Measuring the Evolution of Complex Indicators: Theory and Application to the Poverty Rate in France[*]

Fabien Dell[†] and Xavier d'Haultfœuille[‡]

September 2007

Abstract

This paper addresses the issue of measuring and testing evolutions of complex indicators estimated by surveys. This problem amounts to compute the variance of these complex statistics, but also their covariance over time. Rotating designs may indeed bring important positive correlations between indicators. These correlations are often neglected in testing temporal stability, leading to conclusions deemed to be too conservative. We obtain variance and covariance estimators by adapting the functional delta method to survey sampling. Finally, our method is applied to the evolution of the poverty rate in France.

Keywords : dependent samples, linearization, functional delta method, Hadamard differentiability.

1   Introduction

A major issue in social studies is to assess whether the evolution of an indicator is significant or not. To answer these questions, the statistician faces two problems. Firstly, these statistics are highly non linear, so that usual variance formulas cannot be applied directly. Secondly, samples used in the computation of social indicators are often dependent. Rotating surveys and panels are examples of such dependent samples. This induces a correlation through time between indicators, which is generally positive. Hence, neglecting it leads to conclusions which

1

are too pessimistic. This paper presents a general framework which enables to handle both problems.

The body of literature on the variance of complex statistics is large, and can be divided into two parts. First, many adaptations of the bootstrap to finite populations have been proposed (for recent papers on it, see BOOTH ET AL.[1994], RUST AND RAO[1996], BERTAIL AND COMBRIS[1997], LAHIRI[2003], and for an older survey, RAO ET AL.[1992]). This method is popular because it avoids analytical computations. However, it is seldom theoretically justified for complex surveys (multistage surveys or unequal probabilities designs) and may be hard to implement in such designs. The second strain returns to variances of totals by linearizing complex statistics. This method has been developed for statistics which are solutions of estimating equations (see BINDER[1983], BINDER AND PATAK[1994], DEMNATI AND RAO[2004], for the theory on it and BINDER AND KOVACEVIC[1995], for an application), L-statistics (see SHAO[1994]) or statistics which are differentiable in a particular sense (see DEVILLE[1999]). In this article we rely on the second approach, and present a framework based on the functional delta method. This approach is not new in classical statistics (see for instance GILL[1989]) but, to the best of our knowledge, has been used only once in survey sampling (see BERTAIL ET AL.[2003]). We show that the functional delta method is based on a functional central limit condition on sampling designs, which is stronger than the usual normality condition often admitted by statisticians to make inference. But the scope of applications is also larger, including not only regular functions of means but also maps of the empirical distribution function which are smooth in a weak sense, namely Hadamard differentiable. In that sense it generalizes the paper of DEVILLE[1999], which considers the more restrictive class of Fréchet differentiable statistics (see for instance SHAO AND TU[1995] for a discussion on the difference between these two notions).

The issue of estimating covariances has also been considered in the linear case, see e.g. KISH[1965] or BERGER[2004b]. But the case of nonlinear statistics has not been considered so far. We show that actually, under a straightforward multivariate generalization of the functional central limit condition, linearization can still be applied to compute covariance terms. Moreover, we state an independence condition under which the estimator of the covariance does not depend on the probabilities of response. This is satisfying because these probabilities are unknown and can be consistently estimated only under restrictive assumptions on the response model.

The present paper is organized as follows. Section one is devoted to the theoretical foundations of linearization. The second section considers the computation of covariances when samples are dependent. Finally, the theory is applied in the third section to test the significance of the evolution of the poverty rate in France between 1996 and 2001. Based on

the French "Taxable Income Survey", linearization is used to estimate the variance covariance matrix through time and to test the hypothesis of stable poverty.

2 Linearization based on functional delta method

Whereas computing variances of total was well known from the beginning of survey sampling theory (see for instance HORVITZ AND THOMPSON[1952]), the problem of estimating the accuracy of nonlinear statistics has remained open for years. Indeed, it is generally impossible to obtain exact formulas. This is the reason why linearization is so attractive. The idea is simply to approximate nonlinear statistics by totals. Since asymptotic distributions of totals are often known, linearization gives a first order approximation of the distribution of the statistic and thus enables us to compute confidence intervals (and a fortiori variance). As we will see, the great advantage of linearization method, over bootstrap in particular, stems from the fact that it easily takes complex sample designs, calibration and also (see section 2) temporal dependency into account. This is particularly satisfying in applied statistics.

Let us call $U = \{1, ..., N\}$ the whole population, S the random sample of size n (which is assumed non random) and $(w_k)_{k \in U}$ the sampling weights, $w_k = 1/\pi_k$ where $\pi_k = \Pr(k \in S)$ is the first order inclusion probabilities of k. Linearization is based on the asymptotic behavior of estimators when the population size tends to infinity. Hence, all these quantities depend on N but for the sake of simplicity, we let this dependency implicit. To give sense to these asymptotics, we need regularity conditions on both the population and the sampling design, which will be maintained throughout the paper. More precisely, and in a way very close to DEVILLE[1999], we suppose in this section and the next one that, for any variable of interest $(y_k)_{k \in U}$ and when $N \to \infty$,

$$\begin{cases} \frac{1}{N} \sum_{k \in U} y_k & \longrightarrow \quad E(y) \\ \frac{1}{N} \sum_{k \in S} w_k y_k & \overset{\mathbb{P}}{\longrightarrow} \quad E(y) \\ nV\left(\frac{1}{N} \sum_{k \in S} w_k y_k\right) & \longrightarrow \quad V_{as}^y \\ n\widehat{V}\left(\frac{1}{N} \sum_{k \in S} w_k y_k\right) & \overset{\mathbb{P}}{\longrightarrow} \quad V_{as}^y. \end{cases} \tag{1}$$

Because only the sampling is random, the first and third conditions must be understood in terms of convergence of sequences. In the first, $E(y)$ is simply a notation, which refers to the usual expectation in the case of a superpopulation model. The third condition states that properly normalized variances of means, which depend both on the variable of interest and the survey sampling, also admit a limit. Lastly, we suppose, with the second and fourth assumptions, that their respective estimators become infinitely accurate when $N \to +\infty$. This assumption is usually satisfied when $n \to +\infty$. The estimator of variance considered

here is the Horvitz-Thompson estimator, which is always computable as soon as the second order inclusion probabilities are known and not zero. For more details on other estimators of variance, see for instance BERGER[1996] or BERGER[2004a]. Finally, the following assumption will be used from time to time.

Assumption S $(y_k)_{k \in U}$ are the realizations of a sequence of independent random variables with distribution F. Moreover, $n = o(N/\ln(\ln(N)))$.

The first part of this hypothesis amounts to supposing a superpopulation model. The second part implies that the sampling rate converge to zero as the population size grows. This assumption is restrictive but not fully unrealistic. For instance, the sampling rates of French household surveys never exceed $1/300$.

Now let T be a parameter of interest and $\widehat{T}$ its sample estimation. T will be considered linearizable[1] if there exists variables $(\mathrm{lin}_k T)_{k \in U}$ such as

$$\frac{\widehat{T} - T}{\sqrt{V(\sum_{k \in S} w_k \mathrm{lin}_k T)}} \xrightarrow{d} \mathcal{N}(0, 1). \tag{2}$$

As a variance of a total, $V\left(\sum_{k \in S} w_k \mathrm{lin}_k T\right)$ can be consistently estimated. Hence, if T is linearizable, one can easily make inference on that parameter. We now describe two main situations for which this occurs.

The most simple one arises for differentiable real functions of means, $T = g\left(\frac{1}{N}\sum_{k \in U} y_k\right)$. Suppose that

$$\frac{\sum_{k \in S} w_k y_k - \sum_{k \in U} y_k}{\sqrt{V(\sum_{k \in S} w_k y_k)}} \xrightarrow{d} \mathcal{N}(0, 1) \tag{3}$$

Then the following proposition shows how to adapt the usual delta method to our framework.

Proposition 2.1 *(delta method) if g is differentiable at $E(y)$ with $g'(E(y)) \neq 0$ and*

    1) g' is continuous at $E(y)$,

or 2) assumption S holds,

then T is linearizable with

$$\mathrm{lin}_k T = g'\left(\frac{1}{N}\sum_{k \in S} w_k y_k\right)\frac{y_k}{N}, \tag{4}$$

where it is understood that in the variance of $\sum_{k \in S} w_k \mathrm{lin}_k T$, we do not take into account the variance of $\mathrm{lin}_k T$ itself.[2]

---

[1]The definition given here is slightly different from the usual one (DEVILLE[1999], 1999).

[2]In other terms, $g'\left(\frac{1}{N}\sum_{k \in S} w_k y_k\right)$ is assumed to be non-random in $\mathrm{lin}_k T$. Indeed, we use $\mathrm{lin}_k T$ as an approximation of $g'(E(y))(y_k/N)$, which does not depend on S.

Proof : see appendix A.

Conditions to apply the delta method are more restrictive than in classical statistics, where differentiability at $E(y)$ suffices. This stems from the fact that the parameter of interest is $\frac{1}{N} \sum_{k \in U} y_k$, which changes with N, and not $E(y)$. We propose two ways to circumvent this problem. The first is to impose the continuity of the derivative of g at $E(y)$, while the second relies on a superpopulation model and an asymptotically small sampling rate.

Hence, in the case of smooth functions of means, linearization mainly relies on (3), which is a regularity condition on the sampling design. There is no general answer to the question of which sampling design satisfy this condition, although results have been obtained for some families of sampling designs. See, among others, HʹAJEK[1964], ROSEN[1972], KREWSKI AND RAO[1981] and BERGER[1998] and, for a complete survey on asymptotics in finite population sampling, SEN[1988].

Many complex statistics are in fact smooth functions of means. Among inequalities indicators, the Theil and Atkinson indexes for instance belong to this class. However, quantiles for instance cannot be expressed in such a framework. On the other hand, most statistics T can be written as $T(F_N)$ with

$$F_N(t) = \frac{1}{N} \sum_{k \in U} 1\{y_k \leq t\}, \tag{5}$$

where $1\{A\} = 1$ if A, 0 otherwise. The plug-in estimator of T is then $T(F_S)$ with

$$F_S(t) = \frac{1}{\sum_{k \in S} w_k} \sum_{k \in S} w_k 1\{y_k \leq t\}. \tag{6}$$

This is the so-called Hájek estimator of $F_N(t)$. Let us denote by $F(t)$ the limit of $F_N(t)$ when $N \to +\infty$. Then the linearization of T lies on the simple idea that the delta method can be generalized to maps of functions. This is the principle of the functional delta method (see e.g. VAN DER VAART[1998], chapter 20). Roughly, if T is "smooth" at F and $F_S - F_N$ is small enough, then

$$\sqrt{n}(T(F_S) - T(F_N)) \quad \simeq \quad dT_F(\sqrt{n}(F_S - F_N)), \tag{7}$$

where dT_F is a linear function. The right part of (7) is thus a mean which depends on S through F_S. Hence, as previously, we can expect $T(F_S)$ to be asymptotically normal under a smoothness condition on T and regularity of the sampling design.

