Endogenous Prices in a Riemannian Geometry Framework

François Gardes

Documents de Travail du
Centre d'Economie de la Sorbonne

CES Working Papers

Endogenous Prices in a Riemannian Geometry Framework

François GARDES

2021.26

Maison des Sciences Économiques, 106-112 boulevard de L'Hôpital, 75647 Paris Cedex 13
https://centredeconomiesorbonne.cnrs.fr/
ISSN : 1955-611X

Endogenous Prices in a Riemannian Geometry Framework

July 22, 2021

François Gardes
Paris School of Economics, Université Paris I Panthéon Sorbonne,
Western Catholic University[1]

Abstract

Economic agents decide often under different price conditions (named endogenous prices), and moreover these individual price systems may depend on their previous choices. Such a situation appears for instance for consumers' choices under constraints which imply virtual costs or when non-monetary resources add to the monetary budget constraint. In case where no market exist which regulates these prices and make them converge toward a common price for all agents, observed differences appear between the social distribution of consumer expenditures and their change over time which are modelized here using Riemannian geometry. Observations in a cross-sectional survey are supposed to constitute a Riemannian surface (where each point is associated with a particular price system). Social differences are measured along the geodesics of the Riemannian surface, while changes over time correspond to movements along their tangent spaces (characterized by constant endogenous prices). The Riemannian curvature of the consumption space is thus estimated comparing the derivatives over the surface (corresponding to cross-section differences) to those of the tangent plane (corresponding to the time changes). The Riemannian curvature being shown to be non-null for the Polish consumers surveyed in a four years Polish panel, implies that usual econometric methods based on a unique metric over the (cross-sectional) consumption space are inadequate to estimate geodesics (corresponding

[1]Centre d'Economie de la Sorbonne, 106–112 Boulevard de l'Hôpital, 75647, Paris cedex 13, France. This paper has benefited remarks from seminar participants at University of Québec in Montréal, Caen, Paris X, Clermont-Ferrand, Strasbourg and JMA congress, and from J.L. Arcand, S. Bazen, J. Boelaert, J.F. Caulier, R. Dos Santos Fereira, F. Laisney, P. Merrigan, V. Simmonet, Christophe Starzec and Alain Trognon. Riemannian geometry and its applications to physics have been a locus disputandum since many years with Philip Merrigan (UQAM), implying large benefits for my analysis. The author thanks Professor Gorecki and Christophe Starzec for allowing him to use the Polish panel data set and Nicholas Sowels for revising the text.

to optimal choices) on the Riemannian surface. The curvature of the survey can be linked to changes in latent endogenous prices over the Riemannian surface, which are for instance full prices in a domestic production framework. Finally, the Riemannian structure is used to study the path dependency of consumers' choices.

Keywords: spatial autocorrelation, Riemannian geometry, curvature, virtual price, full price, path dependency.

JEL Classification: C31, C33, C55, D11, D12

Introduction

This article considers the intrinsic geometric properties of a dataset, such as a survey of family budgets, as a tool to get some information on the hidden effect of latent variables in the model. Consider for instance cases whence cross-section estimates differ from time-series, the first one can be biased by the existence of endogenous *permanent* latent variables (such as education level), the effect of which disappears in the time dimension. Another case of permanent latent variables is the relative position of a household in the income distribution, supposed to remain constant through time, which may imply specific constraints or social interaction which determine its economic choices. Comparing two households with different (relative) incomes in the cross-section dimension thus integrates different effects of the relative locations of agents on the income distribution, while the observation of the same household between two periods reveals only the effect of a change of its current income conditional to a constant relative income. With panel data, it is possible to take these difficulties into account and to show that differences still appear between estimates in the temporal and spatial dimensions (see Gardes et al., 2005)[2].

Consider a curve on a surface (x, z) in R^{n+1}, with z a set n explanatory variables and x the explained variable. The curve $x = g(z)$ can be considered as immersed in the smooth surface in R^{n+1} corresponding to all observations of (x, z) given by the cross-section. This surface is thus a hypersurface of dimension n in the space of dimension (n+1) which corresponds to an economic model defined by the equation relating x to z. The *intrinsic geometry*

[2]The difference between cross-section and time-series estimates, recognized early in the literature, is not well accepted by the profession, as it implies abstaining from making dynamic inferences from cross-section estimation (differences between agents observed in the same period does not provide the same information as changes over time for the same individual or the same population). This difference has been suggested to result from aggregation biases, from the time when panels of individuals were not available. Specification biases (whenever the estimates performed on surveys do not take into account dynamic behaviour, such as habits or addiction in consumption functions) and different effects of errors in variables in the two dimensions have also been advanced as potential explanations.

of the surface contains all geometric characteristics which could be observed from within the surface, that is without any information which is not given by the dataset, such as the influence of latent variables. For example of such an intrinsic property, Gauss proved in his Theorema Egrenium, that, for a n-dimensional hypersurface in R^{n+1}, the product of the largest and the smallest among the principal curvatures of the surface (named Gaussian curvature) can be measured from within the surface: this Gaussian curvature is an intrinsic property of the surface. Therefore, although the principal curvatures are not intrinsic, a particular combination of them is.

In this article, surfaces corresponding to a survey are considered as Riemannian manifolds and the tangent planes at each point of the surface are linked to the model conditional to a fixed value of all latent variables which determine x at this point (i.e. for the corresponding values of variables x at this point). In a 3-dimensional space (for instance consumption x determined by income per unit of consumption z_1 and age of the family head z_2), the tangent space at a point (x, z) is defined as corresponding to the variation of (x, z) conditional to a normal to the tangent plane equipolent to the normal at the surface in M (i. e. conditional, in the (n+1)-dimensional space, to the fixity of the basis over the tangent plane). Subsequently, this normal vector will be assimilated to a vector of the shadow prices corresponding to z_1 and z_2, including the influence of (potentially endogenous) latent variables for each x. Therefore, the tangent plane indicates changes in x unrelated to the influence of latent variables (at least independently from the influence of latent variables for a move over the surface). For instance, consider panel data and the difference between the estimate obtained on the cross-section (by a between transform for instance) and the estimate in the time-series dimension (by a within transform). This last estimate can be considered as the one corresponding to changes in x conditionally to no change in a set of permanent latent variables W (such as cohort or education effects which can be supposed to be constant over time). The time-series estimate thus indicates the gradient of x over z conditionally to constant W and are no more biased by potential endogeneity of these variables.

Note that in a Riemannian space all variables play the same role as concerns the curvature of the space. The method in this article measures the particular curvature which is created by the effect of a correlation between the residual in a model and the explanatory variables, this correlation being due to endogenous latent variables which explain part of the residual. It differs from the curvature of the algebraic structure of the model $x = g(z)$ which can be calculated by usual method for a 2 or 3-dimensional space (see Appell, 1900, chapter 9 and section 2.1 in this article) and tensor analysis for larger dimensions (see for instance Jeanperrin, 2000) and which has for instance been analyzed as concerns the power of income in the calculation of the rank of a demand system (Lewbel, 1990). The Riemannian curvature adresses to the particular model g which is analyzed and indicates the remaining change in the hypersurface corresponding to this model, which makes it non-euclidian (meaning that a unique distance cannot be applied to the whole surface).

I consider here that spatial (cross-sectional) relationships correspond to geodesics in a Riemannian space, while temporal relationships are modelled as movements along tangents to these surfaces. Surprisingly, Riemannian geometry has been only little applied to economic problems (it has, however, recently appeared in theoretical statistics). I show that tensor algebra allows us to analyze the difference in the estimations in both dimensions, and to compute the curvature of the Riemannian space. When this curvature is non-zero (i. e. when the integrability conditions, which make it possible to define a common Euclidean metric for all points of the space, do not hold), the space is no longer Euclidean, and the shortest route between two points are not lines but geodesics, which may bring about new econometric problems. Agents situated on a Riemannian space are supposed to follow optimal pathes (along their life-cycle) which are geodesics over the surface. Thus observing the geodesics gives some information on the time structure of cross-sectional data. Geodesics are defined by a system of differential equations the estimation of which allows to recover some of the geometric properties of the Riemannian space. Therefore, considering these equations may permit to estimate dynamic models using only cross-sectional data (see Gardes 2020). Riemannian geometry allows also to test for path dependency, as shown in the last section.

Section 1 and Appendix A present a synthesis of Riemannian geometry and its application to the specification of an economic model. Section 2 proposes an original method to evaluate the Riemannian curvature of Family Budget surveys, and discusses the application of usual econometric methods on cross-sections. Section 3 examines the shadow (endogenous) prices which correspond to the geometric non-euclidian properties of the surface corresponding to a survey and compare them to a system of full prices integrating the value of time spent in domestic activities. Section 4 discusses the test of path dependency due to asymmetric effects of consecutive changes in two explanatory variables. All of these models are derived for households expenditures data but can easily be applied to other types of statistics.

1 The Riemannian Geometry of consumption

1.1 Metric and tangent surfaces

Consider the set of n variables consisting of the n_1 variables at choice (for instance expenditures) x_i^h and the n_2 characteristics z_k^h of an economic agent h. The variables x and z define a point M with coordinates $v = (x, z)$ in $\mathbb{R}^n$, with $n = n_1 + n_2$. A cross-section consists in N such points for all agents surveyed, a time-series in the dynamic configuration of T points for each agent. We suppose that z_i is independent from z_j for all $j \neq i$ and that the characteristics are exogenous ($\frac{\partial z_k}{\partial x_{k'}} = 0,\ k \neq k'$). Suppose N is large and the observations $v = (x, z)$ describe a smooth surface over $\mathbb{R}^n$. At each point, we define a tangent space by means of the gradient of x at this point. As this gradient changes from one point M to another infinitely close point $M + dM$ the derivative corresponding to these two close points on the surface will differ from the gradient at each point. Thus, the gradient at a point is related to the time variations of the variables, while the derivative (named absolute derivative) over the surface describes the cross-section variations. Riemannian geometry consists in connecting together the tangent surfaces corresponding to close points by linear transformations, so that the Riemannian surface can be locally represented by an Euclidean space, the metric of this space defining the scalar product and the distance between two points on the tangent surface. If the integrability conditions are satisfied, these Euclidean metrics can be embedded in an Euclidean metric for the whole Riemannian space, which becomes an Euclidean space. In this case the Riemannian curvature is zero. Conversely, if the curvature differs from zero, the space cannot be considered as Euclidean and one cannot define an Euclidean metric over the whole space.

