

Les statistiques locales : Essai d'estimation localisée directe et indirecte du taux d'emploi à l'aide des méthodes SAE

Local statistics: direct and indirect estimation of the employment rate using SAE methods

MAHALI Kamel

Université Sétif 1, Algérie

kamel.mahali@univ-setif.dz

Date de soumission : 15/08/2019

Date d'acceptance : 02/12/2019

Date de publication : 31/12/2019

Résumé :

Le taux d'emploi permet de donner une bonne indication du potentiel d'un pays en main-d'œuvre à un moment donné. L'Office National des Statistiques (ONS) publie des estimations de ce taux par sexe, par niveau d'instruction, etc. A remarquer que dans les publications de cette institution la dimension locale (wilayas, communes) est absente malgré son importance et sa portée, notre essai consiste à produire des estimations localisées du taux d'emploi à l'aide des méthodes d'estimation sur petits domaines. En termes de conclusion, il était possible de produire des estimations du taux d'emploi par wilayas ; statistiquement, la qualité des estimations produites est acceptable, voire bonne surtout pour l'estimateur composite et l'estimateur appliqué au modèle de Fay et Herriot.

- **Mots-clés :** l'estimation directe ; l'estimateur synthétique post-stratifié ; estimateur composite ; l'estimateur appliqué au modèle de Fay et Herriot ; taux d'emploi.
- **Codes de classification Jel :** C83; C49; C18; J21

Abstract:

The employment rate provides a good indication of a country's potential for labor at a given point in time. The Office for National Statistics (ONS) publishes estimates of this rate by sex, level of education, etc. Note that in the publications of this institution the local dimension (wilayas, communes) is absent despite its importance and scope, our try is to produce locale estimates of the employment rate using small area estimation methods. In terms of conclusion, it was possible to produce estimates of the employment rate by wilayas; statistically, the quality of the estimators is acceptable, even good especially for the composite estimator and the estimator applied to the Fay and Herriot model.

- **Keywords:** direct estimation; the post-stratified synthetic estimator; composite estimator; the estimator applied to the Fay and Herriot model; employment rate.
- **Jel Classification Codes :** C83; C49; C18; J21

I- Introduction :

Par définition, le taux d'emploi est le rapport entre la population occupée et la population en âge de travailler (15 ans et plus). Il constitue l'un des indicateurs les plus pertinents pour évaluer la situation du marché du travail, il permet de donner une bonne indication du potentiel d'un pays en main-d'œuvre à un moment donné et permet aussi de refléter la capacité de mobiliser ce potentiel.

Chaque année, l'Office National des Statistiques, institution centrale de la statistique en Algérie, publie, sur la base des données de l'enquête « Emploi », une estimation du taux d'emploi et publie même des estimations de ce taux par sexe, par niveau d'instruction, par secteur juridique et par secteur d'activité économique. A remarquer que dans ces publications la dimension locale est absente, et puisque l'emploi est avant tout une affaire locale, dans la mesure où l'information à l'échelon local permet d'assurer un contrôle au niveau local de la politique d'emploi, de contrôler de manière à limiter l'expansion du secteur informel, d'impliquer plus les collectivités locales dans la stratégie de l'emploi et de rentabiliser les actions au niveau local. De plus, ce type d'information permet d'étudier les disparités géographiques et de viser l'équilibre territorial en matière d'emploi. Notre essai s'inscrit dans le sens de combler le manque des publications en information par dimension locale (par wilaya), par la production d'estimations localisées du taux d'emploi ou d'estimations du taux d'emploi sur petits domaines où le critère choisi pour définir les domaines est un critère géographique et les domaines ainsi définies sont les wilayas. En d'autres termes, on essaye d'introduire la dimension locale dans certains phénomènes du marché du travail, à savoir l'emploi. Dans cet essai, on utilise les estimateurs direct, synthétique post-stratifié, composite et l'estimateur appliqué au modèle de Fay et Herriot pour produire les estimations localisées du taux d'emploi. Dans le présent article on ne va pas discuter la question segmentation et du cloisonnement relatif des marchés locaux du travail et la signification des données régionalisées sur l'activité.

L'article est présenté en quatre sections et plusieurs sous-sections. Aux sections 2 et 3, respectivement, nous examinons la littérature sur statistiques locales et sur les méthodes d'estimation sur petits domaines, puis nous décrivons la méthodologie utilisée en présentant les différents estimateurs (directs et indirects) et les données mobilisées dans cette application. Ensuite, à la section 3, nous essayons de faire des estimations localisées directe et indirecte du taux d'emploi. A la section 4, nous présentons nos conclusions.

