms that behave competitively. In a competitive market, prices will closely track costs over time. But in a cartel, firms will be much more reluctant to change their price, as doing so may be interpreted by other cartel members as cheating. Moreover, for a cartel it is much easier to make an agreement to literally fix prices rather than to implement a formula that prescribes how the price the cartel sets should fluctuate with underlying costs. Athey et al. (2004) provide a more formal theoretical rationale. Empirical evidence that cartels charge more stable prices can be found in AFGT.

We are not the first to develop a screen to detect anticompetitive behavior. Abrantes-Metz and Bajari (2009) provide an overview of such screens, see also The Economist (2012). 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 statistics that are easily calculated from readily available data, 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.

The identification of suspicious outlets is an important ingredient in our collusion screen. As mentioned above, we look at price variability. Following AFGT, our measure of price variability of station  $i$  is the variation coefficient  $v_i$ , defined as the standard deviation  $\sigma_i$  of  $i$ 's retail price, divided by its mean price  $\mu_i$ . 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 suspicious those stations that have a particularly low  $v_i$ .<sup>2</sup>

Yet, the literature so far does not fully exploit data on the location of outlets.<sup>3</sup> Our contribution is to add a formal test for clustering to the literature, plus an algorithm to determine where these clusters are located. It is important to stress that the price

2 The fact that some local markets may have lower price variability than others may also be explained by the presence of Edgeworth cycles in some markets but not in others (Maskin and Tirole, 1988; Wang, 2009). This is one of the issues an antitrust authority has to take into account when interpreting the outcome of this particular collusion screen. Still, we feel that this is not so much of a problem in the current study: using a data set similar to ours, Faber (2011, p. 4, fn. 3) does not find any evidence for Edgeworth cycles on the Dutch gasoline market.

3 AFGT (2006) use eyeballing to determine that gasoline stations with low price variability in their data set are not clustered. Jimenez and Perdiguero (2012) look at predefined markets.

variance screen is just one possible use 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.

There is an immense statistical literature in numerous fields that studies clustering (see e.g. Waller (2009), and the references therein). Some of this literature focuses on the detection of clustering *per se* (that is, the question whether in general events tend to locate close to each other), whereas other parts focus on the detection of particular clusters (that is, trying to identify locations or areas where events occur particularly often).

Concerning the *detection of clustering*, we closely follow Diggle and Chetwynd (1991), who focus on the possible clustering of diseases. Their test statistic essentially looks at whether ill people are relatively more likely to be located close to other ill people, than healthy people are.<sup>4</sup> We closely follow their approach. The literature on the *detection of clusters* often looks at a *spatial scan statistic*. Kulldorff (1997) introduces what is now the most popular method. He essentially looks for the area with the most unlikely combination of cases and controls, and then uses a Monte Carlo simulation to analyze exactly how unlikely it is that the most unlikely area has the particular likelihood found in the data. We take a somewhat different approach. We first identify clusters. Our clusters are essentially a partition of all events such that, first, each event within a cluster is at most  $h$  kilometer from some other event within the same cluster and, second, any two events in different clusters are more than  $h$  kilometer away. For each cluster, we observe the number of cases and controls in that cluster. The cluster that is the least likely to occur is labeled the most suspicious cluster.

Our proposed method thus consists of two steps. The first step, detecting clustering, largely follows Diggle and Chetwynd (1991). The second step, detecting clusters, is novel and finds the cluster least likely to occur by chance. After finding the least likely cluster, we remove it from our data and test whether there is still significant clustering in the remaining data. If so, we locate the second most suspicious cluster, etcetera.<sup>5</sup> Our main contribution is therefore 2-fold. To the literature of collusion screening, we add a formal test to identify suspicious clusters. To the literature of agglomeration, we add a new method to identify the most prominent clusters.

We apply our method to the Dutch gasoline market, using almost daily prices from 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. We may focus on raw prices, but may also choose to correct prices for station characteristics. Any screen would be of

4 In economics, a related approach appears 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. 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  $h$ . Also 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 estimation.

5 Note that Kulldorff (2010, 21) proposes a similar method to find multiple clusters. He proposes to stop looking, however, after a predetermined ad hoc significance level is no longer crossed by the next most significant cluster.

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 areas close to Hoogeveen and Rotterdam persistently pop up as the most suspicious clusters in our data. Hence, if the Dutch antitrust authority would have used this tool in that period, the advice 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, we find that areas close to Ede, Rotterdam and Den Haag as most suspicious. Hence, there may now be a local cartel near Ede and, if so, it is likely to have formed after 2007. As said, a collusion screen like the one we propose can never serve as legal evidence for the existence of local cartels. Further research to find evidence for collusion will always be necessary.

The article proceeds 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: detecting clustering. We discuss our test statistic and compare it to other methods used in spatial statistics and economics. Section 4 discusses our method to detect clusters. Again, we confront our method with those used in the literature. 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 four steps. Before actually doing the analysis, one has to collect and prepare the necessary data. This is *Step 1: Data preparation*. 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. Also, one has to decide whether to use ‘raw’ price data, or to use prices that are adjusted for e.g. outlet characteristics or area fixed effects. We choose to do the latter.

*Step 2: Testing for clustering* is to determine which outlets are suspicious and which are not. For simplicity, we will refer to suspicious outlets as type 1, and to nonsuspicious 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 nonsuspicious—but we will also run robustness checks using different percentages. 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 labeling, in the sense that the type 1 ‘labels’ are randomly distributed over all existing outlets. This step is described more extensively in Section 3.

If we find evidence for local clustering, we move to *Step 3: Ranking clusters*. 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 least likely to occur under the null hypothesis of random labeling.

*Step 4: Iterative elimination of clusters* consists of eliminating all outlets in the most suspicious cluster from the data. After having done so, we move back to step 2 to test

whether there is evidence for local clustering in the remaining outlets. Steps 3 and 4 are discussed in more detail in Section 4.

### 3. Testing for clustering

In this section, we introduce and motivate our test statistic to determine whether there is evidence for local clustering. Our problem can be stated as follows. We have a set  $N$  consisting of  $n$  outlets.<sup>6</sup> The location of outlet  $i \in N$  is given by  $x_i \in R^2$ . 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  $\gamma$ :  $\gamma \equiv n_1/n$ . 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.<sup>7</sup>

There is an immense statistical literature in numerous fields that studies clustering (see e.g. Waller (2009), and the references therein). Prominent examples include studies in epidemiology that looks at the incidence of certain diseases, studies in botany that analyze whether certain species tend to locate close to each other, and studies in economics that look at industrial agglomeration. Ripley (1976) introduces a test statistic that essentially looks at the average number of events that occurs within  $h$  kilometer of a random event, and compares that to the number that would be expected under complete spatial randomness (CSR). More formally, Ripley’s planar  $K$ -function (Cressie, 1991, 615–619), counts at radius  $h$  the average number of other events within  $h$  of an event:

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

with  $\lambda$  the intensity of the spatial process. With more spatial clustering, events are located closer to each other, hence  $K(h)$  will be higher. Confidence intervals are determined by Monte Carlo simulation.<sup>8</sup> 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), among 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.<sup>9</sup> Diggle and Chetwynd (1991) study a

6 Throughout, we use the convention that uppercase letters refer to the set and lowercase letters denote the cardinality of the set.

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

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

9 As an example, consider an isolated area  $A$  in which four outlets are located, two of which 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

similar problem in the context of possible clustering of rare diseases. Locations where the disease can occur are constrained by locations where people-at-risk live, making this a different problem from Ripley's. Diggle and Chetwynd's (1991) test statistic looks at the average number of ill people (cases) within  $h$  kilometer of a random ill person, relative to the average number of healthy people (controls) within  $h$  kilometer of a random healthy person in a sample of healthy persons of the same size as the population of ill people. If the disease is randomly distributed among the population, both averages should be equal.

Formally, a type 1 event is an occurrence of the disease. 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  $\lambda_1$  the intensity of type 1 events. Now take a sample of controls consisting of  $n_1$  events randomly drawn from the entire population. We can calculate Ripley's  $K$  for this sample:

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

Then, the test statistic  $D(h)$  is defined as

$$D(h) \equiv K_1(h) - K_c(h).$$

Under random labeling,  $D(h) = 0$ . 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, 1157–1158) approximate the true distribution by implementing a Monte Carlo simulation consisting of a number of random permutations of the type 1 labels over the type 1 events and controls. One application in economics that follows a similar approach is Smith et al. (2008), who test the 'spatial void hypothesis' that alternative financial service providers tend to locate in markets where traditional banking services are underprovided.<sup>10</sup>

We closely follow this approach. The only difference is the following. Diggle and Chetwynd (1991) have information on 62 cases diagnosed with childhood leukemia in North Humberside, and 141 controls selected at random from entries on the birth register. For all these cases and controls, they need to find the home address to be able to do their analysis. However, we already have the exact information of all locations of outlets. Therefore, rather than calculating  $K_c(h)$  on the basis of one sample of events, we can be 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.<sup>11</sup>

---

the absolute number of type 1 outlets. For our purposes, an appropriate test statistic should correct for the density of stations and look at the relative number of type 1 outlets in an area, rather than merely at the absolute number.

- 10 An alternative method is provided by Getis and Franklin (1987), who look at the distance between type 1 labels as their unit of analysis. An application in economics that is similar in spirit is 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.
- 11 Note that we have 3035 nonhighway outlets with 151 suspect stations. If we follow Diggle and Chetwynd's approach exactly, we would need to compute the distance between each of the 151 suspect stations and each of the  $3035 - 151 = 2884$  nonsuspect stations. There are approximately 4.2 million pairs of nonsuspect stations (compared to approximately 11,000 pairs of suspect stations). This is a huge

Summarizing, for a given radius of  $h$  we proceed as follows. First we take a sample of  $n_1$  controls, calculate the corresponding  $K_c(h)$ , and repeat this procedure 1000 times to calculate  $\bar{K}_c(h)$ . Next we take a random sample of  $n_1$  events, assign them a type 1 label and calculate the corresponding  $K_1^i(h)$ ,  $i \in \{1, \dots, 1000\}$ . On the basis of that, we calculate  $\bar{D}^i(h) = K_1^i(h) - \bar{K}_c(h)$ . We repeat this procedure 1000 times to calculate the distribution of  $\bar{D}(h)$  under the null of random labeling. Finally, we look at the actual incidence of type 1 labels, calculate the corresponding  $K_1^a(h)$  and  $\bar{D}^a(h)$  and use the derived distribution of  $\bar{D}(h)$  to construct confidence intervals and to test whether the resulting  $\bar{D}^a(h)$  significantly departs from the null of random labeling. If so, we conclude that there are clusters of low price variation at scale  $h$ .