Appendix A recalls briefly some elements of the differential analysis of Riemannian surfaces which are needed for the comparison of cross-section and time-series derivatives.

1.2 Computation of Christoffel symbols

I propose to equate the cross-section marginal propensities with the covariant derivatives $\nabla_k v^i$ and the between-period propensities with the time derivatives $\partial_k v^i$; I therefore compute the Christoffel symbols as

$$\nabla_k v^i - \partial_k v^i = \Gamma^i_{kh} v^h \tag{1}$$

according to the Levi-Civita connection described in Appendix A. The covariant derivatives of these Christoffel symbols $\nabla_s \Gamma^i_{kh} = \Gamma^i_{kh;s}$ can be obtained by estimating the equation:

$$\Gamma^i_{kh} = \Gamma^i_{kh;s} \cdot v^s + \gamma^i_{kh} + \varepsilon^i_{kh} \tag{2}$$

The system of these two equations reduces to:

$$\nabla_k v^i - \partial_k v^i = \Gamma^i_{kh;s} \cdot v^h v^s + \gamma^i_{kh} v^h + \varepsilon^i_k. \tag{3}$$

Another method of estimating the differences between the absolute and partial differentials of v^i would be to use the formulas defining geodesics in Riemannian space: it is well known that the acceleration is null along geodesics, so that:

$$\frac{\mathrm{d}^2 v^i}{\mathrm{d}t^2} + \Gamma^i_{kh} \frac{\mathrm{d}v^h}{\mathrm{d}t} \frac{\mathrm{d}v^k}{\mathrm{d}t} = 0. \tag{4}$$

This equation must be estimated along a geodesics. It thus necessitates defining the path followed by an individual (for instance changes in its consumption and determinants of consumption during its life cycle) taking into account the changing costs along this move.

The Christoffel symbols can therefore be calculated: first, by means of the comparison between absolute and partial derivatives (equation (1)); second, by the estimation of geodesics corresponding to an index of time defined on the cross-section (equation (4)).

1.3 Consequences

Three consequences follow from the calculus of the Chritoffel symbols: first, according to equation (1), the nullity of Christoffel symbols implies the equality of the corresponding covariant and time derivatives[3]: there is no more endogeneity biases which would imply that the cross-section behavior differs from the time series. In that case, the cross-section estimate can be interpreted in terms of changes through time and geodesics are straight lines defined by a null acceleration through time (the first term of equation (4)). A second situation corresponds to non null Christoffel symbols with integrability condition *for points*: in that case, the Cristoffel symbols are symmetrical and there is no path dependency, but the cross-section and time-series estimates may differ. Third, integrability condition *for vectors* is related to no rotation of the basis from one point to another: this situation corresponds to a nul Riemann-Christoffel tensor of curvature. It implies that there exists a system of curviline coordinates defined linearly from the original coordinates, such that a quadratic differential form can be defined as a fixed metric on the whole space.

In a Riemannian space, a metric is attached to each point by the equation which defines locally the distance between two arbitrary close points M and $M' = M + \mathrm{d}M$: $\mathrm{d}(M, M') = g_{ij}\,\mathrm{d}v^i\,\mathrm{d}v^j$. Christoffel symbols are related to the linear element g_{ij} defining the metric by the formula:

$$g_{ih}\Gamma^h_{kj} + g_{jh}\Gamma^h_{ki} = \partial_k g_{ij}, \quad \text{with} \quad \partial_k g_{ij} = \frac{\partial g_{ij}}{\partial v^k} \tag{5}$$

Christoffel symbols thus indicate second-order derivatives and can be used to calculate the curvature of the space. These formula allow us to calculate Γ in terms of the metric:

$$\Gamma^h_{ki} = g^{jh}\Gamma_{kji} = \frac{1}{2}(\partial_k g_{ij} + \partial_i g_{gk} - \partial_j g_{ki}) \tag{6}$$

[3]Note that this equality is also obtained with non nul symbols if the sum $\Gamma^i_{kh} v^h$ is zero.

Therefore, the curvature of the space depends finally on the linear element g defining the metrics which is attached to each point, and which can be represented as a square matrix over indices (i, j) of the different variables.

In the case of a null curvature, the Riemannian surface can therefore be assimilated to an Euclidian space where the minimization of a distance between observations is an adequate method to obtain an estimate of a model. In order to use this minimization, it is necessary to define subspaces of nul curvature where a constant metric can be defined. This metric describes in a sense how a particular situation M is related to its close neighbours. It defines the situations, in terms of variables v, towards which the individual will move when the determinants of its explained variables change. The set of these situations is the geodesic passing through M. Thus, the metric[4] is related to the specific costs (endogenous prices) associated with each situation (see Section 3). These endogenous prices (corresponding to endogeneity effects due to latent variables) are therefore supposed to vary all over the surface.

The Riemannian structure of the consumption space can finally be described as follows:

1. $\Gamma^i_{kh} \neq 0 \iff$ the estimated cross-section behavior differs from the time-series behavior[5].

2. $\Gamma^i_{kh} \neq \Gamma^i_{hk} \iff$ path dependency of choices (here consumption choices). For instance, two consecutive changes can affect a household: first an increase of its income (both adults working on the market), then the birth of a child (implying that only one spouse works on the market), or the reverse two changes. These two inverse sequences, supposed operating during the same length of time, may result in different socio-economic situations for similar households, and therefore to different final position on the surface.

3. $R^h_{irs} \neq 0 \iff$ the differential equations $\mathrm{d}e_i = \omega^j_i e_j$ are not integrable, i. e. there is a torsion of the basis defined on the different points in the space. In that case, the same metric cannot be applied to the whole surface (corresponding to a survey).

4. The sign of R indicates the sign of the Riemannian scalar curvature at each point.

[4]which plays the role of gravitation in physics (describing the properties of the space by the gravity attached to each of its point).

[5]Note that the difference between cross-section and time-series parameters depends on the agent's location on the Riemannian surface, i. e. the endogeneity produced by the curvature of the space is not the same for all agents. The Christoffel symbols are calculated locally, so that endogeneity is also defined locally

2 The Riemannian curvature of the consumption space

2.1 Algebraic curvature of a model

The curvature of a Riemannian surface is usually recovered by two alternative methods: either the metric g_{ij} is calculated using the relationship between a particular basis on the surface (corresponding for instance to polar coordinates) and the Euclidian basis of the space. In that case, the Christoffel symbols can be calculated on each point by equations (5) and (6). Another method calculates directly the curvature of the equation which generates the surface. For instance for a surface in a three dimensional space given by the equation $x = g(z_1, z_2)$, the total curvature is given in terms of the two first order derivatives p, q over z_1 and z_2 and the three second order derivatives r, t and s over (z_1, z_1), (z_2, z_2) and (z_1, z_2): $K = \frac{rt-s^2}{(1+p^2+q^2)^2}$ (see Appell, chapter 16). For instance, a Working specification writes the budget share w_i in terms of log-income per Unit of consumption $\frac{y}{n}$ and log-price:

$$w_i = \alpha_i + \beta_i log(\frac{y}{n}) + \gamma_i log(\pi_i) + \varepsilon_i \tag{7}$$

Its total Gaussian curvature is:

$$K = \frac{-n^2\gamma_i(\beta_i + \gamma_i)}{\left[n^2 + n^2(w_i + \beta_i)^2 + \frac{\gamma_i^2 y^2}{\pi_i^2}\right]^2} \tag{8}$$

which is positive (respectively negative) when $\gamma_i < 0$ and $\beta_i < -\gamma_i$ (respectively $\gamma_i > 0$ and $\beta_i > -\gamma_i$). The estimation of food expenditure on the Polish panel used in this article gives the following estimates for the average values of income, price, food budget share and family size: 588.8 zlotis, 1.15, 0.396 and 2.48. The estimates of the parameters are: $\beta_i = -0.0327$; $\gamma_i = -0.190$ which gives an average total Gaussian curvature $K = -3.63e - 10$ [6].

For the log-linear model $log(z_i) = \alpha_i + \beta_i log(\frac{y}{n}) + \gamma_i log(\pi_i) + \varepsilon_i$, the curvature is:

$$K = \frac{\frac{(\beta_i - 1)\beta_i^2 \gamma i}{\frac{y}{n}^3 \pi i} - (\gamma i - 1)^2 \frac{\gamma i^2}{\pi_i^4}}{\left[1 + \frac{\beta_i^2}{\frac{y}{n}^2 z_i^2 + \frac{\gamma_i^2}{\pi_i^2} z_i^2}\right]^2} \tag{9}$$

[6]Average of the curvatures estimated on each observation. The curvature calculated on the average values is -7.41 e-10

The estimates of the parameters are: $\beta_i = 0.212$; $\gamma_i = -0.192$ which gives an average total curvature $K = -8.79e-6(s.e. -4.97e-6)$ for average values) [7].

The curvature of the model (on the estimated surface) does not indicate the entire curvature of all observations in the survey, since it does not take into account the error term and its possible correlations with the set of explanatory variables (given rise to endogeneity corresponding to endogenous prices). The following method estimates this residual curvature comparing the derivatives estimated by means of the model which defines the Riemannian surface estimated on the survey.

2.2 Empirical curvature of a survey

I propose in this article an original method based on absolute and tangent derivatives in equations (1) and (3). Let us suppose that the survey containing n variables (here the partial expenditure for some good i, its relative price, the household's income and various other covariates such as the age of the head) corresponds to a surface where the explained variable (here the partial expenditure) depends on the other variables. This gives rise to a (n-1) dimension hypersurface the curvature of which can be related to the Christoffel symbols which describe the geometric intrinsic properties of the surface. The survey is supposed to be observed on several periods, constituting a panel or allowing the construction of a pseudo-panel. The relation between these Christoffel symbols and the Lévy-Civita connection is used to evaluate empirically these parameters and subsequently the Riemannian curvature.

