

Arbia, Giuseppa; Espa, Giuseppe; Giuliani, Diego

Article

Measurement errors arising when using distances in microeconomic modelling and the individuals' position is geo-masked for confidentiality

Econometrics

Provided in Cooperation with:

MDPI – Multidisciplinary Digital Publishing Institute, Basel

Suggested Citation: Arbia, Giuseppa; Espa, Giuseppe; Giuliani, Diego (2015) : Measurement errors arising when using distances in microeconomic modelling and the individuals' position is geo-masked for confidentiality, *Econometrics*, ISSN 2225-1146, MDPI, Basel, Vol. 3, Iss. 4, pp. 709-718, <https://doi.org/10.3390/econometrics3040709>

This Version is available at:

<https://hdl.handle.net/10419/171853>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by/4.0/>

Article

Measurement Errors Arising When Using Distances in Microeconomic Modelling and the Individuals' Position Is Geo-Masked for Confidentiality

Giuseppe Arbia ^{1,*}, Giuseppe Espa ² and Diego Giuliani ²

¹ Department of Statistical Science, Catholic University of the Sacred Heart, Rome 00168, Italy

² Department of Economics and Management, University of Trento, Trento 38122, Italy;
E-Mails: giuseppe.espa@unitn.it (G.E.); diego.giuliani@unitn.it (D.G.)

* Author to whom correspondence should be addressed; E-Mail: Giuseppe.arbia@rm.unicatt.it;
Tel.: +39-06-3015-6109.

Academic Editor: Kerry Patterson

Received: 25 August 2015 / Accepted: 20 October 2015 / Published: 29 October 2015

Abstract: In many microeconomic models we use distances. For instance, in modelling the individual behavior in labor economics or in health studies, the distance from a relevant point of interest (such as a hospital or a workplace) is often used as a predictor in a regression framework. However, in order to preserve confidentiality, spatial micro-data are often geo-masked, thus reducing their quality and dramatically distorting the inferential conclusions. In particular in this case, a measurement error is introduced in the independent variable which negatively affects the properties of the estimators. This paper studies these negative effects, discusses their consequences, and suggests possible interpretations and directions to data producers, end users, and practitioners.

Keywords: spatial econometrics; spatial microeconometrics; consistency of estimates; geo-masking; confidentiality; distance evaluation

JEL classifications: C01; C13; C31

1. Introduction

In many microeconomic studies we often use distances. For instance in labor, in schooling, or in health studies, the distance between each observed individual and a conspicuous point (e.g., a hospital or the workplace or a school) is often used as a predictor in a regression model. Furthermore, if our aim is to take into account the interacting effects between individuals using econometrics methods [1], we often build up weight matrices based on some inverse-distance function, using the two by two distances between individuals as the basis for the calculation. When analyzing granular micro spatial data (like e.g., household surveys), in order to preserve confidentiality the coordinates of each individual are often displaced according to some random method. For instance some recent DHS surveys coordinates are first collected in the field using GPS receivers with an accuracy of less than 15 m, but then, in order to ensure that respondent confidentiality is maintained, the observed points are geo-masked along a random distance and a random angle (see, e.g., [2]) using different maximum error distances in urban and rural areas (see [3–5] for details). A second method is the Gaussian geo-masking which consists in randomly reallocating the point in the neighborhood of the true location, with a probability described by a Gaussian bivariate density function centered on the true point. This paper's contribution consists in analyzing the problems induced by geo-masking in econometric models as a problem of measurement error. In Section 2 we will consider the measurement error induced by geo-masking in case the distance is used as a predictor in a regression model. Section 3 concludes with some practical comments and directions to assist the statistics producers to calibrate the geo-masking procedures before delivering the data to the public.

2. Effects of Geo-Masking When We Use the Distance as a Predictor in a Regression

2.1. Generalities

Let us consider the case of the estimation of a linear regression when the true individuals' coordinates are not disclosed for confidentiality and a distance from a relevant point is used as a predictor. For instance, in health economics it is common practice to postulate a relationship between a health outcome for each individual (such as the effect of a health policy), say y , and the individual's distance (say d) from a clinic or a hospital. To illustrate the essence of the problem we will restrict ourselves to the case, admittedly unrealistic, of a simple regression without any further predictors. Furthermore, to simplify the algebra without compromising the generality of the results, we will postulate a relationship between the health outcome and the squared distance from a relevant point. When data are geo-masked, the distance between each individual and a conspicuous point will be upward biased (as shown, e.g., by Arbia *et al.*, 2015 [5] and Elkies, *et al.*, 2015 [6]). This paper shows that, when the masking procedure is disclosed, this information can be taken into account for the benefit of the analysis.