This method is relatively easy to implement and interpret. With the density of type 1 events given by  $\lambda_1$ ,  $\lambda_1 D(h)$  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.<sup>12</sup>

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 first determines the distance  $h$  at which firms still (should) effectively compete with each other. In other words,  $h$  is determined by the size of the relevant geographical market. Alternatively, different distances of  $h$  could be used as a robustness check.

#### 4. Detecting and ranking clusters

Suppose that, using the method described in the previous section, we have found evidence for local clustering at a distance of  $h$  kilometer. In that case, we first determine clusters of type 1 outlets. Second, to judge which cluster is most suspicious, we determine for each cluster the likelihood of observing that number of type 1 outlets under spatial randomness. Finally, we rank them to infer which of these areas is most suspicious.

The literature on the detection of clusters often looks at a *spatial scan statistic*. In its most basic form, such a statistic can be understood as follows. Suppose we have a square region with a number of points. Naus (1965) then considers a rectangular

---

computational burden, especially when we bootstrap to obtain the distribution under the null hypothesis. The double sampling approach that we use is much faster. To be absolutely sure that our double sampling procedure does not affect the results, we have redone one of our analyses without double sampling. In that case, we never find a value of our  $D$ -function that is more than 0.011 different from the value we find with double sampling.

12 An alternative could have been to use the  $M$ -statistic used by Marcon and Puech (2010), see also Marcon et al. (2012). In the context of our application they essentially look—for all type 1 events—at the fraction of type 1 events within a distance  $r$  of that event, take the average of that number over all type 1 events and compare that average to a Monte Carlo simulation. Yet, in our particular application, when implementing this method, we found it to be less stable than the method of Diggle and Chetwynd. The reason may be that in our application the number of cases and controls in an area is often low. Having one extra or one fewer observation can then strongly affect the  $M$ -statistic. One advantage of Marcon and Puech (2010) in other applications is that it is easy to allow for different weights of events. When studying clustering of industries, for example, one can weigh different plants with their level of employment.



scanning window with a fixed size and shape. This window is continuously moved over the study region, covering all possible locations. The spatial scan statistic is the maximum number of points in the scanning window that occurs. The next step is to find the probability of observing at least that many points within the window, under the null hypothesis of CSR. Naus (1965) developed theoretical formulas to obtain upper and lower bounds for those probabilities (see Costa and Kulldorff, 2009).

In related work, Openshaw et al. (1988) develop a graphical method and coined the geographical analysis machine (GAM) to detect clusters. The GAM defines potential clusters as collections of events falling within circles of varying size and midpoints. For each circle, the method evaluates the likelihood of the number of events observed in that circle under the null hypothesis of randomness. Each circle that is significant at 99.8% is reported, and the circle is marked. This yields a map with suspicious circles.

A problem with these methods is that it involves multiple testing. If we look at 10,000 circles, say, and a significance level of 99.8%, we will find some 20 circles that suggest significant clustering. Openshaw et al. (1988) use a significance level of 99.8% rather than 95% to account for this but of course, that is ad hoc. Besag and Newell (1991) and Turnbull et al. (1990) try to get around this problem. Kulldorff (1997) introduces what is now the most popular method. He essentially looks at all possible circles (zones in his terminology) on the area under consideration. For each zone, he observes the number of events in that zone. A Monte Carlo simulation is then used to analyze exactly how unlikely it is that the least likely cluster has the particular likelihood found in the data. Kulldorff's (1997) method allows for rectangles, circles, ellipses and any other prescribed geometric form, and also allows for many data-generating mechanisms, including Bernoulli (as we have in our application), and Poisson. Surveys of spatial scan statistics include Kulldorff (1999) or Costa and Kulldorff (2009). One application in economics is Carlino et al. (2012) who look at different cluster sizes simultaneously to gain further insight in the structure and the nature of R&D clusters in the Northeast corridor of the USA.

Yet, when looking for a potential local cartel, we have no prior concerning its shape. However, adapting the spatial scan statistic to allow for arbitrarily shaped clusters is notoriously hard, as it reintroduces multiple testing problems. As Duczmal et al. (2009) note: 'the collection of all connected zones, irrespective of shape, is very large; the maximum value of the objective function is likely to be associated with 'tree-shaped' clusters, which merely link the highest likelihood ratio cells of the map, without contributing to the discovery of geographically meaningful solutions that correctly delineate the 'true' cluster. In other words, there is much 'noise', against which the legitimate solutions cannot be distinguished.'<sup>13</sup>

We therefore take an approach that is somewhat different from the canonical spatial scan statistic pioneered by Kulldorff (1997). As noted, Kulldorff (1997) looks at all possible circles, and identifies the least likely one as the most suspicious cluster. In our approach, we first identify clusters by selecting the relevant  $h$ . Our clusters are essentially a partition of all events such that, first, each event within a cluster is at most  $h$  kilometer from some other event within the same cluster and, second, any two events

13 Mori and Smith (2014) propose a method to find irregularly shaped clusters when data are available per region, but that method is less suitable when exact locations are available, as is the case in our application.

in different clusters are more than  $h$  kilometer away. For each cluster so formed, we observe the number of events and the number of controls in the convex hull formed by that cluster. Given these, we can calculate the likelihood that such a cluster occurs under random labeling. The most suspicious cluster is then the one that is the least likely to occur.

For our particular application, our method of finding the most suspicious cluster has a number of advantages over that in Kulldorff (1997). First, it is computationally more efficient: our method only requires us to calculate the likelihood of a small number of clusters, while the spatial scan statistic requires the calculation of the likelihood of each possible circle with each possible radius.<sup>14</sup> Second, it allows for clusters of any shape. Cartels do not necessarily come in rectangular, circular or any other predetermined geographical form, hence it is useful to have a cluster detection method to allow for that. Third, for our particular application the method makes more economic sense. Suppose that an antitrust authority concludes, based on observed price responses, that outlets effectively compete with each other at a distance of up to  $h$  km. If there would be a suspicious outlet at a distance smaller than  $h$  km of a cluster of other suspicious outlets, it would be peculiar not to consider that outlet as part of the same potential cartel. Similarly, if in that case a suspicious outlet would be more than  $h$  km from any other suspicious outlet, it would not be an obvious choice to regard it as a potential cartel member. Still, a spatial scan statistic would allow such cases to occur.

Our approach proceeds as follows. 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 this procedure leads to partitioning of all type 1 outlets into clusters. By construction, any type 1 outlet is less than  $h$  kilometer away from at least one other outlet in its cluster, and more than  $h$  kilometer away from outlets in any other cluster.

Consider the set  $N_1$  of type 1 outlets. We consider two type 1 outlets as being *adjacent* if they are located less than  $h$  kilometers from each other. We connect adjacent type 1 outlets. We define clusters as the connected components of the resulting undirected graph of type 1 outlets. That is, a cluster is a subset of  $N_1$  with the adjacency relations restricted to this subset.

Suppose that this procedure yields  $\ell$  clusters  $S_1, S_2, \dots, S_\ell$ . The cardinality of cluster  $S_i$  is denoted  $s_i$ . Without loss of generality, we order clusters from largest to smallest, so  $s_i \geq s_{i+1}, \forall i < \ell$ . 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, whereas  $S_2$  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 in *Step 3 [Ranking clusters]* the area where cluster  $S_i$  is located as the convex hull of the locations of all outlets in  $S_i$ :  $A_i = \text{Conv}(S_i)$ <sup>15</sup>. The

14 Kulldorff (1999) surveys algorithms to make that process more efficient.

15 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

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 overall, a fraction  $\gamma$  of all outlets is of type 1. Under the null hypothesis of random labeling, we can calculate the probability that, given a total of  $s_i + s_i^0$  outlets in  $A_i$ , at least  $s_i$  are of type 1. This probability equals:

$$p(S_i) = \sum_{j=s_i}^{s_i+s_i^0} \frac{\binom{n_1}{j} \binom{n-n_1}{s_i+s_i^0-j}}{\binom{n}{s_i+s_i^0}} \quad (1)$$

which we will refer to as the ‘ $p$ -value’ of cluster  $S_i$ . Note that these are draws from a binomial distribution where  $\binom{n_1}{j}$  etc. denote the binomial coefficients. It is important to note that, since we explicitly focus on clusters of type 1 outlets, the  $p$ -values that we find should not be interpreted as significance levels: also under CSR some clusters will form that are very unlikely to occur when looked at in isolation.

For ease of exposition, we will report the negative of the log of  $p$ . 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. Finally, *Step 4 [Iterative elimination of clusters]* singles out the most suspicious cluster, which is the cluster with the largest value of  $-\log p(S)$ :

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

We remove this cluster from our data and move back to step 2 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. Price data for the gasoline 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. Wang, 2009; Soetevent et al., 2014). Moreover, gasoline markets are often suspected to be prone to anticompetitive price manipulation, and in many countries they are subject to 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).

As noted, we classify as suspect stations that have a particularly low price variability. That variability is measured by the variation coefficient  $v_i$ , defined as the standard deviation of  $i$ 's retail price, divided by its mean price. One might be inclined to argue that, rather, a number of stations are suspect if their prices move in tandem or in other words if the correlation between their prices is high. However, in general price

---

do so, as that would imply that type 0 stations can be part of more than 1 cluster. We also do not believe that this would greatly affect our analysis.

variations are caused by changes in marginal cost and shifts in demand, which have a similar effect on all firms active in the market. Hence, regardless of the industry structure prices tend to be correlated. It is not clear *a priori* that prices in cartels are more heavily correlated than prices in competitive or monopolistic markets. In Appendix A, we discuss this issue, which can be summarized as follows. First, we show that the clusters we identify do not exhibit an unusually high price correlation. Second, we show that there is little relation between the correlation of prices between stations and the distance between those stations. In other words, what we pick up with our method is more than just spatial autocorrelation.

