Projektbereich A
Discussion Paper No. A–390

Family Expenditure Data, Heteroscedasticity and the Law of Demand

by

Werner Hildenbrand *)
Alois Kneip *)

Dezember 1992

*) Wirtschaftstheorie II, Uni Bonn

Institut für Gesellschafts– und Wirtschaftswissenschaften

Abteilung Wirtschaftstheorie II

Rheinische Friedrich-Wilhelms-Universität Bonn

Adenauerallee 24–26

D–5300 Bonn 1

Germany

Telefon (0228) 73 92 42

Telefax (0228) 73 79 40

E–Mail OR226 at DBNUOR1 EARN/BITNET

Printed on HP Laserjet II

at Institut für Gesellschafts- und Wirtschaftswissenschaften

Wirtschaftstheoretische Abteilung II, Bonn

Typeset in TEX by Sabine, Jürgen, Frank, Eckart

1 Introduction

In this paper we present some results of a data analysis of Family Expenditure Surveys for the United Kingdom and France. These surveys contain (among other information) for every household of a large random sample from the whole population of households the *expenditure* (typically per year) on a variety of consumption items, like food (bread, flour, beef and veal, mutton and lamb...), housing, services, transport, etc. The sum of household expenditure on all consumption items that are considered in the survey is called household *total expenditure* .

We shall show that all consumption expenditure data sets which we have analysed so far exhibit a characteristic feature that we shall call "*increasing spread* ".

In this introduction we shall first informally describe the characteristic feature of the consumption expenditure data. Formal definitions of the statistical concepts that are needed are given in section 2. Then we sketch an argument in order to show why a theorist who is concerned about demand aggregation might be interested in the empirical findings of this paper.

It is a well-known empirical fact that the *variance*, $\operatorname{var}(x)$, of the distribution of expenditure on a certain consumption item of all households with total expenditure x is dependent upon the level of total expenditure ("heteroscedasticity") and, further more, this conditional variance $\operatorname{var}(x)$ has a tendency to increase with total expenditure. To be more specific, consider the expression

$$\int (\operatorname{var}(x+\Delta) - \operatorname{var}(x))\, \rho(x) dx\,, \qquad \Delta > 0$$

where ρ denotes the density (histogram) of the total expenditure distribution. Estimates of this expression from family expenditure data typically turn out to be positive.

We extend this notion of "monoscedasticity" of expenditure on a particular consumption item to a comprehensive collection of consumption items. This then leads to the notion of "*increasing dispersion of conditional consumption expenditure*" and to the property of "*average increasing dispersion*" which are defined in the next section. The last property is well supported by the data which have been analysed so far.

Instead of considering the variance $\mathrm{var}(x)$ of the distribution of expenditure on a certain consumption item of all households with total expenditure x, one can consider the *second moment*, denoted by $m^2(x)$. Surely, one expects that the conditional second moment $m^2(x)$ will also depend on x. The question is whether there is also a tendency to increase with total expenditure.

Since $m^2(x+\Delta) - m^2(x) = [\mathrm{var}(x+\Delta) - \mathrm{var}(x)] + [m(x+\Delta)^2 - m(x)^2]$, and since one can expect that $m(x+\Delta)^2 - m(x)^2$ is positive, this increasing tendency of the second moment $m^2(x)$, which we call "increasing spread", will even be more pronounced. Furthermore, if one considers the average over the expenditure distribution, that is to say, the expression

$$\int \left(m^2(x+\Delta) - m^2(x) \right) \rho(x) dx$$

then, estimates of this expression from family expenditure data turn out to be positive.

As before, we extend this notion of "increasing spread" of expenditure on a particular consumption item to a comprehensive collection of consumption items. For this, one considers the *matrix* $m^2(x)$ *of second moments* of the *joint distribution* of expenditure on the collection of consumption items of all households with total expenditure x. Thus, if there are n consumption items then $m^2(x)$ is a $n \times n$ matrix. The definition of the matrix $m^2(x)$ implies that $m^2(x)$ is positive semi-definite. The degree of positive-definiteness can be taken as a measure of "spread" of the joint distribution of expenditure.

The joint distribution of consumption expenditure of all households who spend $x + \Delta$ is said to be more spread than the joint distribution of consumption expenditure of all households who spend x if the matrix

$$m^2(x+\Delta) - m^2(x)$$

is positive semi-definite. This clearly generalises the above one-dimensional notion of increasing spread since it implies that every element on the diagonal of $m^2(x+\Delta)$ is larger or equal to the corresponding element on the diagonal of $m^2(x)$. We define *average increasing spread* by the positive semi-definiteness of the matrix

$$M(\Delta) := \int \left[m^2(x+\Delta) - m^2(x)\right] \rho(x)dx, \qquad \Delta > 0.$$

Estimates of this matrix from consumption expenditure data turn out to be positive definite for all data sets which we have analysed so far.

We shall now sketch an argument[1] which shows that the property of "average increasing spread" is relevant for deriving the Law of Demand of the market demand function.

Let us assume that the demand behavior of a household can be modelled by a *demand function*

$$f^\alpha : (p, x) \mapsto f^\alpha(p, x) \in \mathbb{R}^n$$

where $\mathbb{R}^n$ denotes the commodity space, $p \in \mathbb{R}^n$ denotes the price vector, and x the total expenditure of the household. Different households typically will have different demand functions and total expenditure.

A large population of households is then described by a *joint distribution* μ of $(\alpha, x) \in \mathcal{A} \times \mathbb{R}_+$. The parameterspace $\mathcal{A}$ which is used to parametrize household demand functions can be any separable matrix space.

The *market demand* is defined by

$$F(p) := \int_{\mathcal{A}\times R_+} f^\alpha(p, x) d\mu.$$

The Law of Demand says that for any two price vectors p and q the vector of price changes $(p - q)$ and the vector of demand changes $(F(p) - F(q))$ point in opposite direction, i.e.,

$$(p - q) \cdot (F(p) - F(q)) \leq 0.$$

If the market demand function F is differentiable then this inequality is equivalent with the negative semi-definiteness of the Jacobian matrix $\partial_p F(p) = \int \partial_p f^\alpha(p, x) d\mu$.

The Jacobian matrix $\partial_p f^\alpha$ can be decomposed into two matrices (Slutsky-decomposition):

$$\partial_p f^\alpha(p, x) = S f^\alpha(p, x) - A f^\alpha(p, x)$$

[1] For more details we refer to Hildenbrand (1993).

where $Af^\alpha(p,x) = \partial_x f^\alpha(p,x) f^\alpha(p,x)^\perp$ and $Sf^\alpha(p,x)$ is the Jacobian matrix, evaluated at $q = p$, of the mapping

$$q \mapsto f^\alpha\left(q, q \cdot f(p,x)\right).$$

Consequently, *a positive semi-definite average expenditure effect matrix*

$$\bar{A}(p) = \int_{\mathcal{A}\times \boldsymbol{R}_+} Af^\alpha(p,x) d\mu$$

supports the Law of Demand. Yet, why should the matrix $\bar{A}$ be positive semi-definite?

The expenditure effect matrix $Af^\alpha(p,x)$ is positive semi-definite if, and only if, the vectors $f^\alpha(p,x)$ and $\partial_x f^\alpha(p,x)$ are collinear. There is obviously no *a priori* reason for this to hold. Consequently, if the average expenditure effect matrix $\bar{A}$ is positive semi-definite at all, then it must be the result of an aggregation (averaging) effect.