This intuition can be formalized as follows. Firstly, the smoothness of T is expressed in terms of Hadamard differentiability. A statistic is said to be Hadamard differentiable at F

with respect to the tangent space B if there is a linear and continuous function dT_F such that:

$$\frac{T(F + th_t) - T(F)}{t} - dT_F(h) \longrightarrow 0 \quad \text{as } t \to 0, \ ||h_t - h||_\infty \to 0, \text{ and } h \in B \qquad (8)$$

where $||.||_\infty$ denotes the supremum norm.[3] Moreover T is said to be continuously Hadamard differentiable if

$$\frac{T(F_t + th_t) - T(F_t)}{t} - dT_F(h) \longrightarrow 0 \quad \text{as } t \to 0, \ ||h_t - h||_\infty \to 0 \ (h \in B), \ ||F_t - F||_\infty \to 0 \quad (9)$$

Special directional derivatives will be useful hereafter. As DEVILLE[1999], but in a slightly different way than usually, we define the influence functions by $IT(y, F) = dT_F(g_y)$, where $g_y : x \mapsto 1\{y \leq x\}$.

Hadamard differentiability is stronger than Gateaux differentiability, which corresponds to $h_t = h$, but weaker than the Fréchet differentiability, which is differentiability in the usual sense. Many nonlinear statistics are in fact Hadamard differentiable, so that it has been extensively used, for instance in duration models (see e.g. GILL[1989]). We just display the examples of the poverty rate, which will be used in the application section 3, and of the interquantile ratio. The proofs of the claims are deferred to appendix A.

Example 1 *(poverty rate) A general poverty rate, for a distribution of income F, is defined (see DEVILLE[1999]) by*

$$J_{(\alpha, \beta)}(F) = F(\beta F^{-1}(\alpha)), \qquad (10)$$

for $0 < \alpha, \beta < 1$. The classical poverty rate corresponds to $\alpha = 0,5$ and $\beta = 0,5$ (INSEE) or $0,6$ (EUROSTAT). if F is differentiable everywhere and $F'(F(\alpha)) \neq 0$, $F \mapsto J_{(\alpha, \beta)}(F)$ is Hadamard differentiable tangentially to the set of continuous functions (but it is not continuously Hadamard differentiable). Moreover, its influence function exists for $y > 0$, $y \neq F^{-1}(\alpha)$ and satisfies

$$IJ_{(\alpha, \beta)}(y, F) = 1\{y \leq \beta F^{-1}(\alpha)\} - \frac{\beta F'(\beta F^{-1}(\alpha))}{F'(F^{-1}(\alpha))} 1\{y \leq F^{-1}(\alpha)\}. \qquad (11)$$

Example 2 *(Interquantile ratio) The interquantile ratio R_α is defined by*

$$R_\alpha(F) = \frac{F^{-1}(1 - \alpha)}{F^{-1}(\alpha)}, \qquad (12)$$

for $0 < \alpha < 1/2$. The usual choice of $\alpha = 0.1$ leads to the interdecile ratio. If F is differentiable at $F^{-1}(\alpha)$ and $F^{-1}(1-\alpha)$ with $F'(F(\alpha)) \neq 0$ and $F'(F(1-\alpha)) \neq 0$, $F \mapsto R_\alpha(F)$

---

[3]In fact we could choose another norm, see e.g. BARBE AND BERTAIL[1995] for a discussion on that point.

is Hadamard differentiable tangentially to the set of functions which are continuous at $F^{-1}(\alpha)$ and $F^{-1}(1-\alpha)$ (but it is not continuously Hadamard differentiable). Moreover its influence function exists for $y \notin \{F^{-1}(\alpha),\ F^{-1}(1-\alpha)\}$ and satisfies

$$IR_\alpha(y,F) = \frac{1}{F^{-1}(\alpha)}\left[-\frac{1\{y \le F^{-1}(1-\alpha)\}}{F'(F^{-1}(1-\alpha))} + R_\alpha(F)\frac{1\{y \le F^{-1}(\alpha)\}}{F'(F^{-1}(\alpha))}\right]. \tag{13}$$

Now, the regularity condition on the sampling design writes formally

$$\sqrt{n}(F_S - F_N) \xrightarrow{d} W, \tag{14}$$

where W is a centered continuous gaussian process. In other terms, the empirical process $G_S = \sqrt{n}(F_S - F_N)$ converges weakly to a gaussian process.[4] Eventually, because weak convergence does not imply convergence of moments, the variance estimator can be inconsistent because of "heavy tails" due to the sampling design, just as in bootstrap schemes (see SHAO[1990]).[5] A uniform integrability condition is required to handle this issue.

$$\lim_{M\to+\infty}\ \limsup_{N} E[(dT_F^2(G_S) - M)1\{dT_F^2(G_S) > M\}] = 0. \tag{15}$$

Once these new conditions on T and the sampling design have been introduced, we obtain the following generalization of the usual delta method.

Proposition 2.2 *(functional delta method) If (14) holds, and*

1) T is continuously Hadamard differentiable at F with respect to the tangent space B including the paths of W,

or 2) assumption S holds and T is (simply) Hadamard differentiable at F with respect to the tangent space B including the paths of W,

then

$$\sqrt{n}(T(F_S) - T(F_N)) \xrightarrow{d} \mathcal{N}(0, V(dT_F(W))). \tag{16}$$

Moreover, if (15) holds, $V(dT_F(W)) \neq 0$ and $IT(y_k, F)$ exists for all $k \in \mathbb{N}$, then T is linearizable and

$$lin_k T = \frac{1}{\sum_{k\in S} w_k}\left[IT(y_k, F) - \frac{1}{\sum_{k\in S} w_k}\sum_{k\in S} w_k IT(y_k, F)\right], \tag{17}$$

where, once again, we do not take into account the randomness of $lin_k T$ in the computation of the variance of $\sum_{k\in S} w_k lin_k(T)$.

---

[4]Weak convergence of a sequence of processes G_n is defined by the convergence, for all h continuous and bounded, of $E^*[h(G_n)]$ where E^* is the outer expectation, see VAN DER VAART AND WELLNER[1996] section 1.3 for more details.

[5]This contrasts with classical theory where the asymptotic variance writes (under weak regularity conditions) $V(IT(Y, F))$ (where Y is a random variable with cdf F) and can thus always be estimated consistently by its empirical counterpart.

As in the simple delta method, extra assumptions are needed compared to classical statistics, because the parameter of interest $T(F_N)$ varies with N. BERTAIL ET AL.[2003] emphasize on continuous Hadamard differentiability (see their property 1). This, however, excludes many statistics such as the poverty rate and the interquantile ratio. On the other hand, we show that simple Hadamard differentiability suffices under a superpopulation model and if (roughly) the sampling rate tends to zero. If, furthermore, (15) and regularity conditions on F hold,[6] the poverty rate and the interquantile ratio are then linearizable.

Hence, the functional delta method is a very powerful tool for making inference on complex statistics. However, there is a price to pay for allowing more flexibility on T. Indeed, the regularity condition (14) on the sampling design is more restrictive than (3), since (14) implies (see VAN DER VAART[1998], theorem 18.14):

$$
\begin{cases}
\forall\, t = (t_1, ...t_k)\ \sqrt{n}(F_S(t_1) - F_N(t_1), ..., F_S(t_k) - F_N(t_k)) \overset{d}{\longrightarrow} \mathcal{N}(0, \Sigma_t). \\
\text{The process } \sqrt{n}(F_S - F_N) \text{ is asymptotically tight.}
\end{cases}
\tag{18}
$$

By the Cramér-Wold device, the first condition is satisfied if and only if (3) holds for every variable y'_k which is a linear combination of the $(1\{y_k \leq t\})_{t \in \mathbb{R}}$. But the second condition adds restrictions on the sampling design. The question of how demanding these restrictions may be is beyond the scope of this paper and has not been studied in the literature yet. But we still can see that there is a trade-off between the smoothness imposed on T and the regularity imposed on the sampling design. Theoretical results for complex statistics such as the poverty headcount in complex surveys like the French "Taxable Income Survey" (see next section for a description) are thus hard to assess and often admitted by the statisticians. In the following, (3) will be referred to as a "weak normality condition", and (14) as a "strong normality condition".

Up to now, we have described two main situations in which linearization can be established. However, this linearization is seldom directly applicable. Indeed, F is unknown and thus $IT(y, F)$ cannot be determined in general, as examples 1 and 2 emphasize. However, if $IT(y, F)$ depends on a finite number of unknown parameters, as in the poverty rate or the interquantile ratio, this problem can be handled by the following proposition, which generalizes a result of DEVILLE[1999].

Proposition 2.3 *Let T be a linearized statistic with $lin_k(T) = IT(y_k, F) = v(y_k, \theta)$ where $\theta \in \mathbb{R}^p$ (p independent of N) is unknown but can be estimated consistently by $\widehat{\theta}$. Suppose that*

---

[6]For instance, $F = 0$ on $]-\infty, 0]$ and F is differentiable everywhere. Under assumption S, with probability one, $y_k \notin \{F^{-1}(\alpha), F^{-1}(1 - \alpha)\}$ for all $k \in \mathbb{N}$, so that $IT(y_k, F)$ exists for all k (with probability one).

1) there exists a neighborhood $\mathcal{V}$ of θ such as

$$\sup_{u \in \mathcal{V}} \left| V\left(\int v(y, u) dG_S(y) \right) - V\left(\int v(y, u) dW(y) \right) \right| \to 0. \tag{19}$$

2) $u \to V\left(\int v(y, u) dW(y) \right)$ is continuous at θ.
Then T is also linearizable with

$$lin_k(T) = \frac{1}{\sum_{k \in S} w_k} \left[v(y_k, \widehat{\theta}) - \frac{1}{\sum_{k \in S} w_k} \sum_{k \in S} w_k v(y_k, \widehat{\theta}) \right]. \tag{20}$$

Assumption (19) can be seen as a strong uniform integrability condition, because it enables variance estimators to converge for all $u \in \mathcal{V}$ and not only at $u = \theta$. The second assumption is weaker than Deville's (1999), which assumes continuity of $\theta \to v(y, \theta)$ for all y. This last assumption is indeed too stringent for many applications, as the example of the poverty rate emphasizes (the same applies to the interquantile ratio).