In the remainder of this section, we go through the four steps of our procedure. We first describe our data, discuss how we impute missing data and how we adjust for observable characteristics. Second, we identify the type 1 stations and determine whether there is statistical evidence for local clustering. This turns out to be the case and we rank the clusters in a third step. After removing the most suspicious 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 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 October 2005 and June 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. For simplicity, we will refer to this number in the remainder of this article as ‘cluster distance’. Third, we have to decide on the fraction  $\gamma$  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 adjust these for e.g. local circumstances. Initially we consider adjusted prices. In Section 6, we will do numerous sensitivity analyses to check how sensitive the results are for all of these choices.

## 5.2. Step 1: data preparation

We use a fleet card data set which contains regular price quotes for 3259 gasoline retail outlets in the Netherlands, see Appendix B.1 for more details. For now, we limit attention to the period 1 October 2005–30 June 2007. We only look at prices of regular unleaded gasoline, the most common type and hence the one for which the most data are 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 based on e.g. the zip code, as is often the case in other applications. We follow the method proposed AFGT to impute missing data (see Appendix B.2), and restrict attention to nonhighway stations, for reasons set out in Appendix B.3.

There may be reasons for a high price other than a lack of competitive pressure. For example, stations may offer a better service, they may be located close to the border, close to a highway, they may have higher demand, be located in an area with high land prices, etc. If there are such perfectly valid reasons for structural price differences between stations, then our estimates of the variation coefficient may be biased if we do

not adjust for station heterogeneity.<sup>16</sup> In turn, this may affect the classification of stations that are of type 1, and also the clusters of type 1 stations that our method identifies.

We therefore look at residual price differences between stations after we have taken into account observed heterogeneity in station characteristics. Using these residual 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  $\mu_i$  and the overall average as  $\bar{\mu}$ . Let  $x_i$  denote the vector of station characteristics (including a constant). To estimate the effect of the different site characteristics on the average level of a particular station, we perform the regression

$$\mu_i = x_i\beta + \varepsilon_i. \quad (2)$$

The adjusted average price at station  $i$  is the overall average price plus the unexplained part of its true price:  $\mu_i^a = \bar{\mu} + (\mu_i - x_i\hat{\beta}) = \bar{\mu} + \hat{\varepsilon}_i$ .

High prices may also be caused by high location rents in a particular area. For that reason, we also allow for area fixed effects on the level of two-digit postcodes. The Netherlands consists of 90 such areas. Of course, the choice of the size of areas to consider is tricky. If we choose them too small, we run the risk of creating areas that coincide with local cartels. In that case, high prices in a particular area may wrongly be considered an area fixed effect rather than a cartel. On the other hand, if we choose these areas too large, there may be too much variation in location rents within an area, and hence we do not fully control for those. We feel that, for the Netherlands, taking the two-digit postcode level as our unit of analysis is a reasonable compromise between these two effects.

We could also adjust prices for concentration measures, such as the number of other nonhighway stations within distances of 1, 2, 5, and 10 km. Yet, some care is needed. Standard theory suggests that the existence of local cartels is positively correlated with market concentration, such that adjusting prices for differences in local market concentration may effectively dent the indications for clustering as well.<sup>17</sup> With this caveat in mind, we calculate two sets of adjusted prices: one in which concentration measures are also taken into account, and one in which they are not.

Our baseline estimate use prices adjusted for station characteristics and area fixed effects on the level of two-digit postcodes, but not adjusted for local concentration measures. In Appendix B.4, we describe the characteristics we use in detail, and also study how these affect prices. We will do a robustness check in which we do adjust for local concentration measures.

### 5.3. Step 2: testing for clustering

For each station, we calculate the variation coefficient  $v_i$  and rank these with  $v_{[i]}$  denoting the site with the  $i$ -th lowest variation coefficient. For a given value of  $\gamma$ , the set of type 1 stations that have a  $v_i$  that is among the lowest  $\gamma n$  is denoted

$$N_1(\gamma) \equiv \{i \in S : v_i \leq v_{[\gamma n]}\}.$$

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

17 See e.g. Pepall et al. (2008, section 15.2.1).

As noted, we will use  $\gamma = 0.05$  in this section.

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

In Figure 3, we plot our  $D$ -function for different values of  $h$ . We focus on clusters at a distance of 5 km. At  $h = 5$ , the  $D$ -function shows clear evidence for clustering of type 1 stations. The average type 1 station has almost 0.4 more type 1 neighbors than expected under random labeling. This is a substantial excess, as the average circle with a radius of 5 km only has 0.3 type 1 stations.<sup>18</sup>

#### 5.4. Step 3: ranking clusters

In this step we determine the most suspicious cluster. To do so, we first partition our type 1 stations into clusters. Table 1 gives all clusters with more than two type 1 stations, listing the coordinates of the midpoint of the cluster, the numbers of type 1 and type 0 stations it contains, and the resulting  $p$ -value. The midpoint is given in the RD coordinate system (*Rijksdriehoekskoördinaten*) commonly used in the Netherlands.<sup>19</sup> For ease of reference, we have also included for each cluster the city closest to it.

A cluster of suspicious stations may point to the presence of 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 stations 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.

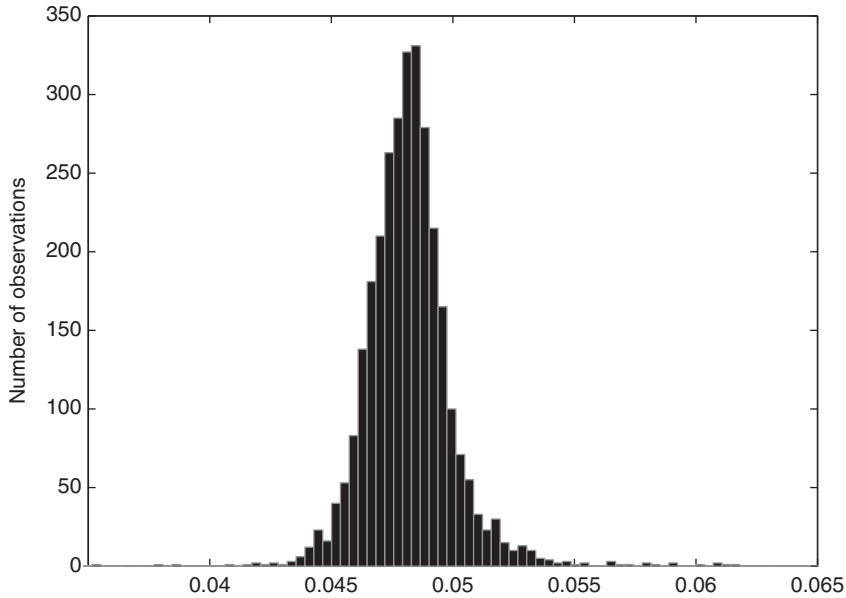
The most suspicious cluster turns out to be an area around the city of Hoogeveen (in the northeast of the country) that consists of 5 type 1 and no type 0 stations, yielding a  $-\log(p)$ -value of 6.5. The values of the HHI for this cluster do not indicate that this is due to high market concentration. Figure 4 shows a map of the Netherlands, with all type 1 (red pluses) and type 0 (gray dots) stations. The suspicious clusters are the convex hulls demarcated by blue lines. The coordinates on the axes correspond to the RD coordinates we also use in Table 1.

#### 5.5. Step 4: iterative elimination of clusters

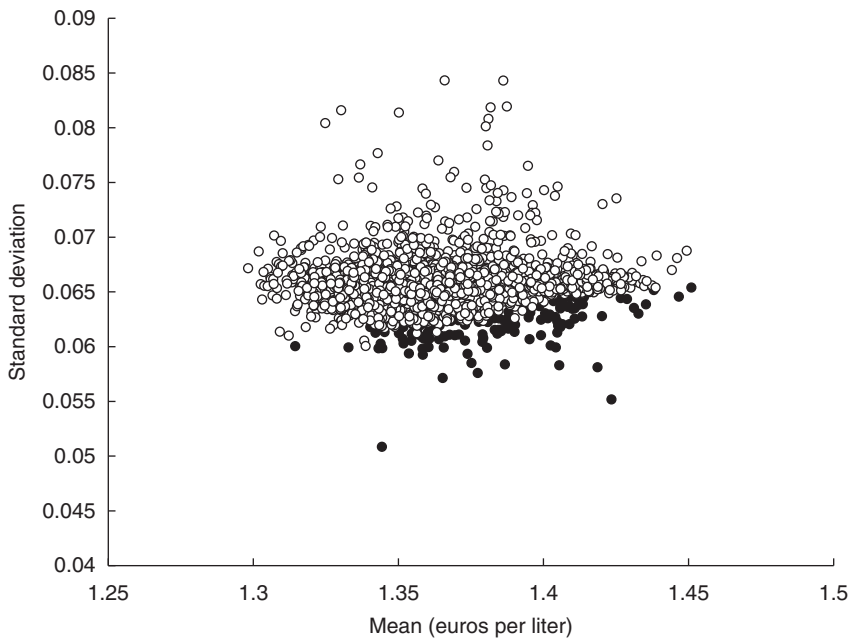
After eliminating this most suspicious cluster, we move back to step 3 to test whether there is evidence for local clustering in the remaining data. Both the number of type 0

18 We have 153 type 1 stations, the size of the Netherlands is roughly 40,000 km<sup>2</sup>. That yields one type 1 station per 261 km<sup>2</sup>; a circle with radius 5 km has an area of 78 km<sup>2</sup>.

