Projektbereich A
Discussion Paper No. A–516

Invariant points of
low dimensional curve families

by

Klaus J. Utikal *)

February 1996

*) Klaus Utikal, Fakultät für Wirtschaftswissenschaften Universität Bonn; e-mail: utikal@track.econ2.de Financial support by Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 303 at the University of Bonn is gratefully acknowledged.

Contents

Abstract

From noisy observations of a finite family of functions an approximation in a lower dimensional space can be constructed using the method of principal components. If certain restrictions are to be satisfied by the approximation, e.g. being densities, this leads to a modified estimation procedure.
It is shown that in certain dimensions this will produce a family of curves which all intersect in the same points. This property may be interpreted in some cases as a characteristic feature of regularity of the data or as an artificial creation by the device in others.

Key words: kernel smoothing, density estimation, principal component analysis, nonparametric functional prediction

JEL C13, C14, C21

1 Introduction

Principal component analysis has been shown to be a useful tool in non-parametric curve estimation, see Silverman (1995), Rice and Silverman (1991). Given observations from a family of curves this method may be used to infer on the dimension of the space they span, see Kneip (1994) and provide an orthonormal basis which is optimal in a certain, well defined sense. For any fixed dimension this basis can be used to predict, cluster or classify the estimated functions Aubrey et al. (1980).

If the underlying family satisfies certain other conditions e.g. being densities or having common moments of some order, these conditions will generally not be conserved by its approximation in principal components. It is shown in the present article that by enforcing such conditions e.g. by renormalization, a common intersection can emerge as a new, prominent feature of the family which is absent when each of its members is estimated separately. It will then be left to the experimenter to draw conclusions on its existence and meaning for the underlying data generating process.

A second aspect is its potential use as data description/reduction technique. Transitions from one curve to another are sketched as deformations of a prototypical curve around certain fixed points. Strength or velocity of the deformation and its nature and direction can be estimated and,in an appropriate graphical representation, are proposed as tools for the study of change among curves successively observed in time.

2 Generalized separable families

We consider a bivariate function that can be written as a sum of products of pairs of univariate functions. These summand functions are called separable since they allow a separation of variables into a product, a property which makes them important in many applications. More precisely we consider a bivariate function of the form

$$f_t(x) = \sum_{j=1}^{p} \theta_t^{(j)} m_j(x) \tag{2.1}$$

a generalized separable family of dimension $p < \infty$ where $t \in T$, $x \in X$ are subsets of the real line. To avoid technicalities we require that no two functions are identical.

Moreover we assume the side conditions

$$l_i(f_t) = c_i \ , \ i = 1, \ldots, l-1, \ \ t \in T \tag{2.2}$$

where l_i are functionals and c_i constants. Denoting by $\underline{l}'$ the transpose of a vector $\underline{l}$, condition (2) can be written as $\underline{l}(f_t) = \underline{c}$, where $\underline{l}' = (l_1, \ldots, l_{l-1})$ and $\underline{c}$ similarly. We assume the constraints to be linear, i.e.

$$l_i \left(\sum_{j=1}^{p} \theta_t^{(j)} m_j \right) = \sum_{j=1}^{p} \theta_t^{(j)} l_i(m_j) \ .$$

We also require the rank condition

$$\text{rank } ([\underline{l}(m_1), \ldots, \underline{l}(m_p)]) = \text{rank } ([\underline{l}(m_1), \ldots, \underline{l}(m_p), \underline{c}]) = l - 1 \ . \tag{2.3}$$

LEMMA 2.1 : If $p = l$ then $f_{t_1}(x_0) = f_{t_0}(x_0)$ for any x_0 , t_0, t_1 implies $f_t(x_0) = f_{t_0}(x_0)$ for all t.

Proof : We may assume without loss of generality that the matrix $[\underline{l}(m_1), \ldots, \underline{l}(m_{p-1})]$ is of full rank. Denote the elements of its inverse by l^{ij}. Therefore the solution of

$$\sum_{j=1}^{p-1} \theta_t^{(j)} \underline{l}(m_j) = -\theta_t^{(p)} \underline{l}(m_p) + \underline{c}$$

can be written in terms of $\theta_t^{(p)}$ as

$$\theta_t^{(j)} = \sum_{i=1}^{p-1} l^{ij} \left(c_i - \theta_t^{(p)} l_i(m_p) \right) = \sum_{l=1}^{p-1} l^{ij} c_i - \theta_t^{(p)} \sum_{i=1}^{p-1} l^{ij} l_i(m_p) \ .$$

