Dual Kernel Density Estimation as a Method for Describing Spatio-Temporal Changes in the Upper Austrian Food Retailing Market

Eva Maria Jansenberger and Petra Staufer-Steinnocher
Department of Economic Geography & Geoinformatics
Vienna University of Economics and Business Administration
Rossauer Lände 23/1
1090 Vienna, Austria

SUMMARY

This article discusses spatio-temporal variations in concentration processes and market dominance in the Upper Austrian food retailing market between 1998 and 2001 using the dual kernel density estimation approach. Dual kernel density estimation seems to be an appropriate instrument for the analysis since this statistical method together with the use of a geographic information system (GIS) not only allows the visualisation of concentration processes and market dominance but also enables the description of small-scale local changes. This is of special interest in the chosen study region since Upper Austria, although consisting mainly of rural areas also comprises urban regions. The result of the study shows an overall increase in market concentration and market dominance in favour of the two market leaders Rewe and Spar. This becomes especially apparent in the urban areas. In rural areas however, it shows that in certain districts the development was in favour of the group of the other food retailers.

KEYWORDS: *Dual Kernel Density Estimation, Spatio-Temporal Analysis, Market Concentration, Market Dominance*

INTRODUCTION

This contribution focuses on spatio-temporal changes of spatial point patterns. For the spatial quantification of such changes the method of kernel density estimation (KDE) is used. KDE is an interpolation technique in spatial point pattern analysis, providing density estimates at any location within a defined study area (Silverman, 1986; Bailey, 1995; Burt, 1996). This research is focused on dual KDE, for which two different data sets are related to each other by arithmetic operations. In order to illustrate this technique examplarily, the changes in spatial concentration of grocery outlets of two groups of retailers in the Austrian province of Upper Austria are analysed for 1998 and 2001, respectively. One group comprises the two market leaders Rewe Austria AG and Spar Österreichische Warenhandels-AG (RS). The other group consists of the other market participants (OMP) in the Upper Austrian market. The resulting maps indicate areas of increasing and decreasing concentration for each of the two groups separately. Concentration is hereby understood as outlet density. Changes in market dominance, too, are analysed using dual KDE by subtracting the estimate for RS's outlet density from OR's outlet density for each observed year.

The paper is organised as follows. In the second section the KDE method is presented. The choice of the kernel function and the issue of the appropriate bandwidth are discussed. The third section describes the application of dual KDE on the Upper Austrian food retailing market and presents the data used for the analysis. The last section sums up the results.

KERNEL DENSITY ESTIMATION

Kernel density estimation is an interpolation technique that generalizes individual point locations or events, $\boldsymbol{s}_i$, to an entire area and provides density estimates, $\lambda(\boldsymbol{s})$, at any location within the study region R. From a visual point of view it can be thought of a three-dimensional sliding kernel function $K(\bullet)$ that 'visits' every location $\boldsymbol{s}$ (see figure 1).

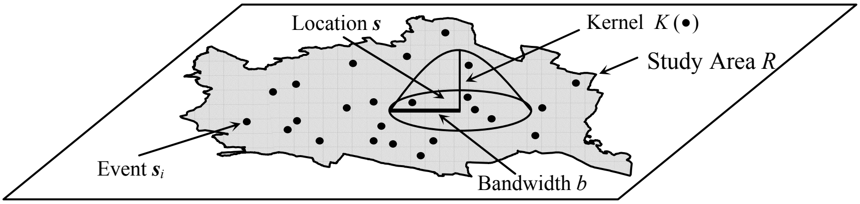

Figure 1: Kernel estimation of a point pattern (Fischer, 2001)

Distances to each observed event $\boldsymbol{s_i}$ that lies within a specified distance b, referred to as the bandwidth, are measured and contribute to the intensity estimate at $\boldsymbol{s}$ according to how close they are to $\boldsymbol{s}$ (Bailey, 1996). This produces a more spatially smooth estimate of variations in $\lambda(\boldsymbol{s})$ than could be attained by using a fixed grid of quadrates. Formally, if $\boldsymbol{s}$ represents a vector location anywhere in R and $\boldsymbol{s_1}, ..., \boldsymbol{s_n}$ are the vector locations of the n observed events, then the intensity $\lambda(\boldsymbol{s})$, at $\boldsymbol{s}$ is estimated as follows (Burt, 1996):

$$\hat{f}(\boldsymbol{s}) = \frac{1}{nb}\sum_{i=1}^{n} K\left(\frac{\boldsymbol{s}-\boldsymbol{s}_i}{b}\right) \qquad [1]$$