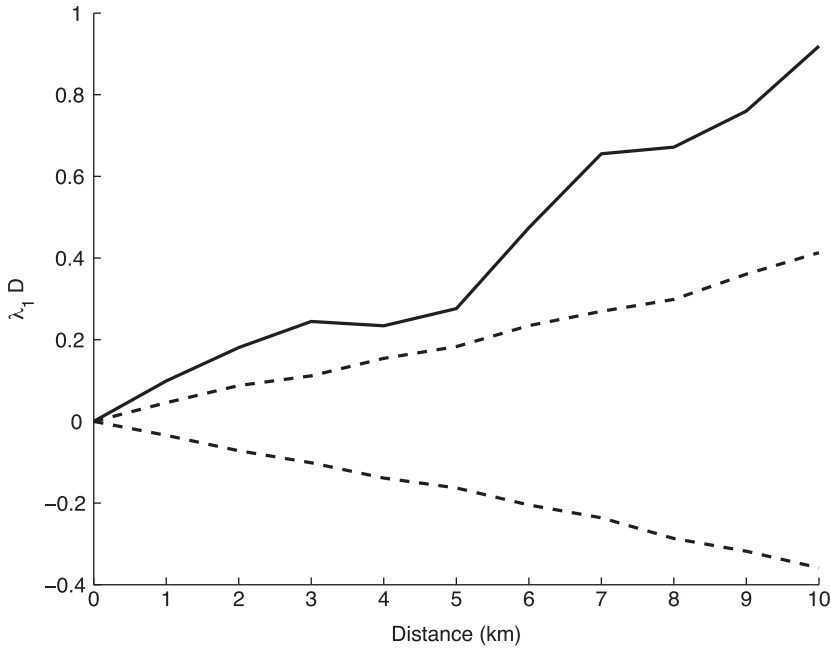
19 In the RD coordinate system, a location with coordinates  $(x,y)$  is situated  $x$  kilometers to the East, and  $y$  kilometers to the North of a fictional origin some 120 km to the southeast of Paris. The main advantage of the system is that coordinates thus represent kilometers, making it much easier to interpret than e.g. latitudes and longitudes. The origin is chosen such that all has points within the Netherlands have coordinates that are strictly positive.



**Figure 1.** Histogram of the variation coefficient. Variation coefficient for the 3035 nonhighway stations (sample: October 2005–June 2007), area fixed effects, adjusted for station characteristics.



**Figure 2.** Relation between mean price and standard deviation. Sample: nonhighway stations, October 2005–June 2007, area fixed effects, adjusted for station characteristics. Mean price (in euro per liter) on the horizontal axis, standard deviation on the vertical axis. Stations with a variation coefficient in the lowest 5% in black.



**Figure 3.** *D*-function. Sample: nonhighway stations, October 2005–June 2007, area fixed effects, adjusted for station characteristics. 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%.

**Table 1.** Clusters in the data

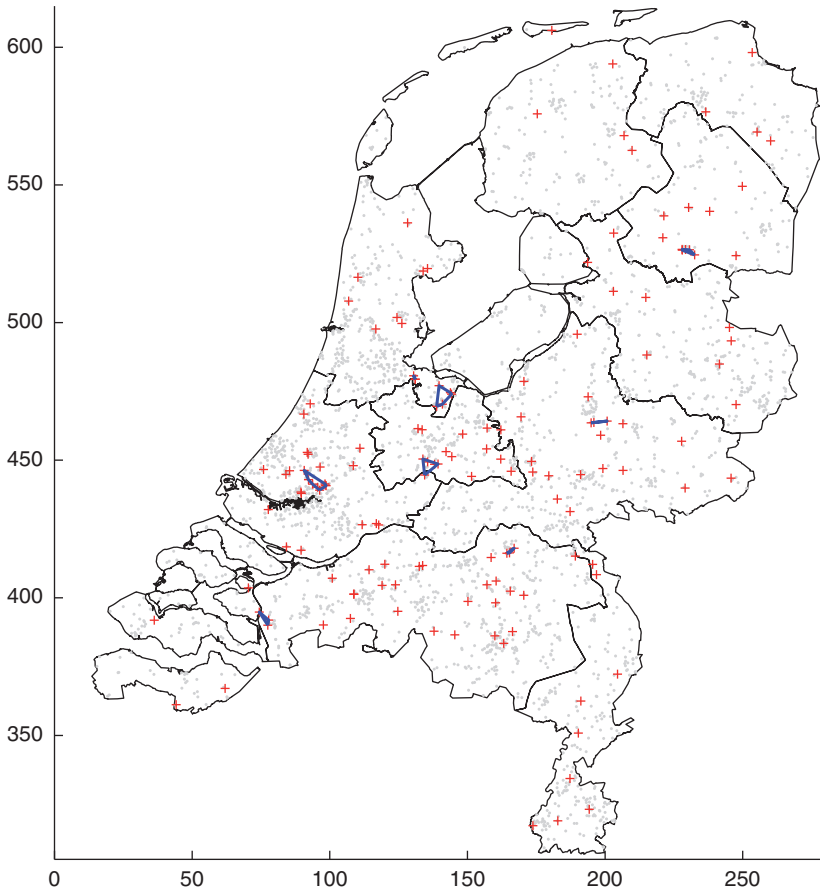
Cluster	Midpoint	Type 1	Type 0	$-\log p$	HHIT	HHIS	Nearest city
$S_1$	(230,526)	5	0	6.5	0.200	0.200	Hoogeveen
$S_2$	(95,442)	8	10	6.0	0.198	0.281	Rotterdam
$S_3$	(136,448)	5	4	4.5	0.210	0.360	Nieuwegein
$S_4$	(166,417)	3	0	3.9	0.556	0.556	Oss
$S_5$	(197,464)	3	0	3.9	0.556	0.556	Apeldoorn
$S_6$	(131,480)	3	0	3.9	0.556	0.556	Weesp
$S_7$	(77,392)	3	1	3.3	0.375	0.556	Bergen op Zoom
$S_8$	(141,473)	5	10	3.2	0.138	0.280	Hilversum

All clusters with more than two type 1 stations. Sample: nonhighway stations, October 2005–June 2007, area fixed effects, adjusted for station characteristics. In all, 5% of stations classified as type 1, cluster distance 5 km. Midpoint of the cluster in RD coordinates (see fn. 19). Type 1 gives the number of stations in the cluster that are classified as suspicious; type 0 the number that are not;  $p$  is the  $p$ -value of the cluster, derived using (1); HHIT is the Herfindahl index for the entire cluster; HHIS that for the subset of suspicious stations within the cluster.

and type 1 stations  $n_0$  and  $n_1$  have now decreased, which has to be taken into account when deriving the new *D*-function.

Figure 5 shows the resulting *D*-function after the elimination of the most suspicious cluster. The function is now no longer significantly different from 0 at  $h = 5$  kilometer.





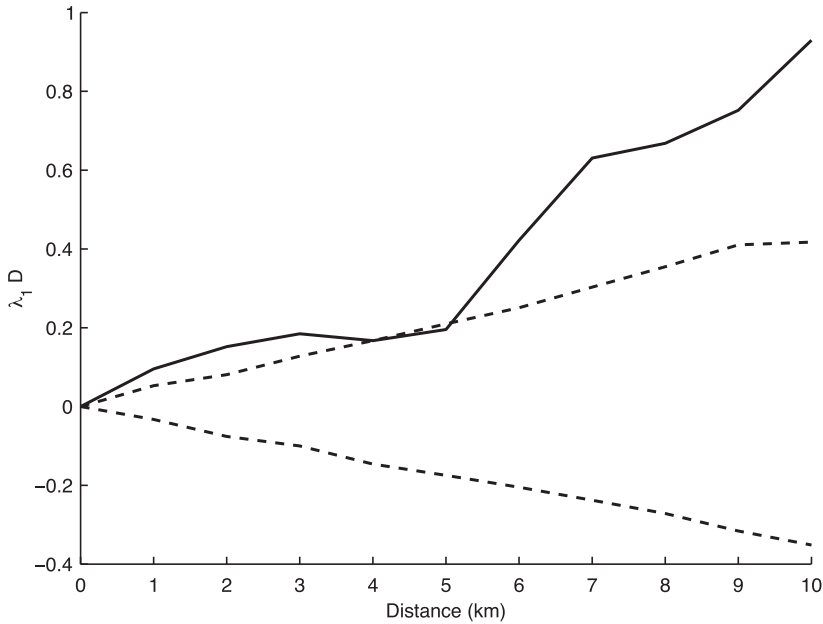
**Figure 4.** Type 1 stations. Map of the Netherlands. Black lines indicate province or country boundaries. Pluses are type 1 stations, gray dots type 0 stations. Polygons reflect convex hulls of clusters of suspicious stations. Sample: nonhighway stations, October 2005–June 2007, area fixed effects, adjusted for station characteristics. In all, 5% of stations classified as type 1, cluster distance 5 km.

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 removing the most suspicious cluster at that distance.

For future reference, the output of our collusion screen in terms of suspicious clusters that are identified, consists of the first line of Table 1. Based on this, the advice to an antitrust authority would be to have a closer look at the area around Hoozeveen. Of course, this in no way 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



**Figure 5.**  $D$ -function after removal of first suspicious cluster. The solid line is the  $D$ -function, 95% confidence interval is indicated by the dashed lines. Sample: nonhighway stations, October 2005–June 2007, area fixed effects, adjusted for station characteristics. In all, 5% of stations classified as type 1, cluster distance 5 km.

on local clustering at 5 km, and choose one particular method for identifying the most suspicious cluster. We used data from 2005 to 2007, rather than focusing on a narrower, wider or different time period. Finally, we chose to adjust prices for station characteristics and area fixed effects, rather than to look at listed prices. 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, we advise not to fully rely on one particular set of choices, but to consider some other choices as well. Moreover, doing so would also make this variance screen harder to beat for a cartel.

### 6.1. An alternative cluster distance

In our baseline, we looked for evidence for local clustering at a distance of 5 km. In this section, we vary this distance by looking at distances of 3 and 7 km, 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) km, while in step 4 we will look at clusters of stations that are at least 3 (7) km from each other.

Figure 3 shows that there is statistical evidence for local clustering at both 3 and 7 km. Table 2 gives the list of suspicious clusters that are generated at a distance of 3 km. Our method now identifies two suspicious clusters, rather than just one, as was the case in our baseline analysis. A number of observations stand out. First, the most

**Table 2.** Identified suspicious clusters at  $h = 3$  km

No.	Midpoint	Type 1	Type 0	$-\log p$	HHIT	HHIS	Nearest city
1	(230,526)	5	0	6.5	0.200	0.200	Hoogeveen
2	(96,441)	7	8	5.5	0.209	0.306	Rotterdam

Sample: nonhighway stations, October 2005–June 2007, area fixed effects, adjusted for station characteristics. In all, 5% classified as type 1. Midpoint of the cluster in RD coordinates (see fn. 19). Type 1 gives the number of stations in the cluster that are classified as suspicious; type 0 the number that are not;  $p$  is the  $p$ -value of the cluster, derived using (1); HHIT is the Herfindahl index for the entire cluster; HHIS that for the subset of suspicious stations within the cluster.