The derivatives $\Gamma^i_{kh;s}$ of Christoffel symbols are computed as indicated in sub-section 1.2 by estimating equation (3). $\nabla_k v^i$ is the derivative on income (i.e. on log-income per UC in a linear Working specification such as equation (7)) estimated in the Between dimension. $\partial_k v^i$ is the corresponding parameter after a within transformation. These parameters are supposed to change from one point on the surface to another. Several estimations of these derivatives could be obtained by grouping the individuals observed in the panel into cells defined by exogenous variables (such as education level, location and age groups). Under some hypothesis on the grouping method, the absolute derivative $\nabla_k v^i$ could be calculated on the cross-section dimension within cells, while the time derivative $\partial_k v^i$ is estimated between the different observations of individual in the same cell and in different

[7] Dividing by x_i (in order to compare the models) gives comparable figures for the normalized curvatures: $-37.7e-9(s.e. -21.3e-9)$ for average values) for the log-linear model and $-9.18e-10(s.e. -1.87e-10)$ for the Working specification. The Working specification is therefore much less curved than the log-linear model, which is an indication of its better modelization of food consumption (not as concerns its better adjustment in terms of minimum residuals but because of its algebraic specification).

periods in order to measure these two derivatives over each cell. A similar method could be applied to a pseudo-panel of repeated cross-sections. These parameters could also be recovered at each point of the space by a local estimation using a non-parametric method.

Another method is used here. We estimate a system of four equations [8]. based on a particular specification (here the linear, log-linear and Working models) on a four years Polish panel (1997-2000) presented in Appendix B: the first four equations relate the budget share w_{ith} of good i in period t for household h to $z_{1t} = lny_{ith} = log(\frac{y_{ht}}{n_{ht}})$ the logarithm of total expenditure per unit of consumption, $z_{2t} = ln\pi_{ht}$ the logarithmic full price changing from one household to another (the definition of which is presented in Appendix C) and $z_{ht}, h > 2$ a vector of covariates describing the households' characteristics.

This set of equation is completed by equations in difference corresponding to equation (3):

$$\partial_k v^i = \nabla_k v^i - (\Gamma^i_{kh;s} \cdot v^h v^s + \Gamma^i_{kh} v^h) + \varepsilon^i_k. \tag{10}$$

which are applied to the data as indicated in Appendix D.

The Riemannian curvature is computed by means of equation (20) in Appendix A:

$$\begin{aligned} R^h_{irs} &= \partial_r \Gamma^h_{is} + \Gamma^h_{hl}\Gamma^l_{is} - \partial_s \Gamma^h_{ir} - \Gamma^l_{ih}\Gamma^h_{sl} = \nabla_r \Gamma^h_{is} - \nabla_s \Gamma^h_{ir} \\ &= (\gamma^i_{ks;h}\nabla_r v^h + \gamma^i_{ks;r}) - (\gamma^i_{kr;h}\nabla_s v^h + \gamma^i_{kr;s}) = \gamma^i_{ks;h}\nabla_r v^h - \gamma^i_{kr;h}\nabla_s v^h \end{aligned}$$

for $h \neq s$ and r, where γ^i_{rsv} is an estimate of $\Gamma^i_{rs;v}$ for good i. Note that in equation (11) the γ are supposed to be symmetrical in (r, s).

The Riemann-Christoffel tensor R^h_{irs} describes the curvature properties of the Riemannian surface. This tensor depends on four indices i, k, h, j and can be contracted to the Ricci tensor by equalizing k with h: $R_{ij} = R^k_{ikj}$. This tensor can finally be contracted equalizing i and j and multiplying by the metric g_{ij} to give rise to the *scalar Riemannian curvature*: $R = g_{ij}R_{ij}$. The $\frac{n(n-1)}{2}$ weights g_{ij} defining the metric depend on the Christoffel symbols by means of $\frac{n^2(n-1)}{2}$ differential equations (6) which cannot be easily resolved. Therefore, this scalar curvature R necessites calibrating the g_{ij} (for instance by the inverse of the variance-covariance matrix of vector u). In case of a surface of two dimensions in a three dimensional space (a variable x, for instance consumption, depending on two determinant variables z_1, z_2, for instance income per unit of consumption and the commodity price), the two dimensional Ricci tensor R_{ij} occurs only two times: for $k = z_1$, $i = z_1$ and $j = z_2$ (or its negative inversing i and j) and for $k = z_2$, $i = z_2$ and

[8] 4n equations for a system of n commodities x_i.

$j = z_1$ (or its negative inversing i and j). All other alternatives (such as $k = z_1$, $i = z_2$ and $j = z_2$) are nul. The two tensors R_{223} and R_{332} thus measure the principal curvatures of the surface (its maximum and minimum curvatures) which allow to calculate the total Gaussian curvature as R_{223} R_{332} and the average curvature as $\frac{R_{223}+R_{332}}{2}$.

Note that the empirical residual curvature has no reason to be the same for all models, since it is calculated conditional to the algebraic curvature incorporated in this model.

2.3 Estimation on a Riemannian space: a variable metric

The consequence of curvature is that the metric changes from one point to another: on a Riemannian space, the differential metric writes $\mathrm{d}s^2 = g_{ij}\,\mathrm{d}v^i\mathrm{d}v^j$, with $\mathrm{d}v^i$ and $\mathrm{d}v^j$ the changes in variables v^i and v^j. Therefore, minimising the Riemannian distance over the whole space corresponds to calculate the usual estimate by Maximum Likelihood or least squares, constraining the optimization by the change of the metric from one point to another: for instance, the metric at each point may change according to the non-monetary resources and the constraints, which characterize this location in a Riemannian space[9]. Therefore it is not possible to recover consumption functions by minimising a unique distance over the whole space: the distance between some point M and its estimate M_1 depends on the coefficients g_{ij} which are attached to this point, if the two points are sufficiently close so that the same metric applies to them. For another remote point M', it will be necessary to take into account the change of the metric when computing the distance from its estimate M_1'. Moreover, if points M and M' are not sufficiently close, the metric changes continuously on every path between them, so that it is necessary to determine the minimum path between them in the Riemannian space, i. e. a geodesic between the two points, by integrating the differential of the metric ds between the two points. This poses the problem of path dependency discussed in section 4.4.

A solution would be to project the Riemannian space on an Euclidian space, where a unique distance is defined. It is well known that every Riemannian manifold of dimension n can be embedded in an Euclidian space of greater dimension. The first embedding theorems were proved by Janet and Cartan, the Euclidian space necessitating at least $\frac{n(n+1)}{2}$ dimensions, but the general theorem was proved by extraordinary methods by John Nash (1956): his Theorem 3 realizes the imbedding of a non-compact n-manifold into an Euclidian space of dimension $\frac{n(n+1)(3n+11)}{2}$. That projection is complex and besides, calculus on the corresponding euclidian space is generally much more difficult than on the Riemannian surface. The only general result

[9]The curvature of the Riemannian space thus indicates the heterogeneity of these conditions of choice over the population, which could be taken into account by changing costs (endogenous prices) over the survey (under some conditions), see section 3.

is the possibility to develop a geodesic curve of the Riemannian surface on an associated euclidian space, such that the metric of the Riemannian surface is projected on an osculatory metric on the euclidian space (such that its linear element and its first derivatives are the same than on the Riemannian space). But that projection is conditional to the particular geodesic curve and does not apply for the whole space.

Geodesic curves are characterized by a nul acceleration of the movement along them: for instance, in case of a three dimensional space (one variable x depending on two variables z_1 and z_2), the tangent lines on the different points of a geodesic are parallel, which means that derivatives along the geodesic do not change because of a rotation of the basis along the curve (no endogeneity bias appear between the time derivatives and the covariant ones along the geodesic curve). A way to obtain time derivative (avoiding the endogeneity bias of the cross-section estmates) is therefore to perform the estimation along geodesic lines between all couples of points.

Geodesics represent the shortest paths from one point to another. So, an agent can be supposed to follow a geodesic over her life cycle. The definition of a time dimension on a cross-section allows estimating Christoffel symbols along geodesics. Considering for instance a survey on households, one solution consists in taking into account the age of the family head as measuring the passage of time in the survey (or, for a survey on firms, the number of years the firm has been active in the market). The past and future consumption of some household h in a dynamic model (representing habit and addiction effects) can then be instrumented by the expenditures of similar households (by age, education, location, family structure) one year younger or older in the survey[10]. One drawback is that age is correlated with cohort effects in cross-sections. I propose (see Gardes, 2020) a formula to correct this cohort effect and find that such instrumentation is very effective compared to usual instruments such as past and future prices.

Once time variations can be calculated using this synthetic index of time, the system of equation corresponding to a geodesic (equation 4) allows estimating Christoffel symbols. These symbols could then be used to correct the covariant derivatives (obtained by the estimation of the model on the Riemannian space) by mean of the Levi-Civita connection (equation 5) in order to calculate the time derivatives independently from endogeneity biases corresponding to the curvature of the survey.

Another method could rely on the consideration of all geodesic between all couples of observations on the survey: $\frac{n(n-1)}{2}$ such couples exist, so that

[10]This correspond to an instrumentation of the past value of some variable by means of its value for a younger similar (matched) agent. Note that it is generally necessary when estimating dynamic models to instrument past values of the variables, even when these past values are observed, so that the lack of information for this method could be similar to usual instrumentation of past values used for the estimation of dynamic models on individual data.

the estimation must be performed on the set of observations corresponding to each of these geodesic lines. Considering two points on the survey, the set of potential successive positions of an economic agent in a survey (such they could be observed through time in a panel) can be supposed to correspond to the most probable track to go from the first position to the second, so that it can be supposed to minimize the cost of displacement between the two points. It is therefore a geodesics corresponding to the cost of this displacement. As will be discussed in the following section, these geodesic lines can also be assimilated to sub-sets where shadow prices (linked to endogeneity biases) are supposed to remain constant, so that all endogeneity biases due to the change of these shadow prices (i.e. to a rotation of basis on the Riemannian space) disappear: the estimates correspond therefore to those which would have been made on the tangent planes (for the three dimensinal case) corresponding to a fixed basis (a constant normal vector to the tangent plane). The development of that method is left to a future research.

3 Endogenous prices

Endogenous latent variables provoke endogeneity biases which imply the curvature of the space containing the survey. In the three dimensional case (for instance an expenditure explained by total expenditure and direct price), the corresponding torsion of the basis from one point to another close point implies the rotation of the normal vector to the tangent plane together with a movement of the two vectors in the basis which define that tangent plane. Suppose M moves over the surface to $M' = M + dM$ according to changes in (z_1, z_2) instead of remaining on the tangent plane linked to the original position M (on a point m' corresponding to the new values of (z_1, z_1). The difference in x between M' and m' can be attributed at the first order to the change in z_1: $dx = \sum_i \frac{\partial x}{\partial z_{1i}} dz_{1i}$, which defines, under a supposed direct price effect, an inverse change in the price π_1 of that variable z_1 keeping constant the value of the other covariates [11]. Moving over the surface can therefore be attributed to a change in the relative prices of covariates z_j. As the normal vector can be associated to the relative price of the two explanatory variables, the torsion of this vector measures the change in prices over the surface. These prices changing from one point to another on the surface are named endogenous prices since they depend both on the agent's characteristics and her previous choices. That change depends on the particular surface which describes the data set, and therefore on the particular model used to describe that survey.

