



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

**ΔΙΕΡΕΥΝΗΣΗ ΤΗΣ ΣΥΜΠΕΡΙΦΟΡΑΣ ΤΩΝ
ΠΡΟΒΛΗΜΑΤΩΝ ΑΥΤΟΣΥΣΧΕΤΙΣΗΣ ΚΑΙ
ΕΤΕΡΟΓΕΝΕΙΑΣ ΣΕ ΓΡΑΜΜΙΚΑ
ΥΠΟΔΕΙΓΜΑΤΑ ΜΕ ΧΩΡΙΚΑ ΔΕΔΟΜΕΝΑ
ΚΑΙ ΜΕΛΕΤΗ ΠΕΡΙΠΤΩΣΗΣ
ΓΟΝΙΜΟΤΗΤΑΣ ΤΟΥ ΠΛΗΘΥΣΜΟΥ ΤΗΣ
ΕΛΛΑΔΑΣ ΣΕ ΕΠΙΠΕΔΟ ΔΗΜΟΥ**

Απόστολος Η. Τσιμπάνος

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Υποβλήθηκε στο

Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης

του Πανεπιστημίου Πειραιώς

Πειραιάς
Ιούνιος 2016



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

**ΔΙΕΡΕΥΝΗΣΗ ΤΗΣ ΣΥΜΠΕΡΙΦΟΡΑΣ ΤΩΝ
ΠΡΟΒΛΗΜΑΤΩΝ ΑΥΤΟΣΥΣΧΕΤΙΣΗΣ ΚΑΙ
ΕΤΕΡΟΓΕΝΕΙΑΣ ΣΕ ΓΡΑΜΜΙΚΑ
ΥΠΟΔΕΙΓΜΑΤΑ ΜΕ ΧΩΡΙΚΑ ΔΕΔΟΜΕΝΑ
ΚΑΙ ΜΕΛΕΤΗ ΠΕΡΙΠΤΩΣΗΣ
ΓΟΝΙΜΟΤΗΤΑΣ ΤΟΥ ΠΛΗΘΥΣΜΟΥ ΤΗΣ
ΕΛΛΑΔΑΣ ΣΕ ΕΠΙΠΕΔΟ ΔΗΜΟΥ**

Απόστολος Η. Τσιμπάνος

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Υποβλήθηκε στο

Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης

του Πανεπιστημίου Πειραιώς

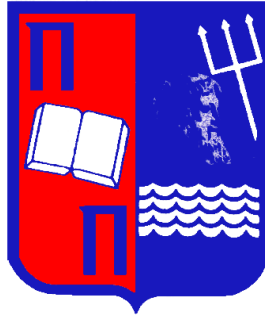
Πειραιάς
Ιούνιος 2016

Η παρούσα Διδακτορική Διατριβή εγκρίθηκε ομόφωνα από την Επταμελή Εξεταστική Επιτροπή που ορίστηκε από τη ΓΣΕΣ του Τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς στην υπ' αριθμ. 4/18.04.2016 συνεδρίασή του σύμφωνα με τον Εσωτερικό Κανονισμό Λειτουργίας του Τμήματος

Τα μέλη της Επιτροπής ήταν:

- Καθηγητής Χρήστος Αγιακλόγλου
- Επίκουρη Καθηγήτρια Γεωργία Βερροπούλου
- Επίκουρος Καθηγητής Σταμάτης Καλογήρου
- Αναπληρωτής Καθηγητής Σωτήριος Καρκαλάκος
- Αναπληρωτής Καθηγητής Κωνσταντίνος Πολίτης
- Αναπληρωτής Καθηγητής Αντώνιος Ροβολής
- Καθηγητής Κλέων Τσίμπος (Επιβλέπων)

Η έγκριση της Διδακτορικής Διατριβής από το Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς δεν υποδηλώνει αποδοχή των γνώμών του συγγραφέα.



UNIVERSITY OF PIRAEUS

**DEPARTMENT OF STATISTICS
AND INSURANCE SCIENCE**

**INVESTIGATING THE BEHAVIOUR OF
AUTOCORRELATION AND
HETEROGENEITY PROBLEMS IN LINEAR
MODELS WITH SPATIAL DATA AND A
CASE STUDY OF FERTILITY OF THE
POPULATION OF GREECE AT LOCAL
AUTHORITY LEVEL**

Apostolos E. Tsimpanos

PhD Thesis

Submitted to
Department of Statistics and Insurance Science of
the University of Piraeus

Piraeus, Greece
June 2016

*Στη μνήμη του αδελφού μου
Παναγιώτη*

Ευχαριστίες

Αισθάνομαι την ανάγκη να ευχαριστήσω τον Καθηγητή κύριο Κλέων Τσίμπο για τη συνεχή καθοδήγησή του και την άψογη συνεργασία μας καθ' όλη τη διάρκεια εκπόνησης της διατριβής μου. Επίσης θερμές ευχαριστίες εκφράζω προς τον Καθηγητή κύριο Χρήστο Αγιακλόγλου και τον Αναπληρωτή καθηγητή κύριο Κωνσταντίνο Πολίτη που ως μέλη της τριμελούς επιτροπής μου προσέφεραν απεριόριστη βοήθεια.

Περίληψη

Η εκτίμηση υποδειγμάτων γραμμικής παλινδρόμησης με χωρικά δεδομένα, δηλαδή με διασπρωματικά δεδομένα που έχουν συλλεχθεί και προέρχονται από διαφορετικές γεωγραφικές περιοχές, εμπεριέχει σοβαρούς κινδύνους εξ' αιτίας της εμφάνισης των χωρικών επιδράσεων οι οποίες εάν αγνοηθούν από τον ερευνητή επηρεάζουν την αξιοπιστία των αποτελεσμάτων. Οι χωρικές επιδράσεις, που οφείλονται στη φύση των δεδομένων, αποτελούνται από τη χωρική εξάρτηση και τη χωρική ετερογένεια. Ειδικότερα, χωρική εξάρτηση σημαίνει ότι οι τιμές μιας μεταβλητής δεν είναι μεταξύ τους ανεξάρτητες αλλά εξαρτώνται σύμφωνα με τη γεωγραφική τους θέση δημιουργώντας συστάδες παρατηρήσεων. Χωρική ετερογένεια σημαίνει ότι η σχέση ανάμεσα στις εξεταζόμενες μεταβλητές δεν παραμένει σταθερή σε ολόκληρη τη γεωγραφική περιοχή αλλά μεταβάλλεται από σημείο σε σημείο. Η αντιμετώπιση των χωρικών επιδράσεων γίνεται με την ενσωμάτωσή τους στο υπόδειγμα μέσω της εξειδίκευσης χωρικών οικονομετρικών υποδειγμάτων.

Η διατριβή αυτή ασχολείται με θέματα που αφορούν τα προβλήματα της χωρικής εξάρτησης και της χωρικής ετερογένειας στο γραμμικό υπόδειγμα παλινδρόμησης. Πιο συγκεκριμένα η δομή της διατριβής έχει ως εξής:

Στο Κεφάλαιο 1 γίνεται μια σύντομη εισαγωγή στη μεθοδολογία της χωρικής ανάλυσης παρουσιάζοντας τις ιδιαιτερότητες των χωρικών δεδομένων και την κατασκευή της μήτρας χωρικών σταθμίσεων. Επιπρόσθετα, παρουσιάζονται οι βασικές χωρικές διαδικασίες και τα βασικά χωρικά οικονομετρικά υποδείγματα.

Στο Κεφάλαιο 2 εξετάζεται, χρησιμοποιώντας προσομοίωση, η συμπεριφορά των LM ελέγχων χωρικής εξάρτησης και εξειδίκευσης που εφαρμόζονται για την ανίχνευση του προβλήματος της χωρικής αυτοσυσχέτισης στα σφάλματα ενός γραμμικού υποδείγματος και την επιλογή του κατάλληλου χωρικού οικονομετρικού υποδείγματος. Τα αποτελέσματα που προέκυψαν από τη διαδικασία της προσομοίωσης επιβεβαιώνουν παλιότερα, γνωστά στη βιβλιογραφία της χωρικής οικονομετρίας, συμπεράσματα για τη συμπεριφορά αυτών των ελέγχων στα μικρά και μέτρια δείγματα και συνεισφέρουν περαιτέρω πληροφορίες για τη συμπεριφορά τους στα μεγάλα δείγματα. Επιπρόσθετα, στο ίδιο κεφάλαιο παρουσιάζεται το φαινόμενο της υποφαινομενικής παλινδρόμησης ως μια άλλη αιτία που οδηγεί σε χωρική αυτοσυσχέτιση στα σφάλματα ενός οικονομετρικού υποδείγματος. Με τη βοήθεια

προσομοίωσης για δυο ανεξάρτητες στάσιμες χωρικές αυτοπαλίνδρομες διαδικασίες πρώτης τάξης, SAR(1) εντοπίζονται οι ομοιότητες στην εμφάνιση του φαινομένου στα χωρικά δεδομένα με τα δεδομένα των χρονοσειρών. Ωστόσο, εάν ο ερευνητής ακολουθήσει τις υποδείξεις από την εφαρμογή των ελέγχων χωρικής εξάρτησης και εκτιμήσει κάποιο χωρικό οικονομετρικό υπόδειγμα δεν εμφανίζεται το πρόβλημα της υποφαινομενικής παλινδρόμησης.

Στο Κεφάλαιο 3 πραγματοποιούνται μια σειρά προσομοιώσεων με σκοπό τη διερεύνηση της συμπεριφοράς της γεωγραφικά σταθμισμένης παλινδρόμησης που συνήθως χρησιμοποιείται στην αντιμετώπιση του προβλήματος της χωρικής ετερογένειας μέσω της εκτίμησης τοπικών υποδειγμάτων καθώς και των ελέγχων που εξετάζουν τη συνεισφοράς της. Τα αποτελέσματα δείχνουν ότι, παρά την καλή εφαρμογή και τα πλεονεκτήματα της μεθόδου, υπάρχουν περιπτώσεις που μπορεί να οδηγήσουν σε παραπλανητικά αποτελέσματα. Στο ίδιο κεφάλαιο εισάγεται η έννοια της τοπικής υποφαινομενικής παλινδρόμησης και μελετάται το πρόβλημα για δυο ανεξάρτητες στάσιμες χωρικές αυτοπαλίνδρομες διαδικασίες πρώτης τάξης, SAR(1). Το βασικό συμπέρασμα είναι ότι ο ερευνητής θα πρέπει να είναι ιδιαίτερα προσεχτικός όταν εκτιμά τοπικά υποδείγματα με τη μέθοδο της γεωγραφικά σταθμισμένης παλινδρόμησης διότι μπορεί να εμφανιστεί σε τοπικό επίπεδο υποφαινομενική συμπεριφορά που δεν μπορεί να αποφευχθεί σε αντίθεση με το ίδιο πρόβλημα στο ολικό επίπεδο που αντιμετωπίζεται.

Στο Κεφάλαιο 4 γίνεται προσπάθεια να ερμηνευθούν εμπειρικά οι χωρικές διαφοροποιήσεις και τα πρότυπα στις σχέσεις μεταξύ του επιπέδου της γονιμότητας του πληθυσμού της Ελλάδας και ενός αριθμού επιλεγμένων κοινωνικό – οικονομικών δεικτών σε τοπικό διοικητικό επίπεδο λαμβάνοντας υπ' όψιν την παρουσία της χωρικής ετερογένειας. Η εκτίμηση του υποδείγματος με τη μέθοδο της γεωγραφικά σταθμισμένης παλινδρόμησης και η χαρτογράφηση των αποτελεσμάτων αποκαλύπτουν ότι σε επίπεδο δήμων υπάρχουν σημαντικές διαφοροποιήσεις τόσο στην κατεύθυνση όσο και στην ένταση των επιδράσεων των κοινωνικών και οικονομικών παραγόντων στα επίπεδα της γονιμότητας.

Abstract

The estimation of linear regression models with spatial data, that is cross-sectional data collected and obtained from different geographical regions, carries serious risks due to the appearance of spatial effects which influence the result reliability in case such effects are ignored by the researcher. Spatial effects, which are attributed to the nature of the data, are distinguished from spatial dependence and spatial heterogeneity outcomes. Spatial dependence means that the values of a variable are not independent of each other but correlated according to their geographical positions creating clusters of observations. Spatial heterogeneity means that the relationship between the variables under consideration does not remain constant throughout the study region but varies from point to point. The treatment of spatial effects is achieved by their incorporation in the model using spatial econometric models.

This PhD thesis deals with topics relating to the problems of spatial dependence and spatial heterogeneity in the linear regression model. More specifically, the structure of the thesis is organized as follows:

Chapter 1 gives a brief introduction to spatial analysis methodology presenting the peculiarities of spatial data and the construction of the spatial weights matrix. In addition, the main spatial processes and spatial econometric models are described.