There are a number of different kernel functions. However, since the result of an analysis is not strongly influenced by the chosen function as long as the function is symmetric (Silverman, 1978; Burt, 1996) there are no binding rules concerning the choice of an appropriate function. The most common kernel function is the normal distribution function:

$$K\left(\frac{d_i}{\boldsymbol{u}}\right) = \frac{1}{2\pi}\exp\left[-\frac{1}{2}\left(\frac{d_i}{\boldsymbol{u}}\right)^2\right] \qquad [2]$$

with

$$\boldsymbol{u} = \left(\frac{\boldsymbol{s}-\boldsymbol{s_i}}{b}\right)$$

Based on this function the density estimate is expressed as (Burt, 1996):

$$\hat{\lambda}(\boldsymbol{s}) = \frac{1}{b^2}\sum_{i=1}^{n} K\left(\frac{d_i}{b}\right) \qquad [3]$$

where d_i is the distance between the point $\boldsymbol{s}$ and the observed event location $\boldsymbol{s_i}$. Because the bandwidth b is the standard deviation of the normal distribution, this function extends to infinity in all directions, i.e., it will be applied to each point in the region (Leitner, 2001). The quartic kernel function on the other hand, has a circumscribed radius b, which is also the bandwidth. The quartic kernel function is therefore applied to a limited area around each event and has the following functional form (Bailey, 1995).

$$K(\boldsymbol{u}) = \begin{cases} \frac{3}{4}\left(1-\boldsymbol{u}^2\boldsymbol{u}\right)^2 & \text{for } \boldsymbol{u}^2\boldsymbol{u} \le 1 \\ 0 & \text{otherwise} \end{cases} \qquad [4]$$

The intensity at $\boldsymbol{s}$ is estimated as

$$\hat{\lambda}(\boldsymbol{s}) = \sum_{i=1,\ d_i \le b}^{n} \frac{3}{\pi \ b^2}\left(1-\frac{d_i^2}{b^2}\right)^2 \qquad [5]$$

Since $d_i \le b$, the summation is only over values of d_i which do not exceed b. The region of influence within which the observed events contribute to $\hat{\lambda}(\boldsymbol{s})$ therefore equals a circle of radius b centred on $\boldsymbol{s}$. In $\boldsymbol{s}$ the weight is $3/\pi b^2$ and drops smoothly to a value of zero at distance b (Gatrell, 1995). This function can be extended by a weighting variable (W_i) as is shown in [6] (Levine, 2002):

$$\hat{\lambda}(\boldsymbol{u}) = \sum_{i=1,\ d_i \le b}^{n} \left[W_i \left(\frac{3}{b^2 \pi}\right)\left(1-\frac{d_i^2}{b^2}\right)\right] \qquad [6]$$

More important than the kernel function is the choice of an appropriate bandwidth b (Silverman, 1986), which has considerable influence on the estimated distribution. The effect of increasing the bandwidth b is to stretch the region around $\boldsymbol{s}$ within which observed events influence the intensity estimate at $\boldsymbol{s}$. For very large b, $\hat{\lambda}(\boldsymbol{s})$ will appear flat and local features will be obscured. If b is small, then $\hat{\lambda}(\boldsymbol{s})$ tends to a collection of spikes centred on $\boldsymbol{s_i}$ (Fischer, 2001).