Consider for instance the case of a curve x dependeing on two covariates z_1 and z_2 the cost of which are noted π_1 and π_2: $x = g(z_1, z_2)$. A movement

[11] or the relative prices of all determinants in the n-case.

on the surface from point M to $M' = M + dM$ creates a rotation of the normal vector n such that: (i) the projection of $n : p(n)$ on an orthogonal basis of the covariates has a different lenght; (ii) $p(n)$ operates a rotation which can be indicated by its angle with the first axis. The first consequence implies an homogenous change of the effect of the costs of the covariates (by means of price-elasticities supposed to remain constant over the surface). The second can be represented by a change of the relative price $\frac{\pi_1}{\pi_2}$, which determines a second change of x. The rotation of the normal vector to the tangent plane thus indicates the price effect of the changes of the price level and of the relative price of the two covariates: these movements over the surface correspond to the effect of shadow prices depending on the position on the surface.

These virtual prices correspond to the effect of latent variables which create the curvature between two close points. An important example of endogenous prices is the shadow price of a constraint: for example, Neary and Roberts (1980) compares demand theory without or under rationing by decomposing the set of commodities into a subset of goods that can be freely chosen and a second subset of rationed goods. They show that, under usual condition for the existence of implicit functions (quasi-concavity of the utility function and differentiability at the optimum), the restricted demand functions can be considered as unrestricted choices with virtual prices defined as those which would induce an unrationed agent to choose the same bundle set as the rationed agent.

The case of full prices

Another example corresponds to non-monetary resources, such as time used in domestic production, which add the cost of time to monetary prices into total (full) costs. This model of the allocation of time implies that the economic costs may differ between households, even if monetary prices are the same (which in fact is not usually the case as prices are known to vary according to the agent's location, sometimes its age or other socio-economic characteristics such as its income class). As a third example, Barten (1964) explained by the public nature of some consumptions, why, in a simple model of the allocation of goods in the family, the relative price of individual consumptions (such as milk) compared to a public consumption such as heating, varies according to the demographic structure of the household (being greater for larger families). These total economic costs are particular example of endogenous prices.

In the case of the allocation of time where consumption depends both on monetary costs and the use of time, Becker's full prices (integrating the value of time used in domestic production with the monetary prices of commodities) correspond to the shadow prices arising from time considered as a non-monetary resource used in the domestic production from market goods and domestic time. Whence the domestic production is supposed to follow a Leontief technology with complementary monetary and time factors (as in the original Becker's model), the full unit price of a commodity (the final good to be consumed) produced with a quantity x_i of a unique market good i and an amount of time t_i writes: $p_i x_i + \omega t_i$ with an opportunity cost of time ω which is usually taken as the agent market wage rate net of taxes or the the minimum wage rate on the labor market. The unit full price of the market good i is therefore: $p_i + \omega \tau_i$ where $\tau_i = t_i / x_i$.

An important difficulty in the estimation of the domestic production model lies in the valuation of time. Gardes (2019, 2020) shows that a Cobb-Douglas specification for both the utility function and the domestic production functions allows estimating locally (for each household) the opportunity cost of time by means of the first order conditions for the substitution between time and monetary resources used for the domestic production[12] The full prices depend on the households'characteristics, both through the household's opportunity cost of time and its domestic production technology. Two definition of full prices are proposed in Appendix C: first under the assumption of substituability between the two factors of production of the activity, time and money (as supposed by Becker and Michael, 1973, and Gronau, 1977), second under complementarity (as in Becker's original model, 1965). With these definitions, it is possible to proxy the changes in the full prices, observing only monetary and full expenditures. It can be expected that the variations of these full prices are correlated with shadow prices corresponding to movements on the Rieamnnian surface and therefore to its Riemannian curvature, since part of the endogeneity biases may be due to the absence of the time resource in the modelization of consumption.

4 Empirical application

4.1 Curvature

The absolute and tangent partial derivatives $\nabla_k v^i$ and $\partial_k v^i$(identified with the cross-section and time-series derivatives) have been estimated on the Polish panel (1997–2000, see Appendix B for a description of the data) containing data over 3052 households. Prices are defined as full prices under the assumption of substituability of the domestic production factors.

The absolute and tangent partial derivatives $\nabla_k v^i$ and $\partial_k v^i$ depend on the specification of the function between the explained variable (partial expenditure) and the set of explanatory variables. Therefore, the estimated empirical curvature indicates whether there remains some unexplained endogeneity in the cross-section estimation, i.e. an effect of endogenous per-

[12]A value of time specific to each type of consumption can also be computed by a model based on a CES specification of both the utility and the domestic production functions, see Gardes (2020).

manent latent variables over the cross-section dimension, which disappears for variations through time. In that case, the specification is not sufficient to take care of this remaining endogeneity. The comparison between the curvatures found for different specifications (for instance the linear, log-linear and Working models used here) can be interpreted as a non-parametric specification test, in the sense that a specification takes care more completely of the endogenous biases created by permanent latent variables. The level of the remaining curvature (once the curvature linked to the specification has been taken into account) indicates the existence of an endogeneity bias when estimating on the surface (i.e. using the dataset given by the survey) instead of estimating on tangent planes where latent variables are constant.

Riemannian curvature	All households	Education 1	Education 2	Education 3	No child	Children
R_{223}	-0.0663	-0.1117	-0.1232	-0.3403	-0.1279	-0.0114
s.e.	(0.0226)	(0.0604)	(0.0413)	(0.0927)	(0.0398)	(0.0355)
R_{332}	0.1850	-0.1409	0.1322	0.7569	0.1925	0.2058
s.e.	(0.0463)	(0.673)	(0.1024)	(0.1895)	(0.0750)	(0.0833)
$\mathrm{R}_{223}R_{332}$	-0.0123	-0.0166	-0.0163	-0.2576	-0.0246	0.0023
s.e.	(0.0015)	(0.0232)	(0.0040)	(0.0260)	(0.0035)	(0.0021)
N	12,133	3,575	7,426	1,132	5,641	6,492

Surveys: four waves of the 1987–1990 Polish panel (3,630 households). *Note:* standard error in parentheses. Bootstrap for the Gaussian curvature $R_{223}R_{332}$.
Specification of the surface: $w_{iht} = \alpha_i + \beta_i \ln m_{ht} + \gamma_i \ln \pi_{iht} + Z_{ht}\delta_i + \varepsilon_{iht}$.
Explanatory variables: lnm_{ht}= logarithmic total expenditure per Consumption Unit, $\ln\pi_{iht}$= logarithmic full price, Z_{ht}= log age of the head and survey dummies.
Data: The Polish panel covering 3,052 households for the period 1997–2000.

Table 1: Curvature for food expenditure (log-linear specification)

Table 1 presents the estimated curvatures corresponding to the endogeneity bias on the income effect (R_{223}) or on the price effect (R_{323}) in the log-linear specification [13]. The Riemannian curvatures are then normalized, being multiplied by the common normalization factor which is, both for R_{223} and R_{323}, equal to the ratio of price over partial expenditure multiplied by the square of income (all these variables being defined in terms of their transformation in the model, for instance in logarithm for the log-linear specification). Estimation over the whole population in all four waves (Table 1, log-linear specification; table D2, linear and Working models) generates, for two third of the commodities, a Riemannian curvature significantly different from zero regarding the derivative k with respect to total expenditure for all consumptions R_{223}. The second principal curvature R_{323} is significantly

[13] The estimates for other expenditures and sub-population defined by age classes or family composition are presented in Appendix. Also, estimates corresponding to the linear model and the Working specification can be found in Appendix (Tables D1 and D2).

different from zero for all commodities (except transport and clothing in the linear specification). This second principal curvature is generally greater (in absolute terms) than the first for the three models. Finally, the Gaussian curvature (obtained as the product of the two principal curvatures) has generally different signs and magnitude for these models, which is normal since that curvature indicates the endogeneity bias due to permanent latent variables the effect of which depends on the specification which is adopted[14].

The estimation for sub-populations (Tables 1 and D1 in Appendix) by age group (20-34, 35-55, over 56), three levels of education, family types (singles, two adult without children, families with children) and income indicates also significant curvatures. The curvatures differ between sub-populations, as concerns their sign as well as their level, which is normal as each expenditure is impacted by specific endogeneity biases due to specific latent variables. In all specifications, the presence of children in the family increases the curvature, which can be explained by the fact that those families have more differentiated conditions of choices (linked to different costs, constraints and information) than families without children: these conditions depend indeed on location, the number of children and perhaps specific characteristics such as education level and religion, which may impart less differentiation for families without children. On the contrary, curvatures are similar for singles and family composed of two adults, which shows that these differences do not depend on the family size (through economies of scale). Also, households in the median age class (head aged 35 to 59 years) are characterized by a greater curvature, i.e. more different conditions of choice. The curvature is also greater for rich households (above the median income per unit of consumption), which shows that insufficient income impose similar economic conditions for poor and inferior middle-class households. The estimates on sub-populations are a little less significant than for the whole population, which could be explained by the greater homogeneity of these sub-populations and a weaker effect of latent variables between similar households.

The curvature for the linear model is significant for almost all commodities (21 over 24 estimations at 5%) and generally greater than for the double-log model (17 only significantly different from zero) which means that the assumption of a constant income elasticity fits better the dataset than supposing the constancy of the marginal propensities[15].

[14]Note that the algebraic curvature of the Working model (+1.39e-5) has the same magnitude as the empirical Gaussian curvature of the survey: 1.38e-5. However, both curvatures have no systematic relation between them.

[15]Note that the empirical curvatures differ from the algebraic curvature corresponding to the specification of the model. They correspond in a sense to the curvature of the dataset once the algebraic curvature of the model has been taken into account by the projection of the empirical dataset on the surface corresponding to the specification.