Chapter 2 examines, using simulation techniques, the behavior of the LM spatial dependence and specification tests which are applied for detecting the spatial autocorrelation problem in the errors of a linear model and for selecting the appropriate spatial econometric model. The results obtained through the simulation process confirm the results presented in the literature concerning test behavior in small and medium samples and contribute to further information for their behavior in large samples. Moreover, the spurious regression phenomenon is also presented in the same chapter as another reason leading to spatial autocorrelation in the errors of an econometric model. Using simulation analysis with two independent stationary spatial autoregressive processes of order one, SAR(1), similarities are recognized in the appearance of the phenomenon for spatial data with time series data. However, if the analyst takes into account the suggestions deriving from the LM spatial

dependence and specification tests and estimates a spatial econometric model the spurious regression problem will not be revealed.

Chapter 3 investigates, employing simulation techniques, the behavior of the geographically weighted regression which is commonly used to remedy the spatial heterogeneity problem through local modeling estimation as well as the performance of the tests which examine its contribution. The findings indicate that, in some cases, despite the good performance and the advantages of this method misleading results can be obtained. Moreover, the concept of spurious local regression is examined for two independent stationary spatial autoregressive processes of order one, SAR(1). The main conclusion of the analysis is that the researcher should be very cautious when estimating local models applying geographically weighted regression as spurious behavior will appear at the local level, a problem that cannot be avoided in contrast to the same problem at the global level which can be treated.

Chapter 4 tries empirically to explain spatial variations and patterns in the relationships between the fertility level of the population of Greece and a number of selected socio-economic indicators at the local authority level taking into account the existence of spatial heterogeneity. The estimation of the model, applying the geographically weighted regression method and mapping the results, reveals that considerable differentials exist at the municipalities level in the direction and intensity of social and economic factors affecting fertility.

Περιεχόμενα

| | |
|-------------------------------|----|
| Κατάλογος Πινάκων | xi |
| Κατάλογος Διαγραμμάτων | xv |

ΚΕΦΑΛΑΙΟ 1

Χωρικά Δεδομένα και Χωρική Οικονομετρία

| | | |
|-------|--|----|
| 1.1 | Εισαγωγή | 1 |
| 1.2 | Οι χωρικές επιδράσεις | 4 |
| 1.3 | Χωρική γειτονία και μήτρες χωρικών σταθμίσεων | 7 |
| 1.4 | Ο συντελεστής χωρικής αυτοσυσχετισης I του Moran | 15 |
| 1.5 | Χωρικές στοχαστικές διαδικασίες | 18 |
| 1.5.1 | Η χωρική διαδικασία λευκού θορύβου | 20 |
| 1.5.2 | Η χωρική αυτοπαλίνδρομη διαδικασία πρώτης τάξης | 20 |
| 1.5.3 | Η χωρική διαδικασία κινητού μέσου πρώτης τάξης | 22 |
| 1.6 | Χωρικά οικονομετρικά υποδείγματα | 23 |
| 1.6.1 | Το υπόδειγμα χωρικής υστέρησης | 23 |
| 1.6.2 | Το υπόδειγμα χωρικού σφάλματος | 27 |
| 1.6.3 | Υποδείγματα χωρικής ετερογένειας | 30 |
| 1.7 | Ανακεφαλαίωση | 32 |

ΚΕΦΑΛΑΙΟ 2

Χωρική Αυτοσυσχέτιση στο Γραμμικό Υπόδειγμα Παλινδρόμησης

| | | |
|-------|---|----|
| 2.1 | Εισαγωγή | 35 |
| 2.2 | Οι συνέπειες της χωρικής αυτοσυσχέτισης στα σφάλματα του γραμμικού υποδείγματος παλινδρόμησης | 36 |
| 2.3 | Έλεγχοι για την ύπαρξη χωρικής αυτοσυσχέτισης στα σφάλματα του γραμμικού υποδείγματος παλινδρόμησης | 43 |
| 2.3.1 | Ο έλεγχος I του Moran για σφάλματα παλινδρόμησης | 43 |
| 2.3.2 | Οι έλεγχοι χωρικής εξάρτησης | 45 |
| 2.4 | Το μέγεθος και η ισχύς των ελέγχων χωρικής εξάρτησης | 50 |
| 2.5 | Έλεγχος για χωρική αυτοσυσχέτιση στο υπόδειγμα χωρικής υστέρησης | 65 |
| 2.5.1 | Η στατιστική $LM_{\lambda p}$ | 66 |

| | | |
|-------|--|----|
| 2.5.2 | Η ισχύς των ελέγχων χωρικής αυτοσυσχέτισης στο υπόδειγμα χωρικής υστέρησης | 67 |
| 2.6 | Χωρική αυτοσυσχέτιση στα σφάλματα του γραμμικού υποδείγματος παλινδρόμησης λόγω ύπαρξης υποφαινομενικής σχέσης | 74 |
| 2.6.1 | Η έννοια της υποφαινομενικής συμπεριφοράς | 74 |
| 2.6.2 | Υποφαινομενική συμπεριφορά με στάσιμες χωρικές αυτοπαλίνδρομες διαδικασίες πρώτης τάξης | 76 |
| 2.7 | Ανακεφαλαίωση | 85 |

ΚΕΦΑΛΑΙΟ 3

Χωρική Ετερογένεια στο Γραμμικό Υπόδειγμα Παλινδρόμησης

| | | |
|-------|--|-----|
| 3.1 | Εισαγωγή | 87 |
| 3.2 | Η γεωγραφικά σταθμισμένη παλινδρόμηση | 89 |
| 3.3 | Έλεγχοι βελτίωσης από την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης | 94 |
| 3.4 | Έλεγχος για χωρική αυτοσυσχέτιση στα σφάλματα του υποδείγματος της γεωγραφικά σταθμισμένης παλινδρόμησης | 99 |
| 3.5 | Η συμπεριφορά της γεωγραφικά σταθμισμένης παλινδρόμησης στην ανίχνευση της χωρικής ετερογένειας και η αποτελεσματικότητα των ελέγχων βελτίωσης | 104 |
| 3.6 | Η συμπεριφορά της γεωγραφικά σταθμισμένης παλινδρόμησης όταν υπάρχουν επιδράσεις χωρικής εξάρτησης | 123 |
| 3.7 | Υποφαινομενική συμπεριφορά σε τοπικό επίπεδο από την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης | 141 |
| 3.7.1 | Τοπική υποφαινομενική συμπεριφορά με στάσιμες χωρικές αυτοπαλίνδρομες διαδικασίες πρώτης τάξης | 142 |
| 3.7.2 | Ανακεφαλαίωση | 152 |

ΚΕΦΑΛΑΙΟ 4

Διερεύνηση των Χωρικών Διαφοροποιήσεων των Επιδράσεων στη Γονιμότητα του Πληθυσμού της Ελλάδας σε Τοπικό Διοικητικό Επίπεδο

| | | |
|------|--|-----|
| 4.1 | Εισαγωγή | 155 |
| 4.2 | Τα δεδομένα και οι μεταβλητές του υποδείγματος | 159 |
| 4.3 | Η εφαρμοζόμενη μέθοδος χωρικής ανάλυσης | 163 |
| 4.4 | Περιγραφική στατιστική παρουσίαση των μεταβλητών | 165 |
| 4.5 | Ανάλυση χωρικής αυτοσυσχέτισης των μεταβλητών | 167 |
| 4.6 | Τα αποτελέσματα της εκτίμησης του απλού υποδείγματος | 168 |
| 4.7 | Τα αποτελέσματα της εφαρμογής της Γεωγραφικά Σταθμισμένης Παλινδρόμησης | 170 |
| 4.8 | Σύγκριση των ολικών και τοπικών επιδράσεων των κοινωνικών και οικονομικών παραγόντων στα επίπεδα της γονιμότητας | 173 |
| 4.9 | Συμπεράσματα | 180 |
| 4.10 | Ανακεφαλαίωση | 183 |

| | |
|--|-----|
| Βασικά Συμπεράσματα της Διατριβής | 185 |
| Βιβλιογραφία | 193 |

Κατάλογος Πινάκων

| | | |
|--------------|---|----|
| Πίνακας 2.1 | Μέσες τιμές I και ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% των ελέγχων χωρικής εξάρτησης για το απλό οικονομετρικό υπόδειγμα για όλα τα μεγέθη δείγματος | 52 |
| Πίνακας 2.2 | Θεωρητικές αναμενόμενες τιμές της στατιστικής I για το απλό οικονομετρικό υπόδειγμα για όλα τα μεγέθη δείγματος και μήτρες χωρικών σταθμίσεων | 53 |
| Πίνακας 2.3 | Μέσες τιμές I και ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% των ελέγχων χωρικής εξάρτησης για όταν ισχύει το υπόδειγμα χωρικής υστέρησης για όλες τις τιμές του ρ και μεγέθη δείγματος Χωρίς διόρθωση τόρος | 55 |
| Πίνακας 2.4 | Μέσες τιμές I και ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% των ελέγχων χωρικής εξάρτησης για όταν ισχύει το υπόδειγμα χωρικής υστέρησης για όλες τις τιμές του ρ και μεγέθη δείγματος Με διόρθωσης τόρος | 56 |
| Πίνακας 2.5 | Μέσες τιμές I και ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% των ελέγχων χωρικής εξάρτησης για όταν ισχύει το υπόδειγμα χωρικού σφάλματος με σφάλματα που ακολουθούν χωρικές αυτοπαλίνδρομες διαδικασίες για όλες τις τιμές του λ και μεγέθη δείγματος Χωρίς διόρθωση τόρος | 59 |
| Πίνακας 2.6 | Μέσες τιμές I και ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% των ελέγχων χωρικής εξάρτησης για όταν ισχύει το υπόδειγμα χωρικού σφάλματος με σφάλματα που ακολουθούν χωρικές αυτοπαλίνδρομες διαδικασίες για όλες τις τιμές του λ και μεγέθη δείγματος Με διόρθωσης τόρος | 60 |
| Πίνακας 2.7 | Μέσες τιμές I και ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% των ελέγχων χωρικής εξάρτησης για όταν ισχύει το υπόδειγμα χωρικού σφάλματος με σφάλματα που ακολουθούν χωρικές διαδικασίες κινητού μέσου για όλες τις τιμές του λ και μεγέθη δείγματος Χωρίς διόρθωση τόρος | 62 |
| Πίνακας 2.8 | Μέσες τιμές I και ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% των ελέγχων χωρικής εξάρτησης για όταν ισχύει το υπόδειγμα χωρικού σφάλματος με σφάλματα που ακολουθούν χωρικές διαδικασίες κινητού μέσου για όλες τις τιμές του λ και μεγέθη δείγματος Με διόρθωσης τόρος | 63 |
| Πίνακας 2.9 | Μέσες τιμές I για τα σφάλματα του υποδείγματος χωρικής υστέρησης με χωρικά αυτοπαλίνδρομα σφάλματα για όλες τις τιμές των ρ και λ και μεγέθη δείγματος Χωρίς διόρθωση τόρος | 69 |
| Πίνακας 2.10 | Μέσες τιμές I για τα σφάλματα του υποδείγματος χωρικής υστέρησης με χωρικά αυτοπαλίνδρομα σφάλματα για όλες τις τιμές των ρ και λ και μεγέθη δείγματος Με διόρθωση τόρος | 69 |