II- Revue de littérature:

L'examen de la littérature sur les statistiques locales (portée et pertinence, sources potentielles et possibilité de produire des estimations localisées), nous montre que de nombreuses études ont été menées depuis les années soixante, les travaux les plus anciens remontent aux Panse et al. 1966 et Singh 1968 (Singh & Sisodia, 2014), ensuite une série de méthodes a été développée dans les années soixante-dix, par Gonzalez et Singh (1977), Purcell et Kish (1979) et Fay et Herriot (1979) ; et dans les années quatre-vingt, par Purcell et Kish (1980) et Drew et Al. (1982) et Rao (1986). Durant les derniers vingt-cinq ans, un progrès important est enregistré pour les techniques d'estimations pour petits domaines, parmi les travaux significatifs qui ont marqués cette époque, on cite Chaudhuri (1992) ; Ghosh et Rao (1994) ; Rao (1999, 2001b, 2014) ; Marker (1999) ; Pfeffermann (2002, 2013) ; Chambers et Tzavidis (2006, 2010) ; Jiang et Lahiri (2006) ; Opsomer et al. (2008) ; Salvati et Pratesi (2009) ; Salvati et al. (2010) ; Molina et Rao (2010) ; etc. Les développements récents dans cette littérature, portent essentiellement soit sur la prédiction empirique, sur la variété des méthodes d'estimation ou sur l'incorporation des structures spatiales : « *Empirical Best Prediction* » (Molina et Rao, 2010) ; « *Dealing with outliers* » (Sinha et Rao, 2009 ; Chambers et al., 2013; Giusti et al., 2013) ; « *Design consistent estimation* » (You et Rao, 2002) ; « *Estimation with M-quantile models* » (Chambers & Tzavidis, 2006; Fabrizi et al., 2013 ; Marchetti et al., 2012) ; « *Non-parametric models* » (Opsomer et al., 2008) ; « *Incorporating spatial structures* » (Salvati et Pratesi, 2009).

Pour notre étude, nous nous sommes appuyés sur deux études empiriques sur les méthodes d'estimation pour petits domaines :

En 2012, dans un essai empirique sur le chômage, Pascal Ardilly a cherché à estimer, au moyen des techniques d'estimation pour petits domaines (estimateur direct, le modèle de Fay et Herriot, le modèle de Poisson, le modèle logistique et l'estimateur EBLUP-B), l'effectif de chômeurs par zone d'emploi au sens du BIT. Selon l'auteur « le choix d'une méthode "optimale" nécessiterait également que l'on s'assure de la stabilité des conclusions dans le temps [...], la question de comparaison des méthodes est en soi assez troublante car sa légitimité ne nous paraît pas si claire : d'une part les différentes estimations ne s'appuient pas sur des aléas de même nature [...], d'autre part elles dépendent toutes d'hypothèses initiales fortes (les "modèles") qui ne sont pas vraiment remises en cause dès lors que les résultats apparaissent acceptables. Juger de la pertinence préalable de ces hypothèses devrait donc faire partie intégrante de l'exercice de comparaison des méthodes. » (Ardilly, 2012a)

Plus récemment M. Guadarrama, I. Molina & J.N.K. Rao (2016), ont publié une étude empirique comparative sur les méthodes d'estimation pour petits domaines. A travers une application sur les indicateurs de pauvreté et la cartographie de la pauvreté (*poverty mapping*), et à travers des études de simulation, les auteurs ont pu montrer que la fiabilité des estimations localisées et les bonnes propriétés dépendent de la validité des hypothèses sous-jacentes (Guadarrama, Molina, & Rao, 2016).

Pour le cas de l'Algérie, en réalité, nous n'avons trouvé que quelques études qui portent sur les sources potentielles d'information sur l'activité et sur les statistiques territoriales. A titre d'exemple, on cite deux travaux menés par Nacer-Eddine Hammouda, le premier article porte sur les sources potentielles de l'information statistique territoriale (Hammouda, 2009), et l'autre sur les statistiques de l'activité (Hammouda, 2005). Dans ces travaux, l'auteur essaye à chaque fois d'énumérer les sources potentielles en citant les points forts et faibles, ainsi que les perspectives et les propositions d'amélioration. Par ailleurs, les études portant sur les statistiques locales (portée, pertinence et possibilité de production des estimations localisées) sont absentes.

III-Données et Méthodologie :

III.1. Données disponibles :

Les données mobilisées dans cette application proviennent de l'enquête Emploi de 2010 et du Recensement Général de la Population et de l'Habitat de 2008 : l'enquête Emploi fournit plusieurs variables à un niveau individuel, parmi lesquelles figure la variable d'intérêt ; le recensement fournit une information auxiliaire à un niveau agrégé.

III.1.1 L'enquête Emploi : L'échantillon de l'enquête Emploi 2010 est tiré de l'échantillon maître élaboré du dernier recensement général de la population et de l'habitat (RGPH 2008), il est tiré des 4194 districts du sondage Wilaya (soit un 1/10^{ème} des districts recensés en 2008). L'ensemble des 4194 districts a constitué la base de sondage à partir de laquelle a été procédé un tirage proportionnel de 339 districts. Le tirage de l'échantillon s'est effectué à deux degrés. En premier temps, les districts ont été stratifiés en huit classes selon la strate géographique (urbaine/rurale) et la taille de l'agglomération. Ensuite, au 1^{er} degré un échantillon de districts a été tiré. Enfin, au deuxième degré, il a été procédé au tirage, à probabilités égales, du tiers des ménages au sein des districts échantillonnés. Au total 14 592 ménages ont été enquêtés (ONS, 2012). Pour notre application, les wilayas "de Tamanrasset, Illizi, Tindouf et Naâma ont été exclues de l'analyse, puisque la taille d'échantillon recoupant ces domaines (wilayas) est nulle (elles ne sont pas couvertes par l'enquête).