**Table 3.** Identified suspicious clusters at  $h = 7$  km

No.	Midpoint	Type 1	Type 0	$-\log p$	HHIT	HHIS	Nearest city
1	(93,444)	15	32	8.6	0.123	0.218	Rotterdam
2	(230,526)	5	0	6.5	0.200	0.200	Hoogeveen

Sample: nonhighway stations, October 2005–June 2007, area fixed effects, adjusted for station characteristics, 5% classified as type 1. Midpoint of the cluster in RD coordinates (see fn. 19). Type 1 gives the number of stations in the cluster that are classified as suspicious; type 0 the number that are not;  $p$  is the  $p$ -value of the cluster, derived using (1); HHIT is the Herfindahl index for the entire cluster; HHIS that for the subset of suspicious stations within the cluster.

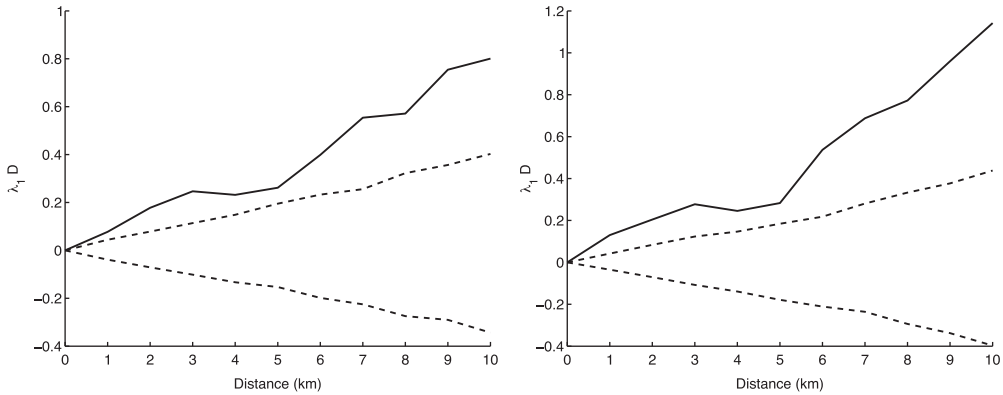
suspicious cluster is the same as that in our baseline case: close to Hoogeveen. Second, the second most suspicious cluster (Rotterdam) was also the second most suspicious in our baseline, as can be seen from Table 1. Yet, in our baseline this cluster was not flagged, as the elimination of the first cluster already yielded lack of statistical evidence for further clustering. That is no longer the case here. By construction, using a distance of 3 km is likely to generate smaller clusters. In our example, Rotterdam is a case in point: the number of stations in the cluster has decreased by 3. This suggests that in the implementation of our screen, it is important not to look at distances that are too small.

Table 3 lists the clusters that are found when looking at a distance of 7 km. The area close to Rotterdam now yields the most suspicious cluster. This cluster is now larger than in the baseline, and also has a higher share of type 0 stations. The second most suspicious cluster is Hoogeveen, which has the same size as in the baseline.

Summing up, choosing the right value of  $h$  implies a trade-off 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. In our application, however, changing the value of  $h$  does not yield substantially different results.

## 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



**Figure 6.** *D*-function. The solid line is the *D*-function, 95% confidence interval is indicated by the dashed lines. Sample: nonhighway stations, October 2005–June 2007, area fixed effects, adjusted for station characteristics. Cluster distance 5 km. In all, 4% (left panel) and 6% (right panel) of stations classified as type 1.

**Table 4.** Identified suspicious clusters, 4% classified as type 1

No.	Midpoint	Type 1	Type 0	$-\log p$	HHIT	HHIS	Nearest city
1	(230,526)	5	0	6.5	0.200	0.200	Hoogeveen

Sample: nonhighway stations, October 2005–June 2007, area fixed effects, adjusted for station characteristics Cluster distance 5 km. Midpoint of the cluster in RD coordinates (see fn. 19). Type 1 gives the number of stations in the cluster that are classified as suspicious; type 0 the number that are not;  $p$  is the  $p$ -value of the cluster, derived using (1); HHIT is the Herfindahl index for the entire cluster; HHIS that for the subset of suspicious stations within the cluster.

4% and the lowest 6% respectively. This may seem a slight change in the number of type 1 stations, but it does imply a change of 20% in the number of type 1 stations that we consider.

In both cases, from Figure 6, we again find statistical evidence for local clustering at 5 km. Table 4 shows that the most suspicious cluster with 4% is again the same cluster in Hoogeveen, while from Table 5 we see that with 6% the most suspicious cluster is the one near Rotterdam, although this cluster is again somewhat different from the Rotterdam cluster we find in the baseline or with  $h = 7$  km.

### 6.3. Accounting for concentration measures

In Section 5.2, we considered the possibility to also adjust prices for local concentration measures. We perform that analysis in this subsection.

Table 6 shows that, similar to the results of some of the previous robustness checks, Rotterdam emerges next to Hoogeveen as a suspicious cluster. The cluster in Hoogeveen is identical to the one identified in previous analyses, while the cluster in Rotterdam again has a somewhat different composition. Since collusion is deemed more likely in markets with higher market concentration, we raised the caveat in Section 5.2

**Table 5.** Identified suspicious clusters, 6% classified as type 1

No.	Midpoint	Type 1	Type 0	$-\log p$	HHI	HHI <sup>1</sup>	Nearest city
1	(94,441)	11	17	6.6	0.145	0.289	Rotterdam

Sample: nonhighway stations, October 2005–June 2007, area fixed effects, adjusted for station characteristics. Cluster distance 5 km. Midpoint of the cluster in RD coordinates (see fn. 19). Type 1 gives the number of stations in the cluster that are classified as suspicious; type 0 the number that are not;  $p$  is the  $p$ -value of the cluster, derived using (1); HHIT is the Herfindahl index for the entire cluster; HHIS that for the subset of suspicious stations within the cluster.

**Table 6.** Identified suspicious clusters, prices adjusted for concentration measures

No.	Midpoint	Type 1	Type 0	$-\log p$	HHIT	HHIS	Nearest city
1	(230,526)	5	0	6.5	0.200	0.200	Hoogeveen
2	(96,442)	8	10	6.1	0.198	0.281	Rotterdam

Sample: nonhighway stations, October 2005–June 2007, area fixed effects, adjusted for station characteristics and concentration measures. In all, 5% of stations classified as type 1, cluster distance 5 km. Midpoint of the cluster in RD coordinates (see fn. 19). Type 1 gives the number of stations in the cluster that are classified as suspicious; type 0 the number that are not;  $p$  is the  $p$ -value of the cluster, derived using (1); HHIT is the Herfindahl index for the entire cluster; HHIS that for the subset of suspicious stations within the cluster.

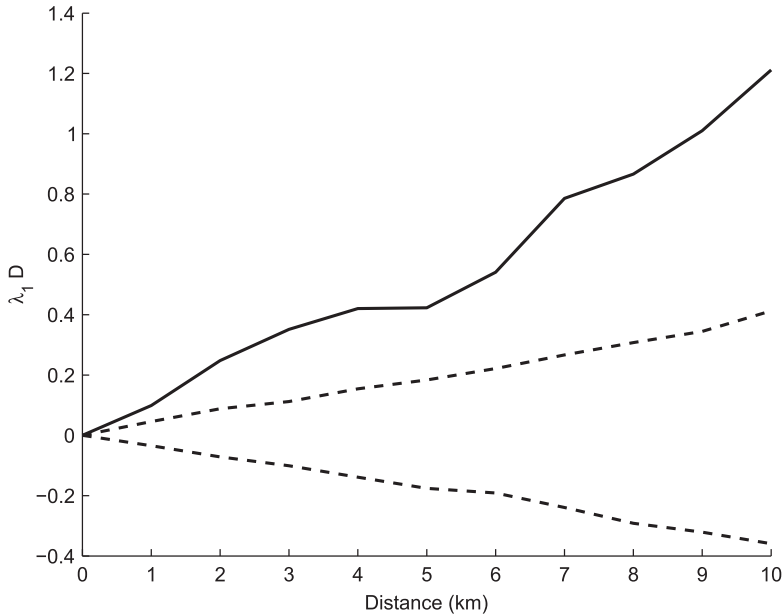
that including measures of market concentration might impact the identification of clusters. However, for our application to the Dutch gasoline market this impact seems limited as we find more or less the same clusters.

#### 6.4. An alternative time period

Next, we investigate how our method is affected when we consider a different time period. In Figure 8, the  $D$ -function is plotted based on price data for the period July 2007 to April 2009. We have ‘treated’ this data in the same manner as described in Section 5.2. Thus, we first follow the method proposed by AFGT to impute the missing data and then adjust the data for area fixed effects and station characteristics. Also for this period, 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 four clusters before there is lack of evidence for further clustering, see Table 7. Ede is flagged as the most suspicious cluster, although Rotterdam still makes it on the list.

One observation that stands out in Table 7 is the extremely high value of the Herfindahl index among type 1 stations in cluster 4. Out of 9 type 1 stations in this cluster, 7 carry the Texaco brand. Among the 21 type 0 stations, there is not a single Texaco station. Hence, rather than a local cartel, this cluster reflects the market dominance of Texaco in this particular area. Although high prices due to a high market concentration are just as detrimental to welfare, they are not illegal and hence leave no scope for an antitrust authority to step in.



**Figure 7.** *D*-function, prices adjusted for concentration measures. The solid line is the *D*-function, 95% confidence interval is indicated by the dashed lines. Sample: nonhighway stations, October 2005–June 2007, area fixed effects, adjusted for station characteristics and concentration measures. In all, 5% of station classified as type 1, cluster distance 5 km.