Hence

$$f_t(x) = \sum_{j=1}^{p-1} \sum_{i=1}^{p-1} l^{ij} c_i m_j + \theta_t^{(p)} m_p - \theta_t^{(p)} \sum_{j=1}^{p-1} \sum_{i=1}^{p-1} l^{ij} l_i(m_p)\, m_j$$

or

$$f_t(x) = \theta_t^{(p)} \xi(x) + \eta(x) \quad \text{where} \tag{2.4a}$$

$$\xi(x) := m_p - \sum_{j=1}^{p-1} \sum_{i=1}^{p-1} l^{ij} l_i(m_p)\, m_j \ . \tag{2.4b}$$

$$\eta(x) := \sum_{j=1}^{p-1} \sum_{i=1}^{p-1} l^{ij} c_j\, m_j \ . \tag{2.4c}$$

Hence, since f_{t_0}, f_{t_1} differ in at least one point, $f_{t_0}(x_0) = f_{t_1}(x_0)$ implies $\xi(x_0) = 0$ and therefore $f_t(x_0) = f_{t_0}(x_0)$ for all t.

□

COROLLARY 1: If a family of densities defined on some subset of the real line spans a two–dimensional space this family will intersect in at least one point. The same is true if all densities have equal mean and span a space of dimension three or if in addition to equal means have equal variance and span a space of dimension 4.

By approximating a given curve family by a generalized separable function satisfying side conditions, a common intersection in the approximating curves can be produced. If the dimension of the approximation is increased and therefore its goodness, this feature vanishes. However it can be forced to reappear if the approximated family is required to satisfy additional properties like the existence of common moments such as can be produced e.g. by a variable transform.

We note that under the conditions of the lemma it follows from (2.4) that f_t can be written as a time invariant function η on which a time varying component is superimposed in a time invariant direction. A simple interpretation in terms of velocity of change is made possible this way as is illustrated in the next section.

A study of the coefficient function $\theta^{(p)}(t)$ may reveal trends and thus be useful for the purpose of predicting future changes.

3 Applications

We outline an application of the proposed method to the study of change in the distribution of income. The data consists of large samples of disposable income of British private households taken cross-sectionally over the years 1968-1986. For a definition of the variables and further details see e.g. Section 2.4 of Hildenbrand (1994). The source of the data is the ESCR Data Archive at the University of Essex, Family Expenditure Survey, Annual Tapes 1968-1986, Department of Employment, Statistics Division, Her Majesty's Stationary Office, London. Densities of mean normalized income were estimated using a power transformed density estimator with power transform $g(x) = (0.1 + x)^{.25}$ proposed by Wand et al (1991) and a Gaussian kernel smoother provided by Splus. This yields the density estimates over nineteen successive years displayed in Figure 3.1 [left]. Assuming model (2.1) with $p = 2$ and the restriction (2.2) that $\int f_t(x)dx = 1$ for $t = 1968, \dots, 1986$ we use the estimators of Section 4 to approximate the curve family. The resulting curves are displayed in Figure 3.1 [right].

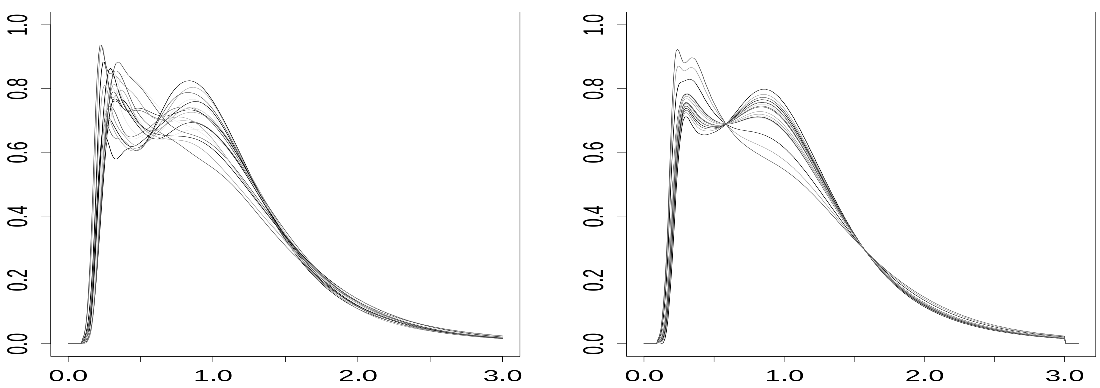

Figure 3.1: Disposable incomes of British households 1968-1986. Kernel density smoothing estimate for each year [left] and two-dimensional approximation [right].