III.1.2 Le Recensement Général : Le recensement constitue la source exhaustive potentielle d'informations, son objectif est d'obtenir une description des caractéristiques sociodémographiques des habitants. Pour notre application, seulement la répartition de la population par groupe d'âge (6 groupes) est mobilisée pour faire l'estimation synthétique post-stratifiée (calage sur les groupes d'âge).

III. 2. Méthodologie :

D'un point de vue méthodologique, pour produire des estimations localisées, nous avons recouru aux techniques d'estimation sur petits domaines. Les estimateurs utilisés sont : l'estimateur direct ; l'estimateur synthétique post-stratifié ; l'estimateur composite et l'estimateur empirique appliqué au modèle de Fay et Herriot. Ici il faut signaler que le choix

d'une méthode d'EPD par rapport aux autres est fait sur la base de quelques contraintes : la disponibilité des données auxiliaires externes (disponible ou non, agrégées ou individuelles) et la faisabilité de la méthode (complexité et logiciel). En ce qui concerne les estimateurs utilisés, en détail, l'estimateur direct est de type « ratio » calé sur taille de la population appartenant au champ ; l'estimateur synthétique est de type post-stratifié ; pour l'estimateur composite, c'est celui dépendant de la taille de l'échantillon qui est utilisé ; enfin, l'estimateur empirique utilisé, est un estimateur sans biais optimum et linéaire (Empirical Best Linear Unbiased Prediction) appliqué au modèle de Fay et Herriot.

D'autre part, concernant le logiciel, les fonctions d'un package SAE ont été utilisées, ce package est conçu sous R par Molina et Marhuenda en 2015 (Molina & Marhuenda, 2015). Pour mesurer la qualité des différents estimateurs, nous avons utilisé le Coefficient de Variation pour mesurer la qualité de l'estimateur direct ; pour les autres estimateurs nous avons calculé l'Erreur Quadratique Moyenne.

IV-Estimation et analyse empirique:

Dans ce qui suit, les estimateurs (direct (Direct) ; synthétique post-stratifié (PsSynthétic) ; composite (ssd) et l'estimateur empirique appliqué au modèle de Fay et Herriot (EBLUPFH)) seront utilisés pour produire des estimations localisées du taux d'emploi.

IV.1. Estimation localisée utilisant l'estimateur direct :

Il s'agit d'un estimateur direct calé sur la taille de la population appartenant au champ (taille estimée). Pour faire l'estimation localisée utilisant cet estimateur, sous R on utilise la fonction suivante : direct (y = EEM2010\$occupes, dom = EEM2010\$wilayalab, sweight = EEM2010\$weight, domsize = Popn) ; Où y : la variable d'intérêt (ici est une variable binaire "1= Occupé ; 0= Non occupé" ; dom: les libellés des domaines ; sweight: les poids w_i associés aux individus ; domsize : $\hat{N}_d = \sum_{i \in S} w_i$, taille estimée de la population du domaine. L'exécution de la fonction a donné les résultats suivants : (tableau 1)

Tableau (1) : Estimation localisée utilisant l'estimateur direct

Domain	n_D	Direct	SD	CV	Domain	n_D	Direct	SD	CV
Adrar	1117	0.409	0.019	4.688	Guelma	1785	0.337	0.014	4.075
Chlef	2006	0.344	0.013	3.825	Constantine	2493	0.350	0.012	3.395
Laghouat	1073	0.380	0.019	4.970	Medea	1179	0.345	0.017	4.967
O.Elbouaghi	876	0.370	0.021	5.565	Mostaganem	965	0.414	0.021	5.040
Batna	1569	0.365	0.015	4.199	M'sila	1144	0.351	0.018	5.014
Bejaia	1511	0.341	0.015	4.418	Mascara	939	0.413	0.021	5.098
Biskra	964	0.353	0.019	5.425	Ouargla	741	0.342	0.021	6.288
Bechar	713	0.425	0.024	5.745	Oran	2606	0.396	0.012	3.128
Blida	1487	0.400	0.016	4.102	El bayadh	516	0.402	0.028	6.941
Bouira	1146	0.341	0.017	5.060	B.B.Arreridj	834	0.335	0.020	5.988
Tebessa	1822	0.339	0.014	4.023	Boumerdes	1871	0.428	0.015	3.537
Tlemcen	1779	0.421	0.015	3.664	Al Taref	343	0.403	0.034	8.482
Tiaret	1642	0.399	0.016	3.920	Tissemsilt	160	0.390	0.050	12.71
Tizi Ouzou	1757	0.360	0.014	3.990	El Oued	1507	0.455	0.018	3.929
Alger	4752	0.369	0.009	2.409	Khenchela	976	0.364	0.019	5.303
Djelfa	1446	0.394	0.017	4.214	Soukahrass	513	0.402	0.028	6.958
Jijel	806	0.348	0.021	5.979	Tipaza	906	0.379	0.020	5.403
Setif	2473	0.376	0.012	3.279	Mila	1864	0.344	0.014	3.953
Saida	324	0.375	0.034	9.094	Ain defla	2296	0.331	0.012	3.640
Skikda	2447	0.349	0.012	3.434	A. Temouchent	812	0.387	0.022	5.662
SidiBelabbes	1583	0.403	0.016	3.977	Ghardaia	678	0.448	0.026	5.811
Annaba	819	0.357	0.021	5.838	Relizane	1350	0.397	0.017	4.337