Exemple 1 (continued) *The influence function of the poverty rate can be written*

$$v(y_k, \theta) = 1\{y \leq \beta \theta_1\} - \frac{\beta \theta_3}{\theta_2} 1\{y \leq \theta_1\}, \tag{21}$$

where $\theta = (\theta_1, \theta_2, \theta_3) = (F^{-1}(\alpha), F'(F^{-1}(\alpha)), F'(\beta F^{-1}(\alpha)))$. Hence $\theta \mapsto v(y_k, \theta)$ is not continuous (because it is not continuous in θ_1). On the other hand, $V(dT_F(W))$ is continuous in θ. This follows from

$$dT_F(W) = \int IT(y, F) dW(y) = W(\beta F^{-1}(\alpha)) - \frac{F'(\beta F^{-1}(\alpha))}{F'(F^{-1}(\alpha))} W(F^{-1}(\alpha)) \tag{22}$$

and the fact that $V(W(t)) \to V(W(s))$ as $t \to s$.[7] Hence if (19) holds, we can choose

$$lin_k T = \frac{1}{\sum_{k \in S} w_k} \left[1\{y \leq \beta F_S^{-1}(\alpha)\} - J_{(\alpha, \beta)}(F_S) - \frac{\beta \widehat{F'}(\beta F_S^{-1}(\alpha))}{\widehat{F'}(F_S^{-1}(\alpha))} \left(1\{y \leq F_S^{-1}(\alpha)\} - \alpha \right) \right], \tag{23}$$

where $\widehat{F'}$ is any locally uniformly consistent estimator of F'.[8]

Lastly, we have supposed until now that the weights were $w_k = 1/\pi_k$. Yet, in most official surveys, the weights are obtained by a calibration procedure and thus depend on S. Variance of such estimators have been studied by DEVILLE AND SÄRNDAL[1992],

---

[7]Indeed by continuity of W, $W(t) \overset{t \to s}{\longrightarrow} W(s)$ almost surely. Thus convergence in distribution holds, whence $E\left(e^{iuW(t)} \right) = e^{-\frac{u^2}{2} V(W(t))} \to e^{-\frac{u^2}{2} V(W(s))}$.

[8]That is, for any x there exists a neighborhood $\mathcal{V}$ of x such as $\sup_{y \in \mathcal{V}} |\widehat{F'}(y) - F'(y)| \to 0$. This assumption is satisfied for instance for the kernel estimator in the i.i.d. case (see for instance SCHUSTER[1969]). In section 3 we use such an estimator, so that we suppose that it still holds in our framework.

STUKEL ET AL.[1996] and by DEVILLE[1999]. In the case of linearization, the result of interest is (DEVILLE[1999]) that $V(T(F_C))$ (where F_C is the estimated distribution function obtained with the calibration weights) can be replaced by $V\left(\sum_{k \in S} \frac{e_k}{\pi_k}\right)$, where e_k are the residuals of the regression of $\mathrm{lin}_k T$ on the auxiliary variables x_k used in the calibration.

3 Computation of covariances when samples are dependent

Under the normality conditions stated above, the linearization method enables to yield confidence intervals and to make tests on a wide range of complex statistics. Unfortunately, the applied statistician (or economist) is often more interested in assessing whether or not evolutions over time are significant. Since many surveys such as the French "Taxable Income Survey" exhibit time dependency (see next section for a precise description of this dependency), calculating confidence intervals without taking into account the covariance term would make us overestimate variance and lower the chances of finding significant evolutions in our data. In this subsection, we show that linearization can be used to estimate this covariance.

Let us consider $(S_1, \ S_2)$ two samples from a same population U, $(w_{1k}, w_{2k})_{k \in U}$ the corresponding weights, and $T = (T_1, T_2)$ a parameter of interest. T_1 and T_2 are two parameters which depend on different variables of interest, for instance the income at t and $t' \neq t$. We suppose that its estimator $\widehat{T} = (\widehat{T}_1, \ \widehat{T}_2)$ is based on (S_1, S_2). Since the aim is to study linear combinations of $(\widehat{T}_1, \ \widehat{T}_2)$ (for example the difference), a bidimensional version of linearization is now required.

$$\left(V\left(\sum_{k \in S_1} w_{1k}\mathrm{lin}_k T_1, \ \sum_{k \in S_2} w_{2k}\mathrm{lin}_k T_2\right)\right)^{-1/2}(\widehat{T} - T) \xrightarrow{d} \mathcal{N}(0, I_2), \qquad (24)$$

where I_2 is the 2×2 identity matrix. If T is a smooth function of a mean of $(y_{1k}, \ y_{2k})$ on $S_1 \times S_2$, equation (24) is satisfied under a weak gaussian condition :

$$V(\widehat{t}_y)^{-1/2}(\widehat{t}_y - t_y) \xrightarrow{d} \mathcal{N}(0, I_2), \qquad (25)$$

where $t_y = (\sum_{k \in U} y_{1k}, \sum_{k \in U} y_{2k})$ and $\widehat{t}_y = (\sum_{k \in S_1} w_{1k}y_{1k}, \sum_{k \in U} w_{2k}y_{2k})$.

If $F_1 \mapsto T_1(F_1)$ and $F_2 \mapsto T_2(F_2)$ are Hadamard differentiable, then this is also the case of $F = (F_1, F_2) \mapsto T(F) = (T_1(F_1), T_2(F_2))$. Consequently the functional delta method can be used under a strong gaussian condition :

$$\sqrt{n}(F_S - F_N) \xrightarrow{d} W \qquad (26)$$

where $F_N = (F_{1,N}, F_{2,N})$, $F_{1,N}$ (resp. $F_{2,N}$) being the cumulative distribution function of the $(y_{1k})_{k \in U}$ (resp. $(y_{2k})_{k \in U}$) in U and $F_S = (F_{1S_1}, F_{2S_2})$ is defined analogously. W is now a

bidimensional continuous gaussian process. It is worth noticing that conditions on (S_1, S_2) to assess (26) (or even (25)) are stronger than the ones for single surveys. But to the best of our knowledge, they have not been studied yet in the literature.

Proposition 3.1 *(multivariate functional delta method) If (26) holds and*

1) T_1 and T_2 are continuously Hadamard differentiable at $F_1(= \lim_{N \to \infty} F_{1,N})$ and F_2 with respect to the tangent spaces B_1 and B_2 with $W \in B_1 \times B_2$,

or

2) $\sqrt{n}(F_{1,N} - F_1, F_{2,N} - F_2)$ admits a limit function h and T_1 and T_2 are simply Hadamard differentiable at F with respect to the tangent space including h and the paths of $W + h$,

then,

$$\sqrt{n}(T(F_S) - T(F_N)) \xrightarrow{d} \mathcal{N}(0, V(dT_F(W))). \tag{27}$$

Moreover, suppose that $dT_{1F_1} \neq 0$, $dT_{2F_2} \neq 0$, that (15) is satisfied for $dT^2_{1F_1}(G_{S_1})$ and $dT^2_{2F_2}(G_{S_2})$, and that the following cross uniform integrability condition holds :

$$\lim_{M \to +\infty} \limsup_N E[(dT_{1F_1}(G_{S_1})dT_{2F_2}(G_{S_2}) - M)1\{dT_{1F_1}(G_{S_1})dT_{2F_2}(G_{S_2}) > M\}] = 0. \tag{28}$$

Then T is linearizable and $lin_k T = (lin_k T_1, lin_k T_2)$ where

$$lin_k T_i = \frac{1}{\sum_{k \in S_i} w_{ik}} \left[IT_i(y_{ik}, F_i) - \frac{1}{\sum_{k \in S_i} w_{ik}} \sum_{k \in S_i} w_{ik} IT_i(y_{ik}, F_i) \right] \quad i \in \{1, 2\}. \tag{29}$$

Proposition (2.3) has also a direct multivariate generalization, so that under regularity conditions, one can compute variables $lin_k T$ satisfying (24) and depending only on (S_1, S_2) and not on (F_1, F_2) which is unknown.

Now, if T is linearizable, $V(\widehat{T}_1)$ (resp. $V(\widehat{T}_2)$) can be approximated by $V(\sum_{k \in S_1} w_{1k} lin_k T_1)$ (resp. $V(\sum_{k \in S_2} w_{2k} lin_k T_2)$) as explained in the first section. Moreover, $\text{Cov}(\widehat{T}_1, \widehat{T}_2)$ can be approximated by

$$\text{Cov}\left(\sum_{k \in S_1} w_{1k} lin_k T_1, \sum_{l \in S_2} w_{2l} lin_l T_2\right) = \sum_{k,l \in U} w_{1k} w_{2l} lin_k T_1 lin_l T_2 \Delta^{12}_{kl}, \tag{30}$$

where $\Delta^{12}_{kl} = Pr(k \in S_1, l \in S_2) - Pr(k \in S_1)Pr(l \in S_2)$. Thus the covariance can be estimated by

$$\widehat{\text{Cov}} = \sum_{k \in S_1, l \in S_2} w_{1k} w_{2l} lin_k T_1 lin_l T_2 \left(1 - \frac{Pr(k \in S_1)Pr(l \in S_2)}{Pr(k \in S_1, l \in S_2)}\right). \tag{31}$$

Note that, similarly to the Horvitz-Thompson estimator of variance, this estimator may be of few practical interest because the second order probabilities $Pr(k \in S_1, l \in S_2)$ are

often hard to compute. Several solutions have been proposed to handle this problem (see e.g. KISH[1965], or BERGER[2004b],). We shall, however, stick to this estimator in our application because the second order probabilities are in fact computable in the French "Taxable Income Survey". Furthermore, despite the complexity of this survey, they lead to quite simple estimators.