The expenditure on commodity h of household α with total expenditure x at the price vector p is denoted by

$$y_h^\alpha(p,x) = p_h \cdot f_h^\alpha(p,x), \qquad h = 1, ..., n$$

It is easy to show that the matrix $\bar{A}$ is positive semi- definite if, and only if, for sufficiently small $\Delta > 0$ the distribution of consumption expenditure $\{y^\alpha(p, x + \Delta)\}$ (i.e., the image distribution of μ under the mapping $(\alpha, x) \mapsto y^\alpha(p,x)$) is more spread than the distribution of consumption expenditure $\{y^\alpha(p,x)\}$. By definition of the partial ordering "more spread" this means that the matrix

$$\tilde{M}(p,\Delta) := m_\mu^2\{y^\alpha(p, x+\Delta)\} - m_\mu^2\{y^\alpha(p,x)\}$$

is positive semi-definite.

Let ρ denote the density of the distribution of total expenditure (the marginal distribution of μ) and $\mu|x$ the *conditional distribution* of α on $\mathcal{A}$ given the total expenditure level x. Then we obtain

$$\tilde{M}(p,\Delta) = \int [m_{\mu|x}^2\{y^\alpha(p, x+\Delta)\} - m_{\mu|x}^2\{y^\alpha(p,x)\}]\rho(x)dx.$$

Now we recall the definition of average increasing spread of conditional expenditure which we discussed in the first part of this introduction. In our new notation this means that the matrix

$$M(p,\Delta) = \int [m^2_{\mu|x+\Delta}\{y^\alpha(p,x+\Delta)\} - m^2_{\mu|x}\{y^\alpha(p,x)\}]\rho(x)dx \ .$$

is positive semi-definite. This matrix $M(p,\Delta)$ looks similar to the matrix $\tilde{M}(p,\Delta)$, yet, in general, they are not identical since the conditional distribution $\mu|x$ might depend on x. However, if this dependence is sufficiently weak, then one can consider the "observable" matrix $M(p,\Delta)$ as a "proxi" for the "unobservable" matrix $\tilde{M}(p,\Delta)$. Since the estimates of the matrix $M(p,\Delta)$ from expenditure surveys turns out to be positive definite one might use these empirical findings as a support for the hypothesis that the matrix $\tilde{M}(p,\Delta)$, and hence, the average expenditure effect matrix $\bar{A}$ is positive semi-definite.

Thus we have shown that the hypothesis of increasing *spread* supports the Law of Market Demand to hold. One can also show that the hypothesis of increasing *dispersion* gives support to a weaker property of the market demand function, that is to say, the Weak Axiom of Revealed Preference. For more details we refer to Hildenbrand (1993).

æ

2 Increasing dispersion and spread

In this section we shall define some simple concepts from descriptive statistics which will be needed for the data analysis of the Family Expenditure Survey in section 3.

Let ν denote a distribution (probability measure) on $I\!R^n$. The *covariance-matrix* , cov ν, of the distribution ν, is defined by

$$(\text{cov } \nu)_{ij} := \int y_i y_j d\nu - \int y_i d\nu \cdot \int y_j d\nu, \quad 1 \leq i,j \leq n$$

The *secondmoment-matrix* , $m^2\nu$, of the distribution ν is defined by

$$(m^2\nu)_{ij} := \int y_i y_j d\nu, \quad 1 \leq i, j \leq n.$$

We only consider distiributions ν for which the covariance and second moment-matrix are finite.

Definition 1

a) The distribution ν_1 on $I\!R^n$ is said to be *more dispersed* than the distribution ν_2 on $I\!R^n$ if the matrix cov ν_1 − cov ν_2 is positive semi-definite.

b) The distribution ν_1 is said to be *more spread* than the distribution ν_2 if the matrix $m^2\nu_1 - m^2\nu_2$ is positive semi-definite.

In order to visualize in the next section the partial orderings of 'more dispersed' and 'more spread' for consumption expenditure distributions we define the *ellipsoid of dispersion* and the *ellipsoid of spread* .

For our purpose, it suffices to consider only the case where the matrices cov ν and $m^2\nu$ are non-singular. In this case we define the *ellipsoid of dispersion* of the distribution ν on $I\!R^n$ by

$$Ell(\text{cov } \nu) := \left\{ z \in I\!R^n \,\middle|\, z \cdot (\text{cov } \nu)^{-1} z \leq 1 \right\},$$

and the *ellipsoid of spread* of the distribution ν by

$$Ell(m^2\nu) := \left\{ z \in I\!R^n \,\middle|\, z \cdot (m^2\nu)^{-1} z \leq 1 \right\},$$

One easily shows that the distribution ν_1 is more dispersed than the distribution ν_2 if and only if

$$Ell(\text{cov } \nu_2) \quad \subset \quad Ell(\text{cov } \nu_1).$$

The distribution ν_1 is more spread than the distribution ν_2 if and only if

$$Ell(m^2\nu_2) \quad \subset \quad Ell(m^2\nu_1).$$

For a distribution ν on $I\!R^n$ we consider the image distribution under the mapping

$$y \mapsto \sum_{h=1}^{n} y_h.$$

We shall assume that this image distribution has a density, which we denote by ρ. For simplicity it is assumed that the support of this density is an intervall $[a,b]$ in $I\!R$.

For every $x \in [a,b]$ we denote by $\nu(x)$ the *conditional distribution of* ν *on the set* $\{y \in I\!R^n | \sum_{h=1}^{n} y_h = x\}$. In the following definition we shall assume that the distribution ν is sufficiently smooth such that cov $\nu(x)$ and $m^2\nu(x)$ are differentiable for $x \in (a,b)$.

Definition 2

The distribution ν *has the property of*

(a.1.) *conditional increasing dispersion*
if for every $x \in (a,b)$ the matrix
$C(x) = \partial_x \,\mathrm{cov}\; \nu(x)$ is positive semi-definite

(a.2.) *average conditional increasing dispersion*
if the matrix
$C_\rho = \int \partial_x \,\mathrm{cov}\; \nu(x)\rho(x)dx$ is positive semi-definite

(b.1.) *conditional increasing spread*
if for every $x \in (a,b)$ the matrix
$M(x) = \partial_x m^2\nu(x)$ is positive semi-definite

(b.2.) *average conditional increasing spread*
if the matrix
$M_\rho = \int \partial_x m^2\nu(x)\rho(x)dx$ is positive semi-definite.

One can show that the property of *conditional increasing spread* for a distribution ν on $I\!R^n_+$ *is stronger* than the property of *conditional increasing dispersion.* The proof of this claim is based on the identity $C(x)\mathbf{1} = 0$, where $\mathbf{1}$ is the vector $(1, \ldots, 1)$, and the following result: the matrix $M(x)$ is positive semi-definite if and only if the matrix $C(x)$ satisfies the inequality

$$v \cdot C(x)v \geq \frac{x}{2}(v \cdot \partial_x m\nu(x))^2$$

for every vector $v \in I\!R^n$ which is orthogonal to $m\nu(x) = \int y d\nu$.

In the next section we consider the expenditure data from the Family Expenditure Survey as a random sample from a hypothetical distribution ν. We shall estimate the matrices $C(x), C, M(x)$ and M, and test whether they are positive definite.

æ

3 Data Analysis

In this section we discuss the empirical evidence of the hypotheses of increasing dispersion and increasing spread of household consumption expenditure. For this we analyse two different family expenditure surveys.