| | | |
|--------------|--|-----|
| Πίνακας 2.11 | Ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% της στατιστικής I για τα σφάλματα του υποδείγματος χωρικής υστέρησης με χωρικά αυτοπαλίνδρομα σφάλματα για όλες τις τιμές των ρ και λ και μεγέθη δείγματος Χωρίς διόρθωση τόρος | 71 |
| Πίνακας 2.12 | Ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% της στατιστικής I για τα σφάλματα του υποδείγματος χωρικής υστέρησης με χωρικά αυτοπαλίνδρομα σφάλματα για όλες τις τιμές των ρ και λ και μεγέθη δείγματος Με διόρθωση τόρος | 71 |
| Πίνακας 2.13 | Ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% της στατιστικής $LM_{\lambda \rho}$ για τα σφάλματα του υποδείγματος χωρικής υστέρησης με χωρικά αυτοπαλίνδρομα σφάλματα για όλες τις τιμές των ρ και λ και μεγέθη δείγματος Χωρίς διόρθωση τόρος | 72 |
| Πίνακας 2.14 | Ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% της στατιστικής $LM_{\lambda \rho}$ για τα σφάλματα του υποδείγματος χωρικής υστέρησης με χωρικά αυτοπαλίνδρομα σφάλματα για όλες τις τιμές των ρ και λ και μεγέθη δείγματος Με διόρθωση τόρος | 72 |
| Πίνακας 2.15 | Ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% της μηδενικής υπόθεσης $\beta_1 = 0$ για το απλό οικονομετρικό υπόδειγμα (2.14) για δύο ανεξάρτητες στάσιμες SAR(1) διαδικασίες με μέσες τιμές στατιστικής I και ποσοστά απόρριψης των ελέγχων χωρικής εξάρτησης για όλες τις τιμές του φ και μεγέθη δείγματος Χωρίς διόρθωση τόρος | 78 |
| Πίνακας 2.16 | Ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% της μηδενικής υπόθεσης $\beta_1 = 0$ για το απλό οικονομετρικό υπόδειγμα (2.14) για δύο ανεξάρτητες στάσιμες SAR(1) διαδικασίες με μέσες τιμές στατιστικής I και ποσοστά απόρριψης των ελέγχων χωρικής εξάρτησης για όλες τις τιμές του φ και μεγέθη δείγματος Με διόρθωση τόρος | 80 |
| Πίνακας 2.17 | Ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% της μηδενικής υπόθεσης $\beta_1 = 0$ για τα χωρικά υποδείγματα (2.15) και (2.16) για δύο ανεξάρτητες στάσιμες SAR(1) διαδικασίες και ποσοστά απόρριψης στατιστικής $LM_{\lambda \rho}$ για τον έλεγχο της απουσίας χωρικής αυτοσυσχέτισης στο υπόδειγμα (2.15) Χωρίς διόρθωση τόρος | 83 |
| Πίνακας 2.18 | Ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% της μηδενικής υπόθεσης $\beta_1 = 0$ για τα χωρικά υποδείγματα (2.15) και (2.16) για δύο ανεξάρτητες στάσιμες SAR(1) διαδικασίες και ποσοστά απόρριψης στατιστικής $LM_{\lambda \rho}$ για τον έλεγχο της απουσίας χωρικής αυτοσυσχέτισης στο υπόδειγμα (2.15) Με διόρθωση τόρος | 83 |
| Πίνακας 3.1 | Συντελεστές προσδιορισμού μεθόδων ελαχίστων τετραγώνων και γεωγραφικά σταθμισμένης παλινδρόμησης και ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% των ελέγχων βελτίωσης για τα τέσσερα υποδείγματα και όλα τα μεγέθη δείγματος | 109 |
| Πίνακας 3.2 | Μέσες, ελάχιστες και μέγιστες τιμές εύρους για τα τέσσερα υποδείγματα και όλα τα μεγέθη δείγματος | 111 |
| Πίνακας 3.3 | Μέσες τιμές στατιστικής I και ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% με την τεχνική των Leung, Mei και Zhang (2000b) του ελέγχου ύπαρξης χωρικής αυτοσυσχέτισης στα εκτιμηθέντα σφάλματα της μεθόδου της γεωγραφικά σταθμισμένης | |

| | | |
|--------------|---|-----|
| | παλινδρόμησης για τα τέσσερα υποδείγματα και όλα τα μεγέθη δείγματος | 111 |
| Πίνακας 3.4 | Συντελεστές προσδιορισμού μεθόδων ελαχίστων τετραγώνων και γεωγραφικά σταθμισμένης παλινδρόμησης και ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% των ελέγχων βελτίωσης για όταν ισχύει το υπόδειγμα χωρικού σφάλματος για όλα τα μεγέθη δείγματος | 125 |
| Πίνακας 3.5 | Μέσες, ελάχιστες και μέγιστες τιμές εύρους για όταν ισχύει το υπόδειγμα χωρικού σφάλματος | 125 |
| Πίνακας 3.6 | Μέσες τιμές στατιστικής I και ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% με την τεχνική των Leung, Mei και Zhang (2000b) και την υπόθεση της τυχαιοποίησης του ελέγχου ύπαρξης χωρικής αυτοσυσχέτισης στα εκτιμηθέντα σφάλματα της μεθόδου της γεωγραφικά σταθμισμένης παλινδρόμησης για όταν ισχύει το υπόδειγμα χωρικού σφάλματος | 127 |
| Πίνακας 3.7 | Συντελεστές προσδιορισμού μεθόδων ελαχίστων τετραγώνων και γεωγραφικά σταθμισμένης παλινδρόμησης και ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% των ελέγχων βελτίωσης για όταν ισχύει το υπόδειγμα χωρικής υστέρησης για όλα τα μεγέθη δείγματος | 133 |
| Πίνακας 3.8 | Μέσες, ελάχιστες και μέγιστες τιμές εύρους για όταν ισχύει το υπόδειγμα χωρικής υστέρησης | 133 |
| Πίνακας 3.9 | Μέσες τιμές στατιστικής I και ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% με την τεχνική των Leung, Mei και Zhang (2000b) και την υπόθεση της τυχαιοποίησης του ελέγχου ύπαρξης χωρικής αυτοσυσχέτισης στα εκτιμηθέντα σφάλματα της μεθόδου της γεωγραφικά σταθμισμένης παλινδρόμησης για όταν ισχύει το υπόδειγμα χωρικής υστέρησης | 134 |
| Πίνακας 3.10 | Συντελεστές προσδιορισμού μεθόδων ελαχίστων τετραγώνων και γεωγραφικά σταθμισμένης παλινδρόμησης και ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% των ελέγχων βελτίωσης για δύο ανεξάρτητες SAR(1) διαδικασίες για όλα τα μεγέθη δείγματος | 144 |
| Πίνακας 3.11 | Μέσες, ελάχιστες και μέγιστες τιμές εύρους για δύο ανεξάρτητες SAR(1) διαδικασίες | 144 |
| Πίνακας 3.12 | Συχνότητα τοπικών ποσοστών απόρριψεων σε επίπεδο σημαντικότητας 5% για δύο ανεξάρτητες SAR(1) διαδικασίες | 148 |
| Πίνακας 3.13 | Μέσες τιμές στατιστικής I και ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% με την τεχνική των Leung, Mei και Zhang (2000b) του ελέγχου ύπαρξης χωρικής αυτοσυσχέτισης στα εκτιμηθέντα σφάλματα της μεθόδου της γεωγραφικά σταθμισμένης παλινδρόμησης για δύο ανεξάρτητες SAR(1) διαδικασίες | 150 |
| Πίνακας 4.1 | Βασικά περιγραφικά στατιστικά μέτρα των μεταβλητών του υποδείγματος | 165 |
| Πίνακας 4.2 | Ανάλυση χωρικής αυτοσυσχέτισης των μεταβλητών του υποδείγματος: Τιμές στατιστικής I του Moran για εναλλακτικό αριθμό γειτόνων | 167 |
| Πίνακας 4.3 | Αποτελέσματα εκτίμησης ολικού υποδείγματος (OLS) | 169 |

| | | |
|-------------|--|-----|
| Πίνακας 4.4 | Στατιστικοί έλεγχοι αυτοσυσχέτισης, ετεροσκεδαστικότητας και κανονικότητας για τα εκτιμηθέντα σφάλματα του ολικού υποδείγματος (OLS) | 169 |
| Πίνακας 4.5 | Αποτελέσματα εκτίμησης τοπικών υποδειγμάτων (GWR) | 171 |
| Πίνακας 4.6 | Έλεγχος ανάλυση διακύμανσης για τη σύγκριση ολικού υποδείγματος (OLS) και τοπικών υποδειγμάτων (GWR) | 171 |

Κατάλογος Διαγραμμάτων

| | | |
|----------------------|--|-----|
| Διάγραμμα 1.1 | Γειτνίαση πύργου (αριστερά) και γειτνίαση βασίλισσας (δεξιά) σε ένα ομαλό τετραγωνικό πλέγμα εννέα χωρικών μονάδων | 10 |
| Διάγραμμα 2.1 | Προσομοιωμένες τιμές των στατιστικών LM-LE και LM-EL για το υπόδειγμα χωρικής υστέρησης με το κριτήριο της βασίλισσας για όλα τα μεγέθη δείγματος και τιμές του ρ Χωρίς διόρθωση τόρος | 57 |
| Διάγραμμα 2.2 | Προσομοιωμένες τιμές των στατιστικών LM-EL και LM-LE για το υπόδειγμα χωρικού σφάλματος με σφάλματα που ακολουθούν χωρικές αυτοπαλίνδρομες διαδικασίες με το κριτήριο της βασίλισσας για όλα τα μεγέθη δείγματος και τιμές του λ Χωρίς διόρθωση τόρος | 61 |
| Διάγραμμα 3.1 | Οι γεωγραφικές συντεταγμένες σε ένα 10x10 ομαλό πλέγμα ($n=100$) | 109 |
| Διάγραμμα 3.2 (A) | Ψηφιδωτά πραγματικών και προσομοιωμένων μέσω τοπικών εκτιμηθέντων τιμών από την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης για όλα τα μεγέθη δείγματος και τις δύο συναρτήσεις σταθμίσεων Υπόδειγμα I: Ολική Σχέση Συντελεστής β_{i1} | 114 |
| Διάγραμμα 3.2 (B) | Ψηφιδωτά πραγματικών και προσομοιωμένων μέσω τοπικών εκτιμηθέντων τιμών από την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης για όλα τα μεγέθη δείγματος και τις δύο συναρτήσεις σταθμίσεων Υπόδειγμα I: Ολική Σχέση Συντελεστής β_{i0} | 115 |
| Διάγραμμα 3.3 (A) | Ψηφιδωτά πραγματικών και προσομοιωμένων μέσω τοπικών εκτιμηθέντων τιμών από την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης για όλα τα μεγέθη δείγματος και τις δύο συναρτήσεις σταθμίσεων Υπόδειγμα II: Χωρικό Καθεστώς Συντελεστής β_{i1} | 116 |
| Διάγραμμα 3.3 (B) | Ψηφιδωτά πραγματικών και προσομοιωμένων μέσω τοπικών εκτιμηθέντων τιμών από την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης για όλα τα μεγέθη δείγματος και τις δύο συναρτήσεις σταθμίσεων Υπόδειγμα II: Χωρικό Καθεστώς Συντελεστής β_{i0} | 117 |
| Διάγραμμα 3.4 (A) | Ψηφιδωτά πραγματικών και προσομοιωμένων μέσω τοπικών εκτιμηθέντων τιμών από την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης για όλα τα μεγέθη δείγματος και τις δύο συναρτήσεις σταθμίσεων Υπόδειγμα III: Χαμηλή Χωρική Ετερογένεια Συντελεστής β_{i1} | 118 |
| Διάγραμμα 3.4 (B) | Ψηφιδωτά πραγματικών και προσομοιωμένων μέσω τοπικών εκτιμηθέντων τιμών από την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης για όλα τα μεγέθη δείγματος και τις δύο συναρτήσεις σταθμίσεων Υπόδειγμα III: Χαμηλή Χωρική Ετερογένεια Συντελεστής β_{i0} | 119 |
| Διάγραμμα 3.5 (A) | Ψηφιδωτά πραγματικών και προσομοιωμένων μέσω τοπικών εκτιμηθέντων τιμών από την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης για όλα τα μεγέθη δείγματος και τις δύο συναρτήσεις σταθμίσεων | |