There are several methods which attempt to optimise the value of b given the observed pattern of event locations. Basically, the bandwidth can either be fixed (fixed kernel estimation) or adaptive (adaptive kernel estimation). In adaptive smoothing sub-areas in which events are more densely packed than in others are visited by a kernel whose bandwidth is smaller than elsewhere in order to avoid smoothing out too much detail. In the fixed choice the bandwidth always stays the same. When using a fixed interval, the appropriate choice of the smoothing parameter will be influenced by the purpose for which the density estimate is to be used. Hereby, several different plots of the data, all smoothed by different amounts, are examined and the bandwidth that seems most in accordance with one's prior ideas about the density is chosen. Silverman (1986) calls this approach "smoothing 'by eye'" as opposed to automatic methods. Within the scope of this study it seems sufficient to rely on a subjective choice of the bandwidth since retail outlets have a certain catchment area that has to be allowed for when analysing market concentration and market dominance. Thus, the choice of the bandwidth has to find a compromise between the requirements of a reasonable density estimate and a catchment area that is not too wide. For every analysis within this study a quartic kernel function and a fixed bandwidth of 3,000 meters is used. Edge correction is not used.

If the KDE is applied to one variable, this is referred to as a single density estimate. If it is applied to two variables, it is called a dual density estimate. In the latter case, a kernel density is estimated for each variable individually and then the two density estimates are related with each other through simple algebraic operations such as the sum, the difference, the quotient and the like. The most commonly used operation in this realm is the quotient. In the present study, however, the absolute difference in densities is used to visualise spatio-temporal changes in the study region.

AN APPLICATION OF THE DUAL KDE METHOD TO THE UPPER AUSTRIAN FOOD RETAILING MARKET

To illustrate the use of dual KDE for retailing applications two different point data sets are used representing the locations of retail outlets in the province of Upper Austria in two different years, namely 1998 and 2001. These data sets are further divided into two groups for each year with the first group representing the outlet locations of the two market leaders Rewe Austria AG and Spar Österreichische Warenhandels-AG (RS), and the second group representing the outlet locations of the other market participants (OMP). In the following sections, market concentration and market dominance will be analysed by comparing these two groups to each other at different points in time (1998 and 2001).

Selected Market Characteristics and Data Representation

The European food retailing industry is characterised by concentration dynamics and massive store size expansion. In Austria, too, the market power of the leading food retailing companies increases steadily. In 2001 the two predominant retail chains had a joint market share of 66 percent. This development is closely linked to an increasing divergence in productivity ratios at the expense of small retailers which are subsequently driven out of the market by the larger chains. The Austrian province of Upper Austria is of special interest since it shows a high degree of market concentration compared to the other provinces. Upper Austria has large rural areas on the one hand but also has a larger urban area along the axis Linz-Wels.

Figure 2 shows the Austrian regions according to ACNielsen. In this definition the province of Upper Austria corresponds to the region I East. In 2001 the total population amounted to 1,380,747 or 16.89 percent of Austria's population. The province accounted for 16 percent (1,824.5 Mio €) of total grocery sales in 2001 (ACNielsen, 2002). The distribution of outlet types in Upper Austria corresponds to the distribution in the other Nielsen regions in Austria with exception of II North with a big number of smaller outlets and comparatively few supermarkets. Looking at the relevance of the outlet types in terms of total revenues, hypermarkets obtain a strong position in the Upper Austrian market being second only to the Vienna region. At the same time the relevance of small stores is very low compared to the other regions except for Vienna. This is interesting, since Vienna is a thoroughly urban region, while Upper Austria is mainly rurally dominated.

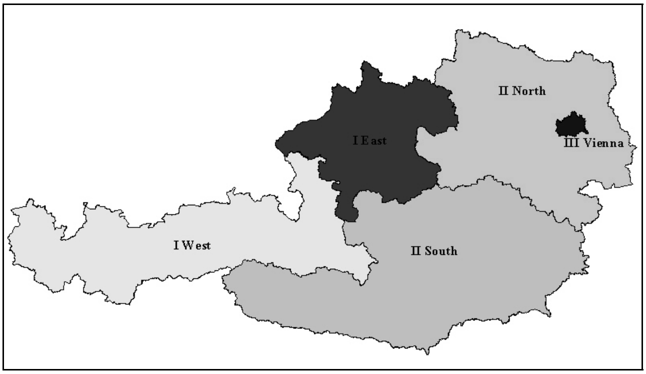

Figure 2: The Nielsen-Regions in Austria (ACNielsen, 2002)

The data employed in the model of the Upper Austrian food retailing market comprise store location data disaggregated by retail chain and outlet type for the years 1998 and 2001. The store location data is derived from the Herold Business Database that comprehends the addresses of all Austrian companies.