One data set is from the United Kingdom Family Expenditure Survey (FES) from 1968 to 1984. For each year the consumption expenditure of approximately 7000 households are reported. Households are selected at random from electorial registers. For details on samples and commodity classification see Family Expenditure Survey (1968 - 1983), Kemsley, Redpath and Holmes (1980). Expenditure on the following nine commodity aggregates are considered (abbreviations used in the subsequent Figures are given in brackets):

1. Housing (Housing)
2. Fuel, light and power (Fuel)
3. Food and non-alcoholic beverages (Food)
4. Clothing and footwear (Clothing)

5. Durable household goods (Durables)
6. Services (Services)
7. Transport (Transport)
8. Alcohol and tobacco (Alcohol and Tobacco)
9. Other goods and miscellaneous (Other Goods)

The second data set is the French "Enquête budget de famille" (EBF) from 1979, 1984-85 and 1989. As for the data expenditure of United Kingdom large samples of randomly selected households are reported. The sample size varies from 9000 to 12000. Details concerning samples and commodity classification can be found in the documents of the French Institut National de la Statistique et des Études Économique. A somewhat finer specification of consumption items than for the FES data is used (abbreviations are given in brackets)

1. Food at home (Food 1)
2. Non-alcoholic beverages (Beverage)
3. Alcoholic beverages (Alcohol)
4. Food outside home (Food 2)
5. Clothing (Clothing)
6. Footwear (Footwear)
7. Housing (Housing)
8. Fuel, light and power (Fuel)
9. Durable household goods and domestic services (Durables)
10. Hygiene and health (Health)
11. Transport and communication (Transport)

12. Tobacco (Tobacco)

13. Culture and leisure (Leisure)

14. Other goods and miscellaneous (Other goods)

The FES and EBF data provide independent samples from the underlying population of households. It is now assumed that household consumption expenditure of each of these populations can be described by a smooth probability distribution ν on $I\!R^N$,which has compact support ($N = 9$ for FES,$N = 14$ for EBF). Of course, ν varies between years.To simplify the interpretation of the results, it is convenient to normalize the data in every year such that the mean total expenditure is equal to one.

We first analyze the hypotheses of *conditonal* increasing dispersion and spread. Consider the diagonal elements of the matrices $cov\,\nu(x)$ and $m^2\,\nu(x)$; let cov_{jj} and $m_{jj}(x)$ denote the $j-$th diagonal element of these matrices. Both $cov_{jj}(x)$ and $m_{jj}(x)$ are functions of total expenditure which can be estimated by non-parametric regression methods. We will call them "variance-curve" and "second moment-curve." Kernel estimators are applied for estimating $cov_{jj}(x)$ and $m_{jj}(x)$. A description of the estimators can be found in the appendix. Figures 1 and 3 show estimated "variance-curves" for FES and EBF data from the years 1974 and 1989. One immediately recognizes that for all consumption items conditional variance has a tendency to increase with total expenditure. Figures 2 and 4 show the same effect for "second moment-curves". One should note that the curve estimates are not very reliable for high values of x. This follows from the theoretical results discussed in the appendix. The hypotheses of conditional increasing dispersion and spread are of course more demanding, and the empirical evidence given in Figures 1 - 4 is incomplete.

A sufficient condition for $C(x)$ be positive semi-definite is that for $x^* > x, \quad cov\,\nu(x^*) - cov\,\nu(x)$ be positive semi-definite. Härdle and Park (1992) developed a method for testing whether $cov\,\nu(x^*) - cov\,\nu(x)$ be positive semi-definite. They applied their statistics to FES data and obtained results which are consistent with the hypothesis of increasing dispersion. Some problem arises from the fact that the difference $|x^* - x|$ can not be very small for the procedure to be reliable.

The matrix $C(x)$ is positive semi-definite if and only if the smallest non-zero eigenvalue of $C(x)$ is larger than zero. Note that by definition at least one eigenvalue of $C(x)$ is equal to zero. We now estimate the matrices $C(x)$ and their eigenvalues by kernel derivative estimation. Let $\hat{\lambda}_{min}$ denote the smallest non-zero eigenvalue of a resulting estimate $\hat{C}(x)$, and let λ_{min} be the smallest non-zero eigenvalue of $C(x)$. Bootstrap techniques now allow to establish confidence bounds c^*_{up}, c^*_{low} such that approximately $P\left(\lambda_{min} - \hat{\lambda}_{min} \geq c^*_{up}\right) \approx 0.05$. and $P\left(\lambda_{min} - \hat{\lambda}_{min} \leq c^*_{low}\right) \approx 0.05$. Details of the statistical procedures can be found in the appendix. Hence, with $c_{up} = c^*_{up} + \hat{\lambda}_{min}, c_{low} = c^*_{low} + \hat{\lambda}_{min}$ we approximately obtain $P\left(\lambda_{min} \geq c_{up}\right) \approx 0.05$ and $P\left(\lambda_{min} \leq c_{low}\right) \approx 0.05$. The values c_{up} and c_{low} can now serve as test statistics for testing positive semi-definiteness

of $C(x)$ with respect to the critical value of $\alpha = 0.05$. The tests

1) reject the hypotheses of conditional increasing dispersion (i.e. of positive semi-definiteness of $C(x)$), if $c_{up} < 0$

2) reject the hypotheses that $C(x)$ is *not* positive semi-definite if $c_{low} > 0$.

Table 1 and Table 2 summarize the results for the FES and EBF data for different values of total expenditure.

x	# $c_{low} < 0$	# $\hat{\lambda}_{min} < 0$	# $c_{up} < 0$
0.25	0	0	0
0.5	5	0	0
0.75	7	4	1
1	6	1	0
1.25	7	1	0
1.5	10	3	1
2	7	2	0
2.5	17	11	1

Table 1. Smallest non-zero eigenvalues of $\hat{C}(x)$, upper and lower confidence bounds: Number of negative values out of 17 estimates (the years 1968-1984), for FES data.

x	# $c_{low} < 0$	# $\hat{\lambda}_{min} < 0$	# $c_{up} < 0$
0.25	0	0	0
0.5	1	0	0
0.75	3	2	0
1	3	1	0
1.25	3	1	0
1.5	3	1	0
2	3	2	0
2.5	3	3	0

Table 2: Smallest non-zero eigenvalues of $\hat{C}(x)$, upper and lower confidence bounds: Number of negative values out of 3 estimates (the years (1979, 1984, 1989))for EBF data.

When interpreting the tables one should keep in mind that for the FES data independent tests for each of the 17 different years were performed with respect to the critical value $\alpha = 0.05$. One rejection out of 17 independent test might occure just by chance. Thus, the above results do not reject the hypothesis of conditional increasing dispersion.

The same approach can be adopted when dealing with the hypothesis of conditional increasing spread. By an analogous statistical method, described in the appendix, an estimate $\hat{\lambda}_{min}$ of the smallest eigenvalue λ_{min} of $M(x)$ can be constructed. With $\alpha = 0.05$ bootstrap leads to upper and lower bounds c_{up} and c_{low} for λ_{min}. Similar to 1 and 2 the values of c_{up} and c_{low} can then serve as statistics for testing the hypothesis of conditional increasing spread. FES and EBF data now lead to the results presentend in Tables 3 and 4.

In contrast to Table 1 and 2, the results of Table 3 and 4 are not in good accordance with the hypothsesis of conditional increasing spread. Too many rejections are obtained for the FES aswell as for EBF data. On the other hand, when considering the tables one might say that rejection of the hypothesis is not very strong, since for a majority of situations $c_{up} > 0$.