From this picture it can be conjectured that while the income distributions change drastically over the years, a certain percentage of households whose incomes lie 40% above or below the mean income remain unaffected by the redistributions. A blowup of the region around 60% reveals a marked trend of household incomes being shifted from the center into lower income classes, from which the existence of an opposite but softer trend towards higher incomes can be concluded around the other fixed point of 1.5, see Figure 3.2. It is to be noted that the mean L^1-distance among the curves of 3.1 [left] and their correspondents in 3.1 [right] is 0.04. Without the restriction (3.2) the corresponding approximation yields a mean L^1-distance is almost the same. Using a three-dimensional unrestricted approximation this will be reduced to 0.02. This gain however is more than offset when two restrictions are to be brought in in order to apply Lemma 2.1. If in addition to the condition of area one we require that the mean of the three-dimensional approximations be equal to one this leads to a rise in the mean L^1-distance to 0.05.

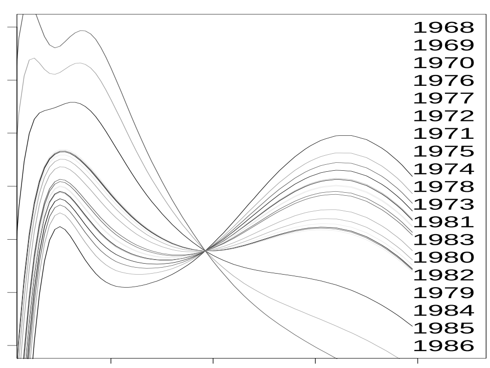

Figure 3.2: Trends in redistribution of incomes

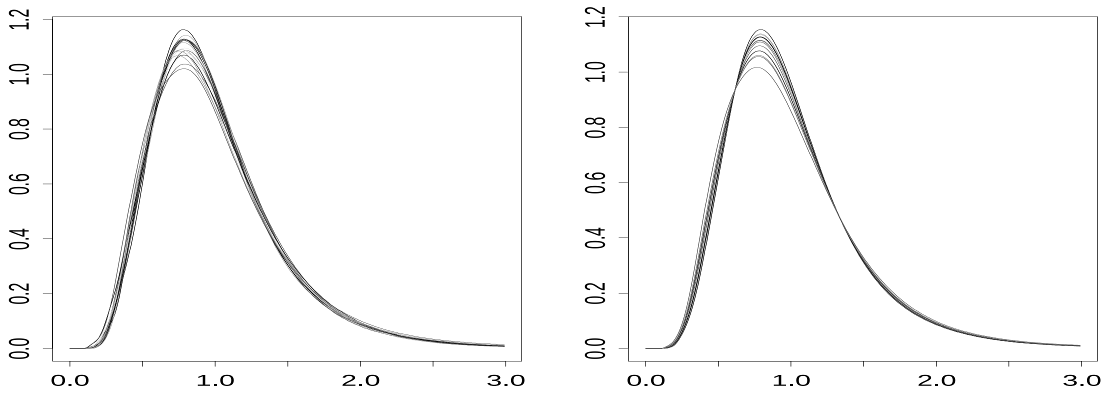

Figure 3.3: Disposable incomes of families whose head is full-time employed. Kernel density smoothing estimates for each year [left] and two-dimensional approximation [right].

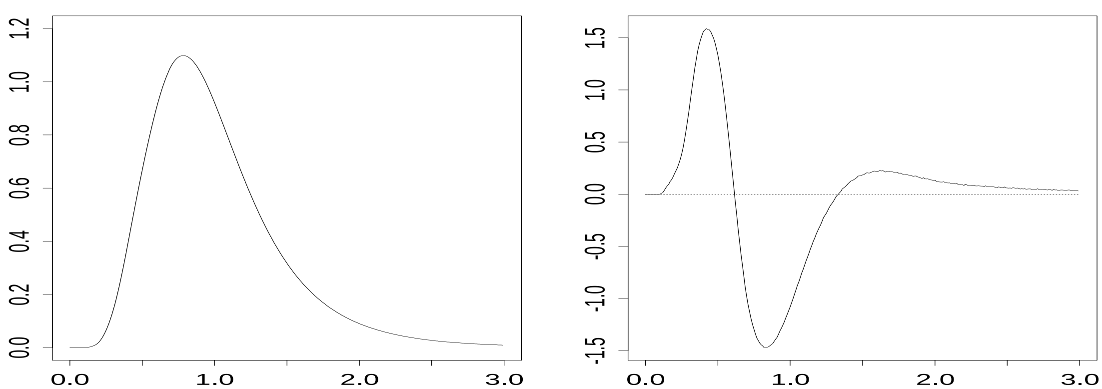

Figure 3.4: Basis functions of (constrained) two-dimensional approximation to Figure 3.3 [left]