Source : estimations de l'auteur.

La pertinence et la qualité de cet estimateur, en comparant les valeurs du coefficient de variation (CV) avec 5 (Ardilly, 2006), on remarque que presque la moitié des domaines ont

des coefficients de variation supérieurs à 5, ce qui conduit à ne pas accepter les estimations directes pour toutes les wilayas.

IV.2. Estimation localisée utilisant l’estimateur indirect avec modélisation implicite :

Pour surmonter le problème qui se présente généralement dans le cas d’estimation sur domaine (la taille faible de l’échantillon dans certains domaines), on peut recourir aux estimateurs indirects qui ont pour principe de base l’emprunt de la force (information).

IV.2.1 Estimation localisée utilisant l’estimateur synthétique :

Si l’estimation synthétique, qui repose en général sur l’hypothèse qu’une relation observée dans un grand domaine, est la même dans le petit domaine, alors dans le cas de post stratification (estimateur synthétique post-stratifié), on suppose que les post-strates vont remplacer les domaines dans l’estimation (surmonter le problème de la taille faible de l’échantillon dans certains domaines). L’estimateur synthétique dans le cas de post-stratification est la somme du rapport entre les taux estimés dans les post-strates \hat{t}_{yh} multiplié par les tailles de la population de l’intersection (Domaine \times post-strate) N_{dh} et la taille estimée de la population dans les post-strates \hat{N}_h :

$$\hat{t}_{y,PsSynthétic} = \sum_{h=1}^H N_{dh} \frac{\hat{t}_{yh}}{\hat{N}_h} \tag{1}$$

Pour faire l’estimation localisée utilisant cet estimateur, les post-strates sont définies en fonction de la variable âge (six groupes d’âges).

Cet estimateur défini sur les post-strates donne des résultats dépendants de la dispersion dans les post-strates (tableau 2), c’est-à-dire, que pour les post-strates à faible S^2 , l’estimateur du taux d’emploi est plus précis en termes d’erreur totale (variance et biais).

Tableau (2) : Estimation localisée utilisant l’estimateur synthétique post-stratifié

Domain	PsSynthetic	Domain	PsSynthetic
Adrar	0.3902	Khenchela	0.3991
Ain defla	0.3960	Laghouat	0.4009
Ain temouchent	0.4152	M’sila	0.3939
Alger	0.4159	Mascara	0.4066
Annaba	0.4147	Medea	0.4002
Batna	0.3958	Mila	0.3953
Bechar	0.4072	Mostaganem	0.4038
Bejaia	0.3997	Oran	0.4146
Biskra	0.3963	Ouargla	0.3975
Blida	0.4108	Oum el bouaghi	0.3999
Bordj bou arreridj	0.3950	Relizane	0.3985
Bouira	0.4003	Saida	0.4055
Boumerdes	0.4068	Setif	0.3990
Chlef	0.3956	Sidi bel abbes	0.4102
Constantine	0.4047	Skikda	0.4004
Djelfa	0.3898	Souk ahras	0.4044
El bayadh	0.3999	Tebessa	0.3981
El oued	0.3863	Tiaret	0.3993
El tarf	0.4126	Tipaza	0.4074
Ghardaia	0.4006	Tissemsilt	0.3919
Guelma	0.4068	Tizi ouzou	0.3985
Jijel	0.3899	Tlemcen	0.4099

Source : estimations de l’auteur.

Sachant que dans la théorie d’estimation sur petits domaines, l’estimateur synthétique post-stratifié donne des résultats d’autant meilleurs que, dans chaque post-strate h, les valeurs de la variable d’intérêt seront peu dispersées autour de leur moyenne ; et la moyenne dans les

intersections (post-strates × domaines) soient proches des moyennes dans les post-strates ($\bar{Y}_{dh} \rightarrow \bar{Y}_h$). L'erreur totale (variance et biais) de l'estimateur post-stratifié prend la forme : (Ardilly, 2006)

$$EQM(\hat{Y}_{PsSynthétic}) = [\sum_{h=1}^H N_{dh}(\bar{Y}_h - \bar{Y}_{dh})]^2 + \sum_{h=1}^H N_{dh}^2 (E(\frac{1}{n_h}) - \frac{1}{N_h})S_h^2 \quad (2)$$

Le calcul de l'EQM permettra de mesurer la qualité des résultats, mais avant de calculer cet indicateur, on passe, à d'autres estimateurs (composite et empirique).