The above results for *conditional* increasing dispersion surely let us expect that the weaker hypothesis of *average* conditional increasing dispersion will be

x	$\# c_{low} < 0$	$\# \hat{\lambda}_{min} < 0$	$\# c_{up} < 0$
0.25	0	0	0
0.5	16	14	2
0.75	17	14	4
1	16	10	0
1.25	16	8	1
1.5	14	8	1
2	16	6	1
2.5	17	14	2

Table 3: Smallest eigenvalues of $\hat{M}(x)$, upper and lower confidence bounds: Number of negative values out of 17 estimates for FES data.

x	# $c_{low} < 0$	# $\hat{\lambda}_{min} < 0$	# $c_{up} < 0$
0.25	0	0	0
0.5	2	1	0
0.75	3	2	0
1	3	3	0
1.25	3	2	1
1.5	3	3	0
2	3	3	1
2.5	3	3	0

Table 4: Smallest eigenvalues of $\hat{M}(x)$, upper and lower confidence bounds: Number of negative eigenvalues out of 3 estimates for EBF data.

in good accordance with the data. Estimates $\hat{C}$ and $\hat{\lambda}_{min}$ of the matrix $C = \int \partial_x cov\nu(x)\rho(x)dx$ and its smallest non-zero eigenvalue λ_{min} can be obtained by average derivative methods. Note that at least one eigenvalue of C has to equal zero. As before for $\alpha = 0.05$ bootstrap allows to establish upper and lower bounds c_{up} and c_{low} for λ_{min}, which can be used as statistics for testing the positive semi-definiteness of C. Interpretation is analogous to 1) and 2). A description of the statistical procedures is given in the appendix. FES and EBF data then lead to the results presented in Tables 5 and 6.

year	sample size	$c_{low} \times 100$	$\hat{\lambda}_{min} \times 100$	$c_{up} \times 100$
1968	7184	0.30	0.42	0.54
1969	7007	0.35	0.44	0.54
1970	6391	0.27	0.33	0.40
1971	7238	0.31	0.40	0.49
1972	7017	0.25	0.36	0.46
1973	7125	0.24	0.35	0.46
1974	6694	0.26	0.39	0.51
1975	7201	0.30	0.43	0.57
1976	7203	0.26	0.38	0.49
1977	7198	0.22	0.32	0.43
1978	7001	0.21	0.28	0.37
1979	6777	0.17	0.22	0.26
1980	6943	0.14	0.32	0.46
1981	7485	0.19	0.26	0.32
1982	7427	0.18	0.27	0.35
1983	6973	0.16	0.21	0.27
1984	7081	0.17	0.30	0.41

Table 5: Smallest non-zero eigenvalues of $\hat{C}$, upper and lower confidence bounds for FES data.

year	sample size	$c_{low} \times 100$	$\hat{\lambda}_{min} \times 100$	$c_{up} \times 100$
1979	10647	0.0094	0.0131	0.0168
1984	11978	0.0090	0.0107	0.0127
1989	9043	0.0094	0.0106	0.0122

Table 6: Smallest non-zero eigenvalues of $\hat{C}$, upper and lower confidence bounds for EBF data.

In all situations c_{low}, $\hat{\lambda}_{min}$ and c_{up} are positive which means that the hypothesis of *average* conditional increasing dispersion *is consistent with the data.* We can even say more. Since $c_{low} > 0$, the opposite hypothesis that C is *not* positive semi-definite is rejected by both FES and EBF data from all years. *In this sense the data even support the hypothesis of average conditional increasing dispersion.*

As explained in the introduction the hypothesis of average conditional increasing *spread* is the most important one of the four hypotheses. Eventhough positive definiteness of the matrices $M(x)$ is rejected, the weaker hypothesis that $M = \int M(x)\rho(x)dx$ is positive definite seems to have a better chance to pass the test.

To visualize the empirical findings we can use ellipsoids of spread. By definition M is positive semi-definite if for all $\Delta > 0$ the difference $\bar{M}(\Delta) - \bar{M}$ is positive semi-definite, where $\bar{M}(\Delta) = \int m^2\nu(x+\Delta)\rho(x)dx$ and $\bar{M}(0) \equiv \bar{M} = \int m^2\nu(x)\rho(x)dx$ This implies that for all $\Delta > 0$ the ellipsoid of spread of the matrix $\bar{M}$ is *contained* in the ellipsoid of spread of the matrix $\bar{M}(\Delta)$. Figures 5 and 6 illustrate that this is in fact the case for all estimates of $\bar{M}(\Delta)$ and $\bar{M}$ for FES and EBF data (statistical tools are described in the appendix).

A more precise analysis of the matrix M is based on the type of approach already described above. Estimates $\hat{M}$ and $\hat{\lambda}_{min}$ of M and its smallest eigenvalue λ_{min} are determined by average derivative methods (we refer again to the appendix for a description of statistical methodology). For $\alpha = 0.05$ bootstrap methods allow to approximate upper and lower confidence bounds c_{up} and c_{low} for λ_{min}, which can be taken as statistics for testing the positive definitnesss of M. Interpretation follows 1) and 2). FES and EBF data then lead to the results given in Tables 7 and 8.

year	sample size	$c_{low} \times 100$	$\hat{\lambda}_{min} \times 100$	$c_{up} \times 100$
1968	7184	0.26	0.34	0.41
1969	7007	0.28	0.36	0.44
1970	6391	0.21	0.26	0.31
1971	7238	0.24	0.32	0.39
1972	7017	0.13	0.34	0.49
1973	7125	0.19	0.29	0.37
1974	6694	0.19	0.31	0.42
1975	7201	0.23	0.36	0.46
1976	7203	0.19	0.31	0.40
1977	7198	0.18	0.26	0.34
1978	7001	0.15	0.21	0.27
1979	6777	0.12	0.16	0.21
1980	6943	0.12	0.28	0.40
1981	7485	0.15	0.20	0.24
1982	7427	0.14	0.21	0.27
1983	6973	0.11	0.14	0.17
1984	7081	0.12	0.22	0.30

Table 7: Smallest eigenvalues of $\hat{M}$, upper and lower confidence bounds for FES data.

year	sample size	$c_{low} \times 100$	$\hat{\lambda}_{min} \times 100$	$c_{up} \times 100$
1979	10647	0.0108	0.0135	0.0167
1984	11978	0.0080	0.0091	0.0101
1989	9043	0.0063	0.0073	0.0083

Table 8: Smallest eigenvalues of $\hat{M}$, upper and lower confidence bounds for EBF data.

As in the analysis of the matrix C, we obtain $c_{low} > 0$ for FES and EBF data for all years. The opposite hypothesis that M is *not* positive definite is thus rejected in all situations. *In this sense, the data give support to the hypothesis of average conditional increasing spread* . Our results are in agreement with those of Härdle, Hildenbrand and Jerison (1991), who first analysed this hypothesis for FES data.

æ

Appendix: Statistical Tools

In this section we describe the statistical procedures that are used to obtain the empirical results described in Section 3.

The data consists of the expenditure on N consumption items for n households. Households are sampled independently such that the expenditure vectors $y_1, ..., y_n$ in $\mathbb{R}^N$ can be considered as independent random variables with common distribution ν. For simplicity we assume that ν has a compact support. The total expenditure on all consumption items of household i is denoted by x_i. The pairs (x_i, y_i) are i.i.d. with a common distribution[2] induced by ν. As in previous sections the marginal density of total expenditure is denoted by ρ.