| | | |
|----------------------|---|-----|
| | Υπόδειγμα IV: Υψηλή Χωρική Ετερογένεια Συντελεστής β_{i1} | 120 |
| Διάγραμμα 3.5 (B) | Ψηφιδωτά πραγματικών και προσομοιωμένων μέσω των τοπικών εκτιμηθέντων τιμών από την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης για όλα τα μεγέθη δείγματος και τις δύο συναρτήσεις σταθμίσεων | |
| | Υπόδειγμα IV: Υψηλή Χωρική Ετερογένεια Συντελεστής β_{i0} | 121 |
| Διάγραμμα 3.6 (A) | Ψηφιδωτά πραγματικών και προσομοιωμένων μέσω των τοπικών εκτιμηθέντων τιμών από την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης στο Υπόδειγμα Χωρικού Σφάλματος με Διτετράγωνη συνάρτηση στάθμισης για όλα τα μεγέθη δείγματος και τιμές του λ | |
| | Συντελεστής β_{i1} | 128 |
| Διάγραμμα 3.6 (B) | Ψηφιδωτά πραγματικών και προσομοιωμένων μέσω των τοπικών εκτιμηθέντων τιμών από την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης στο Υπόδειγμα Χωρικού Σφάλματος με Διτετράγωνη συνάρτηση στάθμισης για όλα τα μεγέθη δείγματος και τιμές του λ | |
| | Συντελεστής β_{i0} | 129 |
| Διάγραμμα 3.7 (A) | Ψηφιδωτά πραγματικών και προσομοιωμένων μέσω των τοπικών εκτιμηθέντων τιμών από την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης στο Υπόδειγμα Χωρικού Σφάλματος με Γκαουσιανή συνάρτηση στάθμισης για όλα τα μεγέθη δείγματος και τιμές του λ | |
| | Συντελεστής β_{i1} | 130 |
| Διάγραμμα 3.7 (B) | Ψηφιδωτά πραγματικών και προσομοιωμένων μέσω των τοπικών εκτιμηθέντων τιμών από την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης στο Υπόδειγμα Χωρικού Σφάλματος με Γκαουσιανή συνάρτηση στάθμισης για όλα τα μεγέθη δείγματος και τιμές του λ | |
| | Συντελεστής β_{i0} | 131 |
| Διάγραμμα 3.8 (A) | Ψηφιδωτά πραγματικών και προσομοιωμένων μέσω των τοπικών εκτιμηθέντων τιμών από την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης στο Υπόδειγμα Χωρικής Υστέρησης με Διτετράγωνη συνάρτηση στάθμισης για όλα τα μεγέθη δείγματος και τιμές του ρ | |
| | Συντελεστής β_{i1} | 135 |
| Διάγραμμα 3.8 (B) | Ψηφιδωτά πραγματικών και προσομοιωμένων μέσω των τοπικών εκτιμηθέντων τιμών από την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης στο Υπόδειγμα Χωρικής Υστέρησης με Διτετράγωνη συνάρτηση στάθμισης για όλα τα μεγέθη δείγματος και τιμές του ρ | |
| | Συντελεστής β_{i0} | 136 |
| Διάγραμμα 3.9 (A) | Ψηφιδωτά πραγματικών και προσομοιωμένων μέσω των τοπικών εκτιμηθέντων τιμών από την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης στο Υπόδειγμα Χωρικής Υστέρησης με Γκαουσιανή συνάρτηση στάθμισης για όλα τα μεγέθη δείγματος και τιμές του ρ | |
| | Συντελεστής β_{i1} | 137 |
| Διάγραμμα 3.9 (B) | Ψηφιδωτά πραγματικών και προσομοιωμένων μέσω των τοπικών εκτιμηθέντων τιμών από την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης στο Υπόδειγμα Χωρικής Υστέρησης με Γκαουσιανή συνάρτηση στάθμισης για όλα τα μεγέθη δείγματος και τιμές του ρ | |
| | Συντελεστής β_{i0} | 138 |
| Διάγραμμα 4.1 | Δείκτης Ολικής Γονιμότητας (TFR) ανά γυναίκα της Ελλάδας της περιόδου 1955 – 2010 | 156 |

| | | |
|-------------------|--|-----|
| Διάγραμμα 4.2 | Δείκτες Ολικής Γονιμότητας (TFR) ανά γυναίκα της Ελλάδας για τις κοορτες γεννήσεων των γυναικών της περιόδου 1940 – 1965, πηγή: Frejka and Sardon (2004) | 156 |
| Διάγραμμα 4.3 | Δείκτες Ολικής Γονιμότητας (TFR) ανά γυναίκα σε επίπεδο νομών της Ελλάδας για τα έτη 1981, 1991 και 2001 | 157 |
| Διάγραμμα 4.4 | Μεταβολές στους Δείκτες Ολικής Γονιμότητας (TFR) ανά γυναίκα σε επίπεδο νομών της Ελλάδας την περίοδο 1981 – 2001 | 158 |
| Διάγραμμα 4.5 | Χάρτες ποσοστιαίων σημείων: Χωρικές κατανομές των μεταβλητών του υποδείγματος | 166 |
| Διάγραμμα 4.6 | Χάρτης τοπικών αριθμών κατάστασης για την ύπαρξη τοπικής πολυσυγγραμμικότητας | 172 |
| Διάγραμμα 4.7 | Χάρτης τοπικών συντελεστών προσδιορισμού | 172 |
| Διάγραμμα 4.8 (Α) | Χάρτες τοπικών συντελεστών και ψευδό- <i>t</i> στατιστικών των τοπικών υποδειγμάτων Σταθερός Όρος | 174 |
| Διάγραμμα 4.8 (Β) | Χάρτες τοπικών συντελεστών και ψευδό- <i>t</i> στατιστικών των τοπικών υποδειγμάτων Ποσοστό Παντρεμένων Γυναικών | 175 |
| Διάγραμμα 4.8 (Γ) | Χάρτες τοπικών συντελεστών και ψευδό- <i>t</i> στατιστικών των τοπικών υποδειγμάτων Ποσοστό Μεταναστών | 176 |
| Διάγραμμα 4.8 (Δ) | Χάρτες τοπικών συντελεστών και ψευδό- <i>t</i> στατιστικών των τοπικών υποδειγμάτων Δείκτης Χαμηλών Κοινωνικών και Οικονομικών Συνθηκών (LSE) | 177 |
| Διάγραμμα 4.8 (Ε) | Χάρτες τοπικών συντελεστών και ψευδό- <i>t</i> στατιστικών των τοπικών υποδειγμάτων Ποσοστό Πυρηνικών Οικογενειών | 178 |

ΚΕΦΑΛΑΙΟ 1

ΧΩΡΙΚΑ ΔΕΔΟΜΕΝΑ ΚΑΙ ΧΩΡΙΚΗ ΟΙΚΟΝΟΜΕΤΡΙΑ

1.1 Εισαγωγή

Η ραγδαία βελτίωση της ισχύς των ηλεκτρονικών υπολογιστών που έχει επιτευχθεί τα τελευταία χρόνια και η ανάπτυξη των γεωγραφικών συστημάτων πληροφοριών έχουν συμβάλλει στην εμφάνιση και στην ανάγκη για μελέτη μιας νέας μορφής στατιστικών δεδομένων στην οποία τα παλιότερα χρόνια δεν είχε δοθεί η δέουσα προσοχή. Ο νέος τύπος δεδομένων αφορά τα χωρικά δεδομένα (*spatial data*) δηλαδή, διαστρωματικά δεδομένα που έχουν συλλεχθεί και προέρχονται από διαφορετικές γεωγραφικές περιοχές.

Η διαφορά των χωρικών δεδομένων από τα υπόλοιπα διαστρωματικά δεδομένα βρίσκεται στο γεγονός πώς κάθε παρατήρηση του δείγματος συνοδεύεται από γεωγραφική αναφορά ως προς τη γεωγραφική περιοχή που έχει συλλεχθεί. Πιο συγκεκριμένα, η παρατήρηση μπορεί να καταγράφεται μαζί με τις γεωγραφικές συντεταγμένες του σημείου του χώρου στο οποίο αναφέρεται ή να ανήκει σε κάποιο πολύγωνο ενός χάρτη. Ειδικότερα, θα μπορούσε να γίνει διάκριση μεταξύ δύο βασικών κατηγοριών δεδομένων. Η πρώτη κατηγορία αφορά τα ονομαζόμενα δεδομένα σημείου (*point data*) όπου κάθε παρατήρηση συλλέγεται σε κάποιο συγκεκριμένο σημείο του χώρου για το οποίο είναι διαθέσιμες και οι γεωγραφικές του συντεταγμένες. Η δεύτερη κατηγορία αφορά τα επιφανειακά ή περιφερειακά δεδομένα (*areal* ή *regional data*). Σε αυτήν την κατηγορία, τα δεδομένα ομαδοποιούνται και θεωρούνται ότι ανήκουν σε ένα πολύγωνο ενός χάρτη όπως είναι οι χώρες, οι νομοί, οι περιφέρειες ή ακόμη και τα απογραφικά τετράγωνα. Εάν υπάρχει η ανάγκη για αναφορά με γεωγραφικές συντεταγμένες τότε συνήθως χρησιμοποιούνται είτε οι συντεταγμένες κάποιας πόλης όπως η πρωτεύουσα της χώρας ή του νομού είτε, τις περισσότερες φορές, το γεωγραφικό κεντροειδές (*geographic centroid*) του πολυγώνου.

Για να γίνει κατανοητή η διάκριση μεταξύ των δύο κατηγοριών δεδομένων θα μπορούσε να αναφερθεί ως παράδειγμα των δεδομένων σημείου το ύψος της βροχόπτωσης και το επίπεδο της ατμοσφαιρικής ρύπανσης. Μια τέτοια επιστημονική μέτρηση θεωρητικά θα

μπορούσε να γίνει από τον ερευνητή με το κατάλληλο επιστημονικό όργανο σε οποιοδήποτε γεωγραφικό σημείο γνωρίζοντας ταυτόχρονα και τις γεωγραφικές του συντεταγμένες. Αντίστοιχα, για τα περιφερειακά δεδομένα παράδειγμα αποτελούν το ύψος της ανεργίας και ο αριθμός των γεννήσεων. Τα δεδομένα αυτά συλλέγονται μέσα σε μία γεωγραφική περιοχή, αθροίζονται και αποδίδονται σε αυτή, χωρίς να υπάρχει η δυνατότητα ταυτοποίησης γεωγραφικών συντεταγμένων για κάθε μεμονωμένη παρατήρηση. Κατά συνέπεια, θα μπορούσε να θεωρηθεί ότι στα δεδομένα σημείου ο γεωγραφικός χώρος υποτίθεται ότι είναι συνεχής ενώ στα περιφερειακά δεδομένα διακριτός. Επιπρόσθετα, είναι φανερό ότι τα περισσότερα οικονομικά και δημογραφικά δεδομένα ανήκουν στην κατηγορία των περιφερειακών δεδομένων.

Ο ιδιαίτερος τρόπος συλλογής των χωρικών δεδομένων έχει αποτέλεσμα την εμφάνιση των χωρικών επιδράσεων (*spatial effects*) που αποτελούνται από τη χωρική εξάρτηση (*spatial dependence*) και τη χωρική ετερογένεια (*spatial heterogeneity*). Χωρική εξάρτηση σημαίνει ότι υπάρχει γεωγραφική εξάρτηση μεταξύ των παρατηρήσεων του δείγματος με συνέπεια να μην είναι ανεξάρτητες αλλά μία παρατήρηση που βρίσκεται σε κάποια συγκεκριμένη περιοχή, όπως αυτή περιγράφεται από τις γεωγραφικές της συντεταγμένες, να συσχετίζεται με κάποια άλλη παρατήρηση που βρίσκεται σε άλλη περιοχή δημιουργώντας συστάδες (*clusters*) παρατηρήσεων. Για το λόγο αυτό ονομάζεται και χωρική αυτοσυσχέτιση παραπέμποντας στον αντίστοιχο όρο της ανάλυσης χρονοσειρών με τον οποίο όμως υπάρχει σημαντική διαφορά καθώς στην ανάλυση χρονοσειρών η εξάρτηση είναι προς μόνο μία κατεύθυνση σε αντίθεση με τη χωρική εξάρτηση που μπορεί να είναι προς πολλές κατευθύνσεις. Χωρική ετερογένεια σημαίνει ότι η σχέση ανάμεσα στις εξεταζόμενες μεταβλητές δεν παραμένει σταθερή σε ολόκληρη τη γεωγραφική περιοχή αλλά μεταβάλλεται από σημείο σε σημείο.

Η παρουσία των χωρικών επιδράσεων εάν αγνοηθεί σε μια στατιστική έρευνα έχει συνέπεια τα αποτελέσματα την ανάλυσης να είναι μη αξιόπιστα. Το πρόβλημα αυτό οδήγησε στην ανάπτυξη ενός καινούριου επιστημονικού τομέα, δηλαδή της χωρικής ανάλυσης (*spatial analysis*) που διαθέτει μεθοδολογικά εργαλεία για την ανάλυση χωρικών δεδομένων. Μέσα στον ευρύτερο τομέα της χωρικής ανάλυσης έχει αναπτυχθεί η χωρική στατιστική (*spatial statistics*) που προσπαθεί να ερμηνεύσει σχέσεις μεταξύ μεταβλητών λαμβάνοντας υπ' όψιν τις γεωγραφικές τους ιδιότητες, δηλαδή τις θέσεις που βρίσκονται οι στατιστικές μονάδες στο χώρο. Στην πραγματικότητα επεκτείνει τις καθιερωμένες τεχνικές της στατιστικής στη μελέτη

χωρικών δεδομένων αντιμετωπίζοντας τα προβλήματα που εμφανίζονται λόγω της ιδιομορφίας τους και ταυτόχρονα εκμεταλλεύεται τις γεωγραφικές τους ιδιότητες για την αποκάλυψη περαιτέρω χρήσιμων πληροφοριών.

