Time as a social distance

FranÃğois Gardes
Paris School of Economics, UniversitÃľ Paris I PanthÃľon Sorbonne

Preliminary version, May 3, 2018

Abstract

This article defines a synthetic time axis in the space of a set of variables which are observed in cross-section. Considering the relative position of two individuals along this time dimension allows us to estimate dynamic models and equations of geodesics on the surface using only cross-section data. An instrumentation using this synthetic time axis is thus proved to be efficient compared to usual instrumentation for dynamic models on panel data. The linear element (i.e. a distance) of the Riemannian space corresponding to this data set can be recovered by means of that synthetic time, which allows to consider the intrinsic properties of that surface.

1 Theory

Considering households, one solution to estimate dynamic models on a cross-section consists in taking into account the age of the family head or, for firms, the number of years the firm has been active in 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 and 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, we 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 *locally* as concerns their influence on savings. For example, a rich household aged 35 may have the same saving behaviour 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.

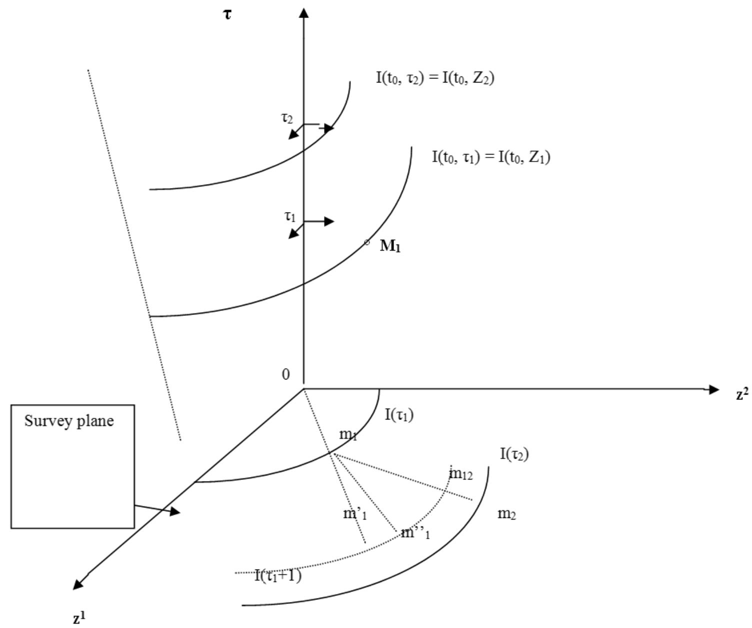
1.1 Substitution between cross-section variables in the time dimension

Suppose an agent h is characterised by a vector of socio-economic variables $Z_h = z_h^k$ (some being time-invariant, noted Z', such as the highest diploma obtained, some varying through time, including age, noted Z''), which influence some explained variable x_h, for instance savings. Under appropriate hypotheses, maximising a utility function depending on x and other determinants of the agent's satisfaction, yields x as a function of individual characteristics and other economic variables such as prices: $x_h = f(Z_h, W)$. In the space of the determinants z^k, a compensating substitution can be effected between two characteristics to maintain the agent in the same savings position (whenever the derivatives of x over these characteristics are non-zero). Whenever such compensation occurs, it is *as if*, under the life cycle saving theory, the two households were at the same point in their life cycle. Income therefore operates some transformation of the time axis such that, considering an individual h with characteristics z_h^k (including age), we can define a hypothetical individual h' with the same characteristics as h, but lower relative income, so that h' behaves (as concerns savings) as if she were one year younger than h. More generally, as long as the Hessian matrix for f is regular, a combination of some of the variables z^k may define a hypothetical time dimension in the cross-section, along which dynamic relations and causality can be defined[1]. This time axis can be interpreted as *a life cycle dimension*, the position on this axis depending not only on age, but on all variables which influence savings, such as income, education or family structure.

Consider the $(K + 1)$ linear space where the vertical axis corresponds to this time dimension τ and other axes correspond to characteristics z^k (all other determinants W being fixed). Consider now the plane Π_{τ_0} intersecting the time axis at τ_0. An isoquant is associated with this time position, since τ corresponds to the age of the typical agent (at point A_0). The isoquant is defined by the substitution between characteristics z_k such that x is constant along it. The ratios of the derivatives of f over z_k and age measure substitution between the agent's age and other characteristics so that synthetic time remains equal to τ_0. The isoquants along the time axis sum up into a cone in (time, Z)-space (this cone depends also on variables W, supposed to be fixed). Each isoquant, corresponding to a definite level on the vertical axis, can be projected onto a plane of characteristics Z, including age, defined by a time position τ_0 (see Figure 1). For some level of all characteristics except age, the age differs for these different isoquants, so that they do not intersect. On the plane Π_{τ_0}, the projected isoquants corresponding to higher τ are posited at higher levels of age and all characteristics positively related to age (i. e. such that the marginal substitution rate between the two, $\frac{\partial x/\partial z^k}{\partial x/\partial \text{age}}$, is negative).

To make it clear, suppose that a household $h = 1$ is posited at point τ_1 of the

[1] If the relationship between the K variables z^k and time is linear, it defines a hyperplane of dimension K. A non-linear continuous relation would define a differential manifold.



$I(\tau_i)$: isoquant corresponding to synthetic period τ_i.