The outlet type that has been allocated to the store location is based on the average outlet size of each store type according to the ACNielsen-classification which distinguishes four types of retailers: hypermarkets (>1,000m^2), supermarkets (400-1,000m^2), large stores (250-400m^2) and small grocery stores (≤250m^2).

Before the two groups of retailers could be compared to each other, a weighting parameter was developed to assign different weights to different outlet types. The weights are based on the average gross sales per outlet based on the average sales area. It became evident, that the ratio of hypermarket : supermarket : large and small outlets equalled about 5 : 3 : 1. Thus, the outlets were weighted with this ratio for further analysis (table 1).

Outlet type	*Average gross sales per outlet* [*Mio €*] *1998*	*2001*	*Weighting ratio* (w)
Hypermarkets (>1,000m^2)	11.42	11.68	5
Supermarkets (400-1,000m^2)	2.14	2.23	3
Large and small grocery outlets	0.67	0.71	1

Table 1: Average gross sales based on outlet types in the Upper Austrian market and weighting ratios (ACNielsen, 2002 and 1999, own calculations)

In addition to the average sales per outlet type, population density is also allowed for in the analysis. Therefore, the final weight parameter (W_i) has the following functional form:

$$W_i = (w)\,(p) \qquad [7]$$

with p representing the population density, i.e. the population per km^2 based on *Zählsprengel* (census tracts corresponding to administrative units). For the analysis of market concentration the outlet locations were weighted with w only. W_i was used for the analysis of market dominance.

The study area is a rectangle with 169,813m east-west-expansion and 148,577m north-south-expansion superimposed over the province of Upper Austria. The rectangle covers an area of about 25,230 km^2.

The dual kernel density estimation: Tasks and results

Using dual KDE, the two groups of retailers (RS and OMP) were compared to each other weighting the individual outlets according to the scheme developed in the previous subsection. The arithmetic operation used for the dual kernel density estimate is the absolute difference, i.e. the absolute density of the second point pattern is subtracted form the absolute density of the first point pattern. To analyse the spatio-temporal changes the arithmetic mean is referred to. The legends in the following figures are to be interpreted as follows:

- „strong decrease“ = decrease above the arithmetic mean.
- „slight decrease“ = decrease below the arithmetic mean.
- „slight increase“ = increase below the arithmetic mean.
- „strong increase“ = increase above the arithmetic mean.

Market concentration is estimated by subtracting the single kernel density estimate of 1998 from the single kernel density estimate of 2001 for each of the groups. The outlet density of the market leaders shows an above average increase in the urban areas (Linz-City, Linz-Land, Wels-City and Steyr-City) while at the same time the other market participants experienced an above average decrease in these areas. This indicates that smaller companies are driven out of the market by the large retailers.

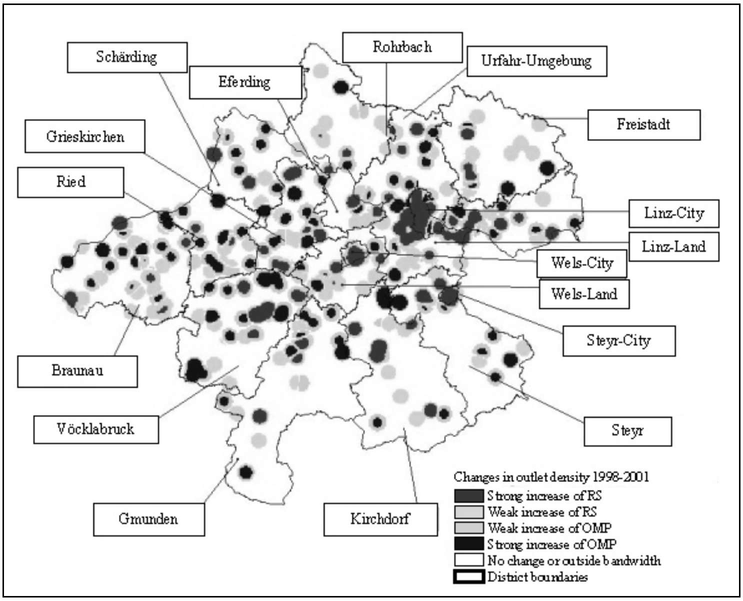

Figure 3: Changes in market concentration in Upper Austria between 1998 and 2001 (ArcAustria Data Program, WIGeoGIS Vienna, 1998 and 2001)