1. Analysis of $M(x)$ and $C(x)$. For every pair of $j, l \in \{1, ..., N\}$ we want to estimate the "second moment-curve"

$$m_{jl}(x) = E(y_{ij} \cdot y_{il} | x_i = x) = \left(m^2 \nu(x)\right)_{j,l}$$

and the "covariance-curve"

$$cov_{jl}(x) = E(y_{ij} \cdot y_{il} | x_i = x) - (E(y_{ij}|x_i = x) \cdot E(y_{il}|x_i = x)) = (cov\, \nu(x))_{j,l}\,.$$

The functions m_{jl} and $cov_{j,l}$ are assumed to be "smooth" functions which are at least three times continuously differentiable. Estimation then relies on non-parametric regression methods; we use kernel estimators to estimate these curves in an interval $[a, b] \subset \mathbb{R}$. For a survey of kernel and other non-parametric curve estimation methods see Härdle (1990). We use the following kernel estimation technique which in the present context has some computational and theoretical advantages compared to the usual methods as proposed by Nadaraya (1964) and Watson (1964) or Gasser and Müller (1984):

First, the data is discretized into bins of the form

$$B_a(k) := \left[a + \frac{(k-1)}{L}, a + \frac{k}{L}\right], \quad k = 1, ..., L$$

for some integer $L < n$. For each bin $B_a(k)$, $k = 1, ..., L$, the average

[2] We always assume that average total expenditure is normalized to 1. Thus, y_i, x_i must be divided by average total expenditure. Clearly, this induces some correlation between different (y_i, x_i). However, since n is large, this effect is negligible.

$$
\begin{aligned}
M_{jl,k} &:= \frac{1}{n_k}\sum_{i=1}^{n} y_{ij}\cdot y_{il}\cdot I\left(x_i\in B_a(k)\right)\\
\text{and}&\\
C_{jl,k} &:= \frac{1}{n_k-1}\Bigg(\sum_{i=1}^{n} y_{ij}\cdot y_{il}\cdot I\left(x_i\in B_a(k)\right)\\
&\qquad -\Big(\sum_{i=1}^{n} y_{ij}\cdot I\left(x_i\in B_a(k)\right)\Big)\cdot\Big(\sum_{i=1}^{n} y_{il}\cdot I\left(x_i\in B_a(k)\right)\Big)/n_k\Bigg)
\end{aligned}
$$

are computed, where $n_k := \#\{X_i|X_i \in B_a(k)\}$. The kernel estimators $\hat{m}_{jl}(x)$ and $\hat{\mathrm{cov}}_{jl}(x)$ are then defined by[3]

$$
\begin{aligned}
\hat{m}_{jl}(x) &= \frac{1}{h}\sum_{k=1}^{n}\int_{a+\frac{(k-1)}{L}}^{a+\frac{k}{L}} W\left(\frac{x-v}{h}\right)dv\cdot M_{jl,k}\\
\hat{cov}_{jl}(x) &= \frac{1}{h}\sum_{k=1}^{n}\int_{a+\frac{(k-1)}{L}}^{a+\frac{k}{L}} W\left(\frac{x-v}{h}\right)dv\cdot C_{jl,k}
\end{aligned}
$$

where W is a kernel function and h denotes the bandwidth. This approach is closely related to the "WARPing" algorithm considered in Hrdle and Scott (1988) and Hrdle (1990) which has been proposed as a method for fast computation of kernel estimators. The number L has to be smaller than n but large enough such that it can be assumed that

$$
\frac{1}{n_k}\sum_{i=1}^{n} m_{jl}(x_i)\cdot I\left(x_i\in B_a(k)\right) \text{ and } \frac{1}{n_k}\sum_{i=1}^{n} cov_{jl}(x_i)\cdot I\left(x_i\in B_a(k)\right)
$$

equal $m_{jl}(t_k)$ and $cov_{jl}(t_k)$ up to a negligible error, where $t_k = a + \frac{2k-1}{2L}$.

To make this more precise let $L = \gamma\, n^{\frac{1}{2}}, 0 < \gamma < \infty$. One can then write

$$
M_{jl,k} = m_{jl}(t_k) + \varepsilon_k, \quad k = 1, ..., L,
$$

where the ε_k are independent with $|E\varepsilon_k| = o(\frac{1}{n})$ as $n \to \infty$. Furthermore, $var(\varepsilon_k) = \sigma^2_{jl}(t_k)/(\frac{n}{L}\rho(t_k)) + o(\frac{n}{L})$ as $n \to \infty$, where $\sigma^2_{jl}(x) := var(y_{ij}y_{il}|x_i = x)$. These properties of the ε_k can easily be derived from standard arguments (note

[3]Estimators have to be modified near the boundary of [a,b]. Possible boundary modifications are, for example, discussed in Müller (1988)

that $En_k \approx \frac{n}{L}\rho(t_k))$. If a second order kernel W is used, then it can be inferred from the arguments of Gasser and Müller (1984) that as $n \to \infty, h \to 0, nh \to \infty$

$$\begin{aligned} E(\hat{m}_{jl}(x) - m_{jl}(x))^2 &= (E\hat{m}_{jl}(x) - m_{jl}(x))^2 + var(\hat{m}_{jl}(x)) \\ &= h^4 \frac{m''_{jl}(x)}{4} d_w + \frac{\sigma^2_{jl}(x)}{nh\rho(x)} c_w + o(h^4 + \frac{1}{n}h) \end{aligned} \tag{1}$$

Here, d_w and c_w are kernel dependent constants. The optimal bandwidth h_{opt} is proportional to $n^{-\frac{1}{5}}$ and results in $E\left(\hat{m}_{jl}(x) - m_{jl}(x)\right)^2 = O\left(n^{-\frac{4}{5}}\right)$. An analogous expression can be established to hold for $\hat{cov}_{jl}(x) - cov_{jl}(x)$. Relation (A1) shows that the method proposed above shows a slightly superior asymptotic behaviour compared to the Nadaraya-Watson and Gasser-Müller kernel estimators. The bias $|E\hat{m}_{jl}(x) - m_{jl}(x)|$ of the Nadaraya-Watson estimator contains an additional term proportional to $\frac{\rho'(x)}{\rho(x)}$ which can lead to strange effects (compare Jennen-Steinmetz and Gasser, 1990). The Gasser-Müller estimator possesses the same bias term as in (A1), but the variance term has to be multiplied by 1.5.

The estimators introduced above were used to generate the curves presented in Figures 1 - 4. They are based on $L = 200$. The Epanechnikov kernel defined by $W(u) = \frac{3}{4}(1 - u^2)$ for $|u| \le 1$ and $W(u) = 0,\ |u| > 1$, was applied which has some optimality properties (compare, e.g., Müller, 1988). Optimal bandwidths were estimated from the data by using a plug-in method similar to the one proposed by Gasser, Kneip and Köhler (1991).