In the case of unit nonresponse, (31) can be rather complex. However, and as we will see in the application, the following device may help to simplify the estimator. Let us define R_1 (resp. R_2) the subsample of respondents in S_1 (resp. S_2). Then, under the hypothesis that response behaviors of two persons at two different time are independent, we have

$$Pr(k \in R_1, \, l \in R_2) = Pr(k \in S_1, \, l \in S_2)P(k \in R_1 | k \in S_1)P(l \in R_2 | k \in S_2), \qquad (32)$$

which remains valid for $k = l$ under the assumption that responding to the first survey does not influence the response to the second one.[9] Hence,

$$\frac{Pr(k \in R_1)Pr(l \in R_2)}{Pr(k \in R_1, \, l \in R_2)} = \frac{Pr(k \in S_1)Pr(l \in S_2)}{Pr(k \in S_1, \, l \in S_2)}, \qquad (33)$$

and therefore the estimator can be rewritten

$$\widehat{\text{Cov}} = \sum_{k \in R_1, \, l \in R_2} w_{1k}w_{2l}\text{lin}_kT_1\text{lin}_lT_2 \left(1 - \frac{Pr(k \in S_1)Pr(l \in S_2)}{Pr(k \in S_1, \, l \in S_2)}\right), \qquad (34)$$

where $w_{1k} = \widehat{Pr}(k \in R_1) = Pr(k \in S_1)\widehat{\text{Pr}}(k \in R_1 | k \in S_1)$. As we will see in subsection 3.2, the advantage of (34) is that $Pr(k \in S_1)Pr(l \in S_2)/Pr(k \in S_1, \, l \in S_2)$ is rather simple, so that this expression can be quite simple. Moreover, under (32), this estimator is unbiased when T_1 and T_2 are linear.

4 Application to testing the evolution of the poverty rate in France

In this last section we apply the results obtained previously to answer a typical question that national statistical offices often address. What can be said about the evolution of a social indicator during a given time span ? On the one hand, reading naively point estimator tables may lead to over-interpret the data, which may not be accurate enough to exhibit statistically significant trends. On the other hand, neglecting dependency between samples may lead to too conservative conclusions.

We focus here on the poverty headcount defined as the proportion of people who earn less than 50% of the median of incomes. The estimators are computed on the basis of the French

---

[9]In other words, we suppose that people do not weary of answering the survey.

"Taxable Income Survey". The details of the sampling design and the computation of the variances are given in the first subsection. As the survey exhibit a strong time dependency, covariances must also be computed, which is the purpose of the second subsection. The third subsection presents a simple solution to linearize an individual indicator like the poverty rate when the sample is based on households. Lastly, the fourth subsection displays the tests of evolution themselves.

4.1 Computation of the variance terms

The French "Taxable Income Survey" is based on the Annual Labour Force Survey, which was conducted by the INSEE until 2002. If we neglect the details of its implementation (triple-stage sampling), the design of this wave survey is basically a stratified area sample design (see ROTH[1991], and also *INSEE Résultats, Enquête Emploi*[1999].

The strata are groups of same urban unit in each French region (where Provence-Alpes-Côte d'Azur and Corsica are gathered together). In each stratum, primary units are area groups, which are made of four areas each.[10] Once an area group is selected, one of its area is surveyed. The year after, this area is still surveyed, with one more area of the group. Each area is surveyed three successive years, so that the third year, three areas are surveyed (the last area is used as a reserve). Lastly, a group can be sampled only once.

We model the "Taxable Income Survey" sample design as a stratified cluster sampling design. In stratum $h \in \{1, ..., H\}$, a sample S_{ah} of m_h areas from M_h is drawn by a simple random sampling without replacement (SRSWOR hereafter) with a sampling rate of $1/300$.[11] Thus we neglect the effect of drawing groups rather than areas directly.

In addition to unit nonresponse of the Labour Force Survey per se, the "Taxable Income Survey" is not based on the full sample of respondents of the Labour Force Survey. Indeed this survey matches respondents from the Labour Force Survey with their income tax returns, which are provided by the French Internal Revenue Service. Sometimes, no tax return can be found for a household in the Annual Labor Force Survey sample. Such failures to match respondents and tax returns are considered as unit nonresponses. We model global nonresponse as a Poisson sampling with a constant per stratum sampling rate f_h. In other words, we suppose that in each stratum h every household has the same probability to respond and to be matched, namely $1/f_h$, and that response (and matching) is independent from a household to another. The sample of respondents in area g of stratum h is denoted R_{hg}.

---

[10] An area theoretically includes 40 households in the case of urban units with fewer than 100,000 inhabitants and 20 for the one with more than 100,000 inhabitants.

[11] For practical reasons, and whereas there is a small variation between strata, we have to assume that the sampling rate is identical in each stratum.

Eventually, adjustments are made for total nonresponse. There are two major calibration stages independent from each other. First, the sample of the Labour Force Survey is calibrated on the demographic structure. Then, another calibration is made after the matching with the data from the tax administration. For the sake of simplicity, and even though it is not equivalent in principle, we act as if there were a single calibration stage and we regress the dependent variable, that is to say the linearized function of the poverty rate, on all of the variables used for either calibration simultaneously.[12] As mentioned in the first section, the variance formulas (and covariance formulas as well) are then computed on the residuals of the regression of the linearized function[13] on these variables (these residuals are denoted z_k hereafter).

To sum up, the French "Taxable Income Survey" can be modeled as a two-stage sampling where the first stage is a stratified random sampling of geographical clusters (area sampling) and the second one is a Poisson sampling design. We thus have in each stratum[14] the classical variance formula for the estimator of the total t_z :

$$V(\widehat{t_z}) = V_{SU} + V_{PU} = \frac{M}{m} \frac{1-f}{f} \sum_{g=1}^{M} \sum_{k=1}^{N_g} z_{gk}^2 + M^2 \left(1 - \frac{m}{M}\right) \frac{S^2}{m}, \qquad (35)$$

where N_g is the number of households in area g, z_{gk} is the variable of interest for household k in area g, $S^2 = \frac{1}{M-1} \sum_{g=1}^{M} \left(t_{zg} - \frac{1}{M} t_z\right)^2$ and t_{zg} is the total of the variable of interest on area g. Then the variance can be estimated by

$$\widehat{V} = \widehat{V}_{SU} + \widehat{V}_{PU} = \frac{M}{m} \frac{1-\widehat{f}}{\widehat{f}^2} \sum_{g \in S_a} \sum_{k \in R_g} z_{gk}^2 + M^2 \left(1 - \frac{m}{M}\right) \frac{s^2}{m}, \qquad (36)$$

where $s^2 = \frac{1}{m-1} \sum_{g \in S_a} \left(\widehat{t}_{zg} - \frac{1}{m} \widehat{t}_z\right)^2$ and $\widehat{f}, \widehat{t}_{zg}$ and $\widehat{t}_z$ are the Horvitz-Thompson estimator of f, t_{zg} and t_z. Note that $\widehat{V}_{PU}$ and $\widehat{V}_{SU}$ are biased, but the sum of the two is unbiased [15] if we ignore the bias stemming from the estimation of f.

4.2 Estimation of the covariances

Figure 1 shows how the French "Taxable Income Survey" is built. S, M and E refer respectively to the exiting, median and entering third of the Labour Force Survey. The entering third

---

[12]These variables used for calibration are the household type (5 modes), the region (8 modes), a dummy variable indicating whether the reference person is a student, an ILO classification of status in employment (4 modes), age combined with gender (total of 10 modes, like in the calibration of the Labour Force Survey) and individual occupational category (7 modes).

[13]See subsection 3.1 for details on the computation of this linearized function.

[14]For the sake of simplicity the index h is omitted thereafter (S_{ah} is replaced by S_a for instance).

[15]See for instance TILLE[2001].

becomes median and then exiting third thus constituting a 3-years long panel. Cells in grey correspond to the subsamples of the "Taxable Income Survey".

Year

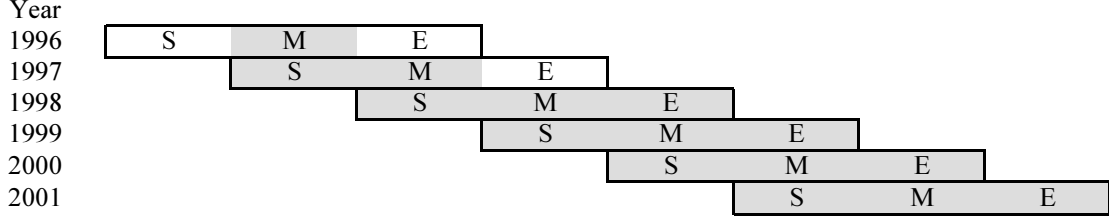|							
1996	S	M	E				
1997		S	M	E			
1998			S	M	E		
1999				S	M	E	
2000					S	M	E
2001						S	M	E

Figure 1: Sample Dependency Structure of the "Taxable Income Survey"

To compute covariances, we first note that, because samples in different strata are independent from one another,

$$\text{Cov} = \sum_{h=1}^{H} Cov_h. \tag{37}$$

Thus we can restrict to covariances in a given stratum. Moreover, we neglect the covariances between samples at t and t' when $|t - t'| > 2$. Theoretically there is yet a (generally negative) covariance due to the fact that areas are drawn only once. However, because of the small sampling rate, these correlations are negligible and we only address the cases $(t, t+1)$ and $(t, t+2)$.

4.2.1 Case of t (sample S_1) and $t+1$ (sample S_2) when $t \geq 1998$

Let us consider the "complete" sample $S_1 \cup S_2$. It is constituted of a SRSWOR of exiting, median and entering areas of date t, and of entering areas of date $t+1$. Once $S_1 \cup S_2$ has been drawn, we reshuffle randomly areas in order to have four sub-sample of equal size. Denoting by π_k^c (resp. π_k) the inclusion probability of k in the "complete" sampling (resp. in only S_1 or S_2), we have

$$\begin{cases} \pi_k^c &= \frac{m^c}{M} \\ \pi_k &= \frac{m}{M} = \frac{3}{4}\pi_k^c. \end{cases} \tag{38}$$

Let us now compute the second order inclusion probability $P(k \in S_1, l \in S_2)$. We face two cases depending on whether k and l are in the same area.