Figure 1: Substitution between individual characteristics and time

time axis, which corresponds to its life cycle position. This point pertains to the isoquant $I(Z_h) = I_1$ defined by the constant life-cycle position: $\tau = \tau_1$, so that the gradient of I_1 is equal to the substitution rate between the characteristics. Suppose that another agent $h = 2$ is posited at point M_2 for a life cycle position $\tau = \tau_2$ on the plane corresponding to τ_2, which differs by k units of time on the life cycle axis ($\tau_2 = \tau_1 + k$). M_2 is on the isoquant $I(Z_h) = I_2$. Projecting M_1 and M_2 on points m_1 and m_2 in the characteristics plane corresponding to some time τ_0, m_1 and m_2 are on distinct isoquants[2] (as long as $\tau_2 \neq \tau_1$). The line between points m_1 and m_2 intersects with the intermediate isoquant corresponding to time $\tau = \tau_1 + 1$ at a point m which indicates the best prediction of the future position of household 1 (compared to agent 2) or the past position of household 2, $(k-1)$ periods before (compared to agent 1)[3]. The calculated position of household h one period later or before depends on the agent h' who is compared to h. This predicted position will be used as a proxy for agent h one year before or later. This instrumented value for $f(Z_{h,\tau-1})$ plays the same role as the instruments for past values of the endogenous variable

[2] Which are the projections of the isoquants on the cone corresponding to τ_1 and τ_2.

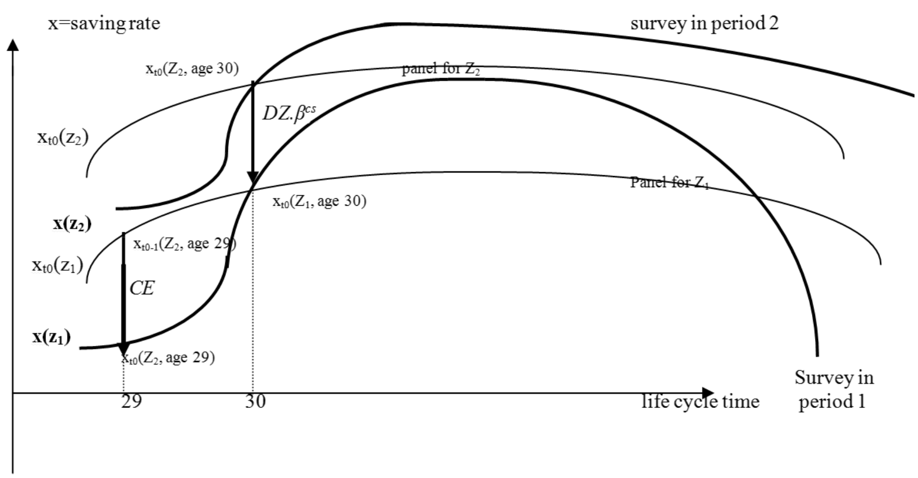
[3] Such a construction does not suppose that time on the life cycle axis is a linear function of the (z^1, z^2), which may require additional hypotheses to estimate position m, such as: the number of periods on the life cycle axis between two vectors of characteristics (z_1, z_2) and (z_1, z_2') on the characteristics plane is a linear function of the difference between the two characteristics vectors.

when estimating dynamic models[4]. Suppose that agent h's level of saving $x_h = f(Z_h, U_h)$ depends on variables z_h^k. The change in characteristics $Z = Z_h - Z'_h$ between two cohorts corresponding to points m_h and m'_h, leads to a first-order change in saving of $\sum_k \frac{\partial f}{\partial z_k} z_k = Z\beta$, with $\beta_k = \frac{\partial f}{\partial z_k}$. Thus, the point on isoquant $I(\tau_h + 1)$ which is paired with household h on isoquant $I(\tau_h)$ can be taken on the line passing through M_h with a direction defined by the β_k (which is Om_h only if $\frac{\beta_{ki}}{\beta_{kj}} = \frac{z^{ki}(h)}{z^{kj}(h)}$, for all k_i, k_j).

To define the agent who is paired with h, we first have to define the changes in characteristics Z between ages τ_h and $\tau_h + 1$. These can be estimated in cross-section data between the averages of each variable for different cohorts in the survey (whenever there is no cohort effect with respect to these characteristics). The calibration of the coefficients β_k must correct for any cohort effect occurring between the comparing two agents aged τ_h and $\tau_h + 1$ in the same cross-section. Considering the variable z_k, we first estimate β_k in the cross-section ($\beta^{\text{c.s.}}$, computed between similar agents aged τ_h and $\tau_h + 1$ in the same survey) and in time-series ($\beta^{\text{t.s.}}$, for the individual in the same cohort aged τ_h in the first wave and $\tau_h + 1$ in the second). The generation effect is thus $\beta = \beta^{\text{c.s.}} - \beta^{\text{t.s.}}$, as a change in z_k over time changes x by $\beta_k^{\text{t.s.}} z_k$ as opposed to the estimated cross-section effect $\beta_k^{\text{c.s.}} z_k$. Taking the example of income, the estimated coefficient in a cross-section regression reveals the correlation between savings and the individual's relative income (i. e. their position in the distribution of income at the time of the survey), controlling for their age, education etc. However, age at the time of the survey may be correlated with relative income, giving rise to the usual bias in the estimated coefficients.

To define the agent who can be paired with h, we first have to define the changes in characteristics Z between ages τ_h and $\tau_h + 1$. These can be estimated in cross-section data between the averages of each variable for different cohorts in the survey (whenever there is no cohort effect with respect to these characteristics). The calibration of the coefficients β_k must correct for any cohort effect occurring between the comparing two agents aged τ_h and $\tau_h + 1$ in the same cross-section. Considering the variable z_k, we first estimate β_k in the cross-section ($\beta^{\text{c.s.}}$, computed between similar agents aged τ_h and $\tau_h + 1$ in the same survey) and in time-series ($\beta^{\text{t.s.}}$, for the individual in the same cohort aged τ_h in the first wave and $\tau_h + 1$ in the second). The generation effect is thus $\beta = \beta^{\text{c.s.}} - \beta^{\text{t.s.}}$, as a change in z_k over time changes x by $\beta_k^{\text{t.s.}} z_k$ as opposed to the estimated cross-section effect $\beta_k^{\text{c.s.}} z_k$. Taking the example of income, the estimated coefficient in a cross-section regression reveals the correlation between savings and the individual's relative income (i. e. their position in the distribution of income at the time of the survey), controlling for their age, education

[4]Other definitions are possible: for instance, the closer point on $I\Big(Z(h)\Big)$ if the isoquant curve is known. The position of some household h one year later may also be defined projecting it (in the space of characteristics Z) on the point on the isoquant of its synthetic age plus one, using the ray Om_h to define its future position in the next period as the intersection between the isoquant of period $(\tau_h + 1)$ and Om_h. This supposes that the change in all characteristics between two age groups is homothetic.