For given x, estimates of the matrices

$$M(x) = (m'_{jl}(x))_{j,l=\{1,\dots,N\}}$$

and

$$C(x) = (\mathrm{cov}'_{jl}(x))_{j,l=\{1,\dots,N\}}$$

were obtained by

$$\hat{M}(x) = (\hat{m}'_{jl}(x))_{j,l=\{1,\dots,N\}}$$

$$\hat{C}(x) = (\hat{\mathrm{cov}}'_{jl}(x))_{j,l=\{1,\dots,N\}} \ ,$$

where $\hat{m}_{jl}(x), \hat{\mathrm{cov}}_{jl}(x)$ were determined as above. A quartic kernel with $W(u) = \frac{15}{16}(1-u^2)^2$, $|u| \le 1$, and $W(u) = 0, |u| > 1$, was used in this estimation step, since this kernel has some optimality properties in estimating derivatives (compare, e.g., Müller, 1988). We are not only interested in estimating the matrices $M(x)$ and $C(x)$, but also in testing positive definiteness. Let us first consider estimates of $M(x)$. If the bandwidth h satisfies $h \to 0, \frac{1}{nh^3} \to 0$ as $h \to \infty$ then it follows from easy computations (compare Gasser and Müller, 1984)

$$\begin{aligned} E\hat{m}_{jl}(x) &= \frac{1}{h}\int W'(u)m_{jl}(x-hu)du + o(\frac{1}{nh}) \\ &= \int W(u)m'_{jl}(x-hu)du + o(\frac{1}{nh}) \\ &= m'_{jl}(x) + \frac{h^2}{2}m'''_{jl}(x)c^*_w + o(h^2 + \frac{1}{nh}) \end{aligned} \tag{2}$$

and

$$\operatorname{var}(\hat{m}'_{jl}(x)) = \frac{\sigma^2_{jl}(x)}{nh^3\rho(x)} \cdot d^*_w + o\left(\frac{1}{nh^3}\right), \tag{3}$$

where c^*_w, d^*_w are constants which can easily be computed from the definition of the quartic kernel. The optimal bandwidth is now proportional to $n^{-\frac{1}{7}}$ yielding

$$E(\hat{m}'_{jl}(x) - m'_{jl}(x))^2 = (E\hat{m}'_{jl}(x) - m'_{jl}(x))^2 + \operatorname{var}(\hat{m}'_{jl}(x)) = O(n^{-\frac{4}{7}}) .$$

Furthermore, by well known central limit theorems the following relation can be shown in a straightforward manner

$$\sqrt{nh^3}(\hat{m}'_{jl}(x) - E\hat{m}'_{jl}(x)) \longrightarrow_{\mathcal{L}} N\Big(0, \frac{\sigma^2_{jl}(x)}{\rho(x)}d^*_w\Big) \tag{4}$$

If $\lambda_{\min}(A)$ denotes the smallest eigenvalue of a matrix A, the techniques developed in Hrdle and Hart (1992) and Hrdle and Park (1992) allow to derive from (A4) that

$$\sqrt{nh^3}\Big(\lambda_{\min}(\hat{M}(x)) - \lambda_{\min}(E\hat{M}(x))\Big) \longrightarrow_{\mathcal{L}} N(0,\gamma), \tag{5}$$

where γ is a suitable constant which can be evaluated from (A4).

When choosing now an undersmoothing bandwidth, i.e., $h = o(n^{-\frac{1}{7}})$, $\frac{1}{nh^3} \to 0$, (A2) implies that $\sqrt{nh^3}(E\hat{m}'_{jl}(x) - m'_{jl}(x)) \longrightarrow_P 0$. Hence, by (A5)

$$\sqrt{nh^3}\Big(\lambda_{\min}(\hat{M}(x)) - \lambda_{\min}(M(x))\Big) \longrightarrow_{\mathcal{L}} N(0,\gamma) . \tag{6}$$

Expressions analogous to (A2) -(A6) can be obtained for $\operatorname{cov}'_{jl}(x)$. The results (A5) and (A6) can in principle be applied for testing $\lambda_{\min}(M(x)) > 0$

(or $\lambda_{\min}(C(x)) > 0$) but this procedure will be complicated (note that one has to plug in estimates of σ_{jl}^2). An attractive alternative consists in a bootstrap approximation of the distribution of $\sqrt{nh^3}(\lambda_{\min}(\hat{M}(x)) - \lambda_{\min}(M(x)))$. We therefore use resamples from the data $\{(Y_i, X_i)\}_{i=1,\ldots,n}$ for the given year and the given country. More precisely bootstrap samples $\{Y_i^*, X_i^*\}_{i=1,\ldots,n}$ are generated by independent sampling (with replacement) from $\{(Y_i, X_i)\}_{i=1,\ldots,n}$. Then estimates $\hat{M}^*(x)$ and $\hat{C}^*(x)$ of $M(x)$ and $C(x)$ are determined from the bootstrap samples. The distribution of $\lambda_{\min}(\hat{M}^*(x))$ or $\lambda_{\min}(\hat{C}^*(x))$ can easily be approximated by Monte Carlo simulations. Adapting arguments of Hrdle and Park (1992) and Hrdle and Hart (1992) it can be shown that the distribution of $\sqrt{nh^3}\Big(\lambda_{\min}(\hat{M}^*(x)) - \lambda_{\min}(E\hat{M}^*(x))\Big)$ is asymptotically close to the distribution of $\sqrt{nh^3}\Big(\lambda_{\min}(\hat{M}(x)) - \lambda_{\min}(E\hat{M}(x))\Big)$. The same result holds when replacing $\hat{M}$ by $\hat{C}$.

A bootstrap test of the hypothesis $\lambda_{\min}(M(x)) \geq 0$ can now be described as follows: One determines the constant $B(x) < 0$ from the bootstrap distribution such that $P\Big(\lambda_{\min}(\hat{M}^*(x)) - \lambda_{\min}(E\hat{M}^*(x)) \leq B(x)\Big) = \alpha$, where α equals, say 0.05. Then one computes the constant c_{up} by $c_{up} = \lambda_{\min}(\hat{M}(x)) - B(x)$. The above asymptotic arguments now imply that

$$P\Big(\lambda_{\min}(M(x)) \geq c_{up}\Big) \approx P\Big(\lambda_{\min}(E\hat{M}(x)) \geq c_{up}\Big) = \alpha \ ,$$

provided that h is small enough such that $|\lambda_{\min}(M(x)) - \lambda_{\min}(E\hat{M}(x))|$ is negligible. The hypothesis is rejected if $c_{up} < 0$. A test of $\lambda_{\min}(\hat{M}(x)) < 0$ can be obtained by evaluating the constants $D(x), c_{low}$ with $P\Big(\lambda_{\min}(\hat{M}^*(x)) - \lambda_{\min}(E\hat{M}^*(x)) \geq D(x)\Big) \leq \alpha$ and $c_{low} = \lambda_{\min}(\hat{M}(x)) - D(x)$. The hypothesis is rejected if $c_{low} > 0$. In an analogous manner one defines tests of $\lambda_{\min}(C(x)) \geq 0$ and $\lambda_{\min}(C(x)) < 0$.