IV.2.2 Estimation localisée utilisant l'estimateur composite :

Cet estimateur est une combinaison linéaire des deux estimateurs précédents (direct et synthétique), l'idée derrière cet estimateur est que, puisque l'estimateur direct est faiblement biaisé et de forte variance ; et l'estimateur synthétique est biaisé mais à faible variance, alors, on combine les deux pour avoir un estimateur meilleur. $\hat{t}_{y,Composite} = \phi \hat{t}_{y,Direct} + (1 - \phi) \hat{t}_{y,PsSynthétic}$ (3)

Le problème pour cet estimateur est le choix du coefficient phi(ϕ). Ce coefficient est compris entre 0 et 1, plus il est proche de 1, plus l'estimateur direct va peser sur l'estimation composite, et plus ϕ est proche de 0, plus l'estimateur synthétique va peser sur l'estimation. Pour notre application, c'est l'estimateur composite dépendant de la taille de l'échantillon qui est utilisé. En détail pour les domaines de taille assez grande, on choisit le coefficient phi, de façon que l'estimateur composite coïncide avec l'estimateur direct ($\phi = 1$). L'estimation a donné les résultats suivants : (tableau 3)

Tableau (3) : Estimation localisée utilisant l'estimateur composite (ssd)

Domain	Composite (ssd)	Compsite Weight	Domain	Composite (ssd)	Composite Weight
Adrar	0.4093	1.0000000	Khenchela	0.3640	0.9999999
Ain defla	0.3308	1.0000000	Laghouat	0.3796	1.0000000
Ain Temouchent	0.3868	1.0000000	M'sila	0.3507	0.9999991
Alger	0.3692	0.9999999	Mascara	0.4130	0.9999992
Annaba	0.3567	1.0000000	Médéa	0.3448	0.9999990
Batna	0.3645	1.0000000	Mila	0.3444	0.9999998
Bechar	0.4249	0.9999985	Mostaganem	0.4142	0.9999994
Bejaia	0.3409	1.0000000	Oran	0.3962	1.0000000
Biskra	0.3531	1.0000000	Ouargla	0.3421	0.9999993
Blida	0.4003	1.0000000	Oum el Bouaghi	0.3698	1.0000000
BBA	0.3353	1.0000000	Rélizane	0.3975	1.0000000
Bouira	0.3411	0.9999999	Saida	0.3755	0.9999997
Boumerdes	0.4283	0.9999995	Sétif	0.3761	1.0000000
Chlef	0.3441	1.0000000	Sidi Belabbes	0.4026	0.9999998
Constantine	0.3498	1.0000000	Skikda	0.3493	0.9999996
Djelfa	0.3938	1.0000000	Souk Ahras	0.4023	0.9999994
El Bayadh	0.4021	1.0000000	Tébessa	0.3393	1.0000000
El Oued	0.4546	0.9999995	Tiaret	0.3986	1.0000000
El Taref	0.4026	1.0000000	Tipaza	0.3786	0.9999993
Ghardaia	0.4478	1.0000000	Tissemsilt	0.3900	0.9999941
Guelma	0.3373	1.0000000	Tizi Ouzou	0.3605	1.0000000
Jijel	0.3481	1.0000000	Tlemcen	0.4206	1.0000000

Source : estimations de l'auteur.

L'exécution de cette fonction donne des résultats intéressants de point de vue statistique (combine entre l'estimation directe et indirecte), le plus remarquable dans ces

résultats est que plus de la moitié des domaines (26 domaines) sont de taille assez grande ($\hat{N}_{DIR} > \delta N_d$; δ constant positif), le coefficient phi choisit est égale à 1 ($\phi = 1$), et l'estimateur composite coïncide avec l'estimateur direct. Pour les 18 restants le coefficient est calculé de la manière suivante: $\phi = \hat{N}_d / \delta N_d$
 (4)

Pour décider la qualité des résultats des deux estimateurs il suffit de calculer un des indicateurs de qualité (EQM par exemple), mais avant de calculer cet indicateur, et pour des raisons de comparaison et de choix d'un estimateur meilleur, on passe, tout d'abord, à un autre estimateur (estimateur eblup-FH).

IV.3. Estimation localisée utilisant l'estimateur indirect avec modélisation explicite :

Par rapport à l'estimation indirecte avec modélisation implicite, les modèles explicites comportent plusieurs différences, dont la plus importante est la présence d'une composante aléatoire de type « domain effects » dans le modèle. Cette différence est un avantage en faveur de la modélisation explicite (Ardilly, 2012b).