– If k and l are in the same area, then

$$Pr(k \in S_1, l \in S_2) = Pr(k \in S_1 \cap S_2) = \frac{1}{2}\pi_k^c = \frac{2}{3}\frac{m}{M}. \tag{39}$$

– If k and l are in two different areas : then, denoting by $P(k \in i)$ the probability of being in the i-th sub-sample (with $1 =$ exiting ones at date t, $2 =$ median ones at date t, $3 =$ entering

15

ones at date t and finally 4 = entering ones at date $t + 1$),

$$Pr(k \in S_1, l \in S_2) = \pi_{kl}^c P(k \notin 4, l \notin 1) \tag{40}$$

$$= \pi_{kl}^c (P(k \notin 4) - P(k \notin 4, l \in 1)) \tag{41}$$

$$= \pi_{kl}^c \left(\frac{3}{4} - \frac{1}{3} P(k, l \in \{1, 2, 3\}) \right), \tag{42}$$

where the third equality stems from the symmetry between samples 1, 2 and 3. Now picking up k and l in $\{1, 2, 3\}$ amounts to picking them both with a SRSWOR in a population of size $\frac{3}{4}m^c = m$ among $m_c = 4/3m$. Therefore,

$$Pr(k \in S_1, l \in S_2) = \pi_{kl}^c \left(\frac{3}{4} - \frac{1}{3} \frac{m(m-1)}{\frac{4}{3}m(\frac{4}{3}m - 1)} \right) \tag{43}$$

$$= \frac{\pi_{kl}^c}{4} \left(3 - \frac{m-1}{\frac{4}{3}m - 1} \right) \tag{44}$$

$$= \frac{\pi_{kl}^c}{4} \frac{3m - 2}{\frac{4}{3}m - 1}. \tag{45}$$

Given the fact that the "complete" sampling is a SRSWOR of m^c areas among M,

$$\pi_{kl}^c = \frac{m^c(m^c - 1)}{M(M-1)} = \frac{\frac{4}{3}m(\frac{4}{3}m - 1)}{M(M-1)}, \tag{46}$$

and thus

$$Pr(k \in S_1, l \in S_2) = \frac{m(3m - 2)}{3M(M-1)}. \tag{47}$$

After rearranging the sums (see appendix B for the detail of computation), this yields

$$\widehat{\mathrm{Cov}} = \left(1 - \frac{3}{2} \frac{m}{M} \right) \sum_{g \in S_{a1} \cap S_{a2}} \widehat{t}_{z_1 g} \widehat{t}_{z_2 g} + \left(1 - \frac{3m(M-1)}{M(3m-2)} \right) \left(\widehat{t}_{z_1} \widehat{t}_{z_2} - \sum_{g \in S_{a1} \cap S_{a2}} \widehat{t}_{z_1 g} \widehat{t}_{z_2 g} \right), \tag{48}$$

where S_{a1} (resp. S_{a2}) denotes the sample of area at t (resp. $t + 1$) and z_1 (resp. z_2) is the variable of interest at t (resp $t + 1$).

4.2.2 Case of t (sample S_1) and $t + 2$ (sample S_2) when $t \geq 1998$.

The reasoning is similar (see appendix B for details) and leads to

$$\widehat{\mathrm{Cov}} = \left(1 - 3 \frac{m}{M} \right) \sum_{g \in S_{a1} \cap S_{a2}} \widehat{t}_{z_1 g} \widehat{t}_{z_2 g} + \left(1 - \frac{3m(M-1)}{M(3m-1)} \right) \left(\widehat{t}_{z_1} \widehat{t}_{z_2} - \sum_{g \in S_{a1} \cap S_{a2}} \widehat{t}_{z_1 g} \widehat{t}_{z_2 g} \right) \tag{49}$$

with the same notation as previously.

In 1996 and 1997, the pattern of the sampling renewal differs from other years. This leads to different covariance estimators, which are given below. [16]

- 1996-1997 :

$$\widehat{Cov} = \left(1 - \frac{2m}{M}\right) \sum_{g \in S_{a1} \cap S_{a2}} \widehat{t}_{z_1 g} \widehat{t}_{z_2 g} + \left(1 - \frac{2m(M-1)}{M(2m-1)}\right) \left(\widehat{t}_{z_1} \widehat{t}_{z_2} - \sum_{g \in S_{a1} \cap S_{a2}} \widehat{t}_{z_1 g} \widehat{t}_{z_2 g}\right) \tag{50}$$

- 1996-1998 : since there is no common subsample between these two years, we assume $Cov = 0$.

- 1997-1998 :

$$\widehat{Cov} = \left(1 - \frac{3m}{M}\right) \sum_{g \in S_{a1} \cap S_{a2}} \widehat{t}_{z_1 g} \widehat{t}_{z_2 g} + \left(1 - \frac{3m(M-1)}{M(3m-1)}\right) \left(\widehat{t}_{z_1} \widehat{t}_{z_2} - \sum_{g \in S_{a1} \cap S_{a2}} \widehat{t}_{z_1 g} \widehat{t}_{z_2 g}\right) \tag{51}$$

- 1997-1999 : for the same reason as for 1996-1998, we assume $Cov = 0$.

4.3 Accounting for clustering specific to households

The linearization of poverty rate has been carried out and implemented within a general framework. But it needs to be adapted to a recent change in the INSEE's estimations of such indices, namely the use of individual distributions derived from household surveys.

More precisely, the variable on which the poverty rates are estimated is the living standard, which is identical for every member of the household.[17] Yet, the INSEE and Eurostat estimate poverty rates based on the population of individuals. To address this issue, we could consider the sample of individuals (i.e. individuals who belong to surveyed households) instead of the sample of households. However, the hypothesis of Poisson sampling for unit nonresponse is not realistic in this case, because in the Labour Force survey the whole household is considered a nonrespondent as soon as one of its members does not answer. Instead, we use the fact that the distribution of living standards on individuals can be easily deduced from the one on households.

---

[16]Detail of the computation of these special cases can be obtained from the authors upon request.

[17]The living standard is the disposable income (*i.e.* income after benefits and taxes) of the household, deflated by an equivalence scale to account more accurately for household size and structure.

Let J_{ind} denote the individual poverty rate, that is

$$J_{\text{ind}(\alpha,\beta)} = F_{\text{ind}}\left(\beta F_{\text{ind}}^{-1}(\alpha)\right), \tag{52}$$

where F_{ind} and F_{ind}^{-1} are defined on the population of individuals. For example,

$$F_{\text{ind}}(x) = \frac{1}{N_{\text{ind}}} \sum_{i \in U_{\text{ind}}} 1\{y_i \leq x\} = \frac{1}{\sum_{k \in U_{\text{hh}}} n_k} \sum_{k \in U_{\text{hh}}} n_k 1\{y_k \leq x\}, \tag{53}$$

where N_{ind} is the number of individuals, U_{ind} (resp. U_{hh}) is the population of individuals (resp. households) and n_k is the size of household k. F_{ind} (and also F_{ind}^{-1}) can therefore be estimated at the household level by using the individual weights instead of the household weights, or, in other words, multiplying the household weights by the number of individuals in the household. Hence, with obvious notations,

$$\widehat{J_{\text{ind}(\alpha,\beta,w)}} = \widehat{J_{\text{hh}(\alpha,\beta,nw)}}. \tag{54}$$

The same reasoning holds for the linearized function of $J_{\text{ind}(\alpha,\beta)}$. Indeed, the variance of $\widehat{J_{\text{ind}(\alpha,\beta)}}$ can be approximated by $V(dJ(F_{\text{Sind}} - F_{\text{ind}}))$ (dJ being the differential of $J_{\text{ind}(\alpha,\beta)}$ at $F_{\text{ind}}^{\infty} = \lim_{N \to \infty} F_{\text{ind}}$). Now, the living standard of every member of the household is identical, so that $\text{lin}_i J_{\text{ind}(\alpha,\beta)}$ is identical for every member of the same household. In other words, the variance of $J_{\text{ind}(\alpha,\beta)}$ can be estimated by the variance of the total of the variable $n_k \text{lin}_k J_{\text{ind}(\alpha,\beta)}$ (weighted by w_k) at the household level. Furthermore,

$$
\begin{aligned}
\text{lin}_k J_{\text{ind}(\alpha,\beta,w)} &= \frac{1}{\sum_{l \in S_{\text{hh}}} w_l n_l} \left[1\{y_k \leq \beta F_{\text{Sind}}^{-1}(\alpha)\} - \widehat{J_{\text{ind}(\alpha,\beta)}} \right. \\
&\quad \left. -\beta \frac{\widehat{F_{\text{ind}}^{\infty\prime}}(\beta F_{\text{Sind}}^{-1}(\alpha))}{\widehat{F_{\text{ind}}^{\infty\prime}}(F_{\text{Sind}}^{-1}(\alpha))} (1\{y_k \leq \beta F_{\text{Sind}}^{-1}(\alpha)\} - \alpha) \right] \tag{55} \\
&= \text{lin}_k J_{\text{hh}(\alpha,\beta,nw)}, \tag{56}
\end{aligned}
$$

where $\widehat{F_{\text{ind}}^{\infty\prime}}$ is a kernel estimator[18] of $F_{\text{ind}}^{\infty\prime}$. Hence, the variable of interest, at the household level, is $y_k = n_k \text{lin}_k J_{\text{hh}(\alpha,\beta,nw)}$. The variable z_k considered in subsections 3.1 and 3.2 is then the residual of the regression of y_k on the variables used in the calibration.

4.4 Testing evolutions

Between 1996 and 2001, while the economic situation in France improved, the poverty rate displayed a decrease (see figure 2). However, the question whether this decrease is significant

---

[18]We use a gaussian kernel and a bandwidth determined by the rule of thumb. The effect of a change of the bandwidth is very small on the final estimator of variance.

or not has remained open. Thanks to the covariance terms calculated (see table 1), we can accurately test the significance of the difference of any pair of poverty rates, taking the sample dependency into account.
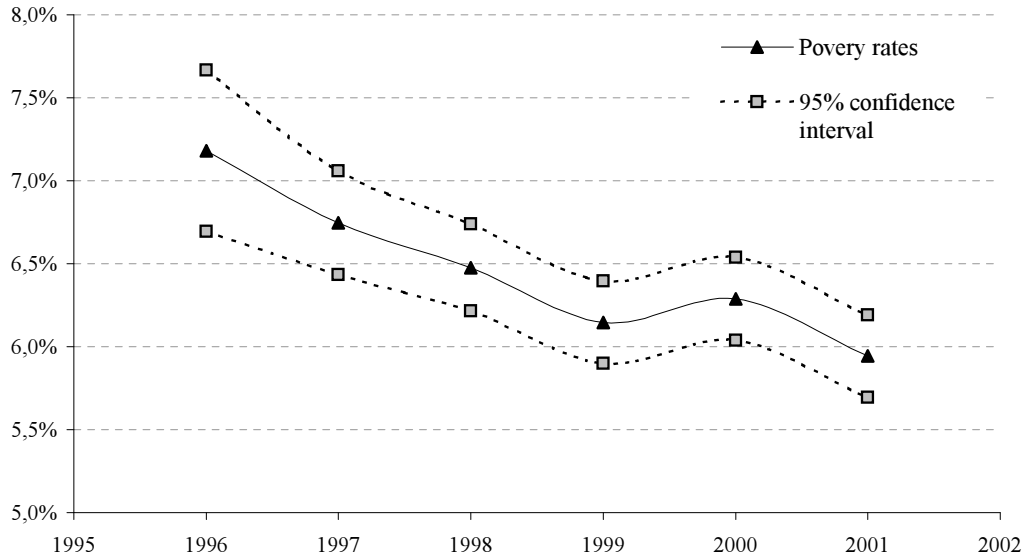

Figure 2: Evolution of the 50% poverty rates

Sources: *Taxable Income Survey* 1996-2001, INSEE-Directorate General of Taxation; "poor field": *i.e.* excluding students and households reporting negative income. Computations by the authors.