This methodology was applied to the given data to obtain the results presented in Tables 1-4. We chose $L = 200$. The bandwidths used at $x = 0.25, 0.5, 0.75, 1.0, 1.25, 1.5, 2.0, 2.5$ where $h = 0.15, 0.2, 0.25, 0.3, 0.35, 0.5, 0.75, 1.0$. For all values of $x < 1.25$ the selected bandwidths are smaller than the (average) optimal bandwidths estimated by a plug-in method similar to the one of Gasser, Kneip and Köhler (1991). They might be considered as undersmoothing bandwidths for which the above theory applies. A difficulty arises for large values of total expenditure. The corresponding values of $\rho(x)$ are comparably small, as is illustrated in Figure 7 which shows a kernel estimated density $\hat{\rho}(x)$ for the French data of the year 1984. The effect generalizes to all years and holds for both data sets considered. Furthermore, $\sigma_{jl}^2(x)$ increases with x. But if $\rho(x)$ becomes small and $\sigma_{jl}^2(x)$ large, then by (A3) the variances $\mathrm{var}\,(\hat{m}'_{jl}(x))$ grow, which will corrupt the estimates $\lambda_{\min}(\hat{M}(x))$ if a small h is selected. This explains the large values of h for large total expenditures. The problem then is that the difference between $E\hat{M}(x)$ and $M(x)$ will not be negligible. However, (A4),(A5) and the bootstrap approximations remain true, and the above tests can still be considered as valid tools for checking the positive semi-definiteness of $E\hat{M}(x), E\hat{C}(x)$. Relation (A2) implies that as $n \to \infty, \frac{1}{nh} \to 0$

$$E\hat{M}(x) \approx \int W(u)M(x - hu)du \ ,$$

and an analogous expression holds for $E\hat{C}(x)$. Since the values of W are positive for the quartic kernel, the hypotheses of conditional increasing spread and dispersion imply that $E\hat{M}(x)$ and $E\hat{C}(x)$ be positive semi-definite asymptotically. Hence, even for large h these hypotheses will have to be rejected if $c_{up} < 0$.

Figure 7: Estimated density of total expenditure for EBF data from 1984

2. Ellipsoids of concentration. Estimates of the matrices $\bar{M}(\Delta) = \int m^2\nu(x+\Delta)\rho(x)dx$ for $\Delta > 0$ can be obtained in a very simple manner. We linearly interpolate the observations $y_{1j} \cdot y_{1l}, ..., y_{nj} \cdot y_{nl}$ to determine a rough estimate $\tilde{m}_{jl}$ of m_{jl}. An estimate $\hat{\bar{M}}(\Delta)$ is then computed by averaging the values $\tilde{m}_{jl}(x_i + \Delta)$, i.e., $\hat{\bar{M}}(\Delta) = (\frac{1}{n}\sum_{i=1}^n \tilde{m}_{jl}(x_i+\Delta))_{j,l=1,...,N}$. Together with $\hat{\bar{M}} \equiv \hat{\bar{M}}(0) = (\frac{1}{n}\sum_{i=1}^n y_{ij} \cdot y_{il})_{j,l=1,...,N}$ these estimates were used to generate the ellipsoids of spread presented in Figures 5 and 6. It is easily seen that as $n \to \infty$

$$\left| \frac{1}{n}\sum_{i=1}^n \tilde{m}_{jl}(x_i+\Delta) - \int m_{jl}(x+\Delta)\rho(x)dx \right| = Op\left(n^{-\frac{1}{2}}\right)$$

holds for all $\Delta \geq 0$. It should be noted, however, that as $\Delta \to 0$
$\Delta \cdot \left(\hat{\bar{M}}(\Delta) - \hat{\bar{M}}\right)$ is **not** a consistent estimator of $M = \partial_\Delta \bar{M}(\Delta)|_{\Delta=0}$.

3. Analysis of M and C. It remains to consider statistical tools for analyzing the matrices

$$\begin{aligned} M &= \int M(x)\rho(x)dx \\ C &= \int C(x)\rho(x)dx \end{aligned}$$

Let us first consider the matrix M.
The basic methodology applied is "average derivative" estimation as introduced by Hrdle and Stoker (1989). The idea is simple. By partial integration one obtains

$$M = \int (m'_{jl}(x))_{j,l=1,...,N}\rho(x)dx = \int (m_{jl}(x))_{j,l=1,...,N}\frac{-\rho'(x)}{\rho(x)}\rho(x)dx$$

Now estimate $L(x) = \frac{-\rho'(x)}{\rho(x)}$ by kernel density estimation. This approach is already described in Hrdle, Hildenbrand and Jerison (1991), who present an analysis of the M-matrix for the FES data. The density ρ can be estimated by a Rosenblatt-Parzen kernel density estimator

$$\hat{\rho}(x) = \frac{1}{nh}\sum_{i=1}^n W\left(\frac{x-x_i}{h}\right)$$

where W is a differentiable kernel function and h is the bandwidth. This results in an estimate $\hat{L}(x) = \frac{-\hat{\rho}'(x)}{\rho(x)}$. We again run into problems if x is large. Then

$\rho(x)$ is very small and $\hat{L}(x)$ might become unstable. The final estimator used in Hrdle, Hildenbrand and Jerison (1991) thus looks as follows.

$$\hat{M} = \frac{1}{n_b} \sum_{i=1}^{n} y_i y_i^T \hat{L}(x_i) \cdot I(x_i \leq b), \tag{7}$$

where I is the indicator function, b denotes the "cut- off" point, and $n_b := \#\{x_i | x_i \leq b\}$. In principle, for the type of densities considered (compare Figure 7) one also has to cut-off very small values of x. This is, however, of minor practical importance and will not be considered in this paper.

Hrdle and Stoker (1989) showed that under weak technical conditions there exists a bandwidth sequence $h = h_n \to 0$ and a sequence $b = b_n \to \infty$ such that as $n \to \infty$

$$\sqrt{n}\left(\hat{M}_{jl} - M_{jl}\right) \longrightarrow_{\mathcal{L}} N(0, \gamma_{jl}), \quad jl = 1, ..., N,$$

where γ_{jl} can be computed from M and ρ. Mammitzsch (1989) showed that the quartic kernel is optimal. Hrdle, Hart, Marron and Tsybakov (1991) analyzed the bandwidth choice and showed that $h \equiv h_n \sim n^{-\frac{2}{7}}$ is optimal. They furthermore give an asymptotic expansion for the optimal bandwidth which allows the construction of plug-in estimators. For further details on average derivative estimation see Stoker (1992).

The cut-off technique in (A7) does not completely solve the problem of instability of $\hat{L}$ at those values of x where $\rho(x)$ is small. A further improvement can be obtained by using variable bandwidths which increase with x. We follow an approach by Wand, Marron and Ruppert (1991). They propose to estimate heavily skewed densities in three steps: First a transformation $x_i \to (\alpha + x_i)^\beta$ $(\alpha > 0, \beta < 1)$ is done to obtain a less skewed density $\rho_{\alpha,\beta}$ of the transformed data $\{(\alpha + x_i)^\beta\}_{i=1,..,n}$. Then $\rho_{\alpha,\beta}$ is estimated by kernel estimation, and finally a back-transformation is done to determine an estimate $\hat{\rho}$ of ρ. The resulting estimator has the form

$$\hat{\hat{\rho}}(x) = \frac{\beta \cdot (\alpha + x)^{\beta - 1}}{nh} \sum_{i=1}^{n} W\left(\frac{(\alpha + x)^\beta - (\alpha + x_i)^\beta}{h}\right) \tag{8}$$

The estimate presented in Figure 7 has been constructed in this way, setting $\alpha = 0.1, \beta = 0.25$ and using a quartic kernel W. This choice of α, β is in line with the results of Wand, Marron and Ruppert (1991) and was used throughout.