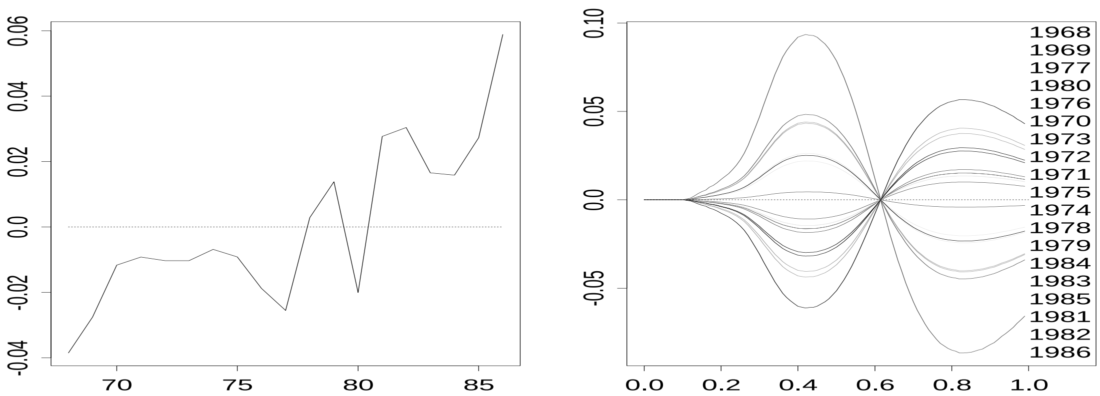

Figure 3.5: Evolution of coordinates [left] of transient component and deformation [right] of time invariant component of Figure 3.4.

A similar analysis is carried out on the income distribution of households whose head is full time employed. Yearly estimates yield a family of mean standardized income densities which appears approximately lognormal with slightly increasing variances though a Kolmogorov-Smirnov-test rejects this hypothesis at an alpha of ten percent in every year. Fitting model (2.1) with $p = 2$ under the restriction (2.2) of equal area one by the methods of Section 4 we obtain a family of densities with fixed points in the same area of relative income as for the total population. To understand the income dynamics of this subpopulation we focus on the basis elements η, ξ from representation (2.4) a graph of which is displayed in Figure 3.4. The curve in 3.4 [left], which can be shown to be a density, is invariant in time. The transient effects are captured by the curve in 3.4 [right] and by the corresponding coordinates as displayed in Figure 3.5 [left]. The transition of densities is therefore characterized by superposition of 3.4 [left] with 3.4 [right] weighted with coordinates from 3.5 [left], this way producing deformations by amounts displayed in 3.5 [right].

4 Estimation

We now present the estimation procedure applied in the previous section. The estimated curve families are defined as parameters of a general regression model. The estimators are described and computed by the algorithm given below, followed by recommendations for its implementation.

No claim is laid on the statistical properties of the estimators which are mostly obvious and part of a well developed theory, see Kneip and Gasser (1988), Kneip (1994). Application of the method to the problem of predicting future curves and observations are part of a general approach described in Engel and Utikal (1996).

Consider the following model. We observe continuous functions $f_1, \ldots, f_\tau$ in the presence of noise in finitely many nonrandom points $x_{t1}, \ldots, x_{t\tau_t},\ t = 1, \ldots, \tau$

$$\tilde{f}_t(x_{ti}) = f_t(x_{ti}) + \varepsilon_{ti}$$

where $\varepsilon_{ti} = \tilde{f}_t(x_{ti}) - f_t(x_{ti})$ are unobserved random variables. We assume that $f_1, \ldots, f_\tau$ are generated by a fixed set of p basis functions $m_1, \ldots, m_p$ satisfying (2.1) and linear constraints of type (2.2).

The problem here is to estimate for fixed p the unknown $\theta_t^{(j)}, m_j$. To assure identifiability we assume that $\{m_1, \ldots, m_p\}$ are eigenfunctions to the p largest eigenvalues $\lambda_1 \geq \ldots \geq \lambda_p$ of the $\tau \times \tau$-matrix $\int \underline{f}(x)\, \underline{f}'(x)\, dx$. In this notation $\underline{f}(x)$ is a columnvector with components $f_t(x)$ for $t = 1, \ldots, \tau$ and the transpose will be denoted by $\underline{f}'(x)$.