Ιστορικά, η πρώτη εφαρμογή χωρικής ανάλυσης έγινε από το γιατρό John Snow το 1854 στο Λονδίνο. Συγκεκριμένα, στο Λονδίνο της Βικτωριανής εποχής υπήρχε έξαρση θανάτων από επιδημία χολέρας. Την εποχή εκείνη ήταν άγνωστες οι αιτίες που προκαλούν την ασθένεια και οι τρόποι με τους οποίους μεταδίδεται. Η κυριαρχούσα άποψη μεταξύ των επιστημόνων της εποχής ήταν ότι η αιτία μετάδοσης της νόσου ήταν αποτέλεσμα μόλυνσης στον αέρα της περιοχής. Ο John Snow (1855) πίστευε ότι το πρόβλημα οφειλόταν στην κατανάλωση μολυσμένου νερού. Για να στηρίξει την άποψή του κατασκεύασε ένα χάρτη της περιοχής όπου υπήρχαν τα κρούσματα της νόσου στον οποίο τοποθέτησε τα σημεία που εντοπίζονταν οι θάνατοι και τις αντλίες ύδρευσης που παρείχαν νερό. Με τον τρόπο αυτό ανακάλυψε ότι στο χάρτη υπήρχαν συμπλέγματα (*clusters*) συμβάντων κοντά σε συγκεκριμένες αντλίες ύδρευσης. Η εξέταση της ποιότητας του νερού από αυτές τις αντλίες αποκάλυψε ότι ήταν μολυσμένο με αποχετευτικά λύματα και η απομόνωσή τους από το δίκτυο ύδρευσης οδήγησε στην αντιμετώπιση της επιδημίας. Ο John Snow συνέλαβε την ιδέα και τη χρησιμότητα του γεωγραφικού συστήματος πληροφοριών (*GIS*) πολλά χρόνια νωρίτερα από την ανακάλυψη των ηλεκτρονικών υπολογιστών και παρατηρώντας την ύπαρξη χωρικής αυτοσυσχέτισης στα δεδομένα του πραγματοποίησε την πρώτη χωρική στατιστική ανάλυση ώστε δικαίως να του αποδίδεται ο χαρακτηρισμός “πατέρας της χωρικής ανάλυσης.”

Πέρασε περίπου ένας αιώνας από την εργασία του John Snow, με εξαίρεση κάποιες μεμονωμένες εργασίες, ώστε να αρχίσει να δίνεται έμφαση στις ιδιαιτερότητες των χωρικών δεδομένων και στην αναγκαιότητα ανάπτυξης εξειδικευμένων τεχνικών.¹ Ο Moran (1948 και 1950) με την παρουσίαση του δείκτη χωρικής αυτοσυσχέτισης *I* γνωστού με το όνομά του έδωσε την ώθηση για την ανάπτυξη της χωρικής στατιστικής. Στις αρχές του 1970 ο Βέλγος οικονομολόγος Jean Paelinck χρησιμοποίησε τον όρο χωρική οικονομετρία (*spatial econometrics*) και μίλησε για την ανάγκη δημιουργίας ενός νέου κλάδου στην οικονομετρία που να ασχολείται με τα προβλήματα που εμφανίζονται όταν εκτιμώνται οικονομετρικά υποδείγματα με περιφερειακά δεδομένα. Στο ομώνυμο βιβλίο (Paelinck και Klaassen, 1979) παρουσίασε τα πρώτα μεθοδολογικά εργαλεία και ακολούθησαν τα βιβλία των Cliff και Ord (1981) και του Anselin (1988). Το τελευταίο θέτει ολοκληρωμένα το μεθοδολογικό πλαίσιο

¹ Μια λεπτομερειακή ιστορική αναδρομή στην ανάπτυξη της χωρικής στατιστικής και της χωρικής οικονομετρίας μπορεί να βρεθεί στους Anselin και Griffith (1988) και Anselin (2010).

της χωρικής οικονομετρίας συγκεντρώνοντας μαζί εξειδικευμένους στατιστικούς ελέγχους και τεχνικές εκτίμησης υποδειγμάτων που βρίσκονταν σε διάφορες εργασίες με αποτέλεσμα να γίνει το βιβλίο σημείο αναφοράς κάθε εργασίας στη χωρική οικονομετρία.

Στο ίδιο βιβλίο ο Anselin αναφέρει τη διαφορά μεταξύ χωρικής στατιστικής και χωρικής οικονομετρίας.² Ειδικότερα, η χωρική οικονομετρία διακρίνεται από τη χωρική στατιστική για τους ίδιους λόγους που γίνεται η διάκριση οικονομετρίας και στατιστικής. Δηλαδή, η χωρική στατιστική οδηγείται από τα χωρικά δεδομένα (*data – driven orientation*) στα πλαίσια της διερευνητικής ανάλυσης δεδομένων (*Exploratory Data Analysis*) χωρίς να υπάρχει εκ των προτέρων κάποιο θεωρητικό υπόδειγμα που να ερμηνεύει τα αποτελέσματα σε αντίθεση με τη χωρική οικονομετρία που υπάρχει το θεωρητικό υπόδειγμα (*model – driven approach*) που πρέπει να εκτιμηθεί και να ελεγχθεί χρησιμοποιώντας τα χωρικά δεδομένα. Όμως, ενώ η χωρική στατιστική χρησιμοποιείται κυρίως για τη μελέτη φυσικών φαινομένων και η χωρική οικονομετρία για την ανάλυση οικονομικών δεδομένων οι μεθοδολογίες τους πολλές φορές αλληλοκαλύπτονται με αποτέλεσμα να μη γίνεται πάντα ορατή η διάκρισή τους.

Στο κεφάλαιο αυτό παρουσιάζονται περιληπτικά βασικές έννοιες και θέματα που αφορούν τη λειτουργία της χωρικής οικονομετρίας. Πιο συγκεκριμένα, εξηγούνται οι ιδιαιτερότητες των χωρικών δεδομένων και η μήτρα χωρικών σταθμίσεων, ο τρόπος ελέγχου για την ύπαρξη χωρικής αυτοσυσχέτισης σε ένα δείγμα, οι βασικές χωρικές διαδικασίες και τα σπουδαιότερα χωρικά οικονομετρικά υποδείγματα. Σκοπός της ανάλυσης είναι να τεθούν οι βάσεις για την υποστήριξη και κατανόηση των επόμενων κεφαλαίων.

1.2 Οι χωρικές επιδράσεις

Το βασικό χαρακτηριστικό των χωρικών δεδομένων το οποίο τα διαφοροποιεί από τα άλλα είδη στατιστικών δεδομένων που χρησιμοποιούνται σε μια οικονομετρική ανάλυση είναι η ύπαρξη των χωρικών επιδράσεων. Οι χωρικές επιδράσεις (*spatial effects*) προέρχονται από τη φύση των χωρικών δεδομένων και διακρίνονται στη χωρική εξάρτηση (*spatial dependence*) και στη χωρική ετερογένεια (*spatial heterogeneity*).

Με τον όρο χωρική εξάρτηση (*spatial dependence*) εννοείται ότι υπάρχει κάποιας μορφής σχέση στη γεωγραφική περιοχή που μελετάται ένα φαινόμενο έτσι ώστε κάτι που συμβαίνει

² Βλέπε Anselin (1988) σελίδα 10.

σε ένα γεωγραφικό σημείο να επηρεάζει το αποτέλεσμα σε ένα άλλο γεωγραφικό σημείο (Anselin, 1988). Η έννοια της χωρικής εξάρτησης, κατά κάποιον τρόπο, εκφράζεται από το γνωστό ως πρώτο νόμο της γεωγραφίας του Tobler (1970) σύμφωνα με τον οποίον “καθετί στο χώρο συσχετίζεται με κάτι άλλο αλλά τα κοντινότερα πράγματα έχουν μεγαλύτερη συσχέτιση από αυτά που είναι πιο απομακρυσμένα.” Μια αιτία που οδηγεί σε χωρική εξάρτηση είναι η ύπαρξη σφάλματος μέτρησης στις τιμές της εξεταζόμενης μεταβλητής. Σφάλμα μέτρησης σε μια χωρική μεταβλητή μπορεί να προκύψει όταν η χωρική μονάδα στην οποία παρατηρείται ένα φαινόμενο δε συμπίπτει με τον πραγματικό χώρο στον οποίον στην πραγματικότητα συμβαίνει. Παραδείγματος χάριν, όταν μελετάται η αγορά εργασίας τα διαθέσιμα στοιχεία απασχόλησης, ανεργίας και εισοδήματος μπορεί να συλλέγονται με βάση τους νομούς ή τα απογραφικά πολύγωνα αλλά η πραγματική οικονομική δραστηριότητα μίας χωρικής μονάδας να επεκτείνεται και σε γειτονικές χωρικές μονάδες. Δηλαδή, σε αυτή την περίπτωση, η χωρική εξάρτηση είναι αποτέλεσμα του τρόπου συλλογής των στατιστικών πληροφοριών, συνέπεια μη συμφωνίας των γεωγραφικών συνόρων με τα πραγματικά γεωγραφικά όρια στα οποία εξελίσσεται το φαινόμενο.

Μια δεύτερη αιτία που προκαλεί χωρική εξάρτηση οφείλεται στην αλληλεπίδραση (*spatial interaction*) μεταξύ των χωρικών μονάδων. Αυτό μπορεί να συμβεί όταν, παραδείγματος χάριν, υπάρχει ανταγωνισμός στις επιχειρηματικές δραστηριότητες ανάμεσα σε γειτονικές χωρικές μονάδες έτσι ώστε όταν ένας επιχειρηματίας προβαίνει στη μείωση της τιμής ενός προϊόντος σε μια περιοχή οι επιχειρηματίες στις γειτονικές περιοχές προβαίνουν στην ίδια ενέργεια με αποτέλεσμα κάτι που παρατηρείται σε ένα γεωγραφικό σημείο να οφείλεται σε κάτι που έχει γίνει σε κάποιο άλλο γεωγραφικό σημείο. Με αυτή τη διαδικασία προκύπτει ένα αποτέλεσμα διάχυσης (*spillover effect*) που επηρεάζει όλες τις χωρικές μονάδες. Η αιτία αυτή χωρικής εξάρτησης είναι πιο ουσιαστική διότι υποδηλώνει την ύπαρξη κάποιας χωρικής διαδικασίας (*spatial process*) που οδηγεί στην εξάρτηση η οποία πρέπει να ληφθεί υπ’ όψιν από τον ερευνητή όταν εξειδικεύει ένα οικονομετρικό υπόδειγμα για την ερμηνεία του φαινομένου.

Η πλήρη μελέτη του χαρακτηριστικού της χωρικής εξάρτησης απαιτεί τη γνώση της από κοινού συνάρτησης κατανομής της εξεταζόμενης μεταβλητής για όλα τα γεωγραφικά σημεία στα οποία παρατηρείται. Επειδή είναι δύσκολη η γνώση της από κοινού συνάρτησης κατανομής συνήθως η χωρική εξάρτηση μελετάται στην ασθενέστερή της μορφή που είναι η χωρική αυτοσυσχέτιση για την οποία χρειάζονται μόνο οι ροπές της κατανομής. Κατά

συνέπεια, μια τυχαία μεταβλητή y θα παρουσιάζει χωρική αυτοσυσχέτιση όταν ικανοποιείται η συνθήκη:

$$Cov(y_i, y_j) = E(y_i, y_j) - E(y_i)E(y_j) \neq 0, \quad \text{για } i \neq j$$

όπου i και j είναι τα σημεία δύο περιοχών στον εξεταζόμενο γεωγραφικό χώρο και y_i και y_j οι τιμές της μεταβλητής σε αυτές τις περιοχές. Αποτέλεσμα της ύπαρξης χωρικής αυτοσυσχέτισης είναι η δημιουργία συστάδων (*clusters*) των τιμών των παρατηρήσεων και όχι η τυχαία κατανομή τους επάνω σε ένα γεωγραφικό χάρτη. Ειδικότερα, όταν οι υψηλές τιμές της μεταβλητής εντοπίζονται κοντά σε υψηλές τιμές και οι χαμηλές τιμές γειτονεύουν με χαμηλές τιμές σχηματίζοντας συστάδες υψηλών και χαμηλών τιμών, τότε υπάρχει θετική χωρική αυτοσυσχέτιση. Αντίθετα, όταν οι υψηλές τιμές της μεταβλητής βρίσκονται κοντά σε χαμηλές και οι χαμηλές τιμές κοντά σε υψηλές τότε υπάρχει αρνητική χωρική αυτοσυσχέτιση. Υπάρχουν ομοιότητες με τη χρονική αυτοσυσχέτιση που εμφανίζεται στην ανάλυση χρονοσειρών όμως στις χρονοσειρές η σχέση εξάρτησης είναι μόνο προς μία κατεύθυνση (η προηγούμενη χρονικά τιμή της χρονοσειράς επηρεάζει την επόμενη) ενώ στο χώρο προς πολλές κατευθύνσεις.