Since $|(0.1+x)^{\frac{1}{4}} - (0.1+x_i)^{\frac{1}{4}}| \ll |x - x_i|$ for large x, (A8) behaves like a kernel estimator which automatically increases its bandwidth for large x_i. An estimator $\hat{M}$ of M can now be constructed by

$$(9) \qquad \hat{M} = \frac{1}{n_b} \sum_{i=1}^{n_b} y_i y_i^T \hat{\hat{L}}(x_i) \cdot I(x_i \leq b),$$

where $\hat{\hat{L}}(x) = -\hat{\hat{\rho}}'(x)/\hat{\hat{\rho}}(x)$. (A9) will be called "power-transformed average derivative estimator". The asymptotic properties derived for (A7) basically generalize.

Besides (A7) and (A9) we considered a further method which might be called a "κ - nearest neighbour average derivative estimator". The idea is to estimate the $m'_{jl}(x)$ directly by determining the slope of a straigth line fit to the κ nearest neighbours of x ($\kappa \in I\!N, \kappa \ll n$).

More precisely, for $i = 1, ..., n$ and $j, l = 1, ..., N$ parameters $\hat{a}_{i,jl}$ and $\hat{b}_{i,jl}$ are obtained by minimizing

$$\sum_{r=1}^{n} (y_{rj} y_{rl} - \alpha x_r - \beta)^2 \cdot I\left(x_r \in J_\kappa(x_i)\right)$$

with respect to α, β, where $J_\kappa(x_i) := \{x_r | x_r$ is one of the κ nearest neighbours of $x_i\}$

Then M is estimated by

$$(10) \qquad \hat{M}_{j,l} = \frac{1}{n} \sum_{i=1}^{n} \hat{a}_{i,jl}, \quad j, l = 1, ..., N.$$

Conceptually (A10) is quite different from (A7) and (A9). In particular, no cut-off point is required. Similar to (A7) it can, however, be shown that $\sqrt{n}\,(\hat{M}_{jl} - M_{jl})$ tends to an asymptotic normal distribution as $n \to \infty$, provided $\kappa \to \infty$ not too fast and not too slow.

All three methods (A7), (A9) and (A10) were applied to estimate the M-matrices for the given data. A quartic kernel was applied in (A7) and (A9), and $b = 3$ resp. $b = 4$ were used. Optimal bandwidths were determined according to asymptotic formulas. For FES as well as for EBF data $h = 0.2$ for (A7) and $h = 0.04$ (for A9) seemed to be appropriate choices over all years. The number κ of nearest neighbours required in (A10) was determined in such a way that with $h = 0.2$ the average number of data points falling into the intervals $[x_i - h, x_i + h]$ was equal to κ. All three methods lead to basically identical results. In particular, (A9) and (A10) were in close coincidence. Estimates proved to be quite robust with respect to different choices of h, κ. Positive semi-definite matrices $\hat{M}$ were found in all situations. Tables 7 and 8 show the results obtained by applying the power-transformed average derivative estimation method (A9).

Tests of positive semi-definiteness of M were based on bootstrap approximations to the distribution of the smallest eigenvalue of $\hat{M}$. This was done in a way completely analogous to the tests of positive semi-definiteness of $M(x), C(x)$. A theoretical justification for the use of bootstrap methods is given by Hrdle and Hart (1992). They only consider estimators (A7), but their results can easily be generalized to (A9) and (A10).

It remains to consider the matrix C. A $\kappa-$nearest neighbour approach yields $\bar{y}_{ij} = \frac{1}{\kappa} \sum_{r=1}^{n} y_{rj} \cdot I(x_r \in J_\kappa(x_i))$ as an estimate of $E(y_{ij}|x_i)$, $\quad j = 1, ..., N; i = 1, ..., n$. One then might define new "observations" z_{ij} by $z_{ij} = y_{ij} - \bar{y}_{ij}$. If κ is chosen appropriately, then $E(z_{ij} z_{il}|x_i) \approx cov_{jl}(x_i)$, and $C \approx \int \partial_x E(z_{ij} z_{il}|x_i = x)\rho(x)dx$. The matrix C can thus be estimated by applying average derivative estimation procedures to the data $\{z_{ij}\}_{i=1,...,n; j=1,...,N}$. More precisely, we use method (A10). The final estimator then looks as follows:

a) Determine parameters $\hat{a}_{i,jl}$ and $\hat{b}_{i,jl}$ by minimizing

$$\sum_{r=1}^{n} \left((y_{rj} - \bar{y}_{rj})(y_{rl} - \bar{y}_{rl}) - \alpha x_r - \beta \right)^2 \cdot I\left(x_r \in J_\kappa(x_i)\right)$$

b) Estimate C by

$$\hat{C}_{j,l} = \frac{1}{n} \sum_{i=1}^{n} \hat{a}_{i,jl}$$

Asymptotic results similar to those of ordinary average derivative estimators can be obtained. Applying this procedure to FES and EBF data lead to the result presented in Figures 5 and 6. The same κ as above was used.

æ

References

Family Expenditure Survey, Annual Base Tapes (1968- 1984): Department of Employment, Statistics Division, Her Majesty's Stationary Office, London. The data utilized in this paper were made available by the ESRC Data Archive at the University of Essex.

Enquête Budget de Famille (1979, 1984-85, 1989): Institut National de la Statistique et des Études Économiques, Division "Condition de Vie des Ménages ", Paris

Gasser, T., Kneip, A. and Kohler, W. (1991) : A flexible and fast method for automatic smoothing; *Journal of the American Statistical Association 86, 643 - 652*

Gasser, T. and Müller, H.G. (1984) : Estimating regression functions and their derivatives by the Kernel method; *Scandinavian Journal of Statistics 11, 171-185*

Härdle, W. (1990) : *Applied nonparametric regression;* Econometric Society Monographs, Cambridge University Press, Cambridge.

Härdle, W. and Hart,J. (1991) : A bootstrap test for a positive definiteness of income effect matrices ; *Econometric Theory, 8, 276-290*

Härdle, W. Hart, J. Marron, J.S. and Tsybakov, A.B. (1992) : Bandwidth choice for average derivative estimation; *Journal of the American Statistical Association 87, 218 -226*

Härdle, W., Hildenbrand, W. and Jerison, M. (1991) : Empirical evidence on the Law of Demand; *Econometrica 59, 1525-1551*

Härdle, W. and Park (1992) : Testing increasing dispersion; *CORE discussion paper, CORE, Université Catholique de Louvain*

Härdle, W. and Stoker, T.M. (Investigating smooth multiple regression by the method of average derivatives) : Journal of the American Statistical Association 84, 986 -995;

Härdle, W. and Scott, D:W: (1988) : Smoothing in low and high dimensions using weighted averages of rounded points; *Manuscript*

Jennen-Steinmetz, C. and Gasser, T. (1986) : A unifying approach to nonparametric regression estimation; *Journal of the American Statistical Association 83, 1084 - 1089*

Kemsley, W.F., Redpath, R.D. and Holmes, M. (1980) : Family Expenditure Survey Handbook; *London: Her Majesty's Stationary Office*

Müller, H.G. (1988) : *Nonparametric regression analysis of longitudinal data;* Springer Verlag, Berlin Heidelberg.

Nadaraya, E.A. (1964) : On estimating regression; *Theory Prob. Appl. 10, 186 -190*

Stoker, T.M. (1992) : Lectures on Semiparametric Econometrics; *CORE Lecture Series, CORE, Louvain-la-Neuve*

Wand, Marron and Ruppert (1991) : Transformations in density estimation; *Journal of the*

American Statistical Association 86, 343 - 353 (with discussion)

Watson, G.S. (1964) : Smooth regression analysis; *Sankhya Series A, 26, 359 - 372*

æ