Les modèles explicites d'estimation localisée se répartissent en deux, les modèles conçus au niveau des domaines (Area Level Model), et ceux conçus au niveau des individus (Unit Level Model). Ici, on rappelle que le choix d'un modèle pour faire l'estimation localisée est fait sur la base de quelques contraintes : la disponibilité des données auxiliaires externes (oui ou non, agrégées ou individuelles), la faisabilité de la méthode (pas trop compliqué et outil logiciel existant). Dans cette application, nous avons mobilisés l'estimation EBLUP appliqué au modèle de Fay et Herriot.

IV.3.1 Estimation localisée utilisant l'estimateur EBLUP-FH :

Avant de construire la base de données définitive au niveau des domaines (Wilayas), qui serve à l'estimation EBLUP appliqué au modèle de Fay et Herriot, il faut déterminer les variables auxiliaires qui peuvent expliquer la variable d'intérêt (ici le taux d'emploi) au niveau des domaines. Dans la recherche des variables explicatives pertinentes au niveau des domaines, nous avons calculé un coefficient de corrélation entre d'une part la variable d'intérêt (estimateur direct) et d'autre part la variable auxiliaire candidate. Le coefficient de corrélation calculé dépend de la nature de la variable explicative : si elle quantitative, le coefficient calculé est le coefficient de Pearson, sinon le coefficient Eta est utilisé. Dans ce dernier cas (variable explicative qualitative), une ANOVA est utilisée pour confirmer le choix. A la fin de cette étape préliminaire à la construction de la base de données définitive, une seule variable est retenue (la région d'appartenance du domaine). Ci-après les résultats de la sélection des variables explicatives.

Tableau (4) : Les résultats de la sélection (la fin de l'étape de sélection)

MESURES DIRECTIONNELLES				Valeur	
Données nominales /intervalle	Eta	Dépendant de taux d'emploi		0.634	
ANOVA					
	Somme des carrés	ddl	Carré moyen	F	Sig.
Variable d'intérêt (taux d'emploi)	181.263	3	60.42	8.9	0.000
	269.553	40	6.739	7	
	450.816	43			

Source : estimations de l'auteur

La base de données définitive construite au niveau des domaines (Wilayas) prend la forme (domaines/variables), en ligne elle contient 44 observations et en colonne les variables suivantes : les étiquêtes des domaines ; les codes des domaines ; la taille des domaines ; l'estimateur direct du taux d'emploi ; l'écart type de l'estimateur direct ; le coefficient de variation ; la covariable (région).

Une fois la base est construite, on passe à l'estimation empirique appliquée au modèle de Fay et Herriot. On rappelle que pour faire cette estimation on doit estimer les paramètres des matrices variance-covariance. Trois méthodes d'estimation peuvent être utilisées à savoir, la méthode des moments, la méthode du maximum de vraisemblance et la méthode du maximum de vraisemblance restreint. Dans le cadre de nos applications on mobilise les trois, ensuite, à chaque fois une sélection de la meilleure prédiction est faite. Cette sélection est basée sur le critère log-likelihood. Sous R, on utilise la fonction suivante : `eblupFH (yi ~ as.factor(region), SD^2, method="FH")` ; Où : y_i : la variable d'intérêt (l'estimateur direct du taux d'emploi) ; region la co-variable ; SD^2 : l'écart type de l'estimateur direct ; method : la méthode utilisée (la méthode des moments, appelée aussi pour ce type d'estimation la méthode de Fay et Herriot «FH», la méthode du maximum de vraisemblance «ML» ou la méthode du maximum de vraisemblance restreint «REML»).

Au Vu des résultats de qualité d'ajustement (tableau 5), et sachant que le modèle est d'autant préférable que le critère de sélection est petit, le critère de sélection étant minimum pour le modèle obtenu par la méthode du maximum de vraisemblance restreint.

Tableau (5) : Mesure de qualité d'ajustement (log-likelihood)

\$est\$fit\$goodness	method	"FH"	"ML"	"REML"
	Log-likelihood	100,9228	101,0517	100,9153

Source : estimations de l'auteur

Après le choix de la méthode qui va servir à estimer les paramètres des matrices variance-covariance. L'estimation a donné les résultats suivants : (tableau 6)

Tableau (6) : Estimation localisée utilisant l'estimateur EBLUP -FH

wilayalab	eblup-FH	wilayalab	eblup-FH
Adrar	0.4063458	Guelma	0.3592925
Chlef	0.3421611	Constantine	0.3937118
Laghouat	0.3973801	Medea	0.3531433
Oum elbouaghi	0.3684201	Mostaganem	0.4075923
Batna	0.3560118	M'sila	0.3551158
Bejaia	0.3605319	Mascara	0.3486341
Biskra	0.4109477	Ouargla	0.4081030
Bechar	0.3515542	Oran	0.3988262
Blida	0.3544176	El bayadh	0.3806264
Bouira	0.3838029	B.b.arreridj	0.3614477
Tebessa	0.3469323	Boumerdes	0.4007312
Tlemcen	0.3533392	Al tarf	0.3981497
Tiaret	0.4010583	Tissemsilt	0.3691312
Tizi.ouzou	0.3519623	El oued	0.4032248
Alger	0.3516438	Khenchela	0.3513096
Djelfa	0.3802704	Soukahras	0.3684563
Jijel	0.4034522	Tipaza	0.3455118
Setif	0.4292610	Mila	0.4009832
Saida	0.3651279	Ain defla	0.3710143
Skikda	0.4184493	Ain temouchent	0.4024417
Sidi bel abbes	0.3443431	Ghardaia	0.3625551
Annaba	0.3525101	Relizane	0.4132387