4.2 Endogenous and full prices

4.2.1 Full prices

Full prices depend on macroeconomic conditions, since the opportunity cost of time diminishes during recessions (also characterized by changes in households' allocation of time, see Alpman and Gardes' analysis of the great recession in the US, 2018). They depend on the opportunity cost of time and on the amount of time consecrated to each activity. Their variation is quite large across the population (for instance, the full price for transport varies from 1.003 to 6.66, with a mean 2.3 and a standard error 2.0). The opportunity cost of time and the time used may both depend on the household's level of being, its situation in its life cycle (family size, age of the head and the spouse), its location. A regression analysis of the logarithm of full prices on these variables shows that their relation to income and age of the head are generally quadratic: decreasing with income for two third of the distribution, except for food the full price of which is increasing with an elasticity 0.17. Full prices increase with age with a minimum around 30 years for food, housing, transport and leisure and increase on the whole distribution of age for clothing and other expenditures. They decrease with the proportion of children in the family, except for clothing where the elasticity to children is -0.4. The variability of full prices is not clearly linked to the Gaussian curvature estimated for the whole population. An analysis using local estimations would be a best test of this potential relationship.

4.2.2 Shadow endogenous prices

Suppose the endogeneity bias which is indicated by the discrepancy between the time-series and the cross-section derivatives can be explained as a shadow price effect, with a matrix of price elasticities $E_{x/\pi}$:

$$A_{k,j} = \nabla_k z^j - \partial_k z^j = E_{x/\pi} dln\pi \quad (11)$$

This equation allows estimating the derivatives (in the cross-section dimension) of the shadow prices π over the explanatory variables z:

$$dln\pi = E_{x/\pi}^{-1} A_{k,j} \quad (12)$$

If the cross-price elasticities are neglected, this derivative over an explanatory variable z_j can be recovered for each commodity x_i as the product of the inverse of the direct price elasticity $E_{x/\pi^{-1}}$ with the difference of the cross-section and time-series derivatives of x_i over z_j:

$$dln\pi_i = \frac{\beta_{j,cs} - \beta_{j,ts}}{E_{\frac{x_i}{\pi_i}}} \quad (13)$$

Finally, a first order approximation of the change in the shadow price over the surface can be calculated as the sum of the product of its derivatives over the set of explanatory variables z_j, with the change of these explanatory variables between two observations (details in Gardes, 2004, 2019).

	Food	Housing	Transport	Other expenditures
Income elasticity of the full price	0.777	-0.673	-1.129	-0.248
Standard error	(0.039)	(0.033)	(0.038)	(0.012)
s.e. of the log-shadow price	0.493	0.349	0.624	0.124
s.e. of the log-full price	0.200	0.293	0.613	0.270
Gaussian curvature (Log-lin model)	138 e-6	-3.96 e-6	-1.90 e-6	4.02 e-6

Surveys: four waves of the 1987–1990 Polish panel (3,630 households).

Table 2: Income elasticities and variances of the full prices, Polish panel 1987-1990

The full price increases with income only for food, since this expenditure is characterized by a cost depending on the opportunity cost of time which is larger for rich households[16].

Boelaert, Gardes and Langlois (2018) applied this analysis of shadow prices (corresponding to the discrepancy between cross-section and time-series estimates of the income effect) to a pseudo-panel of seven Canadian Households Expenditures surveys from 1978 to 2008. The shadow prices are shown to change along the income distribution, increasing with relative income for Food at Home, Alcohol, Tobacco and Private Transport, decreasing for Public Transport, Health expenditures and Education. The recent widening of income inequality in Canada implies that these virtual prices are more unequal in the whole population, which make the conditions of the economic choices more unequal between rich and poor households (a supplementary inequality between them). The virtual prices are also shown to be correlated with full prices (Table 5 in Boelaert et al.), which are indeed an important component of the economic cost.

4.3 Estimating the time derivatives on a cross-section

The tangential (time-series) derivative $\frac{\partial g_{ij}}{\partial v^k}$ can be estimated as the sum of the absolute (cross-section) derivative and of the correction term in the Levi-Civita connection depending on the Christoffel symbols: $\nabla_k v^i = \partial_k v^i + \Gamma^i_{kh} v^h$. Suppose these Christoffel symbols are estimated on the 1997-2000

[16]The same phenomenon appears for full rices, but with a smaller income elasticity: 0.087 (0.009). Income elasticities of Housing and Other Expenditures (Clothing, Leisure, Other) are negatives, while it is positive for Transport, the full cost of which depends also heavily on the opportunity cost of time

panel and applied to the 1987-1990 Polish panel (being multiplied by the variables measured on this panel) in order to estimate the correction factor, the cross-section estimated estimated in Gardes et al. (2005) can be corrected by this correction factor and compared to the estimate performed in the time dimension.

Considering food at home, the following Table shows that the cross-section elasticity (0.494) differs from the time-series one (0.755) which have been estimated on the panel, while the corrected cross-section elasticity (0.888) is close to it. The estimates made on a pseudo-panel of the Polish surveys (supposed to correct errors of measurement on individual data) gives a perfect correction of cross-section estimates.[17] The difference between the cross-section and the time-series income elasticities have been explained in Gardes et al. (2005, see Appendix E) as corresponding to a price effect, the endogeneity bias on the cross-section corresponding to a shadow price correlated to the household's relative position on the income distribution. [18] Dividing this difference (estimated as the correction factor) by a price elasticity (estimated on full prices on the Polish surveys as -0.9) allows to estimate the (relative) income elasticity of this shadow prices as 0.44 (s.e. 0.17) which shows that the full cost of this consumption is indeed greater for the rich, which explains the smaller income-elasticity in the cross-section dimension.[19]

	Cross-section elasticity*	Correction factor***	Corrected elasticity	Time-series elasticity**
Panel estimates	0.494	-0.394	0.888	0.755
	(0.010)	(0.166)	(0.194)	(0.013)
Pseudo-panel estimates	0.589	-0.394	0.983	0.965
	(0.013)	(0.166)	(0.166)	(0.023)

Surveys: four waves of the 1987–1990 Polish panel (3,630 households): Between (cross-section *) and Within (time-series **) estimations (instrumented total expenditures). ***Correction of cross-section parameters: $\mathbb{E}^{\text{c.s.}} - \Gamma^{i}_{kh} v^{h}$, with h total expenditures.

Table 3: Cross-section and time-series income elasticities for food expenditures, Polish panel 1987-1990

This shows that the estimation of the intrinsic geometric properties of

[17]The differences between the actual time-series estimate and the corrected cross-section amount to 18% (individual panel data) and 2% (pseudo-panel data) of the within estimates and are not significantly different from zero

[18]Rich households have a larger opportunity cost for time, which increases the full cost of food at home compared to the poor

[19]Note that the estimate on these surveys of the income elasticity of the full price (under complementarity) for food is also positive: 0.11 for income measured by total expenditure, 0.34 for income given by the survey (with probable errors of measurement)

the consumption survey (measured by the Christoffel symbols) may allow classifying the expenditures according to the endogeneity bias, and to predict the sign of this bias, using only the information given by one survey. These Christoffel symbols can be estimated, either on an independent dataset containing information both on the cross-section and on the time-series dimensions, or by means of the estimation of the equation of geodesics (equation 4) using only the survey on which the cross-section parameters are estimated. [20].Therefore, the estimation of the Christoffel symbols along an average geodesic on one survey gives an information on the difference between cross-section and time-series food consumption laws and may allow to estimate unbiased parameters using only cross-section data.[21]

4.4 A Test of path dependency

The consecutive changes of two determinants of consumption (for instance an increase in income followed by a decrease of the relative price of the commodity, which may both expand consumption) could have a different effect according to the order of these changes. This can be due for instance to a variation of the sensibility to prices after a large increase of the household's level of being, or to the variation of the shadow price of the commodity (for instance its full price if the opportunity cost of time increases with the level of being). In case of a smaller price effect (due to a rise in the household's level of being), the total increase of consumption due to these two consecutive changes may be smaller than the effect of the reverse changes. Changes in household's size (birth of a child) and income offer another example of non-symmetric effects: a birth modifies the necessary expenditures, which can be empirically recovered by the inclusion of family size in the consumption function (or taking into account Barten's shadow prices for each commodity), but it will affect the income and price elasticities of all commodities, because the domestic production technology of the household is modified by this birth (for instance through the increase of the opportunity cost of time or an increasing marginal utility of public family goods).

Vilfredo Pareto has been commonly critizised for his faulty identification of the integrability of consumption functions with the absence of effect of the order of consumptions on utility. In an euclidian geometry, this order cannot be identified looking at the global effect of consumption changes on utility since the zero curvature of the space implies path independency. The same difficulty exist as concerns the order of changes in explanatory variables. On the contrary, this non-symmetry can be recovered in a Riemannian geometry. Path dependency corresponds indeed in a Riemannian

[20] This method implies defining a time scale in the cross-section dataset, for instance by means of the age of the household's head (after a correction of cohort effects linked to this age, see Gardes et al, 2014)

[21] The application of this method is left for a future research.

geometry to the non-symmetry of Christoffel symbols which implies the non-integrability conditions *for points* (i.e. of the positions on the surface, see section 1.3 and Appendix A): in that case, there does not exist an integrand function (defined locally) such that a displacement from one point to another changes this integrand independently from the path followed. Integrability for vectors (i.e. elasticities or marginal propensities) depends actually on the absence of torsion of the basis corresponding to different points on the Riemannian surface (which means that the Riemann-Christoffel tensor of curvature is nul).

Such non symmetric changes must be sizeable in order to be statistically measurable and correspond to long term changes (more than one year for a dataset on households) in order to give rise to different sizeable consequences. For instance, if the family size is impacted by the birth of a new child, the income effect may be estimated on more than one following year in order to test for path dependency[22]. As concerns changes of the family size, our dataset, covering four years, gives rise to only three changes between two consecutive years, which is perhaps insufficient for a robust test of path dependency. Indeed, no significant path dependency appears for our dataset when the two explanatory variables the symmetric effects of which are tested are income and family size. On the contrary, consecutive changes in income and prices are more frequent in the four years panel.

Table 4 presents a Fisher test for symmetrical paths of income and full price changes. The model in equation (11 and D1) with symmetric coefficients γ_{223} and γ_{323} (for symmetric orders of changes in income and prices) is compared to the enlarged model with non-symmetric coefficients. The estimation is made for three cases: a change in income precedes the change in price (case a), compared to the reverse changes (c) and periods with only changes in income defined as case (b). The difference between the symmetric and the non-symmetric models is significant for five commodities over one (Other Expenditures), with the largest significance for Food and Transport. It appears also that the two coefficients, γ_{223} and γ_{323}, change continuously from situation (a) to (b) and finally (c), which proves that income and price elasticities differ more between the case where the variation in income precedes the price change and the reverse case, than compared to the case (b) where no order exists between these changes. Path dependency is thus verified but the non-symmetry for each coefficient and commodity is often small (except for food and transport for which differences are greater than 30%).