The values $\tilde{f}_t(x_{ti})$ were obtained by interpolating and smoothing on a common grid $x_1, \ldots, x_n$ where we suppose that $\tau < n$. In what follows we will not distinguish the notation between $f_t(x_i), \theta_t^{(j)}, m_j(x_i)$ etc. and its corresponding estimates $\tilde{f}_t(x_i), \tilde{\theta}_t^{(j)}, \widetilde{m}_j(x_i)$. We use vector and matrix notation where f is a $\tau \times n$-matrix of row vectors f_t, the rows of m are denoted by m_j and the columns of θ by $\underline{\theta}^{(j)}$. For given p we propose the following estimation procedure for θ, m. The constrained reconstruction $\hat{f}$ of f is now obtained in the following three steps.

A1: Compute the matrix P of eigenvectors corresponding to the p largest eigenvalues $\lambda_1 \geq \ldots \geq \lambda_p$ of the matrix ff'. Denote the diagonal matrix of eigenvalues $\Lambda = Diag(\lambda_1, \ldots, \lambda_p)$. Compute

$$\theta' = P'\Lambda^{\frac{1}{2}}$$

$$m = \Lambda^{-\frac{1}{2}} P' f\,.$$

A2: Then solve

$$\sum_{j=1}^{l-1} \hat{\underline{\theta}}^{(j)} l_i(m_j) + \sum_{j=l}^{p} \underline{\theta}^{(j)} l_i(m_j) = c_i$$

for $\hat{\underline{\theta}}^{(1)}, \ldots, \hat{\underline{\theta}}^{(l-1)}$ and $i = 1, \ldots, l-1$.

A3: The reconstruction of f is now given by

$$\hat{f} = \sum_{j=1}^{l-1} \hat{\underline{\theta}}^{(j)} m_j + \sum_{j=l}^{p} \underline{\theta}^{(j)} m_j \,.$$

The following remark concerns to choice of the dimension p in the case of $l = p - 1$. To trace the essential changes among given curves we use (2.4). While the timeinvariant term $\eta(x)$ is constructed from the first $p = 1$ principal components $\{m_1, \ldots, m_{p-1}\}$ only, it is the coordinate of the p^{th} component that drives the transient term $\theta_t^{(p)} \xi(x)$. As a rule m_p is tracing the low amplitude high frequency fluctuations which are strongly corrupted by noise, which over $\theta_t^{(p)}$ propagates into the reconstructed curve family via (2.4). Therefore, choosing a high dimension p in the interest of getting a good fit to the given curve family in terms of a low L^1-distance therefore can lead to an unexpected deterioration in the characterization of change among its members by using (2.4).

5 Summary

A method is presented by which a family of curves will be approximated by another family whose members all intersect in one or several common points. It is illustrated how these intersections can be interpreted as fixed points around which a prototypical curve pivots. Viewing the curve family as a process this method provides means of studying the trend in the underlying dynamic as well as of predicting future curves. A generalization of the method to two dimensions is possible and should best be applied when fixed points form a closed time invariant curve; other points of the surfaces are transformed in varying degrees which can be studied and graphically displayed by means similar to those illustrated presently in one dimension.

References

[1] Aubrey, D.G., Inman, D.L. and Winant, C.D. (1980). The statistical prediction of beach changes in Southern California. *J. Geophys. Res.*, **85,** 3264-3276.

[2] Engel, J. and Kneip, A. (1995). Model estimation in nonlinear regression. *Proceedings of the* 10^{th} *international workshop on statistical modelling, Innsbruck, Austria.* Springer, New York.

[3] Engel, J. and Utikal, K.J. (1996). Nonparametric functional prediction. *in preparation.*

[4] Hildenbrand, W. (1994). *Market Demand*, Princeton University Press, Princeton, New Jersey ,USA.

[5] Kneip, A. (1994). Nonparametric estimation of common regressors for simultaneous curve data. *Ann. Statist.* **22,** 3, 1386-1427.

[6] Kneip, A. and Gasser, T. (1988). Convergence and consistency results for self-modeling nonlinear regression. *Ann. Statist.*, **16,** 82-112.

[7] Wand, M.P., Marron, J.S, and Ruppert, D.(1991). Transformations in Density Estimation. *J.Amer.statist.assoc.* **86,** 414, pp 343-361.