Z: vector of individual characteristics
$x_{t_0}(Z_j, \text{age} = a)$: observation at t_0 of the saving rate for agents characterized by Z_j, aged a, as observed in the panel
$x(Z_j)$: observation in the survey for individuals with characteristics Z_j
CE: Cohort Effect
$\mathrm{D}Z\beta^{\text{c.s.}}$: change in the saving rate due to change Z in the explanatory variables

Figure 2: Survey and panel observations over the life cycle

etc. However, age at the time of the survey may be correlated with relative income, giving rise to the usual bias in the estimated coefficients.

1.2 Correction of cohort effects

The *first method* consists in defining, for each agent h in a cohort C_h, an agent $S(h)$ in the same survey with the same observed permanent characteristics Z' but one year younger. We then correct for the generation effect associated with these characteristics by computing for each variable of interest x its estimated value for an agent in the same cohort C_h, i. e. having characteristics Z_h in the previous year. Suppose that savings x depend on variables Z, so that, as a first-order approximation:

1. between two periods for individual h: $x(Z_{h,t}) - x(Z_{h,t-1}) = (Z_{h,t} - Z_{h,t-1})\beta^{\text{t.s.}} + \epsilon_{h,t} - \epsilon_{h,t-1}$;
2. between $S(h)$ and h in period t: $x(Z_{h,t}) - x(S(Z_{h,t})) = (Z_{h,t} - Z_{S(h),t})\beta^{\text{c.s.}} + \epsilon_{h,t} - \epsilon_{S(h),t}$.

Now suppose that $Z_{h,t-1}$ is equal to $Z_{S(h),t}$. In order to compare saving by the similar individual $S(h)$ in t to saving by h in $(t+1)$ we correct using the following formula, where the residuals are set to zero:

$$\mathbb{E}\big(x(Z_{h,t-1})\big) = x(Z_{S(h),t}) + (Z_{S(h),t} - Z_{h,t})(\beta^{\text{t.s.}} - \beta^{\text{c.s.}}) \tag{1}$$

The coefficients $\beta^{\text{t.s.}}$ can be estimated on aggregate time-series or on a panel or pseudo-panel containing at least two periods. $Z_{S(h),t}$ can be computed as the average on households having the same permanent characteristics as household h.

1.3 Synthetic time

A *second method* consists in estimating the distance on the time axis between h and each other household of the survey and pairing h with another household or the average of all households distant by one period. The simplest way to define the time distance between two households relies on their age, but this implies, as noted above, cohort effects.

Consider the cross-section difference in some variable x between two households, h and h'. This is related to the change of the vector of all the explanatory variables z_k by the cross-section estimates of the parameters β, and also (through the time-series estimate of β) to their variations between the two positions of agents h and h' on the synthetic time axis:

$$x(Z_{h',t}) - x(Z_{h,t}) = (Z_{h',t} - Z_{h,t})\beta^{\text{c.s.}} + \epsilon_{h'} - \epsilon_h = Z_{h,t}\beta^{\text{t.s.}} + \epsilon,$$

where $Z_t = Z_1(\tau_{h'} - \tau_h)$, Z_1 being the change in explanatory variables for one period over the line defined by $Z(h)$ and $Z(h')$ in the K-dimensional space[5]. This allows us to compute the difference in the positions of h and h' on the time axis:

$$\tau_{hh'} = \tau_{h'} - \tau_h = \frac{\Delta_{\text{c.s.}} Z\beta^{\text{c.s.}}}{Z_1\beta^{\text{t.s.}}} \tag{2}$$

with $\Delta_{\text{c.s.}} Z = Z_{h'} - Z_h$.

As $Z_1\beta^{\text{t.s.}}$ is a first-order measure of the variation in x over one period, $\frac{\Delta_{\text{c.s.}} Z\beta^{\text{t.s.}}}{Z_1\beta^{\text{t.s.}}}$ measures the time τ' necessary to change x from $f(Z)$ to $f(Z + \Delta_{\text{c.s}} Z)$. The difference $(\tau - \tau')$ indicates the additional time for the cross-section comparison between agents differing by $\Delta_{\text{c.s.}} Z$, corresponding to the effect of all non-monetary resources (information, time budget etc.) and constraints (such as the liquidity constraints correlated with Z in cross-sections) which are, in the cross-section dimension, related to this difference in characteristics. This may also be interpreted as the influence of the change in the shadow prices π^v corresponding to these resources and constraints: $(\tau - \tau') = e_p \Delta\pi^v$ with e_p the vector of direct and cross-price propensities. So the distortion of the synthetic price axis depends on the price effect related to the positions of agents in the characteristics space[6].

[5] Z_1 can be calibrated on aggregate time-series or between averages of reference populations using two surveys. For instance, income growth can be calibrated over the whole population (on aggregate time series) or between two surveys for some sub-population. For age, $\ln \text{age} = \ln \frac{\text{age}_h}{\text{age}_{h-1}}$. For the proportion of children in the family, one can calculate $\text{pr} = \text{pr}(\text{age}_h) - \text{pr}(\text{age}_h - 1) + \bar{p}$, with the first term computed on the cross-section and the second $\bar{p}$ is the average variation between t and $(t-1)$ and is computed over the whole population or for the household's reference population.