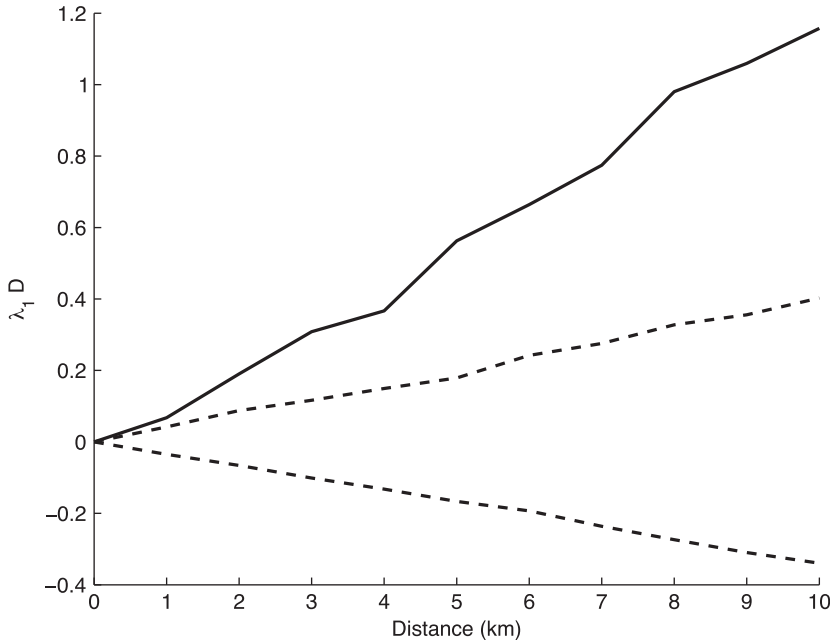
One explanation for the different picture that we see now is that the market environment may have changed substantially; areas that were a cartel in 2005–2007 may not be so anymore in 2007–2009. That is confirmed if we look at the robustness checks that we also did for the period 2005–2007. Changing the cartel distance, the fraction of type 1 stations, or the extent to which we adjust the data consistently yields four or five clusters are flagged as suspicious, with Den Haag being the most suspicious cluster most often, and with substantial overlap among the other clusters that are generated (in particular Ede and Rotterdam) as well.

### 6.5. Using raw data

In Appendix C, we run a number of robustness checks using the raw data, rather than the data adjusted for station characteristics. Rotterdam then consistently pops up as one of the most suspicious clusters, although the exact composition of the cluster varies. Most notably, Hoozeveen is no longer flagged in any analysis. Thus, controlling for station characteristics and area fixed effects is important in identifying Hoozeveen as the most suspicious cluster in our analysis.

### 6.6. Kulldorff's Spatial Scan Statistic

As a final robustness check, we apply Kulldorff's (1997) spatial scan statistic. For this purpose, we use SaTScan, a software package for this purpose that is freely available on the internet. We use SaTScan to look for circular clusters. The results are reported in Table 8.



**Figure 8.**  $D$ -function, July 2007–April 2009). The solid line is the  $D$ -function, 95% confidence interval is indicated by the dashed lines. Sample: July 2007–April 2009, area fixed effects, adjusted for station characteristics. In all, 5% of stations classified as type 1, cluster distance 5 km.

**Table 7.** Identified suspicious clusters, sample period July 2007–April 2009

Cluster	Midpoint	Type 1	Type 0	$-\log p$	HHIT	HHIS	Nearest city
1	(184,445)	5	0	6.5	0.360	0.360	Ede
2	(94,440)	11	28	5.8	0.123	0.201	Rotterdam
3	(137,451)	5	1	5.8	0.361	0.401	Nieuwegein
4	(79,453)	9	21	5.1	0.178	0.630	Den Haag

Sample: nonhighway stations, July 2007–April 2009. In all, 5% of stations classified as type 1, cluster distance 5 km. Midpoint of the cluster in RD coordinates (see fn. 19). Type 1 gives the number of stations in the cluster that are classified as suspicious; type 0 the number that are not;  $p$  is the  $p$ -value of the cluster, derived using (1); HHIT is the Herfindahl index for the entire cluster; HHIS that for the subset of suspicious stations within the cluster.

We first look at our raw data. In that case, SaTScan also identifies a suspicious cluster close to Rotterdam. Yet, with a radius of 8.03 km this cluster is larger and contains many more outlets, and especially more type 0 outlets, than the cluster generated by our own method. SaTScan does not report any other clusters that are significant at 5%.<sup>20</sup>

20 Note therefore that we use the SaTScan algorithm to test for secondary clusters, rather than our own. This algorithm simply entails looking for all clusters that have a  $p$ -value lower than 0.05.

**Table 8.** Using Kulldorff's (1997) spatial scan statistic

Analysis	Max km	No.	Midpoint	Radius	Type 1	Type 0	Nearest city
Raw data	$\infty$	1	(94,444)	8.03	16	51	Rotterdam
	7	1	(94,444)	5.73	11	22	Rotterdam
	5	1	(92,445)	4.55	7	11	Rotterdam
Two-digit postcode	$\infty$	1	(227,532)	10.26	8	3	Hoogeveen
	7	1	(201,464)	5.88	5	3	Apeldoorn
	5	1	(233,525)	4.96	5	4	Hoogeveen

Clusters identified by Kulldorff (1997) using SaTScan, looking for circular clusters. The first three analyses are with our baseline data, using an unrestricted circle size, imposing a maximum radius of 7 km, and a maximum radius of 5 km, respectively. The last three analyses are the same, but accounting for two-digit postcode area fixed effects. The column 'radius' gives the actual radius of the cluster found. In all cases, the reported cluster is the only one that has a  $p$ -value lower than 0.05, according to SaTScan. Midpoint of the cluster in RD coordinates (see fn. 19). Type 1 gives the number of stations in the cluster that are classified as suspicious; type 0 the number that are not;  $p$  is the  $p$ -value of the cluster, derived using (1); HHIT is the Herfindahl index for the entire cluster; HHIS that for the subset of suspicious stations within the cluster.

To find a cluster that matches the one we found in our baseline analysis as closely as possible, we restrict SaTScan to look for circles with a radius of at most 7 km. In that case, we again find a cluster close to Rotterdam. With 11 suspicious sites, it has the same number of such sites as our baseline estimate, but with 22 rather than 11, the number of nonsuspicious sites is again much higher. Also when we restrict clusters to a radius of at most 5 km, the number of nonsuspicious sites remains relatively high. Hence, by allowing the shape of clusters to vary, our method seems to lead to a more precise identification of a potential cartel, in the sense that it yields a much lower number of nonsuspicious sites within a suspicious cluster.

We redo the analysis for our data that allows for area fixed effects at the two-digit postcode level. This confirms what we observed above: the number of type 0 sites found by SaTScan is relatively high compared by the number found by our own method. If we restrict circle sizes to at most 7 km, a cluster around Apeldoorn turns up as the most suspicious. We did not find this particular cluster using our method. In all cases, SaTScan only reports 1 suspicious cluster.

Summing up, one main difference between our method of finding clusters and the method used in Kulldorff's (1997) spatial scan statistic is that the latter returns areas that are somewhat larger and, especially, contain more nonsuspicious stations. An antitrust authority would primarily be interested in the suspicious stations within a cluster. In this particular application, that seems an additional reason to prefer our method.

## 7. Conclusion

In this article, 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 Hoozeveen. In different variations of our method, this area, and one close to Rotterdam consistently emerges as a suspicious region. Naturally, this can never be construed as evidence for collusion, but it suggests that an antitrust authority with limited resources may have a closer look at the stations in that area. For the period 2007–2009, areas around Ede, Rotterdam and Den Haag turn up as most suspicious, depending on the exact method that is used. However, the cluster close to Den Haag is arguably due to high market concentration rather than possible collusion.

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 distance or the fraction of outlets that are classified as suspicious, may also influence its output, although as we saw in Section 6 that it is reasonably insensitive to such choices. Just like any other tool, our collusion screen should be applied with care. Its output serves as a useful starting point for a directed inquiry.

## Acknowledgements

Most of this work was done while Soetevent was affiliated to the University of Amsterdam. The comments of Romy Abrantes-Metz, Joe Harrington, Nick de Roos, two anonymous referees, the editor Kristian Behrens, and seminar participants at EARIE 2009, IIOC 2010, EEA 2011 and 2013 and the University of Sydney are gratefully acknowledged. The usual disclaimer applies.

## Funding

Soetevent's research is supported by the Netherlands Organization for Scientific Research under grant 451-07-010.

## References

- Abrantes-Metz, R., Bajari, P. (2009) Screens for conspiracies and their multiple applications. *The Antitrust Magazine*, 24: 66–71.
- Abrantes-Metz, R., Froeb, L., Geweke, J., Taylor, C. (2006) A variance screen for collusion. *International Journal of Industrial Organisation*, 24: 467–486.
- Athey, S., Bagwell, K., Sanchirico, C. (2004) Collusion and price rigidity. *Review of Economic Studies*, 71: 317–349.