Η δεύτερη μορφή χωρικών επιδράσεων είναι η χωρική ετερογένεια (*spatial heterogeneity*). Ο όρος χωρική ετερογένεια αναφέρεται στην έλλειψη σταθερότητας στη σχέση μεταξύ των μεταβλητών στις διαφορετικές γεωγραφικές περιοχές. Το φαινόμενο αυτό είναι πολύ συνηθισμένο όταν μελετώνται οικονομικές και δημογραφικές μεταβλητές και οφείλεται στις πληθυσμιακές και εισοδηματικές ανισότητες μεταξύ των περιοχών, στη διαφορετική έκταση που έχουν οι απογραφικές περιοχές και γενικότερα στην ανομοιογένεια των μεταβλητών στο χώρο. Στην πιο σύνθετη περίπτωση, η ένταση αλλά και η συναρτησιακή εξάρτηση της σχέσης μεταξύ δύο ή περισσότερων μεταβλητών μπορεί να μην είναι ίδιες σε όλες τις περιοχές. Στα πλαίσια ενός υποδείγματος παλινδρόμησης η χωρική ετερογένεια θα μπορούσε να παρουσιαστεί με τη σχέση:

$$y_i = f_i(x_{1i}, x_{2i}, \dots, x_{ki}, \beta_{1i}, \beta_{2i}, \dots, \beta_{ki}, \varepsilon_i)$$

όπου y είναι η εξαρτημένη μεταβλητή, x_1, x_2, \dots, x_k οι ανεξάρτητες μεταβλητές, $\beta_1, \beta_2, \dots, \beta_k$ οι συντελεστές του υποδείγματος και ε_i τα σφάλματα στην i περιοχή. Στην πιο απλή περίπτωση, η χωρική ετερογένεια παρουσιάζεται μέσω του προβλήματος της ετεροσκεδαστικότητας στα εκτιμηθέντα σφάλματα ενός υποδείγματος παλινδρόμησης.

Στο σημείο αυτό πρέπει να αναφερθεί ότι δεν είναι υποχρεωτικό να εμφανίζονται ταυτόχρονα και οι δύο μορφές των χωρικών επιδράσεων σε μια ανάλυση. Επιπρόσθετα, λόγω του γεγονότος ότι τα χωρικά δεδομένα είναι διαστρωματικά δεδομένα πολλές φορές δεν γίνεται ξεκάθαρο ποια μορφή χωρικών επιδράσεων εμφανίζεται. Από τις δύο μορφές χωρικών επιδράσεων η χωρική εξάρτηση είναι η πιο δύσκολη να αντιμετωπιστεί και εάν αγνοηθεί μπορεί να έχει δραματικές συνέπειες στην αξιοπιστία των αποτελεσμάτων μίας οικονομετρικής ανάλυσης. Στην κατεύθυνση αντιμετώπισης των προβλημάτων που εμφανίζονται από τη χωρική εξάρτηση συμβάλλουν αποφασιστικά οι εξειδικευμένες τεχνικές της χωρικής οικονομετρίας. Η δεύτερη μορφή χωρικών επιδράσεων, δηλαδή η χωρική ετερογένεια, είναι πιο εύκολο να αντιμετωπιστεί στα πλαίσια των μεθόδων της κλασσικής οικονομετρίας, ιδιαίτερα όταν εμφανίζεται το πρόβλημα της ετεροσκεδαστικότητας. Εντούτοις, και σε αυτή την περίπτωση η χωρική οικονομετρία μπορεί να βοηθήσει στο να προκύψουν πιο αξιόπιστα αποτελέσματα και περισσότερες πληροφορίες στον ερευνητή.

1.3 Χωρική γειτονία και μήτρες χωρικών σταθμίσεων

Το πρώτο στάδιο σε κάθε ανάλυση χωρικών δεδομένων είναι ο καθορισμός της χωρικής γειτονίας μεταξύ των γεωγραφικών περιοχών που εμπλέκονται στην ανάλυση. Ο ορισμός της χωρικής γειννίας είναι απαραίτητος για να οριστεί η δομή της χωρικής εξάρτησης που ενσωματώνεται σε κάθε χωρικό οικονομετρικό υπόδειγμα και γίνεται από τον αναλυτή εξωγενώς. Ο τρόπος ορισμού της χωρικής εξάρτησης πολλές φορές επηρεάζει σημαντικά τα αποτελέσματα της ανάλυσης χωρίς να υπάρχει κάποιο συγκεκριμένο στατιστικό κριτήριο για την ορθότητα της επιλογής παρά μόνο η προηγούμενη εμπειρία του αναλυτή από παλιότερες αναλύσεις στην ίδια γεωγραφική περιοχή.

Αυστηρά στατιστικά, η χωρική γειτονία ορίζεται σύμφωνα με την από κοινού και την υπό συνθήκη συναρτήσεις πυκνότητας του συνόλου των τυχαίων μεταβλητών που παρατηρούνται στην εξεταζόμενη γεωγραφική περιοχή. Σύμφωνα με τον Anselin (1988), όταν μια γεωγραφική περιοχή αποτελείται από n χωρικές μονάδες στις οποίες παρατηρείται μια τυχαία μεταβλητή y , τότε το σύνολο των γειτόνων \mathbf{J} για την i μονάδα ορίζεται ως οι χωρικές μονάδες j για τις οποίες η y_j περιλαμβάνεται στη συνάρτηση πιθανότητας της y_i υπό συνθήκη το διάνυσμα \mathbf{y} σε όλες τις άλλες περιοχές, δηλαδή όταν ισχύει ότι:

$$P[y_i | \mathbf{y}] = P[y_i | \mathbf{y}_J]$$

όπου \mathbf{y} είναι το $(n \times 1)$ διάνυσμα όλων των τιμών της y και \mathbf{y}_J είναι ένα διάνυσμα που αποτελείται από ένα υποσύνολο των παρατηρήσεων για την y_j για τις οποίες ισχύει ότι $j \in \mathbf{J}$. Κατά συνέπεια, η υπό συνθήκη κατανομή της y_i δεδομένων όλων των άλλων μεταβλητών είναι η ίδια με την υπό συνθήκη κατανομή δεδομένου του υποσυνόλου των γειτόνων. Ο ορισμός της γειτονίας δεν είναι εύκολο να γίνει με τις από κοινού και υπό συνθήκη συναρτήσεις κατανομής και για το λόγο αυτό ορίζεται με τη βοήθεια των ροπών που υπολογίζονται μέσω του διαθέσιμου δείγματος. Με βάση τις ροπές, οι γείτονες της χωρικής μονάδας i θα αποτελούνται από το σύνολο $j \in \mathbf{J}$ για το οποίο ισχύει ότι:

$$\text{Cov}(y_i, y_j) \neq 0, \quad \text{για } i \neq j$$

Όπως είναι φανερό, ο υπολογισμός των συνδιακυμάνσεων δεν είναι δυνατός, καθώς για κάθε χωρική μονάδα υπάρχει μόνο μία διαθέσιμη παρατήρηση και οι συνολικές συνδιακυμάνσεις που πρέπει να υπολογιστούν είναι $n \times (n-1)/2$. Δηλαδή, ακόμη και εάν χρησιμοποιούταν ολόκληρο το δείγμα όταν $n > 3$, οι συνδιακυμάνσεις είναι περισσότερες από τις υπάρχουσες παρατηρήσεις. Υπό προϋποθέσεις, το πρόβλημα θα μπορούσε να επιλυθεί στην περίπτωση που αναλύονταν δεδομένα πάνελ. Στην πράξη, αυτό που συνήθως γίνεται είναι ο καθορισμός της χωρικής γειτονίας με την εφαρμογή μίας μήτρας χωρικών σταθμίσεων (*spatial weights matrix*) που δηλώνει την αλληλεπίδραση ανάμεσα στις διαφορετικές χωρικές μονάδες. Ειδικότερα, η μήτρα χωρικών σταθμίσεων είναι η $(n \times n)$ θετικά ορισμένη και συμμετρική μήτρα \mathbf{W} :

$$\mathbf{W} = \begin{bmatrix} w_{11} & w_{12} & w_{13} & \dots & w_{1j} & \dots & w_{1n} \\ w_{21} & w_{22} & w_{23} & \dots & w_{2j} & \dots & w_{2n} \\ w_{31} & w_{32} & w_{33} & \dots & w_{3j} & \dots & w_{3n} \\ \vdots & \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ w_{i1} & w_{i2} & w_{i3} & \dots & w_{ij} & \dots & w_{in} \\ \vdots & \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ w_{n1} & w_{n2} & w_{n3} & \dots & w_{nj} & \dots & w_{nn} \end{bmatrix}$$

για την οποία το στοιχείο w_{ij} έχει μη μηδενική τιμή όταν οι χωρικές μονάδες i και j θεωρούνται γειτονικές και μηδενική τιμή όταν δεν γειτονεύουν. Επιπρόσθετα, θα ισχύει ότι $w_{ii} = 0$, δηλαδή η διαγώνιος της μήτρας θα αποτελείται από μηδενικές τιμές, έτσι ώστε να

ισχύει $tr[\mathbf{W}] = 0$ εννοώντας ότι μια χωρική μονάδα δεν μπορεί να γειτονεύει με τον εαυτό της.

Για να κατασκευαστεί μια μήτρα χωρικών σταθμίσεων επιλέγεται από τον ερευνητή κάποιο κριτήριο χωρικής γειτονίας και με την εφαρμογή του κριτηρίου εντοπίζονται για κάθε χωρική μονάδα όλες οι γειτονικές της μονάδες. Στην αρχική της μορφή, η μήτρα χωρικών σταθμίσεων περιέχει τις τιμές ένα και μηδέν και αποτελεί μια μήτρα δυικής χωρικής γειτονίας (*binary contiguity matrix*) \mathbf{B} με στοιχεία που ορίζονται ως:

$$w'_{ij} = \begin{cases} 1 & \text{εάν οι μονάδες } i \text{ και } j \text{ είναι γείτονες} \\ 0 & \text{διαφορετικά} \end{cases}$$

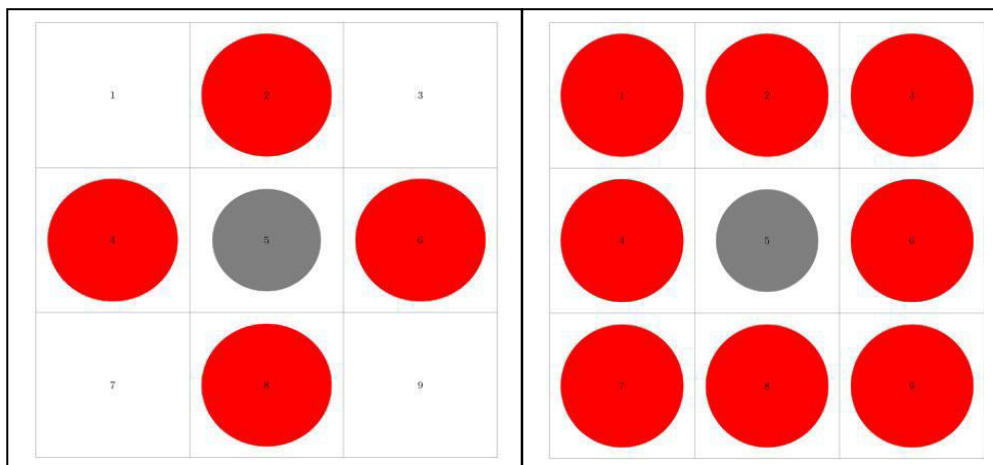
Στη συνέχεια, κατασκευάζεται η μήτρα χωρικών σταθμίσεων τυποποιώντας τα στοιχεία κάθε γραμμής στη μήτρα \mathbf{B} ώστε να αθροίζονται στη μονάδα σύμφωνα με τη σχέση:

$$w_{ij} = \frac{w'_{ij}}{\sum_{j=1}^n w'_{ij}}$$

και να εξασφαλίζεται ότι για κάθε στοιχείο στη μήτρα χωρικών σταθμίσεων θα ισχύει $0 \leq w_{ij} \leq 1$. Με την τυποποίηση αυτή επιτυγχάνεται η χρησιμοποίηση της μήτρας χωρικών σταθμίσεων για τον υπολογισμό χωρικών υστερήσεων για τις τιμές μιας μεταβλητής. Η χωρική υστέρηση (*spatial lag*) συσχετίζει την τιμή μιας μεταβλητής σε μία συγκεκριμένη χωρική μονάδα με τις τιμές της στις γειτονικές της μονάδες. Ουσιαστικά για κάθε χωρική μονάδα αποτελεί ένα σταθμισμένο μέσο όρο των τιμών της μεταβλητής στις γειτονικές της χωρικές μονάδες με συντελεστές στάθμισης τα αντίστοιχα στοιχεία της μήτρας χωρικών σταθμίσεων. Ο υπολογισμός του $(n \times 1)$ διανύσματος $\mathbf{W}\mathbf{y}$ των χωρικών υστερήσεων για τη μεταβλητή \mathbf{y} γίνεται πολλαπλασιάζοντας τη μήτρα χωρικών σταθμίσεων \mathbf{W} με το διάνυσμα της μεταβλητής. Από τη διαδικασία αυτή συμπεραίνεται ότι η χωρική υστέρηση της \mathbf{y} στην i χωρική μονάδα θα δίνεται από το άθροισμα $\sum_{j=1}^n w_{ij} y_j$. Η χωρική υστέρηση θυμίζει τη γνωστή

από την ανάλυση χρονοσειρών χρονική υστέρηση (*time lag*) και κατά κάποιον τρόπο επιτρέπει τη μετακίνηση μεταξύ των χωρικών μονάδων. Η ουσιαστική διαφορά που υπάρχει είναι ότι στις χρονοσειρές η κίνηση που μπορεί να γίνει είναι χρονικά μόνο προς τα πίσω ή προς τα εμπρός, ενώ στο χώρο οι κινήσεις είναι προς πολλές κατευθύνσεις.