[6] Note that equation (**??**) can be interpreted along the same lines.

Formula (1) shows corrected savings for a similar agent observed in the same survey, while formula (2) allows us to calculate (under a hypothesis defining Z_1) the movement on the time axis between the two agents and to pair agents according to their time position, for instance such that $\tau_{hh'} = 1$[7].

The time scale is independent of agents h and h' who are being compared: first, the time lag $\tau_{h,h'}$ is symmetric, as is clear from the symmetry of $\Delta_{\text{c.s.}}Z$ in formula (2). Second, it is additive: $\tau_{h,h''} = \tau_{h,h'} + \tau_{h',h''}$—as is also clear from the linearity of (2). These properties are sufficient to define uniquely a time scale up to the choice of the origin.

Suppose for example that only the age of the head changes between two periods or two households, with the same coefficient in the two dimensions: $\beta^{\text{c.s.}}_{\text{age}} = \beta^{\text{t.s.}}_{\text{age}}$. In this case, $\mathbb{E}(\tau_{hh'}) = \frac{\Delta(Z_{h'}-Z_h)}{Z_1} = \text{age}_{h'} - \text{age}_h$. If $\beta^{\text{c.s.}}_{\text{age}} > \beta^{\text{t.s.}}_{\text{age}}$, the cohort effect is positive and the difference between h and h' on the time axis is greater than their age difference because of this cohort effect. The effect of a difference in income between two households on the time axis can be analysed similarly. For example, for food at home and considering only income elasticities (which can, for Poland, be calibrated at 0.5 in cross-sections and 0.8 in time-series, see gardes and Starzec, 2002), $\tau = 6$ years when comparing h aged 30 with income y_h and other characteristics Z'_h and household h' aged 30 with income $y_{h'} = 2y_h$ and the same characteristics: $\tau_{hh'} = \frac{-0.1\Delta_{\text{c.s.}}y}{-0.25g}$ (we suppose that income increases by $g = 5\%$ each year at this age). Thus, the time distance between households increases when g decreases, because it will take longer for h to attain the income position of h'. Note that, due to the correction by the cross-section and time series elasticities, 6 years is less than the ratio 14 necessary to double income with an increase of 5% per year (i. e. for the same income elasticity on cross-section and time-series).

The synthetic time scale depends on the endogenous variable being analysed. Nevertheless, we can imagine relationships between the time scales corresponding to different expenditures because of the additivity constraint. When considering for instance different expenditures $i = 1, \dots, n$, with coefficients β_i estimated under the additivity constraint, one obtains from equation (**??**) if only z_k changes or if all variables change proportionally: $Z_1 \sum_i \beta_i^{\text{t.s.}} \tau_i = \Delta_{\text{c.s.}}Z \sum_i \beta_i^{\text{c.s.}} = 0 \sum_i \beta_i^{\text{t.s.}} \tau_i = 0$, so that for $n = 2$: $\beta_1 = -\beta_2 \tau_1 = \tau_2$, and for $n = 3$: $\tau_3 = \tau_1 \frac{\beta_1}{\beta_1+\beta_2} + \tau_2 \frac{\beta_2}{\beta_1+\beta_2}$.

Finally, the *first method* can be applied to all similar agents aged one year less than household h, correcting the cohort effect by (1), then estimating a dynamic model by instrumenting past values of the variable to which (1) applies, either by the average corrected x for similar agents or by one of the set of similar agents chosen by minimising some distance.

The *second method* consists in estimating the time distance between agents, thus pairing agent h with some h' (or all h') at unit time distance. A dynamic relation can also be estimated over all agents ordered along the synthetic time dimension (with appropriate modelling of the partial adjustment according to the time distance between two consecutive agents).

[7] These pairings may be compared to simple pairing by age.

1.4 Defining a global time index over the whole space

As the cross-section and time-series propensities change from one point to another over the consumption space, the time index defined by equation (2) is in fact attached to a particular point or to a sub-space where the propensities are constant. It is thus necessary to examine how the point-time measures can be joined together over the whole space or a sub-set of the space. A natural idea consists in associating the time indices along a continuous curve corresponding to a geodesic, since in a Riemannian space, metrics can be continuously associated along each continuous curve in the space: in this way, each situation can be time related to those which correspond to other periods in the agent's life cycle. This is left for future work.

2 Empirical application:

2.1 Synthetic time

The synthetic time is estimated for 3052 households surveyed in the four years Polish panel (1997-2007). Its span covers 105 years in 1997 (respectively 84, 94 and 96 for years 1998, 1999 and 2000) with a standard error of the logarithm (measuring the inequality of this time variable between households) equal to 0.295 (compared to 0.466 for income per UC). This standard error can be interpreted as a social distance between households. It increases sharply for households the head of which is older than 40 (0.172 and 0.199 before 30 and between 31 and 40; .32 to .33 after 40, while income per UC stay at the same level of inequality within the same age groups. It decreases from 0.34 to 0.25 beween the first and the last quartiles of the income per UC distribution.

2.2 Estimation of dynamic models on cross-sections

\section{Empirical applications} \label{sec:empirical}

In this section, we use formula~\eqref{eq:expected} to correct for cohort effects between households pertaining to different generations and estimate a dynamic model on the Polish panel (1997-2000). Then, the equations of geodesics defined in Gardes (2018) are estimated along the synthetic time defined over the same data set.