- Carlino, G. A., Hunt, R. M., Carr, J. C., Smith, T. E. (2012) The agglomeration of R&D labs Working Paper 12-22, Philadelphia Fed.
- Costa, M. A., Kulldorff, M. (2009) Applications of spatial scan statistics: a review. In J. Glaz, V. Pozndyakov, S. Wallenstein (eds) *Scan Statistics: Methods and Applications*, pp. 129–152. Boston, MA: Birkhäuser.
- Cressie, N. (1991) *Statistics for Spatial Data*. New York: Wiley and Sons.
- Diggle, P., Chetwynd, A. (1991) Second-order analysis of spatial clustering for inhomogeneous populations. *Biometrics*, 47: 1155–1163.
- Dixon, P. (2002) Ripleys *K* function. In A. H. El-Shaarawi, W. W. Piegorsch (eds) *Encyclopedia of Environmetrics*, vol. vol. 3, pp. 1796–1803. Chichester: John Wiley & Sons.
- Duczmal, L., Ribeiro Duarte, A., Tavares, R. (2009) Extensions of the scan statistic for the detection and inference of spatial clusters. In J. Glaz, V. Pozndyakov, S. Wallenstein (eds) *Scan Statistics: Methods and Applications*, pp. 153–177. Boston, MA: Birkhäuser.
- Durantou, G., Overman, H. (2005) Testing for localisation using micro-geographic data. *Review of Economic Studies*, 72: 1077–1106.
- Ellison, G., Glaeser, E. L. (1997) Geographic concentration in US manufacturing industries: a dartboard approach. *Journal of Political Economy*, 105: 889–927.
- European Union (1999) Merger Procedure Case No IV/M 1383 - Exxon/Mobil, Regulation (EEC) No 4064/89L.
- Faber, R. (2011) More new evidence on asymmetric gasoline price responses. Mimeo.
- Froeb, L., Cooper, J., Frankena, M., Pautler, P., Silvia, L. (2005) Economics at the FTC: cases and research, with a focus on petroleum. *Review of Industrial Organization*, 27: 223–252.
- Getis, A., Franklin, J. (1987) Second-order neighborhood analysis of mapped point patterns. *Ecology*, 68: 473–477.
- 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 P. Buccrossi (ed.) *Handbook in Antitrust Economics*, chap. 6, pp. 213–258. Cambridge, MA: MIT Press.
- Jimenez, J. L., Perdiguero, J. (2012) Does rigidity of prices hide collusion? *Review of Industrial Organization*, 41: 223–248.
- Kulldorff, M. (1997) A spatial scan statistic. *Communications in Statistics: Theory and Methods*, 26: 1481–1496.
- Kulldorff, M. (1999) Spatial scan statistics: models, calculations and applications. In N. Balakrishnan, J. Glaz (eds) *Recent Advances on Scan Statistics and Applications*, pp. 303–322. Boston, MA: Birkhäuser.
- Kulldorff, M. (2010) *SaTScan User Guide for version 9.0*. Available online at: <http://www.satscan.org>. [Accessed 14 April 2014].
- Marcon, E., Puech, F. (2010) Measures of the geographic concentration of industries: improving distance-based methods. *Journal of Economic Geography*, 10: 745–762.
- Marcon, E., Puech, F., Traissac, S. (2012) Characterizing the relative spatial structure of point patterns. *International Journal of Ecology*, 1–11.
- Maskin, E., Tirole, J. (1988) A theory of dynamic oligopoly II: price competition, kinked demand curves, and edgeworth cycles. *Econometrica*, 56: 571–599.
- Mori, T., Smith, T. E. (2014) A probabilistic modeling approach to the detection of industrial agglomerations. *Journal of Economic Geography*, 14: 547–588.
- Naus, J. (1965) Clustering of random points in two dimensions. *Biometrika*, 52: 263–267.
- NMa (2006) Benzinescan 2005/2006 Discussion paper.
- Openshaw, S., Craft, A., Charlton, M., Birch, J. (1988) Investigation of leukaemia clusters by use of a geographical analysis machine. *Lancet*, 331: 1533–1575.
- Pepall, L., Richards, D., Normann, G. (2008) *Industrial Organization: Contemporary Theory and Empirical Applications*, 4th edn. Malden, MA: Blackwell Publishing.
- Picone, G. A., Ridley, D. B., Zandbergen, P. A. (2009) Distance decreases with differentiation: strategic agglomeration by retailers. *International Journal of Industrial Organization*, 27: 463–473.
- Ripley, B. (1976) The second-order analysis of stationary point processes. *Journal of Applied Probability*, 13: 255–266.
- Rysman, M., Greenstein, S. (2005) Testing for agglomeration and dispersion. *Economics Letters*, 86: 405–411.

- Smith, T. E., Smith, M. M., Wackes, J. (2008) Alternative financial service providers and the spatial void hypothesis. *Regional Science and Urban Economics*, 66: 274–279.
- Soetevent, A., Haan, M., Heijnen, P. (2014) Do auctions and forced divestitures increase competition? Evidence for retail gasoline markets. *Journal of Industrial Economics*, 62: 467–502.
- Stoyan, D., Penttinen, A. (2000) Recent applications of point process methods in forestry statistics. *Statistical Science*, 51: 61–78.
- The Economist, (2012) The scam busters - How antitrust economists are getting better at spotting cartels.
- Turnbull, B., Iwano, E., Burnett, W., Howe, H., Clark, L. (1990) Monitoring for clusters of disease: application to leukemia incidence in upstate New York. *American Journal of Epidemiology*, 132: S136–S143.
- Waller, L. A. (2009) Detection of clustering in spatial data. In A. S. Fotheringham, P. A. Rogerson (eds) *The SAGE Handbook of Spatial Analysis*, pp. 299–320. London: SAGE Publications.
- Wang, Z. (2009) Mixed strategy in oligopoly pricing: evidence from gasoline price cycles before and under a timing regulation. *Journal of Political Economy*, 117: 987–1030.

## Appendix

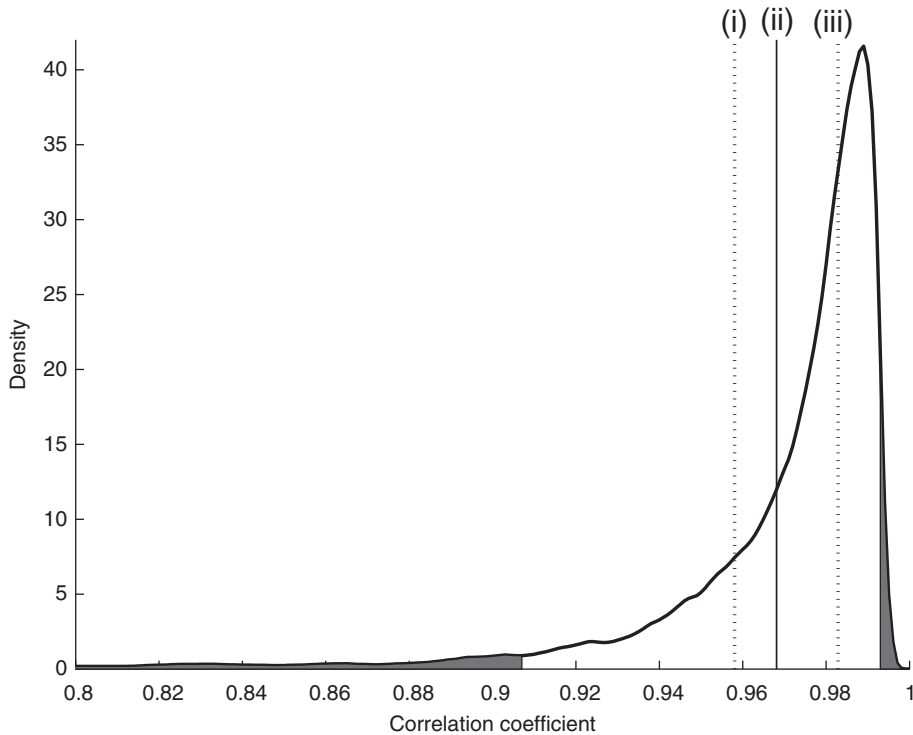
### A. Correlation between prices

To study the extent to which suspect clusters have highly correlated prices, we focus on the two most prominent suspect clusters in our data set: the Hooegeveen-cluster (Table 1) and the Rotterdam-cluster (Table B2, first line). For each cluster, we compute the price correlation between the suspected stations in this cluster, the correlation between the suspect stations and the nonsuspect stations, and the correlation between nonsuspect stations (using the raw data for all comparisons). Then we compute averages of the correlation coefficients. See Table A1 for the results.

Note that the price correlation is high with the average price correlation in the sample being close to 97%. Hooegeveen is slightly above the average and Rotterdam is slightly below. To get an idea of how the averages for the Hooegeveen-cluster and the Rotterdam-cluster compare to the distribution of the correlation coefficients, Figure A1 shows a kernel density estimate of the full distribution of pairwise correlation coefficients for all station pairs. Note that for both Rotterdam and Hooegeveen the averages stay way clear from the tails of the distribution. Therefore, there is no indication that within our suspected cartels prices are correlated more strongly than between nonsuspect firms.

**Table A1.** Average correlation between stations

Hooegeveen	
Suspect–suspect	0.9829
Nonsuspect–nonsuspect	0.9682
Suspect–nonsuspect	0.9589
Rotterdam	
Suspect–suspect	0.9581
Nonsuspect–nonsuspect	0.9681
Suspect–nonsuspect	0.9667



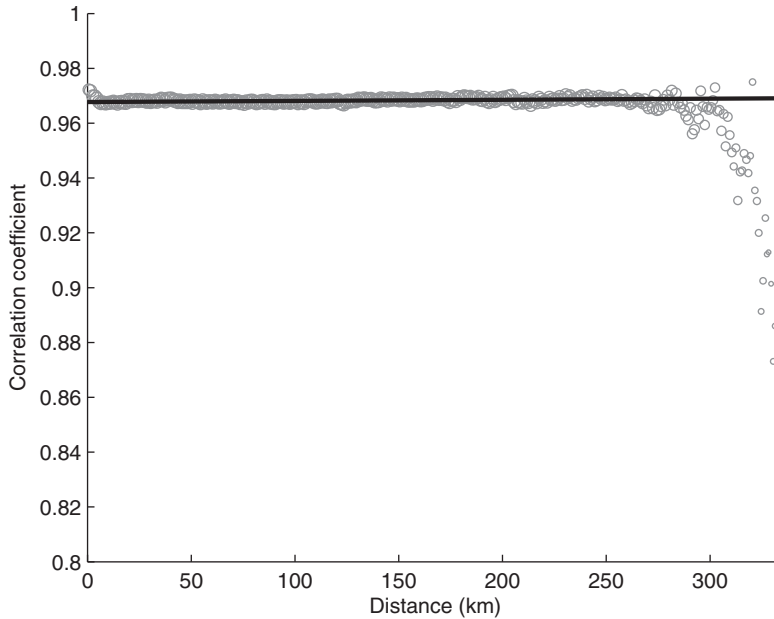
**Figure A1.** Distribution of the pairwise correlation coefficients of all station pairs. The thick black line is a kernel estimate of the distribution of the correlation coefficients, the grey area are the 5% lowest and the 5% highest correlation coefficients, line (i) is the average correlation in the Rotterdam-cluster, line (ii) is the average correlation in the full sample and line (iii) is the average correlation in the Hooerveen cluster.

In Figure B1, we have plotted the relationship between the distance between any two stations on the one hand, and their price correlation on the other. The black line is the regression line, the grey circles represent the underlying data. For that purpose, we have put all correlations in bins of 1 km and calculated the average for each bin. The larger the circle, the larger the number of observations in that bin.

The relation is flat, except for very large distances where one of the two stations of the pair is influenced either by a foreign market because it is located close to the German or Belgian border, or by being relatively isolated as it is located on one of the Dutch Wadden islands. For very short distances, the relation displays a negative slope, but this may well be due to area-specific co-movements of unobservables. This indeed is an additional justification for using ‘adjusted prices’ in the analysis.

## B. Data

In this Appendix we describe our data (Appendix B.1), how we dealt with missing data (Appendix B.2), our choice of stations to include (Appendix B.3), and how we adjust for station characteristics (Appendix B.4).



**Figure B1.** Relation between distance between stations and their price correlation. Black line is regression line, gray circles represent observations. Each gray circle gives the average of all correlations between stations at a similar distance, the size of the circle reflects the number of stations it represents.