To start with, let us recall that the classical error measurement theory (e.g., [7]) defines the true model as:

$$y_{ij} = \alpha + \beta d_{ij}^2 + v_{ij} \quad (1)$$

for each individual observed in the point of coordinates (i,j) , with d_{ij} the distance between point (i,j) and the point of interest and $v_{ij} \approx n.i.d.(0, \sigma_v^2)$. The distance is observed with an error due to geo-masking and the measurement error is defined as:

$$u_{ij} = \bar{d}_{ij}^2 - d_{ij}^2$$

with \bar{d}_{ij}^2 the squared distance observed after geo-masking. Following the classical theory, as it is known, u_{ij} should be assumed to be such that: $E(u_{ij}) = 0$, $Var(u_{ij}) = \sigma_u^2 = \text{constant}$, and u_{ij} independent of v_{ij} and of d_{ij}^2 . Normality of u is also often assumed. In these conditions, having called $\hat{\beta}$ the OLS estimator of β , such estimator will be still unbiased, but less efficient, since its variance can now be expressed as:

$$Var(\hat{\beta}) = \frac{\beta^2 \sigma_u^2 + \sigma_v^2}{n \sigma_{d^2}^2} > \frac{\sigma_v^2}{n \sigma_{d^2}^2} \tag{2}$$

The estimator will also be inconsistent with a downward asymptotic bias towards zero (called *attenuation*) quantified by the expression [7]:

$$p \lim \hat{\beta} = \beta \left(\frac{\sigma_{d^2}^2}{\sigma_u^2 + \sigma_{d^2}^2} \right) < \beta \tag{3}$$

with $\sigma_{d^2}^2$ the variance of the squared uncontaminated distances.

However, in the case of a measurement error induced by geo-masking, the results are quite different from the classical, as we will illustrate in the next sections.

2.2. Gaussian Geo-Masking within a Circle

Let us start considering the effect of geo-masking when using a distance as a regressor in an econometric model, in the case of Gaussian geo-masking, that is when the true individuals' location is perturbed with a bivariate Gaussian distribution centered on the true point. More formally let us consider the point of coordinates (i,j) and let us geo-mask this point by disclosing, instead, the coordinates $\bar{i} = i + \varepsilon_i$ and $\bar{j} = j + \varepsilon_j$ with $\varepsilon_i \approx n.i.d.(0, \sigma_\varepsilon^2)$. Let us further consider, for each individual point, its distance from a conspicuous point that, without loss of generalities, we can allocate at the origin of the Cartesian system. In this case the true squared distance of the point of coordinates (i,j) from the conspicuous point is given by:

$$d_{ij}^2 = i^2 + j^2$$

while, after geo-masking we observe instead

$$\bar{d}_{ij}^2 = \bar{i}^2 + \bar{j}^2 = d_{ij}^2 + (\varepsilon_i^2 + \varepsilon_j^2) + 2(i\varepsilon_i + j\varepsilon_j) = d_{ij}^2 + u_{ij} \tag{4}$$

because of our definitions.

So the term u defines the measurement error on the independent variable of the model as in the classical theory. However, in contrast with the classical theory, in this case we have:

$$E(u_{ij}) = 2\sigma_\epsilon^2 \neq 0 \quad \forall i, j \tag{5}$$

and

$$Var(u_{ij}) = 4\sigma_\epsilon^2 (\sigma_\epsilon^2 + d_{ij}^2) \tag{6}$$

Thus the measurement error has non-zero mean and non-constant variances. (See Appendix A for the proof). The non-zero mean does not affect the point estimate of the parameter β , but only the constant term. Equation (6) shows that the procedure of geo-masking also induces heteroscedasticity.