\subsection{Estimating addiction effects on cross-sections} \label{sec:addiction}

In the simple version of the Becker—Murphy model, the consumer is supposed to maximise the present value of her inter-temporal additive utility $\Sigma_t \beta^{t-1} U(C_t, C_{t-1}, Y_t, e_t)$ under the inter-temporal budget constraint $\sum_t \beta^{t-1}(Y_t + P_t C_t) = A_0$ with C fixed at $t = 0$. This yields, for quadratic utility, a dependency of current expenditures on past and future consumption, as well as on current prices and current and future values of the variables e_t entering directly in the utility function (and not through their effects on current consumption) [p. 89, eq. 5.4]becker1996:

$$C_t = \theta C_{t-1} + \beta\theta C_{t+1} + \theta_1 P_t + \theta_2 e_t + \theta_3 e_{t+1} + u_t.$$

In this specification, past and future expenditures must be instrumented (because of the autocorrelation of the residual). The instruments used (past income and prices) are typically very inefficient in applications on aggregate time-series. With individual-level data, three waves are required to estimate the reduced equation, with the usual difficulties for estimating dynamic models on panel data. In previous estimations, either on aggregate or individual data, an important parameter: the inter-temporal substitution rate is poorly estimated, with highly implausible values (such as 300% per year for Chalupka's 1998 estimation on individual data).

We have proposed gardes2002evidence the use of a new type of instrumentation, considering the *next or previous cohorts* defined by the same permanent characteristics as h, $H^-_{h,t}$ and $H^+_{h,t}$, to individual h observed in the same period t (i. e. all households with the head one year younger or older than h). Agent h's past and future expenditures $C_{h,t-1}$ and $C_{h,t+1}$ are instrumented by the average expenditures at t of $H^-_{h,t}$ and $H^+_{h,t}$:

$$\mathbb{E}_t(C_{h,t-1}) = \sum_{h' \in H^-_{h,t}} \frac{1}{H^-_{h,t}} C_{h',t}$$

with the same holding for future consumption. We then correct for the *generation effect* associated with these characteristics by means of equation eq:mainfull[8].

This estimation is carried out on Polish panel data for tobacco expenditures (see the Appendix 2.3 for a description of the data-set, and gardes2002evidence, for details of the estimation strategy). The *first classic instrumentation* uses past, present and future prices (Ia) eventually combined with household income (Ib), and is estimated on the first differences to cancel out the individual fixed effects. No selection bias appears. The coefficients (Table **??**) of past and future tobacco expenditures are positive and very significant, as rational addiction would imply. On the contrary, the rate of time preference β is the same for estimations in levels and first differences.

The *second instrumentation* based on cohorts observed in the same year and corresponding to the household (aged one year less or more), gives similar but much more precise results. The implied yearly rate of time preference β is around 32%. This is a very encouraging result, as this instrumentation allows us to estimate dynamic models on cross-sections without any retrospective questions (often imprecisely recorded, as shown for instance in PSID data).

[8] This correction amounts to 8% of expenditures.

	TS	TS	CS	CS
Total Expenditure per UC	0.137	0.107	0.053	0.049
s.e.	0.008	0.007	0.019	0.019
Past Food Expenditure	0.742	0.428	0.476	0.311
s.e.	0.008	0.010	0.035	0.040
Future Food Expenditure	-	0.451	-	0.321
s.e.	-	0.017	-	0.015
N	6095	6095	2811	2811
R^2	0.645	0.734	0.068	0.090

Table 1: Dynamic model of food consumption (logarithmic specification)

TS: estimation on 1998 and 1999 surveys, past and future expenditures observed in 1997 and 2000 for the same household

CS: estimation for 1997 survey, Cohort instrumentation of past and future expenditures

2.3 Estimation of the Christoffel symbols along geodesics

In this section, I use formula (1) to correct for cohort effects between households pertaining to different generations and estimate the equations of geodesics.

Consider the curve in the consumption space of different positions of a household in his life cycle (as indicated by the curves $x_{\tau_0}(z_k)$ in Figure 2). We can identify this curve with a geodesic between its extreme points, as the household is supposed to maximize its satisfaction over the life cycle. Indeed, this maximization is made under all the constraints which apply to the household at each period of its life cycle, and considering both the monetary and the non-monetary resources of the household. These constraints and non-monetary resources are represented by the shadow prices on each point of this geodesic. Therefore, optimizing its behaviour under these constraints defines the optimal path as the geodesic over the life cycle of the household.

A typical household can be defined by a set of characteristics, some of which are permanent and other change according to the normal evolution along the life cycle. On a survey, the different positions of such a typical household can be observed as the average of all households at each point of the life cycle path. Between two period along this life cycle path, the changes in the bundle of goods are biased by generation effects, which can be removed by the formula which relates these two types of partial derivatives: $\nabla_k v^i = \partial_k v^i + \Gamma^i_{kh} v^h$ whenever we have some information on the cross-section and time-series consumption laws. If we estimate on a single survey, we do not have information on time-series parameters, but we can estimate the Christoffel symbols Γ^i_{kh} using the equation of geodesics:

$\frac{d^2 v^i}{dt^2} + \Gamma^i_{kh} \frac{dv^h}{dt} \frac{dv^k}{dt}$ (3)

with v^i equal to the budget share and h logarithmic total expenditure per Unit of Consumption and logarithmic number of U.C.

Thus, compute time-series estimates of the derivatives $\frac{\partial g_{ij}}{\partial v^k}$ using the cross-section estimates of the covariant derivatives and the equation which relates these two types of partial derivatives: $\nabla_k v^i = \partial_k v^i + \Gamma^i_{kh} v^h$. A convergence process on the Γ^i_{kh} thus allows to compute the time derivatives $\partial_k v^i$, knowing the cross-section derivatives and using the geometric properties of the consumption space described by the Christoffel symbols.