[22]Expressed in terms of shadow prices, some time is necessary before their variation is perceived and taken into account by economic agents.

Commodity	Food	Housing	Transport	Clothing	Leisure	Other
γ_{223a}	-0.3270	0.2606	0.0236	0.2957	-0.1512	-0.4105
s.e.	(0.0694)	(0.0835)	(0.0140)	(0.0665)	(0.0334)	(0.0805)
γ_{223b}	-0.3142	0.2889	-0.0103	0.3138	-0.1429	-0.4053
s.e.	(0.0692)	(0.0828)	(0.0116)	(0.0636)	(0.0334)	(0.0802)
γ_{223c}	-0.2567	0.3306	-0.0524	0.3585	-0.1269	-0.4013
s.e.	(0.0692)	(0.0831)	(0.0116)	(0.0648)	(0.0334)	(0.0801)
γ_{323a}	-0.2285	-0.5718	-0.0975	0.0752	0.0315	-0.2776
s.e.	(0.0872)	(0.0822)	(0.0072)	(0.0368)	(0.0146)	(0.0423)
γ_{323b}	-0.2626	-0.5889	-0.0926	0.0752	0.0273	-0.2931
s.e.	(0.0873)	(0.0827)	(0.0063)	(0.0356)	(0.0146)	(0.0404)
γ_{323c}	-0.3361	-0.6221	-0.0784	0.0772	0.0233	-0.2818
s.e.	(0.0872)	(0.0830)	(0.0064)	(0.0366)	(0.0146)	(0.0405)
Fisher	(22.97)	(6.30)	(22.84)	(3.30)	(9.22)	(1.57)

Surveys: four waves of the 1987–1990 Polish panel (3,630 households). *Note:*standard error in parentheses. γ_{223b} are the income coefficients, γ_{323b} the price coefficients. Fisher limit=3.11. (H0)$(\gamma_{223a}, \gamma_{323a})$=$(\gamma_{223borc}, \gamma_{323borc}$, (H1) $\gamma(a) \neq \gamma(borc)$.

Table 4: Test of path dependency between income and price

Datasets giving information over long-term changes (containing consecutive periods characterized by the change in the first variable followed by consecutive periods with changes in the second variable) would give more robust test of path dependency. Perhaps pseudo-panel data could be used for this purpose.

Conclusion

We have shown[23] that the consumption space has a Riemannian structure, which allows us to connect the time dimension of consumption laws to the social distribution of consumption expenditure in the population. Riemannian curvature of consumption data means that there exists, for the social distribution of consumption, path-dependency with respect to the order of the changes in the different variables influencing consumption choices. For instance, comparing a couple of two adults to a family with children in a cross-section allows computing an equivalence scale which may depend not only on income levels, but also on income changes, in the sense that an increase in family size $\mathrm{d}S$ for a family of size S and income $y_1 = y_0 + \mathrm{d}y$ may not give rise to the same levels of expenditure as an increase in income $\mathrm{d}y$ for a family with size $S + \mathrm{d}S$ and income y_0 (whenever the condition for point integrability, i. e. the symmetry of Christoffel symbols, does not hold).

[23]All these analyses can be generalized to the case of a system of explained variables depending on a set of more than two explanatory variables.

Thus, considering the Riemannian structure of the expenditure space may be useful for the classic identification problem in equivalence scale models[24].

Second, the impossibility of defining a unique metric for the whole population means that usual econometric estimations of consumption laws on cross-section data are misleading. This is due to the fact that local conditions of choice, which correspond to local shadow prices, are not taken into account in the estimation. In a sense, the Riemannian curvature of the consumption space can be related to social heterogeneity, and the change of the basis from one point to another in this space indicates the variation in the various constraints and non-monetary resources which influence consumer choice by means of the associated shadow prices. The conditions of choice depend on the situation of the agent, i. e. her location in the Riemannian space. Barten's (1964) discussion of the change in relative monetary prices due to changes in family composition can be considered as an example of this relationship.

Finally, the geometric structure of a survey could be considered to estimate dynamic models on a cross-section. For a survey on households, it could consists in taking into account the age of the family head or, for firms, the number of years the firm has been active on the market, as measuring the passage of time in the survey. For instance, the past and future consumption of some household h in a dynamic model (representing habit or addiction effects) can be instrumented by the expenditures of similar households (by education, location, family structure) one year younger or older in the survey. One drawback is that age is correlated with cohort effects in cross-sections. To correct this cohort effect, one can use the fact that in theories such as the life cycle, both the dependent variable and its determinants are related to the time dimension. For instance, considering savings, income per unit of consumption increases early in the life cycle, then falls as household size increases, and finally increases to a constant level at the end of the life cycle, while the savings rate also varies systematically over the life cycle. Therefore, income and age can substitute each other *locally* as concerns their influence on savings. For example, a rich household aged 35 may have the same saving behavior as a poor household aged 50, so that income compensates for age in determining savings. Similar relations as those considered with the head's age can be found for other household's characteristics (such as its demographic composition), so that a combination of all may be more efficient to date each household on the time axis. Two methods to correct for cohort effects are presented in Gardes (2020). These corrections have been applied to the estimation of a dynamic model of consumption and to the estimation of Christoffel symbols using the equation of geodesics (which relates the second order derivatives over time of some variable at choice with a linear combination of first order derivatives

[24] I am grateful to Alain Trognon for this suggestion.

with coefficients depending on the Christoffel symbols). Defining such a synthetic time index, the estimation of Christoffel symbols can be used to correct for the endogeneity biases of cross-section estimates and make them comparable to time-series estimations (cancelling the endogeneity biases due to permanent latent variables).

Appendix A: Elements of Differential Geometry and Riemannian surfaces

Definitions

Consider that $v = (x, z)$ pertains to a n-dimensional manifold V_n, associated with a domain E^* of $\mathbb{R}^n$ (the atlas at point (x, z)) by a C^p-diffeomorphism $(p \geq 2)$:

$$m = (v^1, \ldots, v^n) \in \mathrm{E}_n^* \subset \mathbb{R}^n \to M = f(m) \in V_n \tag{14}$$

Here v^i are the coordinates of M on the chart E^*. This means that every neighborhood of a point M_0 in V_n can be represented by the system of n coordinates $\{v^i\}_{i=1}^n$. V_n is said continuously differentiable at order $p \geq 2$ if other systems of coordinates, $\{u\}$, are related to $\{v\}$ by a C^2-diffeomorphism $u^i = g(v)$. A natural basis $\{e_i\}$ corresponds to these coordinates for point M_0 such that $OM_0 = \sum_i v^i e_i = v^i e_i$, using the Einstein convention (summation runs over the indexe which appears both as a subscript and a superscript). So $\frac{\partial M}{\partial v^i} = e_i$. As E_n^* is Euclidean, we can define its metric at each point M_0 by the quadratic form :

$$\mathrm{d}s^2 = g_{ij}\mathrm{d}v^i\mathrm{d}v^j \tag{15}$$

with $g_{ij} = e_i e_j$. The g_{ij} are arbitrary continuously differentiable (at order p) functions of the y^i. So far, V_n is a topological manifold covered by compatible C^p coordinate charts, with a "Riemannian metric" g at each point which is any smooth positive definite matrix[25].

We can now define the tangent surface at m_0 (the corresponding point of $M_0 \in V_n$ in E^*) by the equation:

$$m_0 m = \left[(v^i - v_0^i) + \Psi^i(v^i - v_0^i)\right] e_i$$

with Ψ^i a second order function with respect to $(y^i - y_0^i)$. This equation defines the first-order representation in the neighbourhood of M_0. Hence the system of coordinates v^i also applies to point m in the neighbourhood of M_0. It can be shown that any change in the system of coordinates from $\{y\}$

[25]Note that this is not really a more general setting than $\mathbb{R}^n$, as proved by the Nash theorem—every abstract Riemannian manifold can be *isometrically* embedded in some $\mathbb{R}^m$, m>n, but the calculus is usually easier in V_n.

to another system $\{v'\}$, with $A_i^k = \frac{\partial v^i}{\partial v'^k}$, changes the metrics by the formula $g_{ij} = A_i^k A_j^l g_{kl}$, so that the metric does not truly depend on the system of coordinates: it is an intrinsic notion. A second-order representation of the tangent surface can also be defined in M_0 so that the Euclidean metric on E^* and the Riemannian metric on V_n have the same coefficients with the same derivatives (they are said to be connected metrics for $v^i = v_0^i$).

From point M in V_n to another point $M + \mathrm{d}M$ (with $\mathrm{d}M = \{\mathrm{d}v^i\}$) which is infinitely close to M, the basis $\{e_i\}$ changes according to the formula:

$$e_i + \mathrm{d}e_i = e_i + \omega_i^j e_j \tag{16}$$

with ω_i^j defined as a linear function of the changes $\mathrm{d}y^i$ of the coordinates of M:

$$\omega_i^j = \Gamma_{ki}^j \mathrm{d}v^k \tag{17}$$

This change is related to the change of the metric, since:

$$g_{ij} = e_i e_j = g_{ij} = \frac{\partial x}{\partial v^i} \partial x \partial v^j \tag{18}$$

Coefficients Γ are named Christoffel symbols.

If the metric g satisfies the *integrability conditions*, there exists a system of coordinates in E_n^* such that the metric in E_n^* takes the form of equation (**??**) on all points of E_n^*. This means that g is a continuous function on E_n^*. These conditions, which apply to the differential equations relating the two points and the associated basis ($\mathrm{d}M = \mathrm{d}v^i e_i$ and $\mathrm{d}e_i = \omega_i^j e_j$) require the symmetry of the second derivatives of M and e_i, and can be written: $\Gamma_{ki}^j = \Gamma_{ik}^j$ for all i, k on all points $\{v^i\}$. They allow us to calculate the n^3-scalar Γ_{ki}^j knowing the $\frac{n(n+1)}{2}$ values g_{ij} at each point.