Source : estimations de l'auteur

Il faut signaler ici que les taux estimés à partir du modèle Fay Herriot doivent être considérés comme des prédictions provenant d'un modèle. D'autre part, le modèle définitif donne aussi les résultats d'ajustement suivants : (tableau 7)

Tableau (7) : Résultats d’ajustement "modèle de Fay et Herriot, méthode REML"

\$EST\$FIT\$METHOD [1] “REML”				
\$EST\$FIT\$CONVERGENCE [1] TRUE				
\$EST\$FIT\$ITERATIONS [1] 3				
\$est\$fit\$estcoef	beta	Std.error	tvalue	pvalue
X(intercept)	0.3656	0.0069	53.0433	0.0000
Xas.factor(region) 2	- 0.0100	0.0092	- 1.0884	0.2764
Xas.factor(region) 3	0.0384	0.0102	3.7804	0.0002
Xas.factor(region) 4	0.0399	0.0152	2.6347	0.0084
\$est\$fit\$refvar [1] 0.0002987057				
\$est\$fit\$goodness				
loglike 100.9153	AIC - 191.8307	BIC - 182.9097	KIC - 186.8307	

Source: estimations de l’auteur

On constate, tout d’abord, que la convergence a eu lieu très rapidement (après trois itérations seulement) et la variance estimée des effets aléatoire dans ce modèle vaut 2.98E-04. On constate aussi que, à part le deuxième facteur (la deuxième modalité de la covariable), la constante et tous les autres facteurs sont significatives.

IV.4. Qualité et pertinence des estimateurs (EQM) :

Concernant la qualité des estimateurs (synthétique, composite et EBLUP-FH), l’estimation de l’erreur quadratique moyenne (EQM) nous permet de mesurer la qualité et de faire la comparaison entre les estimateurs. Cet indice de qualité mixte deux autres concepts de qualité « biais et variance » (l’EQM est égale à la variance plus le carré du biais).

Pour l’estimateur synthétique et composite, les calculs sont basés sur la proposition de Ghosh et Rao (Ghosh & Rao, 1994). Il faut signaler ici que, pour quelques domaines (Wilayas), nous avons obtenu des valeurs négatives de l’EQM de l’estimateur synthétique, ces valeurs négatives ont été remplacées par la moyenne des valeurs positives.

Les résultats de la pertinence et de la qualité des différents estimateurs, montrent que l’estimateur synthétique est meilleur que l’estimateur direct (les valeurs de son erreur quadratique varient entre 0.0000944 et 0.00464), toutefois, la qualité s’est améliorée encore mieux pour l’estimateur composite (on remarque que, à part les wilayas « El Tarf, Saida et Tissemsilt », les valeurs de l’erreur quadratique moyenne oscillent entre 0.0000791 et 0.000141). On remarque que les estimations localisées gagnent plus de qualité dans le cas de l’estimateur empirique appliquée au modèle de Fay et Herriot (les valeurs de Erreur Quadratique Moyenne se sont améliorées, et elles ne dépassent pas 0.000320). (Tableau 8)

Tableau (8) : L’Erreur Quadratique Moyenne de l’estimateur synthétique, composite et EBLUP-FH

EQM Domain	<i>Psynthetic</i>	<i>Comp_ssd</i>	<i>eblupFH</i>	EQM Domain	<i>Psynthetic</i>	<i>Comp_ssd</i>	<i>eblupFH</i>
Adrar	1.98E-3	3.68E-4	1.96E-4	Khenchela	8.58E-4	3.73E-4	1.92E-4
Ain defla	4.10E-3	1.45E-4	1.10E-4	Laghouat	9.44E-5	3.56E-4	2.31E-4
A.Temochent	3.24E-4	4.80E-4	2.21E-4	M’sila	1.56E-3	3.09E-4	1.75E-4
Alger	2.10E-3	7.91E-5	6.83E-5	Mascara	1.98E-3	4.43E-4	2.14E-4