Furthermore from Equation (4) we have:

$$u_{ij} = \bar{d}_{ij}^2 - d_{ij}^2 = (i + \epsilon_i)^2 + (j + \epsilon_j)^2 - d_{ij}^2 \tag{7}$$

As a consequence (since d_{ij}^2 , i and j are constant terms and $\epsilon_i \approx n.i.d.(0, \sigma_\epsilon^2)$) $u \approx \chi_2^2(\lambda)$ is a non-central Chi-squared with 2 degrees of freedom and non-centrality parameter $\lambda = \frac{d_{ij}^2}{\sigma_\epsilon^2}$. The proof

is left to Appendix B.

Following the classical theory, the OLS estimator will be less efficient and inconsistent recalling Equations (2) and (3). In particular, in the case of Gaussian geo-masking, the variance of the OLS estimator will be:

$$Var(\hat{\beta})_{ij} = \frac{4\beta^2 \sigma_\epsilon^2 (\sigma_\epsilon^2 + d_{ij}^2) + \sigma_v^2}{n\sigma_{d^2}^2} > \frac{\sigma_v^2}{n\sigma_{d^2}^2} \tag{8}$$

Thus, the larger are σ_ϵ^2 and the larger the distance from the conspicuous point is, the lower the precision of the estimate will be. The precision also depends on the square of the true value of β . Furthermore, using Equation (6) to evaluate the attenuation, we have:

$$p \lim \hat{\beta} = \beta \left(\frac{\sigma_{d^2}^2}{\sigma_u^2 + \sigma_{d^2}^2} \right) = \beta \left(\frac{\sigma_{d^2}^2}{4\sigma_\epsilon^2 (\sigma_\epsilon^2 + d_{ij}^2) + \sigma_{d^2}^2} \right) \tag{9}$$

which shows that the attenuation effect on the OLS estimator is greater in the presence of a larger geo-masking variance of higher distances.

In practical cases, to communicate with practitioners, it is useful to introduce the Gaussian geo-masking mechanism with reference to a maximum displacement distance which is easier to interpret than a variance for non-specialists. Since in a Gaussian distribution $P(-3\sigma \leq x \leq 3\sigma) = 0.9973$, with a probability close to 1 we can assume that the maximum displacement distance is 3σ . If we call θ^* such maximum distance, we have that $3\sigma = \theta^*$. So the expected measurement error is $E(u) = 2\sigma_\epsilon^2 = \frac{2}{9}\theta^{*2}$ and the bias can be seen as a fraction of the maximum squared displacement distance. Furthermore its variance can be expressed as $Var(u) = \frac{4}{81}\theta^{*4} + \frac{8}{9}\theta^{*2}d_{ij}^2$ which shows that uncertainty increases with the maximum displacement distance and with the absolute position of the individual with respect to the conspicuous point. By using this alternative expression, the variance of the OLS estimator can be expressed as:

$$Var(\hat{\beta})_{ij} = \frac{\beta^2 \left(\frac{4}{81} \theta^{*4} + \frac{4}{9} \theta^{*2} d_{ij}^2 \right) + \sigma_v^2}{n\sigma_{d^2}^2} \tag{10}$$

and the attenuation effect as:

$$p \lim \hat{\beta} = \beta \left(\frac{\sigma_{d^2}^2}{\sigma_u^2 + \sigma_{d^2}^2} \right) = \beta \left(\frac{\sigma_{d^2}^2}{\frac{4}{81} \theta^{*4} + \frac{4}{9} \theta^{*2} d_{ij}^2 + \sigma_{d^2}^2} \right) \tag{11}$$

which shows more intuitively the negative effects of geo-masking on the OLS estimates. The greater the maximum displacement distance is, the larger both the loss in efficiency and the attenuation effect will be.

2.3. Uniform Geo-Masking within a Circle

Let us now turn to analyze the effects of a uniform geo-masking (such as the one employed, e.g., by DHS, 2013 [4]), that is a mechanism which transforms the coordinates displacing them along a random angle (say δ) and a random distance (say θ) both obeying a uniform probability law. The mechanism can be formally expressed through the following hypotheses:

HP1: $\theta \stackrel{iid}{\sim} U(0, \theta^*)$ and $\delta \stackrel{iid}{\sim} U(0, 360^\circ)$, with θ^* the maximum distance error, and

HP2: θ and δ are independent.