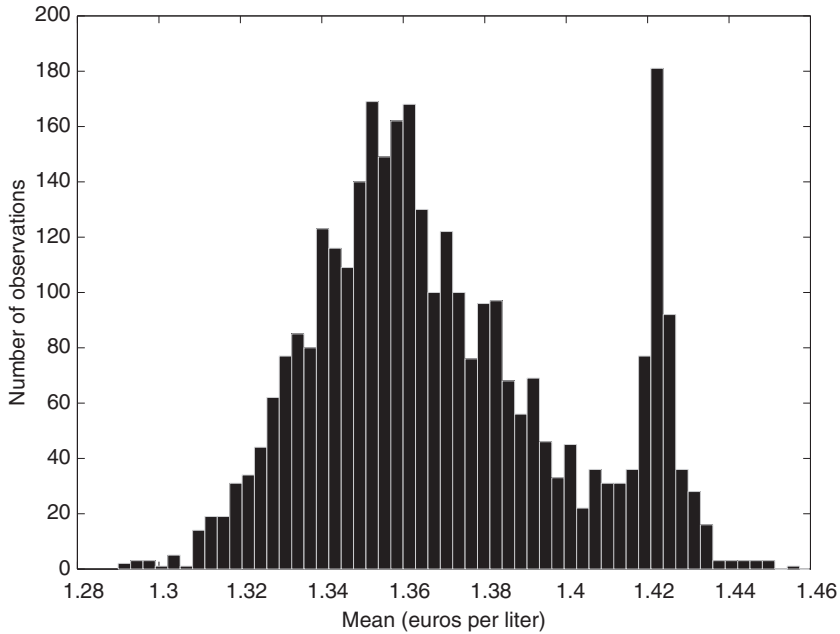
### B.1 Data description

We use a fleet card data set which contains regular price quotes for 3259 gasoline retail outlets in the Netherlands. For comparison, the Dutch competition authority NMa (2006a, pg. 8) cites a total number of 3625 outlets in the Netherlands in 2004. An estimate of Bovag (the Dutch industry association for the automotive sector) mentions 4319 outlets in 2005.

Price data were downloaded on a daily basis from the web site of Athlon, the largest independent car leasing company in the Netherlands with a fleet of over 125,000 cars. One of the main advantages of this type of fleet card data is that it is *not* likely to be biased. All our data come from drivers of lease cars that use the car on behalf of their employer and hence do not have to pay for their own gasoline bill. Therefore, they have little incentive to seek out the cheapest provider of gasoline, and the probability that a station is sampled on any particular day should be uncorrelated with its price.

### B.2 Data imputation

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 any given day, we observe a price quote for on average 37.5% of all stations. 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



**Figure B2.** Histogram of average price, (station level). Average prices for 3259 stations (sample period: October 2005–June 2007).

problems, we follow the method proposed AFGT to impute the missing data. 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.<sup>21</sup> Missing prices can now be replaced by a draw from the posterior distribution. We then proceed with the analysis using the imputed data.

### **B.3 Choice of stations**

Figure B2 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.<sup>22</sup> In our analysis, we therefore exclude the 224 highway stations and limit attention to the 3035 remaining nonhighway stations.

### **B.4 Adjusting for station characteristics**

As station characteristics, we include the number of pumps; the plot size; the size of shop area, and dummies for being close to the German or Belgian border, being

21 Full details can be found in AFGT, pp. 475–478.

22 See e.g. European Union, 1999, where it is argued in the Exxon/Mobil case that ‘in some countries, it is possible to consider fuel retailing on motorways as a separate product market’ (point 436).

**Table B1.** Regression of average price on explanatory variables including two-digit zip-code fixed effects (sample: nonhighway stations; October 2005–June 2007)

Local competition measures	(1)		(2)	
	Excluded		Included	
Sample mean	Coefficient	S.E.	Coefficient	S.E.
	1.3613		1.3613	
Geographical characteristics				
German border	-0.0031	(0.0029)	-0.0064*	(0.0029)
Belgian border	0.0032	(0.0035)	0.0000	(0.0035)
Site characteristics				
Company owned	-0.0142**	(0.0010)	-0.0126**	(0.0010)
Major brand	0.0092**	(0.0010)	0.0096**	(0.0010)
No. of pumps	-0.0002	(0.0004)	0.0000	(0.0004)
Express	-0.0041**	(0.0014)	-0.0045**	(0.0014)
Hot drinks	0.0042**	(0.0014)	0.0040**	(0.0014)
Car wash	0.0012	(0.0010)	0.0014	(0.0010)
Plot size (area)	-0.7084	(0.3613)	-0.7605*	(0.3570)
shop area	44.2255 <sup>†</sup>	(21.8997)	47.2853*	(21.6376)
Local demand				
No. of priv. owned cars $\leq 20$ km	-0.0029 <sup>†</sup>	(0.0018)	0.0009	(0.0022)
Local market concentration				
<i>ln(# non-highway stations+1) at...</i>				
$\leq 1$ km			-0.0034**	(0.0008)
1 – 2 km			-0.0048**	(0.0074)
2 – 5 km			-0.0007	(0.0008)
5 – 10 km			-0.0023 <sup>†</sup>	(0.0014)
<i>ln(# highway stations+1) at...</i>				
$\leq 1$ km			0.0076	(0.0050)
1 – 2 km			0.0025	(0.0022)
2 – 5 km			0.0018	(0.0011)
5 – 10 km			0.0001	(0.0010)
Zip-code fixed effects		Yes		Yes
R <sup>2</sup>	0.2826		0.3031	
Obs.	3035		3035	

Plot size area and shop area in square kilometer; privately owned cars in '000.000.

<sup>†</sup>Significant at the 10% level; \*significant at the 5% level; \*\*significant at the 1% level.

company owned, carrying one of the four major brands,<sup>23</sup> serving hot drinks, having a car wash and being fully automated ('express').<sup>24</sup> We also include the log of the numbers of cars owned by private households within 20 km of the station as a measure of local demand.<sup>25</sup> Inclusion of these variables is motivated by Soetevent, Haan and Heijnen (2014), where we find that these indeed affect gasoline prices.

23 Esso, Shell, Texaco and BP.

24 Data on the characteristics of each gasoline station were obtained from Experian Catalist Ltd.

25 These data are available for over 98% of all stations in our sample.

**Table B2.** Using raw data

Analysis	No.	Midpoint	Type 1	Type 0	$-\log p$	HHIT	HHIS	Nearest city
Baseline	1	(95,443)	11	13	8.3	0.163	0.174	Rotterdam
$h = 3$ km	1	(95,442)	9	11	6.8	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
$h = 7$ km	1	(93,444)	16	31	9.6	0.112	0.180	Rotterdam
	2	(163,406)	12	17	8.4	0.119	0.190	Veghel
4% type 1	1	(137,449)	4	0	5.2	0.375	0.375	Nieuwegein
6% type 1	1	(95,443)	11	13	8.3	0.163	0.174	Rotterdam
Site chars	1	(91,443)	15	25	9.7	0.135	0.173	Rotterdam
Site + conc	1	(91,443)	15	25	9.7	0.135	0.173	Rotterdam
2007–2009	1	(160,377)	9	12	6.6	0.152	0.259	Eindhoven
	2	(103,490)	5	1	5.8	0.222	0.280	Haarlem
	3	(80,453)	10	25	5.4	0.171	0.820	Den Haag
	4	(94,440)	10	29	4.9	0.120	0.180	Rotterdam
	5	(92,464)	4	1	4.6	0.440	0.625	Leiden

Using data uncorrected for station characteristics and area fixed effects, in respectively the baseline case (nonhighway stations, 5% classified as type 1, cluster distance 3 km, sample period October 2005–June 2007); when looking for clustering at 3 km; when looking at clustering at 7 km; with 4% of outlets classified as suspicious; with 6% of outlets classified as suspicious; when controlling for site characteristics; when controlling for site characteristics and concentration measures; when using sample period July 2007–April 2009. Midpoint of the cluster in RD coordinates (see fn. 19). Type 1 gives the number of stations in the cluster that are classified as suspicious; type 0 the number that are not;  $p$  is the  $p$ -value of the cluster, derived using (1); HHIT is the Herfindahl index for the entire cluster; HHIS that for the subset of suspicious stations within the cluster.

We also adjust prices for two-digit postcodes. Each address in the Netherlands has a postcode that consists of four digits and two letters (e.g. 9743 BE). The areas where these postcodes are located are nested: for example, the area with addresses that have a four-digit postcode of the form 9743 xx is a subset of the area with addresses that have a three-digit postcode of the form 974x xx, which in turn is a subset of the area with addresses that have a two-digit postcode of the form 97xx xx, etcetera). The Netherlands consists of 90 two-digit postcode areas, that have an average size of 388 km<sup>2</sup>, and an average of 181,425 inhabitants.<sup>26</sup>

The estimates in column (1) of Table B1 show that *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 1.4% lower. Prices at fully automated stations are 0.4% lower on average. Column (2) also includes local concentration measures. The estimates show that the presence of other nonhighway stations within 2 km 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.

26 These numbers apply to 2011. *Source*: Statistics Netherlands.



### **C. Further robustness checks: using raw data**

In this section, we look at our raw data, that is, the data that are not adjusted for area fixed effects or station characteristics. On the basis of those prices, we determine which stations are type 1, and which are type 0. On the basis of that classification we determine whether there is evidence for clustering and, if so, where the most suspicious cluster is located. We first perform this robustness check for our baseline analysis; for the period October 2005–June 2007, with 5% of stations labeled as suspicious, and a cluster distance of 5 km. Rotterdam is now the most suspicious cluster.

We also do the other robustness checks we performed above. The results are reported in Table B2. When we look at a distance of 3 km., Rotterdam is again the most suspicious, but three other clusters are also flagged. In the other robustness checks, Rotterdam consistently pops up as one of the most suspicious clusters, although the exact composition of the cluster varies. Most notably, Hoogeveen is no longer flagged in any analysis. Thus, controlling for station characteristics and area fixed effects is important in identifying Hoogeveen as the most suspicious cluster in our analysis. Conditional on geographical, zip code and site characteristics, the Hoogeveen cluster is identified as a suspicious cluster, whereas the area does not stand out when these factors are not taken into consideration. In 2007–2009, Eindhoven is now the most suspicious cluster, followed by Haarlem. The third most suspicious cluster, Den Haag, again has a high level of local concentration among suspicious stations.