At the same time, there is a tendency towards an increase in outlet density in favour of the other market participants in less populated rural areas. In the entire province however, the outlet density of the group OMP showed a strong decrease. This group reached an increase in outlet density in only 39.59 percent of the area in which changes became evident, while RS increased their outlet density in 61.62 percent of this area. Taking all this into consideration, the analysis of outlet density shows a distinct development in favour of RS.

The direct comparison of the two groups (figure 3) for which the dual kernel estimate of OMP was subtracted from the dual kernel estimate of RS confirms this result. The west-east decomposition that has already become apparent in the distinct analysis of the two groups becomes more evident. For the OMP the highest values in density can be observed in the (rural) west of the province, while on the other hand, RS display the highest increases in central Upper Austria extending to the east, especially in Linz and the surrounding areas, in the district Perg and in Steyr-City and its immediate surroundings. Though the direct comparison shows an above average increase of about 14 percent for each of the two groups, the outlet density of RS with 60 percent increased considerably stronger than the outlet density of OMP (39.75 percent). This indicates a significant increase in market concentration in favour of the market leaders in the observed period of time.

Market dominance is estimated by subtracting the density estimate of RS from the density estimate of OMP for each of the two observed years separately. The results (figures 4 and 5) correspond to a large extent to the results of the analysis of the outlet density. In 2001 the group of the OMP still accounted for 53.57 percent of the areas in which market dominance was observed. However, the analysis shows, that RS could improve their situation in the observed period of time. Especially in the city of Linz RS could increase their dominance, in the cities Wels and Steyr, too, a development in favour of RS becomes evident. Though OMP could increase their dominant position in selected rural areas, their market position was weakened in the observation period.

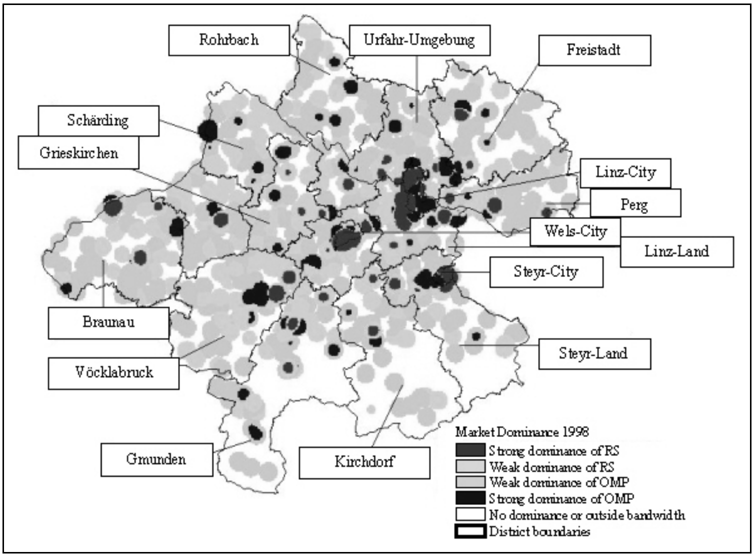

Figure 4: Market dominance in Upper Austria in 1998
(ArcAustria Data Program, WIGeoGIS Vienna, 1998)