Assuming again, without loss of generality, that the conspicuous point is located in the origin, the true squared distance between point of coordinates (i, j) and the conspicuous point before geo-masking is measured by $d_{ij}^2 = (i^2 + j^2)$, while, after geo-masking, it can be expressed, using the polar coordinates, as:

$$\bar{d}_{ij}^2 = (i + \theta \cos \delta)^2 + (j + \theta \sin \delta)^2 \tag{12}$$

Expanding Equation (12) we obtain:

$$\bar{d}_{ij}^2 = (i^2 + 2i\theta \cos \delta + \theta^2 \cos^2 \delta + j^2 + 2j\theta \sin \delta + \theta^2 \sin^2 \delta)$$

so that we can now express the measurement error as:

$$u_{ij} = \bar{d}_{ij}^2 - d_{ij}^2 = 2\theta(i \cos \delta + j \sin \delta) + \theta^2 (\cos^2 \delta + \sin^2 \delta) \tag{13}$$

Similarly to the Gaussian case we have a non-zero mean and a non-constant variance, given by:

$$E(u_{ij}) = \frac{\theta^*}{3} \neq 0 \tag{14}$$

and

$$Var(u_{ij}) = \frac{17}{180} \theta^{*4} + \frac{2}{3} \theta^{*2} d_{ij}^2 \tag{15}$$

Again the proofs are left to the appendices, specifically to Appendix C.

So, consistently with the results obtained with a Gaussian geo-masking and according with the intuition, the measurement error increases its variance as the maximum displacement distance θ^* increases and as we move away from the conspicuous point.

If we use this result again to provide an explicit expression to the estimation variance and to the attenuation effect, we have, respectively:

$$Var(\hat{\beta})_{ij} = \frac{\beta^2 \left(\frac{17}{180} \theta^{*4} + \frac{2}{3} \theta^{*2} d_{ij}^2 \right) + \sigma_v^2}{n \sigma_{d^2}^2} \tag{16}$$

and

$$p \lim \hat{\beta} = \beta \left(\frac{\sigma_{d^2}^2}{\sigma_u^2 + \sigma_{d^2}^2} \right) = \beta \left(\frac{\sigma_{d^2}^2}{\frac{17}{180} \theta^{*4} + \frac{2}{3} \theta^{*2} d_{ij}^2 + \sigma_{d^2}^2} \right) \tag{17}$$

which lead to very similar conclusions to those found for the Gaussian geo-masking (see Equations (10) and (11)). The greater the maximum displacement distance is, the lower the precision and the larger the attenuation effects are.

3. Discussion and Conclusions

In this paper we examined the measurement error on distances introduced by the procedure of geo-masking the individuals' true location to protect their confidentiality, when such distances are used as predictors in a linear regression. The formal expressions that we derived for the loss in efficiency and for the attenuation in the case of Gaussian and uniform geo-masking, are very important under the practical point of view. In fact, the true location of each of the observed individuals is known to the producer of the official statistics before the displacement is introduced and so it is the variance term $\sigma_{d^2}^2$. So, in principle, the data producers could calculate the appropriate expression before geo-masking the data when choosing the maximum location error (θ^*) so as to limit the negative consequences on the subsequent analysis. Furthermore, the data producer could disclose to the end users and to the practitioners the level of attenuation which is expected given the chosen geo-masking procedure. In fact, for any given dataset, Expressions (11) and (17) are just functions of the maximum displacement distance θ^* .

To illustrate this point, suppose, for instance, that $n = 100$ individuals have been observed in a unitary squared study area as it is shown in Figure 1.

Taking these points as given we have that $d_{ij}^2 = 0.520151$ (considering for operational reasons the mean of all squared distances from the origin) and $\sigma_{d^2}^2 = 0.1592879$. Figure 2 reports the behavior of the attenuation effect for Gaussian and uniform geo-masking for the values of the maximum displacement distance θ^* ranging between 0 and 1.44 (1.44 being the theoretical maximum possible distance in a unitary square).

Two features emerge from the inspection of the graph. First, the attenuation increases dramatically already at small levels of θ^* . Secondly, the Gaussian geo-masking, other things being constant,

produces more severe consequences on the estimation of β than the uniform geo-masking. This graph could be used by the data producers to calibrate the optimal value of θ^* and to communicate to the practitioners the resulting level of attenuation they should expect from a regression analysis.