The derivatives $\frac{v^k}{t}$ can be estimated between two close points M and $M+M$ if an index of time is defined between these points. Suppose the difference on the time scale discussed in Section **??** between points M_0, M_1 and M_2 are noted $\Delta\tau(M_0, M_1) = \Delta\tau_1$ and $\Delta\tau(M_0, M_2) = \Delta\tau_2$. The time index being additive, we have: $\Delta\tau(M_0, M_i) + \Delta\tau(M_i, M_i') = \Delta\tau(M_0, M_i')$, $j \in \{1; 2\}$. So, with $v_j = v(M_j)$ and $v_j' = v(M_j')$,

$$\frac{v_2' - v_1}{\Delta\tau(M_1, M_2')} - \frac{v_2 - v_0}{\Delta\tau(M_0, M_2')} \quad \text{and} \quad \frac{v_1' - v_2}{\Delta\tau(M_2, M_1')} - \frac{v_1 - v_0}{\Delta\tau(M_0, M_1')}$$

are discrete measures of $\frac{{}^2 v^i}{t^2}$ which can be used to estimate Γ^i_{kh} in equation (**??**)[9].

Households are grouped by age classes of two years in the 1990 Polish survey. Each cell contains 140 households on average. The series of these 25 cells are considered as describing a geodesic for the average Polish household. The derivatives $\frac{v^k}{t}$ and the acceleration $\frac{{}^2 v^k}{t^2}$ are computed between consecutive points on the average geodesic, which allows to estimate the system of equations (**??**). The cross-section income elasticities for Food at home expenditures and other expenditures are thus corrected using the estimated Γ^i_{kh}, and compared to time-series parameters computed on the four years panel.

For food at home, which represents one half of total expenditures, the indices T_1 (calculated for all variables) and T_{123} (calculated only for the term corresponding to the product of income and family size) have the same sign as the difference between the cross-section and time-series elasticities (Table **??**). Their magnitude is similar to this difference, which is curious as this difference must be compared to $\frac{T_1}{w}$ and $\frac{T_1 23}{w}$, with w the average budget share. For the other expenditures, there is a strong correlation between the difference of the elasticities and the indices, specially index T_1: for five items over seven (for which the difference between the elasticities is significant) the signs are the same, and the two expenditures with different sign are Food away (an expenditure which is very particular in Poland) and alcohol and tobacco expenditures (which are not well measured in households surveys). Moreover, the order of the eleven expenditures according to the difference between the elasticities is very similar to the order according to the index T_1: the average difference of the rank is 1.24, compared to a random difference which would be 3.75. The average levels of T_1 and the elasticity difference are also similar (and for six cases, the elasticity difference is very close to T_1), which is not expected since the elasticity difference corresponds to the ratio of T_1 and the budget shares.

Considering only food at home and all other expenditures, Table 2.3 shows

[9] Note that this method also uses the true time dimension (between periods), as the time structure of cross-section is defined according to the true time gradient of variables v^k.

	$E^{\text{c.s.}} - E^{\text{t.s.}}$	T_1	1^{st} CF	T_{123}	2^{nd} CF	Sign $\frac{(1)}{(2)}$	Sign $\frac{(1)}{(3)}$
Food at Home	−0.35	−0.55	−1.135	−0.241	−0.498	Yes	Yes
Food Away	(+0.11)	−0.026	−6.5	0.111	27.765	No	Yes
Alcohol-Tobacco	0.42	−0.049	−1.55	0.123	3.878	No	Yes
Housing	(+0.06)	−0.24	−2.44	0.477	4.851	No	Yes
Energy	−0.58	−0.112	−2.82	-0.095	−2.389	Yes	Yes
Clothing	(+0.01)	−0.071	−0.76	0.112	1.194	No	Yes
Transportation	0.97	1.211	18.69	0.427	6.594	Yes	Yes
Health	(−0.07)	−0.144	−5.48	0.029	1.099	Yes	No
Culture	0.35	0.518	6.82	−0.307	−4.046	Yes	No
Various expend.	(+0.20)	0.044	1.41	0.066	2.120	Yes	Yes
Financial Goods	0.74	1.006	19.88	−1.034	−20.425	Yes	No

Estimation on the Polish Panel, 1987–1989.
Between parentheses: difference between elasticities is not well estimated
$T_1 = \Gamma^i_{kh} v^h$ with i budget share, k logarithmic total expenditure and h logarithmic total expenditure per Unit of Consumption and logarithmic number of U.C. and the budget share.
$T_{123} = \Gamma^i_{kh} v^h$ with i budget share, k logarithmic total expenditure and h logarithmic number of U.C.

Table 1: Income Elasticities and Correction Factors

that the corrected (by T_{123}) cross-section elasticity are close to those which have been estimated on the panel.

	Cross-section elasticity*	Correction factor**	Corrected elasticity	Time-series elasticity*
Food at Home	0.451	0.498	0.949	0.805
[-0.5ex]	(...)	(0.079)	(...)	(...)
Other expenditures	1.475	−0.47	1.005	1.169
[-0.5ex]	(0...)	(0.079)	(...)	(...)

Population: Households with head aged more than 20, with positive expenditure on food at home for all four years.
Surveys: * four waves of the 1987–1990 Polish panel (3,630 households): Between and Within estimations (instrumented total expenditures). ** 1990 survey.
Equation of the geodesic: $\frac{2v^i}{t^2} + \Gamma^i_{kh} \frac{v^h}{t} \frac{v^k}{t} = 0$, with v^i budget share and h logarithmic total expenditure per Unit of Consumption and logarithmic number of U.C.
Correction of cross-section parameters: $\mathbb{E}^{\text{c.s.}} - \Gamma^i_{kh} v^h$, with h total expenditures.