Annaba	2.93E-3	4.34E-4	2.05E-4	Medea	2.77E-3	2.93E-4	1.73E-4
Batna	7.51E-4	2.34E-4	1.50E-4	Mila	2.41E-3	2.85E-4	1.30E-4
Bechar	1.98E-3	5.96E-4	2.40E-4	Mostaganem	1.98E-3	4.36E-4	2.12E-4
Bejaia	3.23E-3	2.27E-4	1.49E-4	Oran	1.87E-4	1.54E-4	1.16E-4
Biskra	1.50E-3	3.67E-4	1.91E-4	Ouargla	2.61E-3	4.63E-4	2.64E-4
Blida	1.98E-3	2.70E-4	1.65E-4	O.elBouaghi	4.79E-4	4.24E-4	2.03E-4
BBA	3.16E-3	4.03E-4	1.99E-4	Relizane	1.98E-3	2.97E-4	1.76E-4
Bouira	3.20E-3	2.98E-4	1.74E-4	Saida	1.98E-3	1.17E-3	2.87E-4
Boumerdes	2.34E-4	2.30E-4	1.50E-4	Setif	3.73E-4	1.52E-4	1.13E-4
Chelf	2.48E-3	1.73E-4	1.25E-4	SidiBelAbbes	1.98E-3	2.56E-4	1.62E-4
Constantine	2.88E-3	1.41E-4	1.07E-4	Skikda	2.47E-3	1.44E-4	1.09E-4
Djelfa	1.98E-3	2.75E-4	1.67E-4	Souk Ahras	1.98E-3	7.82E-4	2.51E-4
El Bayadh	1.98E-3	7.79E-4	2.61E-4	Tebessa	3.27E-3	1.86E-4	1.30E-4
El Oued	4.34E-3	3.19E-4	2.17E-4	Tiaret	1.98E-3	2.44E-4	1.58E-4
El Taref	1.98E-3	1.17E-3	2.75E-4	Tipaza	4.08E-4	4.19E-4	2.06E-4
Ghardaia	1.55E-3	6.77E-4	3.11E-4	Tissemst	1.98E-3	2.46E-3	3.20E-4
Guelma	4.64E-3	1.89E-4	1.31E-4	Tizi Ouzou	1.24E-3	2.07E-4	1.41E-4
Jijel	1.31E-3	4.33E-4	2.05E-4	Tlemcen	1.98E-3	2.37E-4	1.55E-4

Source : estimations de l'auteur

V- Conclusion:

En conclusion, pour cette application, il était possible de produire des estimations localisées (estimation du taux d'emploi par wilayas), par l'utilisation, simplement, de certaines méthodes statistiques d'estimation (les méthodes d'Estimation sur Petits Domaines). Ces méthodes donnent habituellement de bons résultats (et ceci est le cas dans cette application). Pour certaines de ces méthodes, elles augmentent la taille d'échantillon globale et dans les domaines (wilayas), pour d'autres elles empruntent de l'information disponible en dehors des domaines considérés (*borrow strength*).

Statistiquement, la qualité des estimations produites est acceptable, voire bonne surtout pour l'estimateur composite et l'estimateur empirique appliqué au modèle de Fay et Herriot. De plus le coût derrière la production de ces estimations localisées, est presque nul, seulement nous avons exploité la base de données individuelles de l'Enquête Emploi 2010, avec quelques données auxiliaires externes issues du Recensement Général de la Population et de l'Habitat (RGPH 2008) et à l'aide de quelques méthodes d'estimation sur petits domaines.

-Références:

1. Ardilly, P. (2006). Panorama des principales méthodes d'estimation sur les petits domaines. *INSEE, Méthodologie Statistique*, (M0602).
2. Ardilly, P. (2012a). Estimation localisée du chômage: une application des techniques d'estimation sur petits domaines. *Actes Des 11e Journées de Méthodologie Statistique*. Paris: INSEE.
3. Ardilly, P. (2012b). Estimation sur petits domaines à l'Insee: premier bilan et pistes de progrès. *Actes Des 7e Colloque Francophone Sur Les Sondages*. Rennes: ENSAI.
4. Ghosh, M., & Rao, J. N. K. (1994). Small area estimation: an appraisal. *Statistical Science*, 9(1), 55–76.
5. Guadarrama, M., Molina, I., & Rao, J. N. K. (2016). A comparison of small area estimation methods for poverty mapping. *Statistics in Transition*, 17(1). <https://doi.org/10.21307/stattrans-2016-005>
6. Hammouda, N. (2005). Le système d'information sur le marché du travail: Bilan et perspectives. *Actes Du Séminaire National "Le Système National d'Information Economique : Etat et Perspectives*. Alger: CERIST.

7. Hammouda, N. (2009). Quelles statistiques pour quels territoires ? *Actes Du Séminaire de Lancement Sur "L'ingénierie Territoriale Au Service de l'attractivité, de La Compétitivité et Du Développement Durable Des Territoires*. Alger: Ministère de l'Aménagement du Territoire et de l'Environnement.
8. Molina, I., & Marhuenda, Y. (2015). sae: An R package for small area estimation. *The R Journal*, 7(1), 81–98.
9. ONS. (2012). Enquête emploi auprès des ménages 2010. *Collections Statistiques*, 170, 97.
10. Singh, B., & Sisodia, B. V. S. (2014). Application of Ratio-type Estimators for Small Area in Repeated Surveys. *Pakistan Journal of Statistics and Operation Research*, 10(1), 73–85.