The Levy-Civita linear connection

Consider now the change from M to $M + \mathrm{d}M$. The natural basis $\{e_i\}$ assigned to point M changes to $\{e_i + \mathrm{d}e_i\}$ on point $M + \mathrm{d}M$. Thus, any vector $v = v^i e_i$ is differentiated as

$$\mathrm{d}v = \mathrm{d}v^i e_i + v^i \,\mathrm{d}e_i = \mathrm{d}v^i e_i + v^h \omega_h^i e_i$$

The components of the vector $\mathrm{d}v$ write $\nabla v^j = \mathrm{d}v^i + \omega_h^i v^h$ and are denoted the *absolute differential* of v^i. The corresponding partial derivatives[26] are:

$$\nabla_k v^i = \partial_k v^i + \Gamma_{kh}^i v^h. \tag{19}$$

[26] The absolute (or covariant) derivatives are the component of a tensor, i. e. they change according to certain formulas when the basis changes, while the ∂_k cannot be considered as tensors. Thus, only the ∇_k correspond to the change from M to $M + \mathrm{d}M$ on the Riemannian surface.

This defines the Levy-Civita connection between the tangent spaces.

We identify the *absolute derivative*, which includes the change of coordinates along the tangent surface and the change of the natural basis, with the cross-section marginal propensities on the Riemannian surface. These marginal propensities correspond to the overall change between two points M and M + dM on the surface, while the derivatives $\partial_k v^i$ are associated with the change of the coordinate v^i along the tangent surface (i.e. for a constant basis)[27].

The estimation of the cross-section and time-series parameters for each point of the surface thus allows us to calculate $\Gamma^i_{kh} v^h$ by means of equations (11). Supposing that these parameters are constant over a neighbourhood of M_0 allows us to calculate the coefficients Γ^i_{kh} as the parameters of the coordinates $v^h(M)$ for M in this neighbourhood.

In the application, we consider $v = (x, z)$ as containing the expenditures x_i of households on n commodities, and the characteristics z_k of the household, such as its income (or total expenditure), family size, education level of the head, location, monetary prices and so on.

Path dependency and Cartan's method

Consider now Cartan's *quasi-parallelogram* (Cartan () and Ivey and Haudsberg (): an initial change from M_0 to $M_1 = M_0 + \mathrm{d}M_0$ is followed by a second change from $M_0 + \mathrm{d}M_0$ to $M_1' = [M_0 + \mathrm{d}M_0 + \delta(M_0 + \mathrm{d}M_0)]$. The two derivatives d and δ can be commuted:

$$M_0 \;\Rightarrow\; M_2 = M_0 + \delta M_0 \;\Rightarrow\; M_2' = M_0 + \delta M_0 + \mathrm{d}(M_0 + \delta M_0).$$

Cartan shows that points M_1' and M_2' coincide if and only if the Christoffel symbols are symmetrical:

$$\mathrm{d}\delta m_0 - \delta \mathrm{d} m_0 = (\Gamma^h_{ki} - \Gamma^h_{ik})\, \mathrm{d}v^k\, \delta v^i e_h = 0 \iff \Gamma^h_{ki} = \Gamma^h_{ik},$$

which is the *integrability condition for points*, with m_0 being the point corresponding to M_0 on the related Euclidean surface.

As M varies on the Riemannian surface, the basis also changes and is submitted to a rotation which is measured by the Riemann—Christoffel Tensor of Curvature. Suppose the two differentiations defined by Cartan are used one after another. The difference of the change of the basis according to

[27] The variation of the basis could be related to changes of costs (including monetary prices and shadow prices) over the surface. These costs are specific to each observation, i.e. each point on the surface, and can be named endogenous prices. On the contrary, all costs are supposed to be constant on the tangent plane. This plane thus indicates all possible positions (for different values of the explanatory variables z) conditional to constant costs.

the order of differentiation can be computed as: $\mathrm{d}\delta e_i - \delta \mathrm{d}e_i = R^h_{irs}\, \mathrm{d}y^r\, \delta y^s e_h$, and this difference is null if the Tensor of Curvature disappears:

$$\mathrm{d}\delta e_i - \delta \mathrm{d}e_i = R^h_{irs}\, \mathrm{d}y^r\, \delta y^s e_h = 0 \iff R^h_{irs} = \partial_r \Gamma^h_{is} + \Gamma^h_{rl}\Gamma^l_{is} - \partial_s \Gamma^h_{ir} - \Gamma^h_{rl}\Gamma^l_{ir} = 0. \tag{20}$$

R^h_{irs} is called the Riemann-Christoffel Tensor of Curvature. This condition (20) defines the *integrability for vectors*, i. e. basis, and the absence of torsion in the Riemannian space[28]. The Riemann—Christoffel Tensor of Curvature can also be defined in terms of the covariant derivatives of the Christoffel symbols using the Ricci identities, $R^h_{irs} v^s = \nabla_{rs} v^h - \nabla_{sr} v^h$, so that:

$$R^h_{irs} = \nabla_r \Gamma^h_{is} - \nabla_s \Gamma^h_{ir}. \tag{21}$$

The test of the Riemannian versus the Euclidean structure of the space relies on the nullity of the scalar Riemann curvature (which is obtained by a contraction[29] of the curvature tensor): the space is locally Euclidean only if the scalar Riemann curvature is zero at each point.

Appendix B: The Polish Data (1997-2000)[30]

Household budget surveys have been conducted in Poland for many years. In the period analysed, the annual total sample size was about 30 thousand households, which represent approximately 0.3% of all households in Poland. The data were collected by a rotation method on a quarterly basis. The master sample consists of households and persons living in randomly selected dwellings. This was generated by a two-stage, and in the second stage, two-phase sampling procedure. The full description of the master sample generating procedure is given by Kordos and Kubiczek (1991).

On every annual sample it is possible to identify households participating in the surveys during four consecutive years. For the four years panel from 1997 to 2000, 3052 households remain in the dataset after deleting a few number of households with missing values. The available information is as detailed as in the cross-section surveys: the usual socio-economic characteristics of households and individuals, as well as information on income and expenditures. A large part of this panel containing demographic and income variables is included in the comparable international data base of panels in the framework of the PACO project (Luxembourg) and is publicly available. The four surveys from 1997 to 2000 have been matched with the 2000 Time use Polish survey by Rubin's method, which assigns an adult observed in the Time use survey to each adult of a household [31]. The usual

[28]Note that this torsion measures the change of endogenous prices over the surface.

[29]Contraction consists in summing up a tensor over the same index which is in both high and low position: $R_{ij} = R^h_{ihj}$.

[30]Dataset prepared by C. Starzec.

[31]Matching performed in collaboration with Kristof Starzec and Anil Alpman

matching method by regression has two shortcomings: the reduction of the variance of the imputed values and the conditional independency assumption of the variables which are imputed with those which are observed in the dataset. Both problems can be solved to a great extent by the statistical matching procedure proposed by Rubin (1986). This procedure allows its user to assume a partial correlation value between the two variables that are jointly unobserved. This method conserves the individual distribution of the matched surveys and from this point of view is better than the usual regression method which diminishes drastically the variance of the matched variables compared to their distribution in the survey where they are observed. This matching procedure is discussed in Alpman and Gardes (2015), Alpman (2016) and Alpman, Gardes and Thiombiano (2017).

The 1997-2000 panel corresponds to the post transition high economic growth period with relatively low inflation, decreasing unemployment and generally improved socio-economic situation in the context of almost totally liberalized economy.

Appendix C: Full prices in a Becker's domestic production framework (based on Gardes (2019)

Full prices under substitutability assumption

Following Becker and Michael (1975) and Gronau (1977), the full expenditure can be written as the sum of its monetary and time components:

$\pi_i^1 z_i = p_i x_i + \omega t_i$

with ω the oportunity cost of time, π and p the full and the monetary prices corresponding to the quantities z and x of the activity and of the corresponding market good (or bundle of goods), for activity i (note that prices and the opportunity cost of time depend on household and time, which indices are removed).

The full price is the derivative of the full expenditure over z [32] , which gives rise for the Cobb-Douglas specification of the domestic production functions to:

$$p_i^{f1} = p_i \frac{\partial x_i}{\partial z_i} + \omega \frac{\partial t_i}{\partial z_i} = \frac{1}{a_i} p_i^{\alpha_i} \omega^{\beta_i} \left\{ \left(\frac{\beta_i}{\alpha_i} \right)^{\alpha_i} + \left(\frac{\alpha_i}{\beta_i} \right)^{\beta_i} \right\} \tag{22}$$

The calculation of full prices for CES specifications is presented in Gardes (2020).

[32]Procedure suggested by Anil Alpman

Full prices under complementarity assumption

Becker's full price for one unit of activity i can be written: $p_i^{f_2} = p_i + \omega\tau_i$ with τ_i the time necessary to produce one unit of the activity i. Under a Leontief technology (as in Becker's model) the full price can be proxied by the the ratio of the full expenditure over its monetary component:

$$\pi_i = \frac{(p_i + \omega\tau_i)\, xi}{p_i x_i} = \frac{1}{p_i} p_i^{f_2} \tag{23}$$

There exists a simple relation between these two definitions of the full prices. Using equations (9) we obtain:

$$p_i^{f_1} = \frac{1}{a_i} p_i^{\alpha_i} \left(\frac{m_i}{\omega_{ht} t_i}\right)^{\beta_i} \left\{1 + \frac{\omega t_i}{pi}\right\}$$

So that their logarithmic transforms differ only by $\beta_i \log(m_i/t_i)$ on a cross-section:

$$\log\left(p_i^{f_1}\right) = constant + \beta_i \log\left(\frac{m_i}{ti}\right) + \log(\pi_i)$$

with prices p_i set to one.

Appendix D: Empirical curvature: results

The system of equations corresponding to equation (12) for the Working model [33] writes (conditionally to symmetrical Christoffel symbols, i.e. in absence of path dependency):

$$\begin{aligned} w_{ith} - w_{i,t-1,h} = dw = \alpha_{it}^0 + \beta_i dlny_{ht} - \Big(&\gamma_{211} w_{ith}^2 + \\ &\gamma_{212} w_{ith} lny_{ht} + \gamma_{213} w_{ith} ln\pi_{ht} + \gamma_{222} lny_{ht}^2 + \\ &\gamma_{223} lny_{ht} ln\pi_{ht} + \gamma_{233} ln\pi_{ht}^2\Big) dlny_{ht} - \Big(\gamma_{311} w_{ith}^2 + \\ &\gamma_{312} w_{ith} lny_{ht} + \gamma_{313} w_{ith} ln\pi_{ht} + \gamma_{322} lny_{ht}^2 + \\ &\gamma_{323} lny_{ht} ln\pi_{ht} + \gamma_{333} ln\pi_{ht}^2\Big) dln\pi_{ht} + \eta_{ith} \end{aligned}$$

where γ_{rsv} is an estimate of $\Gamma_{rs;v}^i$ for good i, w is the budget share for good i, y total expenditure, π the price of good i and η the error term. Similar specification are easily written for the linear and the log-linear models.