Table 1: Variance-Covariance Matrix for the Poverty Rate (50% of the median)

Year	1996	1997	1998	1999	2000	2001
1996	6.15E-06	1.53E-06	0	0	0	0
1997		2.55E-06	6.84E-07	0	0	0
1998			1.78E-06	6.35E-07	2.60E-07	0
1999				1.61E-06	6.112E-07	2.21E-07
2000					1.63E-06	5.73E-07
2001						1.59E-06

Sources: *Taxable Income Surveys* 1996-2001, INSEE- French Internal Revenue Service; "poor field": *i.e.* excluding students and households reporting negative income. Computations by the authors.

Lecture: since, the matrix is symmetric, we only fill the upper triangle for the ease of reading.

Table 2 gives the p-values of tests conducted on pairs of years, the null hypothesis being the equality of the two poverty rates. If dependency is neglected, no year-to-year evolution is significant at the level of 5%. But taking dependency into account allows us to accept the

1998-1999 and 2000-2001 changes as real decreases. On the other hand, the decreases of the beginning of the period cannot be considered significant, as well a the slight increase between 1999 and 2000 which appears as a mere blip.

Table 2: Tests of significance of the evolution of the poverty rate

	taking dependence into account		neglecting dependence	
Years	t-value	p-value	t-value	p-value
1996-1997	1.82	0.07	1.47	0.14
1997-1998	1.57	0.12	1.30	0.19
1998-1999	2.27	0.02	1.79	0.07
1999-2000	-1.00	0.32	-0.79	0.43
2000-2001	2.39	0.02	1.92	0.05
1998-2000	1.11	0.27	1.02	0.31
1999-2001	1.23	0.22	1.14	0.26

Sources: *Taxable Income Surveys* 1996-2001, INSEE-French Internal Revenue Service; "poor field": *i.e.* excluding students and households reporting negative income. Computations by the authors.

5 Conclusion

In this paper we address the issue of measuring evolutions for complex statistics when surveys are dependent. We present a linearization framework which generalizes the paper of DEVILLE[1999] and enables to handle non regular statistics such as quantiles. This method can be used to compute variances but also, under a multivariate functional theorem, covariance terms. We then apply the theory to the evolution of the poverty rate in France between 1996 and 2001. As expected, neglecting the covariance terms lead to too conservative conclusions.

References

BARBE P. and BERTAIL P. (1995), *The weighted bootstrap*, Springer.

BERGER Y. G. (1996), "Asymptotic variance for sequential sampling without replacement with unequal probabilities", *Survey Methodology,* **22**, p. 167–173.

BERGER Y. G. (1998), "Rate of convergence to normal distribution for the Horvitz-Thompson estimator", *Journal of Statistical Planning and Inference,* **67**, p. 209–226.

BERGER Y. G. (2004*a*), "A simple variance estimator for unequal probability sampling without replacement", *Journal of Applied Statistics,* **31**, p. 305–315.

BERGER Y. G. (2004*b*), "Variance estimation for measures of change in probability sampling", *The Canadian Journal of Statistics,* **32**, p. 451–467.

BERTAIL P., BOIZOT C. and COMBRIS P. (2003), "Evaluation de la précision d'estimateurs de fonctionnelles: l'exemple de la consomation alimentaire", *Cahiers d'économie et de sociologie rurale,* **67**, p. 72–101.

BERTAIL P. and COMBRIS P. (1997), "Bootstrap généralisé d'un sondage", *Annales d'Economie et Statistique,* **46**, p. 49–83.

BINDER D. A. (1983), "On the variances of asymptotically normal estimators from complex surveys", *International Statistical Review* **51**, p. 279–292.

BINDER D. A. and KOVACEVIC M. S. (1995), "Estimating some measures of income inequality from survey data: an application of the estimating equations approach", *Survey Methodology,* **21**, p. 137–145.

BINDER D. A. and PATAK Z. (1994), "Use of estimating functions for interval estimation from complex surveys", *Journal of the American Statistical Association,* **89**, p. 1035–1043.

BOOTH J. G., BUTLER R. W. and HALL P. (1994), "Bootstrap methods for finite populations", *Journal of the American Statistical Association,* **89**, p. 1282–1289.

DEMNATI A. and RAO J. N. K. (2004), "Linearization variance estimators for survey data", *Survey Methodology,* **30**, p. 17–34.

DEVILLE J. C. (1999), "Variance estimation for complex statistics and estimators : linearization and residual techniques", *Survey Methodology,* **25**, p. 193–204.

DEVILLE J. C. and SÄRNDAL C. E. (1992), "Calibration estimators in survey sampling", *Journal of the American Statistical Association,* **87**, p. 376–382.

GILL R. D. (1989), "Non- and semi-parametric maximum likelihood estimators and the von-mises method (i)", *Scandinavian Journal of Statistics,* **16**, p. 97–128.

HÁJEK J. (1964), "Asymptotic theory of rejective sampling with varying probabilities from a finite population", *The Annals of Mathematical Statistics,* **35**, p. 1491–1523.

HORVITZ D. G. and THOMPSON D. J. (1952), "A generalization of sampling without replacement from a finite universe", *Journal of the American Statistical Association,* **47**, p. 663–685.

INSEE Résultats, Enquête Emploi (1999).

KISH L. (1965), *Survey Sampling,* Wiley.

KREWSKI D. and RAO J. N. K. (1981), "Inference from stratified samples, properties of the linearization, jackknife and balanced repeated replication methods", *The Annals of Statistics,* **9**, p. 1010–1019.

LAHIRI P. (2003), "On the impact of bootstrap in survey sampling and small-area estimation", *Statistical Science,* **18**, p. 199–210.

RAO J. N. K., WU C. F. J. and YUE K. (1992), "Some recent work on resampling methods for complex surveys", *Survey Methodology,* **18**, p. 209–217.

ROSEN B. (1972), "Asymptotic theory for successive sampling with varying probabilities from a finite population, i.", *The Annals of Mathematical Statistics,* **43**, p. 373–397.

ROTH N. (1991), "L'enquête emploi: échantillon 1992 et années suivantes", *INSEE méthodes,* **29-30-31**.

RUST K. and RAO J. N. K. (1996), "Variance estimation for complex surveys using replication techniques", *Statistical Methods in Medical Research,* **5**, p. 283–310.

SCHUSTER E. (1969), "Estimation of a probability density function and its derivatives", *The Annals of Mathematical Statistics,* **40**, p. 1187–1195.

SEN P. K. (1988), Asymptotics in finite population sampling,, *in* P. R. Krishnaiah and C. R. Rao, eds, "Handbook of statistics", Vol. 6, Elsevier Science Publishers, pp. p. 291–331.

SHAO J. (1990), "Bootstrap estimation of the asymptotic variances of statistical functionals", *Annals Instit. Stat. Math.,* **42**, p. 737–752.

SHAO J. (1994), "L-statistics in complex survey problems", *The Annals of Statistics,* **22**, p. 946–967.

SHAO J. and TU D. (1995), *The Jackknife and Bootstrap*, Springer.

STUKEL D. M., HIDIROGLU M. A. and SÄRNDAL C. E. (1996), "Variance estimation for calibration estimators: A comparison of jackknifing versus taylor linearization", *Survey Methodology,* **22**, p. 117–125.

TILLE Y. (2001), *Théorie des Sondages*, Dunod.

VAN DER VAART A. W. (1998), *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics.

VAN DER VAART A. W. and WELLNER J. (1996), *Weak Convergence and Empirical Process*, Springer.

Appendix

A proofs of section 1

A.1 Proposition 2.1

1) By the fourth assumption of (1) and (3),

$$\sqrt{n}\left(\frac{1}{N}\sum_{k\in S}w_k y_k - \frac{1}{N}\sum_{k\in U}y_k\right) \xrightarrow{d} \mathcal{N}(0, V_{\text{as}}^y). \tag{57}$$

If g is continuously differentiable at $E(y)$, by the uniform delta method (see VAN DER VAART[1998], page 37),

$$\sqrt{n}\left[g\left(\frac{1}{N}\sum_{k\in S}w_k y_k\right) - g\left(\frac{1}{N}\sum_{k\in U}y_k\right)\right] \xrightarrow{d} \mathcal{N}(0, g'^2(E(y))V_{\text{as}}^y). \tag{58}$$

The result follows because $g'^2(E(y))V_{\text{as}}^y$ can be consistently estimated by $g'^2(\frac{1}{N}\sum_{k\in S}w_k y_k)nV\left(\frac{1}{N}\sum_{k\in S}w_k y_k\right)$.

2) First, by assumption S and the law of the iterated logarithm (see VAN DER VAART[1998], section 2.7), with probability one,

$$\limsup\sqrt{\frac{N}{2\ln(\ln(N))}}\left(\frac{1}{N}\sum_{k\in U}y_k - E(y)\right) = -\liminf\sqrt{\frac{N}{2\ln(\ln(N))}}\left(\frac{1}{N}\sum_{k\in U}y_k - E(y)\right) = 1. \tag{59}$$

Since $n = o\left(N/\ln(\ln(N))\right)$, we get, with probability one,

$$\sqrt{n}\left(\frac{1}{N}\sum_{k\in U}y_k - E(y)\right) \longrightarrow 0. \tag{60}$$

Hence, by Slutski's lemma,

$$\sqrt{n}\left(\frac{1}{N}\sum_{k\in S}w_k y_k - E(y)\right) \xrightarrow{d} \mathcal{N}(0, V_{\text{as}}^y). \tag{61}$$

Then by the delta method applied at $E(y)$,

$$\sqrt{n}\left[g\left(\frac{1}{N}\sum_{k\in S}w_k y_k\right) - g(E(y))\right] \xrightarrow{d} \mathcal{N}(0, g'^2(E(y))V_{\text{as}}^y). \tag{62}$$

Now, by definition of differentiability, with probability one,

$$\sqrt{n}\left[g\left(\frac{1}{N}\sum_{k\in U}y_k\right) - g(E(y))\right] \to 0. \tag{63}$$

By applying once more Slutski's lemma, we obtain (58). The conclusion follows $\square$

A.2 Poverty rate

To prove Hadamard differentiability, we use the chain rule (see VAN DER VAART[1998], section 20.2) by noting that the poverty rate can be written as the composed map $\psi_2 \circ \psi_1$ where

$$\begin{cases} \psi_1(F) & = & (F, \beta F^{-1}(\alpha)) \\ \psi_2(F, x) & = & F(x). \end{cases} \tag{64}$$