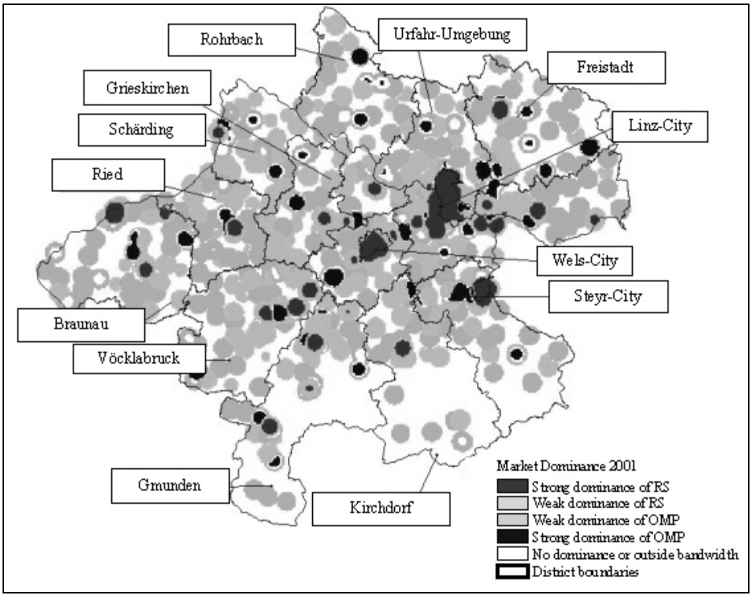

Figure 5: Market dominance in Upper Austria in 2001
(ArcAustria Data Program, WIGeoGIS Vienna, 2001)

SUMMARY

In this paper the dual KDE is applied to estimate spatio-temporal changes in spatial point patterns. The example used to demonstrate the technique is the development of the highly competitive Upper Austrian food retailing market between 1998 and 2001. While traditional approaches for the analysis of concentration processes and market dominance analyse a region as a whole without allowing for local variations, dual KDE delivers interesting results concerning small scale-variations in the development of concentration processes and market dominance.

Summarising the results of the performed density analyses, a decline in outlet density as well as in market dominance can be observed for the group OMP and an overall development in favour of RS becomes evident. This confirms the assumption of increasing market concentration in the Upper Austrian grocery industry as well as the crowding-out of smaller retailers from the market.

BIBLIOGRAPHY

ACNielsen, 1999, *Statistisches Jahrbuch 1998*. Vienna: ACNielsen

ACNielsen, 2000, *Statistisches Jahrbuch 1999*. Vienna: ACNielsen

ACNielsen, 2001, *Statistisches Jahrbuch 2000*. Vienna: ACNielsen

ACNielsen, 2002, *Statistisches Jahrbuch 2001*. Available from http://www.acnielsen.at, accessed 1.9.2002

Bailey, T.C. and Gatrell, A.C., 1995, *Interactive Spatial Data Analysis*. Essex: Longman

Burt, J.E. and Barber, G.M., 1996, *Elemantary Statistics for Geographers*. New York: The Guilford Press

Fischer, M.M., Leitner, M. und Staufer-Steinnocher, P., 2001, Spatial Point Pattern Analysis – Some Useful Tools for Analysing Locational Data. In: Wohlschlägl, H. (ed.):*Geographischer Jahresbericht aus Österreich*, vol. 58, pp.49-66. Vienna: Institute for Geography and Regional Research. Vienna University.

Gatrell, A.C., Bailey, T.C., Diggle, P.J. and Rowlingson, B.S., 1996, *Spatial Point Pattern Analysis and its Application in Geographical Epidemiology*. Available from http://www.dpi.inpe.br/cursos/ser301/referencias/gattrel_paper.pdf, accessed 7.5.2002

Leitner, M. and Staufer-Steinnocher, P., 2001, Kernel Density Estimation - An Innovative Exploratory Tool for Geomarketing. In Strobl, J., Blaschke, T. & Griesebner, G. (eds.): *Angewandte Geographische Informationsverarbeitung XIII*, pp.122-129. Heidelberg: Herbert Wichmann Verlag

Leitner, M. and Staufer-Steinnocher, P., 2002, Spatio-Temporal Concentration Processes in the Vienna Food-Retailing Market, Draft. Department of Econonmic Geograhpy & Geoinformatics, Wirtschaftsuniversität Wien

Levine, N. (2002): *CrimeStat Manual – Chapter 8 Kernel Density Interpolation*. Available from http://www.icpsr.umich.edu/NACJD/crimestat/CrimeStatChapter.8.pdf, accessed 8.3.2003

Silverman, B.W. (1978): Choosing the Window Width When Estimating a Density. *Biometrika* 65, pp.1-11

Wieser, R., Aiginger, K. & Wüger, M. (1999): *Marktmacht im Einzelhandel*. Vienna: Österreichisches Institut für Wirtschaftsforschung