The two tensors R_{223} and R'_{332} can finally be written for the Working model:

$R_{223}(i,h,t) = \gamma_{213} \nabla_{ly} w_{ith} + \gamma_{233} \nabla_{ly} ln\pi_{ht} - \gamma_{211} \nabla_{l\pi} w_{iht} - \gamma_{211} \nabla_{l\pi} lny$
$R_{332}(i,h,t) = \gamma_{321} \nabla_{l\pi} w_{ith} + \gamma_{322} \nabla_{l\pi} lny_{ht} - \gamma_{331} \nabla_{ly} w_{iht} - \gamma_{333} \nabla_{ly} ln\pi_{ht}$

[33] Similar equations for the Linear and the Log-linear models.

	Food	Housing	Transport	Clothing	Leisure	Other
R_{223}, Whole survey	0.0437	0.0238	0.0931	-0.0096	-0.0280	-0.0043
s.e.	(0.0093)	0.0123	0.0286	0.0074	0.0100	0.0020
R_{332}, Whole survey	0.0418	-0.3643	-0.9361	0.0298	0.0159	-0.0312
s.e.	(0.0541)	0.2176	0.4717	0.0109	0.0240	0.0024
$R_{223}R_{332}$, Whole survey	18.27x10^{-4}	-86.70x10^{-4}	-0.0872	-2.86x10^{-4}	-4.45x10^{-4}	1.34x10^{-4}
s.e.	(6.82x10^{-4})	(2.03x10^{-4})	0.0148)	0.714x10^{-4})	2.02x10^{-4})	0.18x10^{-4})
$\frac{R_{223}+R_{332}}{2}$ Whole survey	0.0427	-0.1703	-0.4213	-0.0048	-0.0061	-0.0178
s.e.	(0.0274)	(0.0274)	(0.0274)	(0.0274)	(0.0274)	(0.0274)
Age 20-34: R_{223}	-0.0967*	-0.0020	-0.0064	0.01461*	-0.1951*	0.0274
Age 20-34: R_{332}	-0.0902	-0.2347	0.0061	-0.0379	0.1066	-0.0121
Age 35-54: R_{223}	0.0620*	0.0796*	0.0135*	0.0051	-0.0149	-0.0047*
Age 35-54: R_{332}	0.0181	-0.4818	0.0029	0.0059	0.0727	-0.0298*
Age 55- : R_{223}	0.0139	0.0139	0.0338*	0.0061	-0.0573*	-0.0155*
Age 55- : R_{332}	0.0599	0.0599	0.0032	0.0864*	-0.3343*	-0.0022
Education 1 : R_{223}	-0.1146*	-0.0731	0.0351*	0.0059	-0.0200*	0.0025
Education 1 : R_{332}	-0.0231	-0.0548	0.0022	0.0583*	-0.1590	-0.0066*
Education 2 : R_{223}	-0.0749*	0.0731*	0.0165*	-0.0192	0.0131	-0.0012
Education 2 : R_{332}	0.0299	0.4845	0.0083	0.0263	0.0595	-0.0227*
Education 3 : R_{223}	-0.0136	0.1370*	-0.0887*	-0.0117	-0.0413	0.0062*
Education 3 : R_{332}	0.02946	-0.0582	-0.0039	0.0225	-0.3646	0.0025
Singles : R_{223}	0.1416*	0.0418*	0.1841*	0.0355	-0.1733*	0.0095
Singles : R_{332}	0.3134	-0.4342	0.0382	0.1763	-0.2466*	-0.0170*
2 adults : R_{223}	-0.0037	0.0611*	0.0166*	-0.0142	-0.1348*	-0.0438*
2 adults : R_{332}	0.1314	-0.4636	0.0039	-0080	-0.4021*	-0.0255*
Family with children : R_{223}	0.0443*	0.0273	-0.0255*	-0.0411*	-0.0829*	0.0067
Family with children : R_{332}	0.0197	-0.4344	0.0220	0.0114*	0.1290*	-0.0120*

Surveys: four waves of the 1987–1990 Polish panel (3,630 households). * significant at 5%; $R_{12}R_{12}^{'}$ multiplied by 10^4

Table D1. Gaussian Riemannian curvature: $\mathbf{R_{223}R_{323}}$ for the Working specification, Polish Panel, 1997–2000.

		Food	Housing	Transport	Clothing	Leisure	Other
Linear	R_{223}	-4.48e-7*	-0.0000357	-0.0000339	-4.92e-7	-2.64e-7	-0.000010
	R_{332}	-0.000840	0.002554	0.0002027*	8.31e-7**	2.92e-6	-0.0007885
	$R_{223}.R_{332}$	3.76e-10	-9.13e-8	-6.87e-9	-4.09e-13	-7.70e-13	7.94e-9
Log-linear	R_{223}	-0.0663	0.1031	0.06925	0.0209	0.0673	0.1971
	R_{332}	0.1850	0.2031	0.1072	0.0569	0.0262	0.0374
	$R_{223}.R_{332}$	-0.0123	0.0209	0.00742	0.00119	0.00176	0.00737
Working	R_{223}	-0.01028	-0.000203**	0.00100*	0.000287**	0.00205	0.000625
	R_{332}	-0.1347	0.01932	-0.001895	0.000535	-0.0002914	0.00158
	$R_{223}.R_{332}$	0.000138	-3.96e-6	-1.90e-6	1.54e-7	-5.97e-7	9.85e-7

Surveys: four waves of the 1987–1990 Polish panel (3,630 households).

Table D2. Gaussian Riemannian curvature: $\mathbf{R_{223}R_{323}}$ for the Linear, Log-linear and Working specifications, Polish Panel, 1997–2000.

References

Alpman, A. (2016) Implementing Rubin's Alternative Multiple-Imputation Method for Statistical Matching in Stata." Stata Journal, 16, 717-739.

Alpman, A. et Gardes, F., 2015, Time Use during the Great Recession - A Comment, Cahiers de recherche du CES, 2015.12.

Alpman, A., Francois Gardes, and Noel Thiombiano (2017) Statistical Matching for Combining Time-Use Surveys with Consumer Expenditure Surveys: An Evaluation on Real Data." CES Working Paper, 2017.24.

Angrist, Joshua D. and Alan B. Krueger (1999). Empirical strategies in labor economics, Labor Economics, vol. 3A. Ed. by O. C. Ashenfelter and D. E. Card. Vol. 3. North-Holland, pp. 1277–1397.

Appell, P. (1905) Cours d'Analyse Mathématique, Gauthier-Vilard ed.

Barten, A.P. (1964). Family composition, prices and expenditure patterns. In: Economic analysis for national economic planning. Ed. by Gordon Mills Peter E. Hart and John K. Whitaker. London: Butterworths.

Becker G.S. and Michael (1973) On the New Theory of Consumer Behavior." Swedish Journal of Economics, 75.4: 378-396.

Bettendorf L., Barten A.P. Rationnement dans les systèmes de demande : calcul des prix virtuels. Economie & Prévision. 121, 1995-5.

Boelaert, J., Gardes, F., Langlois, S., 2017, Consommation marchande et contraintes non monétaires au Canada (1969-2008), L'Actualité Economique-Canadian Review of Economic Analysis.

Cartan, E., 1946, Lecons sur la géométrie des Espaces de Riemann, second ed., Gauthier-Villars, Paris.

Christensen, B., 2014, Heterogeneity in Consumer Demands and the Income Effect: Evidence from Panel Data, The Scandinavian Journal of Economics, 116, 2, 335-355.

Cardoso, N., Gardes, F., 1996b, Estimation de lois de consommation sur un pseudo-panel d'enquêtes de l'INSEE (1979, 1984, 1989) , Economie et Prévision, 5 : 111–125.

Collet, R., Gardes, F., Starzec, C., 2011, Does Addiction Matters in Transportation, w.p. CES, Université Paris I.

Delachet, A. (1969). Le Calcul Tensoriel. QSJ 1336. Presses Universitaires de France.

Diaye, M.A., Gardes, F., and Starzec, C. (2004). Non-parametric Tests of Demand Theory and Rational Behavior. Annals of Economics and Statistics, 135, September, 89-120.

Gardes, F., 2019, The estimation of price elasticities and the value of time in a domestic framework: an application on French micro-data, Annals of Economics and Statistics, 135, September, 89-120.

Gardes, F., 2018, Time as a social distance, w.p. PSE, University Paris I, May.

Gardes, F., G.J. Duncan, P. Gaubert, C. Starzec, 2005, A Comparison of Consumption Laws Estimated on American and Polish Panel and Pseudo-Panel Data, Journal of Business and Economic Statistics, April.

Gardes, F., Langlois, S., Richaudeau, D., 1996, Cross-section versus Time Series Income Elasticities of Canadian Consumption, Economics Letters, 51, 169-175.

Gardes, F. and Starzec, C., 2014, Evidence on Addiction Effects from Households Expenditure Surveys: the Case of the Polish Panel, w.p. CES, Université Paris I.

Gronau R. Leisure, Home Production, and Work – The theory of the Allocation of Time revisited. Journal of Political Economy 85 (1977): 1099-1123.

Jeanperrin, C., 2000, Utilisation du Calcul Tensoriel dans les Géométries Riemanniennes, Ellipse, Paris.,

Heckman, James L. (1999). Causal parameters and policy analysis in economcs: A twentieth century retrospective. Tech. rep. National Bureau of Economic Research.

Hicks, John Richard (1979). Causality in economics. Basic Books New York.

Ivey, Thomas A. and J. M. Landsberg (2003), Cartan for beginners, in: Graduate Studies in Mathematics 61.

Kordos, J. and A. Kubiczek (1991), Methodological Problems in the Household Budget Surveys in Poland, Technical Report GUS, Warsaw.

Lichnerowicz, André (1987). Eléments de calcul tensoriel. Jacques Gabay.

Morgan, Frank (1998). Riemannian Geometry: A Beginners Guide. AK Peters/CRC Press.

Mundlak, Yair (1978), On the pooling of time series and cross section data, in: Econometrica: Journal of the Econometric Society, pp. 69–85.

Neary, J.P., Roberts, K.W.S., 1980, The Theory of Household Behaviour under rationing, European Economic Review, 19, pp. 25-42.