If F is differentiable at $F^{-1}(\alpha)$, ψ_1 is Hadamard differentiable at F tangentially to the set of functions that are continuous at $F^{-1}(\alpha)$, because quantiles are (see VAN DER VAART[1998], section 21.2). Let us show that ψ_2 is also Hadamard differentiable at $(F, \beta F^{-1}(\alpha))$ tangentially to the set $\mathcal{C} = \{(h,x)/h$ is continuous at x, F is differentiable at $x \}$. Let $\varepsilon_t \in \mathbb{R}$, $\varepsilon \to \varepsilon$ and $h_t \in \mathbb{R}^{\mathbb{R}}$, $||h_t - h||_\infty \to 0$. As F is differentiable at x,

$$\frac{\psi_2(F + th_t, x + t\varepsilon_t) - \psi_2(F, x)}{t} - \varepsilon F'(x) - h(x) = \left(\frac{F(x + t\varepsilon_t) - F(x)}{t} - \varepsilon F'(x) \right)$$
$$+ (h_t(x + t\varepsilon_t) - h(x)). \tag{65}$$

The result follows because both terms tend to zero when $t \to 0$ (the second by continuity of h at x and uniform convergence of h_t). The last point to check to apply the chain rule is that $d\psi_{1|F}(h) \in \mathcal{C}$ for all continuous h. This stems from the fact that

$$d\psi_{1|F}(h) = \left(h, \frac{-\beta h\left(F^{-1}(\alpha)\right)}{F'\left(F^{-1}(\alpha)\right)} \right). \tag{66}$$

Indeed, by assumption h (resp. F) is continuous (resp. differentiable) everywhere, thus it is continuous (resp. differentiable) at $-\beta h\left(F^{-1}(\alpha)\right) / F'\left(F^{-1}(\alpha)\right)$.

To obtain the influence function, first note that $dJ_{(\alpha,\beta)|F}(g_y)$ is well defined for all $y > 0$, $y \neq F^{-1}(\alpha)$ although g_y is not continuous at y. Indeed, we have shown that h must be continuous at $F^{-1}(\alpha)$ (to belong to the tangent space of ψ_1) and at $-\beta h\left(F^{-1}(\alpha)\right) / F'\left(F^{-1}(\alpha)\right)$ (to belong to $\mathcal{C}$). This requirement is fulfilled here, because $y \neq F^{-1}(\alpha)$ and $-g_y\left(F^{-1}(\alpha)\right) / F'\left(F^{-1}(\alpha)\right) \in \{0, -\beta/F'\left(F^{-1}(\alpha)\right)\}$, and both are inferior to $y > 0$.

Now apply the chain rule :

$$dJ_{(\alpha,\beta)F} = d\psi_{2|(F,\beta F^{-1}(\alpha))} \circ d\psi_{1|F} \tag{67}$$

Furthermore, $d\psi_{2|(F,x)}(h, \varepsilon) = \varepsilon F'(x) + h(x)$, so that

$$dJ_{(\alpha,\beta)|F}(h) = d\psi_{2|(F,\beta F^{-1}(\alpha))} \left(h, \frac{-\beta h\left(F^{-1}(\alpha)\right)}{F'\left(F^{-1}(\alpha)\right)} \right) \tag{68}$$

$$= -\beta \frac{F'\left(\beta F^{-1}(\alpha)\right)}{F'\left(F^{-1}(\alpha)\right)} h\left(F^{-1}(\alpha)\right) + h\left(\beta F^{-1}(\alpha)\right). \tag{69}$$

25

The result follows directly by putting $h = g_y$.

To conclude, note that the poverty rate is not continuously Hadamard differentiable. Indeed, it is not even continuously Gateaux differentiable, because $||F_t - F||_\infty \to 0$ does not imply $F'_t(u) \to F'(u)$ for a given u $\square$

A.3 Interquantile ratio

The interquantile ratio $R_\alpha(F)$ can be written as the composed map $\psi_4 \circ \psi_3(F)$ where

$$\begin{cases} \psi_3(F) & = & (F^{-1}(\alpha), F^{-1}(1 - \alpha)) \\ \psi_4(x, y) & = & \frac{x}{y}. \end{cases} \tag{70}$$

ψ_3 is Hadamard differentiable tangentially to the set of functions which are continuous at $F^{-1}(\alpha)$ and $F^{-1}(1 - \alpha)$. ψ_4 is also differentiable (tangentially to $\mathbb{R}^2$). Thus by the chain rule R_α is also Hadamard differentiable tangentially to the set of functions which are continuous at $F^{-1}(\alpha)$ and $F^{-1}(1 - \alpha)$.

g_y is continuous at every $x \neq y$, thus it is continuous at $F^{-1}(\alpha)$ and $F^{-1}(1 - \alpha)$ if $y \notin \{F^{-1}(\alpha), F^{-1}(1 - \alpha)\}$. Hence the influence function exists for these values of y. It is easily obtained by applying the chain rule and noting that

$$d\psi_{4|(x,y)}(u, v) = \frac{1}{y}\left[u - \psi_4(x, y)v\right]. \tag{71}$$

Eventually, the influence function depends on $F'\left(F^{-1}(\alpha)\right)$ and $F'\left(F^{-1}(1 - \alpha)\right)$, thus it is not continuously Gateaux differentiable and a fortiori Hadamard differentiable $\square$

A.4 Proposition 2.2

1) By the uniform functional delta method (see VAN DER VAART AND WELLNER[1996], section 3.9.1), since $W \in B$,

$$\sqrt{n}(T(F_S) - T(F)) \xrightarrow{d} dT_F(W). \tag{72}$$

The result follows because dT_F is a linear functional and W is a centered Gaussian process.

2) By assumption S and the functional law of the iterated logarithm (see for instance VAN DER VAART[1998], section 19.1), with probability one,

$$\sqrt{n}\left(F_N - F\right) \longrightarrow 0 \tag{73}$$

Now, by Slutski's lemma (see VAN DER VAART[1998], section 18.2), with probability one,

$$\sqrt{n}(F_S - F) = \sqrt{n}(F_S - F_N) + \sqrt{n}(F_N - F) \xrightarrow{d} W. \tag{74}$$

Because $W \in B$, we can apply the simple functional delta method (see VAN DER VAART[1998], section 20.2):

$$\sqrt{n}(T(F_S) - T(F)) \xrightarrow{d} dT_F(W). \tag{75}$$

Furthermore, by definition of the Hadamard differentiability, with probability one

$$\sqrt{n}(T(F_N) - T(F)) \longrightarrow 0. \tag{76}$$

Hence, by linearity of dT_F,

$$\sqrt{n}(T(F_S) - T(F_N)) \xrightarrow{d} dT_F(W). \tag{77}$$

The first part of the proposition follows because W is a centered gaussian process and dT_F is linear.

For the second part of the proposition, we first prove that

$$V(dT_F(G_S)) \rightarrow V(dT_F(W)), \tag{78}$$

First note that $dT_F(G_S)$ is well defined. Indeed,

$$G_S = \frac{\sqrt{n}}{\sum_{k \in S} w_k} \sum_{k \in S} w_k \left(g_{y_k} - \frac{1}{N} \sum_{l \in U} g_{y_l} \right), \tag{79}$$

and by assumption $IT(y_k, F)$ exists for all $k \in \mathbb{N}$. Now write $h_m = dT_F^2 \wedge m$. Then, for every m,

$$
\begin{aligned}
|E[dT_F^2(G_S)] - E[dT_F^2(W)]| &\leq |E[dT_F^2(G_S)] - E[h_m(G_S)]| + |E[h_m(G_S)] - E[h_m(W)]| \\
&\quad + |E[h_m(W)] - E[dT_F^2(W)]|.
\end{aligned}
\tag{80}
$$

Fix $\varepsilon > 0$. Because $h_m \rightarrow dT_F^2$, the third term tends to zero by dominated convergence. Hence there exists m_1 such as $|E[h_m(W)] - E[dT_F^2(W)]| < \varepsilon$ for every $m \geq m_1$. By assumption (15), there also exists m_2 such as for every $m \geq m_2$

$$\limsup_N |E[dT_F^2(G_S)] - E[h_m(G_S)]| < \varepsilon. \tag{81}$$

Now take $m_3 = m_1 \vee m_2$. There exists N_1 such as, for every $N \geq N_1$, $E[|dT_F^2 - h_{m_3}|(G_S)] < 2\varepsilon$. Moreover h_{m_3} is continuous and bounded, hence the second term of (80) tends to zero when $N \rightarrow +\infty$ by definition of weak convergence. In other terms, there exists N_2 such as for all $N \geq N_2$, $|E[h_{m_3}(G_S)] - E[h_{m_3}(W)]| < \varepsilon$. Thus, for every $N \geq N_1 \vee N_2$, $|E[dT_F^2(G_S)] - E[dT_F^2(W)]| \leq 4\varepsilon$. To see why this implies (78), note that assumption (15) implies

27

$$\lim_{M \to +\infty} \limsup_N E[(|dT_F(G_S)| - M)1\{|dT_F(G_S)| > M\}] = 0. \tag{82}$$

Indeed for $M \geq 1/4$, $1\{|dT_F(G_S)| > \sqrt{M}\} \leq (|dT_F(G_S)| + \sqrt{M})1\{|dT_F(G_S)| > \sqrt{M}\}$, so that

$$(|dT_F(G_S)| - \sqrt{M})1\{|dT_F(G_S)| > \sqrt{M}\} \leq (dT_F^2(G_S) - M)1\{dT_F(G_S)^2 > M\}. \tag{83}$$

Hence, using (82) similarly than (15) yields $E[dT_F(G_S)] \to E[dT_F(W)] = 0$ and consequently (78) holds. Now, by linearity,

$$dT_F(G_S) = \frac{\sqrt{n}}{\sum_{k \in S} w_k} \sum_{k \in S} w_k \left(IT(y_k, F) - \frac{1}{N} \sum_{l \in U} IT(y_l, F) \right). \tag{84}$$

Thus

$$\frac{T(F_S) - T(F_N)}{\sqrt{V \left(\frac{1}{\sum_{k \in S} w_k} \sum_{k \in S} w_k u_k \right)}} \xrightarrow{d} \mathcal{N}(0, 1), \tag{85}$$

where $u_k = IT(y_k, F) - \frac{1}{N} \sum_{l \in U} IT(y_l, F)$. Note that the left term is well defined because $V(dT_F(W)) \neq 0$, so that $V \left(\frac{1}{\sum_{k \in S} w_k} \sum_{k \in S} w_k u_k \right) \neq 0$ for all N large enough.