Table 2: Cross-section and time-series income elasticities

Finally, the estimation of geodesics along the life-cycle, although it has been performed very roughly, has allowed to classify the expenditures according to the endogeneity bias, and to predict the sign of this bias, using only the information given by one survey. For expenditures characterized by high values of this bias (very different cross-section and time-series parameters), the magnitude of the bias is close to the index T_1. 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 one survey.

Another application of the instrumentation of past and future was made (Collet, Gardes and Starzec, 2014) for the estimation of habit and addiction effects for tobacco, alcohol and transport expenditures in the Becker-Murphy farmework. This model conclude to a reduced equation where present consumtion depends on past and future expenditures, which must be instrumented (because of the autocorrelation of the residual). The instruments used (past values income and prices) are typically very inefficient in applications on aggregate time-series. With individual-level data, three waves are required to estimate the reduced equation, with the usual difficulties for estimating dynamic models on panel data. In previous estimations, either on aggregate or individual data, an important parameter: the inter-temporal substitution rate is poorly estimated, with highly implausible values (such as 300% per year for Chalupka's 1998 estimation on individual data). Instruments for these variables correspond to the values for the *next or previous cohorts* defined by the same permanent characteristics as the observed household, correcting for the *generation effect* associated with these characteristics by means of equation [10]. The coefficients of past and future tobacco expenditures are positive and very significant, as rational addiction would imply. Moreover, the implied yearly rate of time preference β is around 32%, which is a very encouraging result compared to the results in the litterature.

Conclusion

We have shown that the substitution between time and the socio-economic determinants of household behaviour makes it possible to estimate dynamic models on cross-sectional individual data. The position of individuals on the synthetic time axis also provides a natural distance between them, which can be related to the differences which are observed between their expenditure patterns. The linear element corresponding to that distance allows to estimate the curvature of the Riemaniann surface using the Riemann-Christoffel tensor of curvature empirically estimated in Gardes (2018) by means of the definition of the scalar curvature corresponding to a contraction of the Ricci tensor: $R = g^{ij}R_{ij}(g^{ij}$ being the inverse matrix of the metric g_{ij}). It also provides a natural parametric representation of the surface corresponding to the virtual move of households on the surface.

Appendix A: The Polish Data (1997-2000)

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

[10]This correction amounts to 8% of expenditures.

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 data set 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. The usual matching method by regression has two shortcomings: the reduction of the variance of the imputed values and the conditional independency assumption of the variable 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 times 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. The matching procedure is discussed in Alpman (2016), Alpman and Gardes (2015) 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 B: Estimation of the opportunity cost of time

To estimate the shadow price of time, I assume (Gardes, 2017) that the consumer combines time with monetary expenditures to produce activities that generate utility in a model where the market work time is valued by the consumer's wage rate while the remaining time (e.g., time allocated to leisure or non-market work) is valued by the shadow price of time that may differ from the wage rate. It is assumed that the consumer's utility function is given by u(Q) = Q i aiQ i i where ai is a positive parameter and Qi is the quantity of the activity i produced by the combination of monetary and time inputs denoted mi and ti, respectively: $Q_i = b_i m_i^{\alpha_i} t_i^{\beta_i}$ where $m_i = x_i p_i$ with x_i the quantity of the market goods i, p_i its price, and b_i a positive parameter. The domestic production Q is assumed to depend on expenditures m_i (rather than quantities x_i) because the dataset informs only expenditures. This approach yields consistent results when all households face the same prices.Expenditures are divided by a yearly price index and the regressions include State and year dummies.

The choice of the Cobb-Douglas forms allows the parameters to be identiable. Note that the estimation is made locally, either over sub-population or locally, for each household, using a non-parametric method similar to the Lowess estimator. The Cobb-Douglas specification implies simply constant substitution between time and monetary resources only in the neighborhood of each individual's equilibrium point. Combining the utility and the production functions allows to write the utility in terms of inputs:

$$u(m,t) = (\prod_i a_i^{\gamma_i} b_i^{\gamma_i})(\prod_i m_i^{\frac{\alpha_i\gamma_i}{\sum \alpha_i\gamma_i}})^{\sum \alpha_i\gamma_i}(\prod_i t_i^{\frac{\beta_i\gamma_i}{\sum \beta_i\gamma_i}})^{\sum \beta_i\gamma_i} = Am'^{\sum \alpha_i\gamma_i} t'^{\sum \beta_i\gamma_i}$$

where m' and m' are geometric weighted means of the monetary and time inputs and $A = \prod a_i^{\gamma_i} b_i^{\gamma_i}$. In this framework, the consumer is subject to an income constraint, $\Sigma m_i = wt_w + V$, and to a time constraint, $\Sigma t_i + t_w = T$, where t_w is the time allocated to market work, V is other income and Y and T total income and available time (excluding sleeping time which is not included in the time uses t_i). Utility maximization implies that the shadow price of time, denoted ω, is given by $\omega = \frac{\frac{\partial u}{\partial t'}\frac{\partial t'}{\partial(\Sigma t_i)}}{\frac{\partial u}{\partial m'}\frac{\partial m'}{\partial Y}} = \frac{m'\Sigma\beta_i\gamma_i\frac{\partial t'}{\partial(\Sigma t_i)}}{t'\Sigma\alpha_i\gamma_i\frac{\partial m'}{\partial Y}}$ (A1). The shadow price of time differs from the market wage rate when, for instance, there exists some market imperfections, transaction costs, and constraints on the labor market or in the home sector. The shadow price of time can be estimated provided that estimates of α_i, β_i and γ_i are available. Under the assumed functional forms, the optimal combination of the inputs yields $\alpha_i$and β_i: $\alpha_i = \frac{m_i}{\omega t_i + m_i}$ and $\beta_i = \frac{\omega t_i}{\omega t_i + m_i}$. The optimal allocation of an input across the activities i and j implies:

$\frac{\gamma_i}{\gamma_j} = \frac{\beta_j t_i}{\beta_i t_j} = \frac{a_j mi}{a_i m_j}$ and the system of $(n-1)$ equations: $m_i\gamma_j = m_j\gamma_i + \omega\gamma_i t_j - \omega\gamma_j t_i$ which allow to estimate an average opportunitry cost of time $\omega$and all parameters γ.