Όπως αναφέρθηκε, για τον ορισμό της χωρικής γειτνίασης με τη βοήθεια της μήτρας χωρικών σταθμίσεων είναι αναγκαία η υιοθέτηση κάποιου κριτηρίου χωρικής γειτονίας. Τα κριτήρια χωρικής γειτονίας αποτελούν ένα σημαντικό ερευνητικό τομέα στη χωρική ανάλυση και απασχολούν μεγάλο μέρος της βιβλιογραφίας. Ένα από τα πρώτα κριτήρια που έχουν προταθεί βασίζεται στην ύπαρξη κοινών φυσικών γεωγραφικών συνόρων μεταξύ των χωρικών μονάδων θεωρώντας ότι δύο περιοχές στο γεωγραφικό χάρτη είναι γείτονες εάν εφάπτονται τα σύνορά τους. Η προσέγγιση αυτή είναι ιδιαίτερα εύχρηστη όταν οι εξεταζόμενες χωρικές μονάδες αποτελούν στοιχεία ενός ομαλού πλέγματος (*regular grid*) και αντιγράφοντας τις κινήσεις των πιονιών επάνω σε μια σκακιέρα ορίζει τη γειτονία με δυικό τρόπο. Ειδικότερα, επάνω σε ένα ομαλό πλέγμα μπορεί να οριστεί το κριτήριο του πύργου (*rook contiguity*) σύμφωνα με το οποίο δύο περιοχές θα είναι γείτονες εάν έχουν κοινή ακμή (*common edge*) και το κριτήριο της βασίλισσας (*queen contiguity*) που ορίζει δύο περιοχές ως γείτονες εάν έχουν κοινή ακμή (*common edge*) ή κοινή κορυφή (*common vertex*). Με τον ίδιο τρόπο ορίζεται και το κριτήριο του αξιωματικού (*bishop contiguity*) όπου δύο περιοχές θα είναι γείτονες εάν έχουν κοινή κορυφή. Επειδή το κριτήριο αυτό σπάνια ανταποκρίνεται στις συνθήκες ενός χάρτη δεν χρησιμοποιείται συχνά και δεν θα γίνει αντικείμενο περεταίρω συζήτησης.



Διάγραμμα 1.1
Γειτνίαση πύργου (αριστερά) και γειτνίαση βασίλισσας (δεξιά) σε ένα ομαλό τετραγωνικό πλέγμα εννέα χωρικών μονάδων

Στο Διάγραμμα 1.1 παρουσιάζονται για ένα ομαλό τετραγωνικό πλέγμα (*regular grid*) διαστάσεων 3x3 που αποτελείται από εννέα χωρικές μονάδες οι εφαρμογές των κριτηρίων

του πύργου και της βασίλισσας. Όπως φαίνεται στο αριστερό πλέγμα του διαγράμματος, η μεσαία χωρική μονάδα θα έχει γείτονες, σύμφωνα με το κριτήριο του πύργου, τις χωρικές μονάδες 2, 4, 6 και 8, δηλαδή τις χωρικές μονάδες με τις οποίες έχει κοινή ακμή. Αντίστοιχα, στο αριστερό πλέγμα του διαγράμματος φαίνεται ότι με βάση το κριτήριο της βασίλισσας η μεσαία χωρική μονάδα θα έχει γείτονες τις χωρικές μονάδες με τις οποίες έχει κοινή ακμή ή κοινή κορυφή, δηλαδή τις μονάδες 1, 2, 3, 4, 6, 7, 8 και 9. Κατά συνέπεια, διαπιστώνεται ότι σε ένα ομαλό πλέγμα κάθε μη ακραία χωρική μονάδα θα έχει τέσσερις γείτονες όταν εφαρμόζεται το κριτήριο του πύργου και οκτώ γείτονες όταν εφαρμόζεται το κριτήριο της βασίλισσας. Αυτό δεν ισχύει για τις χωρικές μονάδες που βρίσκονται στις άκρες του πλέγματος κάτι που μπορεί να προκαλέσει κάποιου είδους μεροληψία ιδιαίτερα σε δείγματα που βασίζονται σε μικρό αριθμό χωρικών μονάδων. Για το λόγο αυτό, σε θεωρητικές εφαρμογές (κυρίως προσομοιώσεις) για να αντιμετωπιστούν οι πλευρικές επιδράσεις (*edge effects*) εφαρμόζεται η λεγόμενη διόρθωση τόρου (*torus correction*) με την οποία οι χωρικές μονάδες της πρώτης στήλης θεωρούνται ότι συνορεύουν με τις αντίστοιχες χωρικές μονάδες της τελευταίας στήλης και οι χωρικές μονάδες της πρώτης γραμμής με τις αντίστοιχες χωρικές μονάδες της τελευταίας γραμμής ώστε σε κάθε μονάδα του πλέγματος να υπάρχει σταθερός αριθμός γειτόνων.

Η μήτρα δυικής χωρικής γειτονίας \mathbf{B} με την εφαρμογή του κριτηρίου του πύργου για το πλέγμα των εννέα χωρικών μονάδων προκύπτει βάζοντας στα στοιχεία του πίνακα που εκφράζουν τις χωρικές μονάδες με κοινή ακμή τη μονάδα και το μηδέν σε διαφορετική περίπτωση, δηλαδή:

$$\mathbf{B} = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$

Τυποποιώντας κατά γραμμή τα στοιχεία της προηγούμενης μήτρας υπολογίζεται η αντίστοιχη μήτρα χωρικών σταθμίσεων \mathbf{W} , που θα είναι:

$$\mathbf{W} = \begin{bmatrix}
 0,00 & 0,50 & 0,00 & 0,50 & 0,00 & 0,00 & 0,00 & 0,00 & 0,00 \\
 0,33 & 0,00 & 0,33 & 0,00 & 0,33 & 0,00 & 0,00 & 0,00 & 0,00 \\
 0,00 & 0,50 & 0,00 & 0,00 & 0,00 & 0,50 & 0,00 & 0,00 & 0,00 \\
 0,33 & 0,00 & 0,00 & 0,00 & 0,33 & 0,00 & 0,33 & 0,00 & 0,00 \\
 0,00 & 0,25 & 0,00 & 0,25 & 0,00 & 0,25 & 0,00 & 0,25 & 0,00 \\
 0,00 & 0,00 & 0,33 & 0,00 & 0,33 & 0,00 & 0,00 & 0,00 & 0,33 \\
 0,00 & 0,00 & 0,00 & 0,50 & 0,00 & 0,00 & 0,00 & 0,50 & 0,00 \\
 0,00 & 0,00 & 0,00 & 0,00 & 0,33 & 0,00 & 0,33 & 0,00 & 0,33 \\
 0,00 & 0,00 & 0,00 & 0,00 & 0,00 & 0,50 & 0,00 & 0,50 & 0,00
 \end{bmatrix}$$

Με παρόμοιο τρόπο παράγονται η μήτρα δυικής χωρικής γειτονίας και η μήτρα χωρικών σταθμίσεων με την εφαρμογή του κριτηρίου της βασίλισσας. Οι μήτρες αυτές είναι, αντίστοιχα:

$$\mathbf{B} = \begin{bmatrix}
 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\
 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\
 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\
 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 \\
 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\
 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 \\
 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0
 \end{bmatrix}$$

και

$$\mathbf{W} = \begin{bmatrix}
 0,00 & 0,33 & 0,00 & 0,33 & 0,33 & 0,00 & 0,00 & 0,00 & 0,00 \\
 0,20 & 0,00 & 0,20 & 0,20 & 0,20 & 0,20 & 0,00 & 0,00 & 0,00 \\
 0,00 & 0,33 & 0,00 & 0,00 & 0,33 & 0,33 & 0,00 & 0,00 & 0,00 \\
 0,20 & 0,20 & 0,00 & 0,00 & 0,20 & 0,00 & 0,20 & 0,20 & 0,00 \\
 0,12 & 0,12 & 0,12 & 0,12 & 0,00 & 0,12 & 0,12 & 0,12 & 0,12 \\
 0,00 & 0,20 & 0,20 & 0,00 & 0,20 & 0,00 & 0,00 & 0,20 & 0,20 \\
 0,00 & 0,00 & 0,00 & 0,33 & 0,33 & 0,00 & 0,00 & 0,33 & 0,00 \\
 0,00 & 0,00 & 0,00 & 0,20 & 0,20 & 0,20 & 0,20 & 0,00 & 0,20 \\
 0,00 & 0,00 & 0,00 & 0,00 & 0,33 & 0,33 & 0,00 & 0,33 & 0,00
 \end{bmatrix}$$

Όπως συμπεραίνεται, οι μήτρες που προκύπτουν από την εφαρμογή του κριτηρίου της βασίλισσας είναι πιο πυκνές σε σχέση με το κριτήριο του πύργου δηλώνοντας ότι η εξάρτηση μεταξύ των χωρικών μονάδων του πλέγματος είναι πιο έντονη. Τα κριτήρια του πύργου και

της βασίλισσας πολλές φορές εφαρμόζονται με τον ίδιο τρόπο και στις πραγματικές γεωγραφικές περιοχές (πολύγωνα) ενός χάρτη. Η αξιοπιστία τους όμως εξαρτάται κάθε φορά από τις γεωγραφικές ιδιαιτερότητες της περιοχής που μελετάται. Πιο συγκεκριμένα, μπορεί να προκύψουν χωρικές μονάδες με πολύ μικρό αριθμό γειτόνων και άλλες με πολύ μεγάλο αριθμό γειτόνων. Σε περιπτώσεις που η γεωγραφική περιοχή περιλαμβάνει μονάδες χωρίς φυσικά σύνορα όπως, παραδείγματος χάριν, συμβαίνει με τα νησιά τότε αυτές οι μονάδες δεν θα έχουν κανένα γείτονα. Επίσης, εάν σε κάποιες περιοχές του χάρτη υπάρχουν πολλές χωρικές μονάδες με μικρή γεωγραφική έκταση η κάθε μία, όπως για παράδειγμα σε κάποια πυκνοκατοικημένη έκταση σαν το νομό Αττικής, τότε θα είναι αναμφίβολο εάν τα αποτελέσματα των κριτηρίων μπορούν να ανταποκριθούν στις πραγματικές σχέσεις εξάρτησης στην περιοχή.

Παρά την ευκολία των κριτηρίων που βασίζονται στην ύπαρξη κοινών γεωγραφικών συνόρων μεταξύ των χωρικών μονάδων τα μειονεκτήματά τους έχουν οδηγήσει στη ανάπτυξη εναλλακτικών κριτηρίων. Το κριτήριο των k κοντινών γειτόνων θεωρεί για κάθε χωρική μονάδα ως γείτονες τις k μονάδες που βρίσκονται πιο κοντά σε αυτή ανεξάρτητα της θέσης τους και της απόστασής τους, δηλαδή:

$$w'_{ij} = \begin{cases} 1 & \text{εάν η μονάδα } j \text{ είναι ένας από τους } k \text{ κοντινούς γείτονες της } i \\ 0 & \text{διαφορετικά} \end{cases}$$

Το αποτέλεσμα αυτού του κριτηρίου είναι όλες οι μονάδες να έχουν τον ίδιο αριθμό γειτόνων και να μην εμφανίζονται μονάδες χωρίς κανένα γείτονα αλλά οι μήτρες χωρικών γειτόνων που κατασκευάζονται μπορεί να είναι μη συμμετρικές. Αντί για αριθμό κοντινών γειτόνων θα μπορούσε να οριστεί μια γεωγραφική απόσταση εντός της οποίας κάθε χωρική μονάδα j να θεωρείται γείτονας της i , δηλαδή:

$$w'_{ij} = \begin{cases} 1 & \text{εάν } d_{i,j} < d \\ 0 & \text{διαφορετικά} \end{cases}$$

όπου με $d_{i,j}$ συμβολίζεται η απόσταση μεταξύ των μονάδων i και j και με d η απόσταση σύγκρισης. Το κριτήριο αυτό οδηγεί σε μήτρες συμμετρικές αλλά με μη σταθερό αριθμό γειτόνων. Για να αντιμετωπιστεί το πρόβλημα εμφάνισης μονάδων με απουσία γειτόνων, ο ερευνητής θα πρέπει να καθορίσει την απόσταση σύγκρισης d σε τέτοιο σημείο ώστε να εξασφαλίζεται ένας γείτονας τουλάχιστον σε κάθε μονάδα κάτι που όμως μπορεί να εμφανίσει χωρικές μονάδες με εξαιρετικά μεγάλο αριθμό γειτόνων.