Hence we are reduced to the variance of a ratio. An application of the simple delta method shows that, since u_k is centered, the variance is in fact asymptotically equivalent to the variance of the mean $\frac{1}{N} \sum_{k \in S} w_k u_k$. To conclude, let $\theta = \frac{1}{N} \sum_{k \in U} IT(y_k, F)$. Because any (consistent) estimator $\widehat{V}$ of $V(\frac{1}{N} \sum_{k \in S} w_k u_k)$ is a quadratic form of θ, it is continuous in θ. Thus replacing θ by
$\widehat{\theta} = \frac{1}{\sum_{k \in S} w_k} \sum_{k \in S} w_k IT(y_k, F)$ in $\widehat{V}$ does not affect its consistency. By the same reasoning, we can also replace $1/N$ by $1/\sum_{k \in S} w_k$. Whence the second statement of the proposition $\square$

A.5 Proposition 2.3

It suffices to show that

$$\left| V \left(\int v(y, \widehat{\theta}) dG_S(y) \right) - V \left(\int v(y, \theta) dW(y) \right) \right| \xrightarrow{\mathbb{P}} 0, \tag{86}$$

where, in the first variance, we do not take into account the variance due to $\widehat{\theta}$. By continuity of $u \mapsto V \left(\int v(y, u) dW(y) \right)$ at θ, there exists a neighborhood $\mathcal{V}'$ of θ such as for all $u \in \mathcal{V}'$,

$$\left| V \left(\int v(y, u) dW(y) \right) - V \left(\int v(y, \theta) dW(y) \right) \right| < \varepsilon. \tag{87}$$

Furthermore, since $\widehat{\theta}$ is consistent, there exists N_0 such as for all $N \geq N_0$, $\widehat{\theta} \in \mathcal{V} \cap \mathcal{V}'$ with probability $1 - \eta$ for all $\eta > 0$. Now,

$$
\begin{aligned}
&\left| V\left(\int v(y,\widehat{\theta}) dG_S(y) \right) - V\left(\int v(y,\theta) dW(y) \right) \right| \\
\leq\ &\left| V\left(\int v(y,\widehat{\theta}) dG_S(y) \right) - V\left(\int v(y,\widehat{\theta}) dW(y) \right) \right| \\
&+ \left| V\left(\int v(y,\widehat{\theta}) dW(y) \right) - V\left(\int v(y,\theta) dW(y) \right) \right|.
\end{aligned}
\tag{88}
$$

With probability $1 - \eta$, the first term is inferior to the supremum on $\mathcal{V}$. Hence by assumption (19), there exists $N_1 \geq N_0$ such as the first term is inferior to ε with probability $1 - \eta$. With the same probability, the second term is also bounded by ε. Whence the result $\square$

B proofs of section 3

B.1 Case of t (sample S_1) and $t+1$ (sample S_1)

We start from (34) to compute the covariances. In the double sum, we must distinguish depending on whether k and l belong to the same area or not. In the second case, the formula of single and double order inclusion probabilities lead to

$$
\frac{Pr(k \in S_1)Pr(l \in S_2)}{Pr(k \in S_1, l \in S_2)} = \frac{m^2/M^2}{m(3m-2)/(3M(M-1))} = \frac{3m(M-1)}{M(3m-2)}.
\tag{89}
$$

Otherwise,

$$
\frac{Pr(k \in S_1)Pr(l \in S_2)}{Pr(k \in S_1, l \in S_2)} = \frac{m^2/M^2}{2m/3M} = \frac{2m}{3M}.
\tag{90}
$$

Let S_{a1} (resp S_{a2}) denote the sample of areas drawn at t (resp at t+1). Then the covariance estimator writes

$$
\begin{aligned}
\widehat{\text{Cov}} &= \left(1 - \frac{3}{2}\frac{m}{M} \right) \sum_{g \in S_{a1} \cap S_{a2}} \sum_{k,l \in R_g} w_k z_k\, w'_l z'_l \\
&+ \left(1 - \frac{3m(M-1)}{M(3m-2)} \right) \sum_{\substack{g_1 \in S_{a1} \\ g_2 \in S_{a2} \\ g_1 \neq g_2}} \sum_{\substack{k \in R_{g_1} \\ l \in R_{g_2}}} w_k z_k\, w'_l z'_l
\end{aligned}
\tag{91}
$$

$$
= \left(1 - \frac{3}{2}\frac{m}{M} \right) \sum_{g \in S_{a1} \cap S_{a2}} \widehat{t}_{z_1 g}\widehat{t}_{z_2 g} + \left(1 - \frac{3m(M-1)}{M(3m-2)} \right) \sum_{\substack{g_1 \in S_{a1} \\ g_2 \in S_{a2} \\ g_1 \neq g_2}} \widehat{t}_{z_1 g_1}\widehat{t}_{z_2 g_2}
\tag{92}
$$

$$
= \left(1 - \frac{3}{2}\frac{m}{M} \right) \sum_{g \in S_{a1} \cap S_{a2}} \widehat{t}_{z_1 g}\widehat{t}_{z_2 g} + \left(1 - \frac{3m(M-1)}{M(3m-2)} \right)
$$
$$
\left[\sum_{g_1 \in S_{a1} \cap S_{a2}} \widehat{t}_{z_1 g_1} \left(\sum_{\substack{g_2 \in S_{a2} \\ g_2 \neq g_1}} \widehat{t}_{z_2 g_2} \right) + \sum_{g_1 \in S_{a1}/S_{a2}} \widehat{t}_{z_1 g_1} \left(\sum_{\substack{g_2 \in S_{a2} \\ g_2 \neq g_1}} \widehat{t}_{z_2 g_2} \right) \right]
\tag{93}
$$

$$
= \left(1 - \frac{3}{2}\frac{m}{M}\right) \sum_{g \in S_{a1} \cap S_{a2}} \widehat{t}_{z_1 g}\widehat{t}_{z_2 g} + \left(1 - \frac{3m(M-1)}{M(3m-2)}\right) \left(\sum_{g_1 \in S_{a1} \cap S_{a2}} \widehat{t}_{z_1 g_1}(\widehat{t}_{z_2} - \widehat{t}_{z_2 g_1}) \right.
$$
$$
\left. + \sum_{g_1 \in S_{a1} \backslash S_{a2}} \widehat{t}_{z_1 g_1}\widehat{t}_{z_2} \right) \tag{94}
$$
$$
= \left(1 - \frac{3}{2}\frac{m}{M}\right) \sum_{g \in S_{a1} \cap S_{a2}} \widehat{t}_{z_1 g}\widehat{t}_{z_2 g} + \left(1 - \frac{3m(M-1)}{M(3m-2)}\right)
$$
$$
\left(\widehat{t}_{z_2} \left(\sum_{g_1 \in S_{a1} \cap S_{a2}} \widehat{t}_{z_1 g_1} + \sum_{g_1 \in S_{a1} \backslash S_{a2}} \widehat{t}_{z_1 g_1} \right) - \sum_{g \in S_{a1} \cap S_{a2}} \widehat{t}_{z_1 g}\widehat{t}_{z_2 g} \right) \tag{95}
$$
$$
= \left(1 - \frac{3}{2}\frac{m}{M}\right) \sum_{g \in S_{a1} \cap S_{a2}} \widehat{t}_{z_1 g}\widehat{t}_{z_2 g}
$$
$$
+ \left(1 - \frac{3m(M-1)}{M(3m-2)}\right) \left(\widehat{t}_{z_1}\widehat{t}_{z_2} - \sum_{g \in S_{a1} \cap S_{a2}} \widehat{t}_{z_1 g}\widehat{t}_{z_2 g} \right)_{\square} \tag{96}
$$

B.2 Case of t (sample S_1) and $t+2$ (sample S_2)

We follow the same path, the "complete" sample $S_1 \cup S_2$ being now constituted of the entering, median and exiting subsamples of date t, the entering ones of date $t+1$ and the entering ones of date $t+2$. Now

$$
m^c = \frac{5}{3}m; \quad \pi_k^c = \frac{5}{3}\pi_k. \tag{97}
$$

As previously, we distinguish two cases.

– If k and l are in the same area, that area has to be in the entering sample at t and thus

$$
Pr(k \in S_1, l \in S_2) = \frac{1}{3}\frac{m}{M}. \tag{98}
$$

– If k et l are in two different areas, the probability can be written as previously, that is

$$
Pr(k \in S_1, l \in S_2) = \pi_{kl}^c P(k \notin \{4,5\}, l \notin \{1,2\}) \tag{99}
$$
$$
= \pi_{kl}^c (P(k \notin \{4,5\}) - P(k \notin \{4,5\}, l \in \{1,2\})) \tag{100}
$$
$$
= \pi_{kl}^c \left(\frac{3}{5} - \frac{2}{3}P(k,l \in \{1,2,3\}) \right). \tag{101}
$$

where the third equality stems for the symmetry of 1, 2 and 3. Now picking up k and l in $\{1,2,3\}$ amounts to picking them both according to a SRSWOR of $\frac{3}{5}m^c = m$ among

$m_c = 5/3m$. Therefore,

$$Pr(k \in S_1, l \in S_2) = \pi_{kl}^c \left(\frac{3}{5} - \frac{2}{3} \frac{m(m-1)}{\frac{5}{3}m(\frac{5}{3}m-1)} \right) \tag{102}$$

$$= \frac{\pi_{kl}^c}{5} \left(3 - \frac{2(m-1)}{\frac{5}{3}m - 1} \right) \tag{103}$$

$$= \frac{\pi_{kl}^c}{5} \frac{3m-1}{\frac{5}{3}m - 1}. \tag{104}$$

Besides, the "complete" sample is actually a SRSWOR of m^c areas among M, so that

$$\pi_{kl}^c = \frac{m^c(m^c - 1)}{M(M-1)} = \frac{\frac{5}{3}m(\frac{5}{3}m - 1)}{M(M-1)}, \tag{105}$$

and thus

$$Pr(k \in S_1, l \in S_2) = \frac{m(3m-1)}{3M(M-1)}. \tag{106}$$

Hence, as previously,

$$\widehat{\text{Cov}} = \left(1 - 3\frac{m}{M} \right) \sum_{g \in S_{a1} \cap S_{a2}} \widehat{t}_{z_1 g} \widehat{t}_{z_2 g} + \left(1 - \frac{3m(M-1)}{M(3m-1)} \right) \left(\widehat{t}_{z_1} \widehat{t}_{z_2} - \sum_{g \in S_{a1} \cap S_{a2}} \widehat{t}_{z_1 g} \widehat{t}_{z_2 g} \right) \quad \square \tag{107}$$