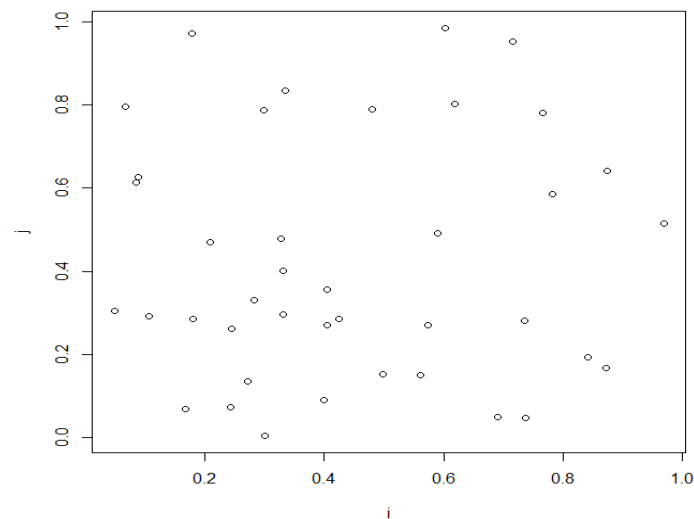


Figure 1. Spatial coordinates of $n = 100$ individuals located in a unitary square.

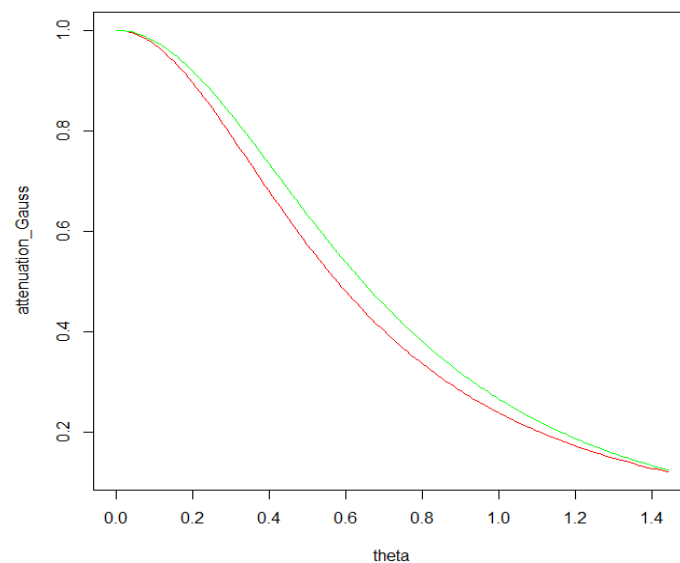


Figure 2. Attenuation effect in the presence of geo-masking as a function of the maximum displacement distance θ^* . Gaussian geo-masking (red line). Uniform geo-masking (green line).

Author Contributions

Authors have equally contributed to this work.

Conflicts of Interest

The authors declare no conflicts of interest.

Appendix A

Proof: Expression (5) in the text can be proved as follows. First of all, consider that:

$$E(u_{ij}) = E[(\epsilon_i^2 + \epsilon_j^2) + 2(i\epsilon_i + j\epsilon_j)] = 2E(\epsilon^2) + 2E(\epsilon)(i + j) = 2E(\epsilon^2)$$

since $E(\epsilon_i) = 0$. Furthermore, $Var(\epsilon^2) = E(\epsilon^2) = \sigma_\epsilon^2$

Q.E.D.

Similarly Equation (6) can be proved by considering that:

$$Var(u) = Var[(\epsilon_i^2 + \epsilon_j^2) + 2(i\epsilon_i + j\epsilon_j)] = 2Var(\epsilon^2) + 4Var(\epsilon_i)(i^2 + j^2) = 2Var(\epsilon^2) + 4\sigma_\epsilon^2 d_{ij}^2$$

due to independence. For the term $Var(\epsilon^2)$ in this expression, consider that, since $\epsilon_i \approx n.i.d.(0, \sigma_\epsilon^2)$,

$\frac{\epsilon}{\sigma_\epsilon} \approx n.i.d.(0, 1)$, and $\frac{\epsilon^2}{\sigma_\epsilon^2} \approx \chi_1^2$. So $Var\left(\frac{\epsilon^2}{\sigma_\epsilon^2}\right) = 2$ and, hence, $Var(\epsilon^2) = 2\sigma_\epsilon^4$ which substituted in the

previous expression, proves the result.