This system is overidentied and the unknown parameters, that is, ω and each γ_i, can be estimated as a system of $(n-1)$ equations under the constraint $\sum_i \gamma_i = 1$. Using the shadow price of time estimated from the system, we obtain the estimates of α_i and β_i for each individual using equation (4) that allow to derive estimates of the shadow price of time at the individual level by equation (A1).

Appendix C: Full prices in a Becker's domestic production framwork

Full prices under substitutability assumption

Following 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 ÏĂ and p the full and the monetary prices corresponding to the quantities z and x of the activity and of the corresponding market good, 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, which writes for the Cobb-Douglas specification of the domestic production functions:

$\mathrm{p}_i^{f1} = p_i \frac{\partial x_i}{\partial z_i} + \omega \frac{\partial t_i}{\partial z_i}$

The optimization program gives rise to the first order condition: $\frac{t_1}{x_1} = \frac{p_1}{\omega} \frac{\beta_1}{\alpha_1}$ (note that all variables, including Îś and Îš, are measured or estimated at the household level). Writing the quantity of the activity z$_i$ in terms, either of t or x, gives:

$t_i = \frac{1}{a_i} z_i (\frac{p_i \beta_i}{\alpha_i})^{\alpha_i}$and $x_i = \frac{1}{a_i} z_i (\frac{\omega \alpha_i}{p_i \beta_i})^{\beta_i}$

So that the full price becomes:

$p_i^{f_1} = \frac{1}{a_i} p_i^{\alpha_i} \omega^{\beta_i} \left\{ \left(\frac{\beta_i}{\alpha_i}\right)^{\alpha_i} + \left(\frac{\alpha_i}{\beta_i}\right)^{\beta_i} \right\}$

This derivation of ÏĽ,Îś and Îš at the individual level allows identifying the full price for each household (a_i being supposed constant across the population).

Full prices under complementarity assumption

BeckerâĂŹs full price for one unit of activity i can be written: $\mathrm{p}p_i^{f_2} = p_i + \omega \tau_i$ with τ_i the time use necessary to produce one unit of the activity i. Suppose (as in BeckerâĂŹs model) that a Leontief technology allows the quantities of the two factors to be proportional to the activity:

$x_i = \xi_i z_i$ and $t_i = \tau_i z_i$with τ_i=$\frac{\theta_i}{\xi_i}$

This case corresponds to an assumption of complementarity between the two factors in the domestic technology, which allows calculating a proxy for the full price of activity i by the ratio of full expenditure over its monetary component:

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

Note that under the assumption of a common monetary price pi for all households in a survey made in period t, this ratio contains all the information on the differences of full prices between households deriving from their opportunity cost for time ÏĽ_h and the home production technology represented by the coefficient of production ÏĎ_i. If the monetary price p changes between households or periods, the full price can be computed as the product of this proxy π_{ih} with p_i: $p_i^{f_2} = p_i \pi_i$. With these definitions, it is possible to measure the full prices, observing only monetary and full expenditures.

The first definition of prices corresponds to a complete substitution between the two factors in the model (which is used in section 1 to estimate the opportunity cost of time), since the Cobb-Douglas domestic production functions are characterized by a unitary elasticity of substitution between the two factors. Its relies on the estimation of three parameters: Îś,Îš and ÏĽ. On the other hand, the second definition supposes no substitution between the two domestic production factors but it may give a more robust measure of the full prices since it depends only on the estimation of the householdsâĂŹ opportunity cost for time ÏĽ. Both definitions of the full prices will be used in the estimation. However, 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 Îš_i logâĄąãĂŰm_i/t_i ãĂŮ on a cross-section:

$log(p_i^{f_1}) = constant + \beta_i log(\frac{m_i}{ti}) + log(\pi_i)$ with prices p_i set to one.

Two hypotheses were necessary to derive full prices from monetary and time expenditures: first, the domestic production functions are supposed to be Leontief functions with constant production coefficients (for BeckerâĂŹs prices) or Cobb-Douglas functions (for GronauâĂŹs prices); second no joint production exists, which may be more easily verified for broad categories of activities such as housing and food.

References

[1] References

Alpman, Anil (2016) \Implementing Rubin's Alternative Multiple-Imputation Method for Statistical Matching in Stata." Stata Journal, 16, 717ôĂĂĂ739.

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

Alpman, Anil, 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.

Delachet, AndrÃľ (1969). Le Calcul Tensoriel. Vol. 1336. Presses universitaires de France.

Gardes, F., 2016, The estimation of price elasticities and the value of time in a domestic framework: an application on French micro-data, under revision in Annals of Economics and Statistics.

Gardes, F., 2018, Endogenous prices in a Riemaniann framework, w.p. PSE, University Paris I.

Jeanperrin, Calcul Tensoriel,

Ivey, Thomas A. and J. M. Landsberg (2003). âĂIJCartan for beginnersâĂİ. In: Graduate Studies in Mathematics 61.

Kordos, J. and A. Kubiczek (1991). âĂIJMethodological Problems in the Household Budget Surveys in PolandâĂİ. In: GUS, Warsaw.

Lichnerowicz, AndrÃľ (1987). ElÃľments de calcul tensoriel. Jacques Gabay.

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

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