Q.E.D.

Appendix B

Proof: To prove that the distribution of the measurement error is a non-central Chi square, in Expression (7) consider that, if we divide both terms by σ_ϵ^2 we have:

$$\frac{u_{ij}}{\sigma_\epsilon^2} = \left(\frac{i + \epsilon}{\sigma_\epsilon}\right)^2 + \left(\frac{j + \epsilon}{\sigma_\epsilon}\right)^2 - \frac{d_{ij}^2}{\sigma_\epsilon^2} = \left(\frac{i}{\sigma_\epsilon} + \frac{\epsilon}{\sigma_\epsilon}\right)^2 + \left(\frac{j}{\sigma_\epsilon} + \frac{\epsilon}{\sigma_\epsilon}\right)^2 - \frac{d_{ij}^2}{\sigma_\epsilon^2}$$

The last term in this expression is a constant while each of the term in the brackets is $N\left(\frac{1}{\sigma_\epsilon}; 1\right)$.

Hence u is the sum of two independent squared normal distributions with non-zero mean and, as a consequence, it is distributed as a non-central Chi-squared with 2 degrees of freedom and

non-centrality parameter $\lambda = \left(\frac{i^2}{\sigma_\epsilon^2} + \frac{j^2}{\sigma_\epsilon^2}\right) = \frac{d^2}{\sigma_\epsilon^2}$.

Q.E.D.

Appendix C

Proof: The result reported in Equation (14) can be proved as follows. Consider that Expression (13) can be written as:

$$E(u_{ij}) = E[2\theta(i\cos\delta + j\sin\delta) + \theta^2(\cos^2\delta + \sin^2\delta)]$$

and, due to Hp2:

$$E(u_{ij}) = 2iE(\theta)E(\cos\delta) + 2jE(\theta)E(\sin\delta) + E(\theta^2)E(\cos^2\delta + \sin^2\delta)$$

Furthermore, we have that due to HP1:

$$E(\text{Cos}\delta) = \int_0^{360} \text{Cos}(\delta)f(\delta)d\delta = \frac{1}{360} \int_0^{360} \text{Cos}(\delta)d\delta = \frac{1}{360} [\text{Sin}\delta]_0^{360} = 0 = E(\text{Sin}\delta)$$

We also have that:

$$E(\theta^2) [E(\text{Cos}^2\delta + \text{Sin}^2\delta)] = E(\theta^2)$$

from the Pythagorean identity.

Finally, since, from HP1, $\theta \stackrel{iid}{\approx} U(0, \theta^*)$, then $E(\theta) = \frac{\theta^*}{2}$ and $Var(\theta) = \frac{\theta^{*2}}{12}$, so that $E(\theta^2) = Var(\theta) + E(\theta)^2 = \frac{\theta^{*2}}{12} + \frac{\theta^{*2}}{4} = \frac{\theta^{*2}}{3}$ which proves Expression (11).

Q.E.D.

Similarly, to prove Expression (15) consider that:

$$Var(u_{ij}) = 4i^2Var(\theta\text{Cos}\delta) + Var(\theta^2\text{Cos}^2\delta) + 4j^2Var(\theta\text{Sin}\delta) + Var(\theta^2\text{Sin}^2\delta)$$

since all cross terms have zero expectation.

Let us examine the various terms in this expression. First of all, we have that:

$$Var(\theta\text{Cos}\delta) = E(\theta^2\text{Cos}^2\delta) - E(\theta \text{Cos} \delta)^2$$

and, due to Hp2:

$$Var(\theta\text{Cos}\delta) = E(\theta^2)E(\text{Cos}^2\delta) - E(\theta)^2 E(\text{Cos}\delta)^2 = E(\theta^2)E(\text{Cos}^2\delta)$$

because $E(\text{Cos}\delta) = 0$:

Furthermore:

$$E(\text{Cos}^2\delta) = \int_0^{360} \text{Cos}^2(\delta)f(\delta)d\delta = \frac{1}{360} \int_0^{360} \text{Cos}^2(\delta)d\delta = \frac{1}{360} \left[\frac{360}{2} + \frac{\text{Cos}360\text{Sin}360 - \text{Cos}0\text{Sin}0}{2} \right]_0^{360} = \frac{1}{2}$$

and, as already seen $E(\theta^2) = \frac{\theta^{*2}}{3}$. So, eventually:

$$Var(\theta\text{Cos}\delta) = \frac{\theta^{*2}}{6} = Var(\theta\text{Sin}\delta).$$

Secondly, we have that

$$Var(\theta^2\text{Cos}^2\delta) = E(\theta^4\text{Cos}^4\delta) - E(\theta^2\text{Cos}^2\delta)^2 = E(\theta^4)E(\text{Cos}^4\delta) - E(\theta^2)E(\text{Cos}^2\delta)^2$$

In this expression we have that:

$$\begin{aligned} E(\theta^4) &= \int_0^{\theta^*} \theta^4 f(\theta)d\theta = \frac{1}{\theta^*} \int_0^{\theta^*} \theta^4 d\theta = \frac{1}{\theta^*} \left[\frac{\theta^5}{5} \right]_0^{\theta^*} = \frac{1}{\theta^*} \frac{\theta^{*5}}{5} = \frac{\theta^{*4}}{5}, \\ E(\text{Cos}^4\delta) &= \int_0^{360} \text{cos}^4 \delta f(\delta)d\delta = \frac{1}{360} \left\{ \left[\text{cos}^3 \delta \text{sin} \delta \right]_0^{360} + \frac{3}{4} \int_0^{360} \text{cos}^2 \delta d\delta \right\} = \\ &= \frac{1}{360} \left\{ \frac{3}{4} \int_0^{360} \text{cos}^2 \delta d\delta \right\} = \frac{1}{360} \left\{ \frac{3}{4} \frac{360}{2} \right\} = \frac{3}{8} \end{aligned}$$

and, as already seen:

$$E(\text{Cos}^2\delta) = \frac{1}{2}$$

As a consequence we have:

$$\text{Var}(\theta^2 \text{Cos}^2\delta) = \frac{\theta^{*4}}{5} \frac{3}{8} - \frac{\theta^{*4}}{9} \frac{1}{4} = \frac{17}{360} \theta^{*4} = \text{Var}(\theta^2 \text{Sin}^2\delta)$$

So eventually:

$$\begin{aligned} \text{Var}(u) &= \frac{4i^2\theta^{*2}}{6} + \frac{4j^2\theta^{*2}}{6} + \frac{34}{360} \theta^{*4} \\ &= \frac{2\theta^{*2}}{3} (i^2 + j^2) + \frac{17}{180} \theta^{*4} \end{aligned}$$

and finally

$$\text{Var}(u) = \frac{2\theta^{*2}}{3} d^2 + \frac{17}{180} \theta^{*4}$$

Q.E.D.

References

1. Arbia, G. *A Primer for Spatial Econometrics*; Palgrave Mac Millan: London, UK, 2014.
2. Collins, B. *Boundary Respecting Point Displacement*; Python Script; Blue Raster, LLC: Arlington, VA, USA, 2011.
3. Burgert, C.R.; Colston, J.; Roy, T.; Zachary, B. Geographic displacement procedure and georeferenced data release policy for the Demographic and Health Surveys. In *DHS Spatial Analysis Report No. 7*; ICF International: Calverton, MA, USA, 2013.
4. DHS. *Spatial Analysis Report 8. Guidelines on the Use of GPS DHS Data*; ICF Interactional: Calverton, MD, USA, September 2013.
5. Arbia, G.; Espa, G.; Giuliani, D. Dirty spatial econometrics. *Ann. Reg. Sci.* **2015**, submitted.
6. Elkies, N.; Fink, G.; Baernighausen, F.G. Scrambling geo-referenced data to protect privacy induces bias in distance estimation. *Popul. Environ.* **2015**, *37*, 83–98, doi:10.007/s11111-014-0225-0.
7. Verbeek, M. *A guide to Modern Econometrics*; Wiley: Chichester, UK, 2004.

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).