Μια άλλη κατηγορία κριτηρίων χωρική γειτονίας ορίζει τις σταθμίσεις ως αποτέλεσμα κάποιας συνάρτησης της απόστασης μεταξύ των χωρικών μονάδων παρακάμπτοντας το στάδιο ορισμού της μήτρας δυικής χωρικής γειτονίας. Με τον τρόπο αυτό, οι χωρικές μονάδες που είναι πιο απομακρυσμένες από την i μονάδα θα λαμβάνουν μικρότερη στάθμιση από αυτές που βρίσκονται πιο κοντά σε αυτή αντανακλώντας σε κάποιο βαθμό τον πρώτο νόμο της γεωγραφίας του Tobler (1970). Οι προκύπτουσες μήτρες χωρικών σταθμίσεων μπορούν και σε αυτή την περίπτωση να τυποποιηθούν κατά γραμμή. Για παράδειγμα, το w_{ij} στοιχείο της μήτρας χωρικών σταθμίσεων, όπως αναφέρουν οι Anselin και Bera (1998), μπορεί να είναι αποτέλεσμα εφαρμογής της συνάρτησης:

$$w_{ij} = \frac{1}{d_{i,j}^\alpha}$$

όπου $d_{i,j}$ είναι η απόσταση μεταξύ των μονάδων i και j και α είναι ένας συντελεστής που επιλέγεται από τον ερευνητή για να εκφράζει τη μείωση της εξάρτησης μεταξύ των μονάδων καθώς αυξάνεται η γεωγραφική απόσταση. Εναλλακτικά, θα μπορούσε να χρησιμοποιηθεί η αντίστροφη εκθετική συνάρτηση που ορίζεται ως:

$$w_{ij} = e^{-\frac{d_{i,j}}{a}}$$

Οι Cliff και Ord (1981) έχουν προτείνει μια συνάρτηση στάθμισης που συγκρίνει το σχετικό μήκος των κοινών συνόρων που μοιράζονται δύο μονάδες με την αντίστροφη απόστασή τους, δηλαδή:

$$w_{ij} = \frac{\beta_{i,j}^b}{d_{i,j}^\alpha}$$

όπου $\beta_{i,j}$ είναι η αναλογία των κοινών συνόρων της j με την i μονάδα ως προς τη συνολική περίμετρο των συνόρων της i , $d_{i,j}$ η απόσταση μεταξύ των μονάδων και α και b συντελεστές που ορίζονται από τον ερευνητή. Η επιλογή θετικών τιμών για τις δύο παραμέτρους έχει αποτέλεσμα να λαμβάνουν μεγαλύτερες σταθμίσεις οι μονάδες που έχουν μακρύτερα κοινά σύνορα.

Τα διαθέσιμα κριτήρια χωρικής γειτονίας και οι μέθοδοι ορισμού της μήτρας χωρικών σταθμίσεων δε εξαντλούνται στην παραπάνω παρουσίαση. Επιπρόσθετα, αξίζει να αναφερθεί ότι στις εμπειρικές εφαρμογές της χωρικής οικονομετρίας η χρησιμοποίηση των κριτηρίων και οι απαραίτητοι υπολογισμοί για την παραγωγή της μήτρας χωρικών σταθμίσεων γίνονται

με τη βοήθεια ψηφιακών χαρτών μέσα από γεωγραφικά συστήματα πληροφοριών και εξειδικευμένων λογισμικών χωρικής ανάλυσης. Χωρίς τη βοήθεια των ηλεκτρονικών υπολογιστών δεν είναι δυνατόν να οριστούν οι μήτρες λόγω της πολυπλοκότητας των συγκρίσεων και της δυσκολίας των μαθηματικών υπολογισμών που απαιτούνται.

1.4 Ο συντελεστής χωρικής αυτοσυσχετίσης I του Moran

Ο συντελεστής χωρικής αυτοσυσχετίσης I έχει προταθεί από τον Moran (1948 και 1950) για τη μέτρηση της χωρικής αυτοσυσχετίσης σε μια μεταβλητή y . Ορίζεται ως:

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

όπου n είναι ο αριθμός των γεωγραφικών περιοχών, \bar{y} η μέση τιμή της μεταβλητής y και w_{ij} το αντίστοιχο στοιχείο της μήτρας χωρικών σταθμίσεων \mathbf{W} . Όταν η μήτρα χωρικών σταθμίσεων χρησιμοποιείται στην τυποποιημένη κατά γραμμή μορφή της θα ισχύει ότι

$\sum_{i=1}^n \sum_{j=1}^n w_{ij} = n$ και ο συντελεστής I γράφεται ισοδύναμα:

$$I = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Εάν υποθεθεί ότι δεν υπάρχει χωρική αυτοσυσχετίση τότε ο συντελεστής I θα έχει αναμενόμενη τιμή:

$$E(I) = \frac{-1}{n-1}$$

η οποία εξαρτάται αποκλειστικά από το μέγεθος του δείγματος.

Εάν $I > E(I)$, τότε υπάρχει θετική χωρική αυτοσυσχετίση.³ Αυτό σημαίνει ότι οι παρατηρήσεις της μεταβλητής y σχηματίζουν συστάδες στο χώρο έτσι ώστε οι μεγάλες τιμές να βρίσκονται κοντά στις μεγάλες τιμές και οι μικρές τιμές κοντά στις μικρές. Όταν $I < E(I)$

³ Πολλές φορές στη βιβλιογραφία θεωρείται ότι υπάρχει θετική χωρική αυτοσυσχετίση όταν $I > 0$ και αρνητική αυτοσυσχετίση όταν $I < 0$. Η προσέγγιση αυτή στην πραγματικότητα δεν διαφέρει από την παρουσίαση αυτής της ενότητας καθώς είναι φανερό ότι ουσιαστικά ακόμη και για μικρά n το $E(I)$ θα είναι πολύ κοντά στο μηδέν.

υπάρχει αρνητική χωρική αυτοσυσχέτιση και στο χώρο θα σχηματίζονται συστάδες μεγάλων και μικρών τιμών και αντίθετα εμφανίζοντας στο χάρτη ένα πρότυπο που μοιάζει με σκακιέρα. Τέλος, όταν $I = E(I)$ δεν υπάρχει χωρική αυτοσυσχέτιση και οι παρατηρήσεις κατανέμονται στο χώρο με τυχαίο τρόπο.

Οι Cliff και Ord (1981) υπολόγισαν τις ροπές της I υποθέτοντας τη συνθήκη της κανονικότητας. Ειδικότερα, εάν η μεταβλητή y ακολουθεί την κανονική κατανομή, τότε η διακύμανση της I θα είναι:

$$Var(I) = E(I^2) - E(I)^2$$

όπου

$$E(I^2) = \frac{n^2 S_1 - n S_2 + 3 S_0^2}{S_0^2 (n-1)(n+1)}$$

με

$$S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij},$$

$$S_1 = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (w_{ij} + w_{ji}),$$

και

$$S_2 = \sum_{i=1}^n \left(\sum_{j=1}^n w_{ij} + \sum_{j=1}^n w_{ji} \right)^2$$

Δηλαδή, η διακύμανση θα εξαρτάται μόνο από το μέγεθος του δείγματος και τα στοιχεία της μήτρας χωρικών σταθμίσεων. Κατά συνέπεια, υποθέτοντας κανονική κατανομή, ο έλεγχος της μηδενικής υπόθεσης της απουσίας χωρικής αυτοσυσχέτισης μπορεί να γίνει με τον μετασχηματισμό:

$$Z = \frac{I - E(I)}{\sqrt{Var(I)}}$$

που θα ακολουθεί την τυποποιημένη κανονική κατανομή.

Όταν η μεταβλητή y δεν ακολουθεί κανονική κατανομή ο έλεγχος της μηδενικής υπόθεσης για την απουσία χωρικής αυτοσυσχέτισης θα μπορούσε να πραγματοποιηθεί κάνοντας την υπόθεση της τυχαιοποίησης (*randomization*), δηλαδή θεωρώντας την εκτιμηθείσα τιμή της I σε σχέση με όλες τις δυνατές τιμές που θα μπορούσε να λάβει εάν υποθέτονταν όλες οι δυνατές μεταθέσεις της y σε όλες τις γεωγραφικές περιοχές. Ο Ord (1980) έδειξε ότι εάν

υποθεθεί το σύνολο των $n!$ ισοπίθανων τυχαίων μεταθέσεων της y τότε υπό τη μηδενική υπόθεση της απουσίας χωρικής αυτοσυσχέτισης και δεδομένου του παρατηρηθέντος δείγματος αλλάζει η δεύτερη ροπή της I η οποία θα είναι:

$$E(I^2) = \frac{n \left[(n^2 - 3n + 3) S_1 - n S_2 + 3 S_0^2 \right] - b_2 \left[(n^2 - n) S_1 - 2n S_2 + 6 S_0^2 \right]}{(n-1)(n-2)(n-3) S_0^2}$$

με

$$b_2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^4}{\left[\sum_{i=1}^n (y_i - \bar{y})^2 \right]^2}$$

Δηλαδή, στην περίπτωση αυτή η διακύμανση θα εξαρτάται και από τις τιμές του δείγματος και από τα στοιχεία της μήτρας χωρικών σταθμίσεων. Εφόσον το δείγμα είναι αρκετά μεγάλο η I μετασχηματίζεται και πάλι σε τυποποιημένη κανονική μεταβλητή και διεξάγεται ο έλεγχος.

Ο συντελεστής I του Moran θα μπορούσε να ερμηνευθεί σαν ένας συντελεστής γραμμικής συσχέτισης. Αυτό προκύπτει παρατηρώντας την αλγεβρική του μορφή όπου ο αριθμητής του κλάσματος αποτελεί τη συνδιακύμανση της εξεταζόμενης μεταβλητής με τη χωρική της υστέρηση και ο παρανομαστής τη διακύμανση της μεταβλητής. Όμως, σε αντίθεση με ότι ισχύει για το συντελεστή γραμμικής συσχέτισης, ο συντελεστής I μπορεί να λάβει τιμές εκτός του εύρους $[-1, 1]$. Συγκεκριμένα, το εύρος των τιμών της I εξαρτάται αποκλειστικά από τη μήτρα χωρικών σταθμίσεων \mathbf{W} που χρησιμοποιείται στην ανάλυση και όχι από τις τιμές της μεταβλητής. Οι de Jong, Sprenger και Veen (1984) έδειξαν ότι οι τιμές της I βρίσκονται στο διάστημα:

$$\frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \lambda_{\min} \leq I \leq \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \lambda_{\max}$$

όπου λ_{\min} και λ_{\max} είναι η ελάχιστη και η μέγιστη ιδιοτιμή της μήτρας:

$$\left(\mathbf{I} - \frac{1}{n} \mathbf{1}\mathbf{1}' \right) \mathbf{W} \left(\mathbf{I} - \frac{1}{n} \mathbf{1}\mathbf{1}' \right)$$

στην οποία με \mathbf{I} συμβολίζεται η $(n \times n)$ μοναδιαία μήτρα και με $\mathbf{1}$ ένα $(n \times 1)$ διάνυσμα από μονάδες. Επιπρόσθετα, ο Anselin (1996) παρουσιάζει το συντελεστή I του Moran, για την περίπτωση που η μήτρα χωρικών σταθμίσεων είναι τυποποιημένη ανά γραμμή, σαν το

συντελεστή της κλίσης μίας παλινδρόμησης με εξαρτημένη μεταβλητή τη χωρική υστέρηση και ανεξάρτητη μεταβλητή την εξεταζόμενη μεταβλητή σε αποκλίσεις από τη μέση τιμή της. Αυτό φαίνεται όταν υπολογιστεί ο I με τη βοήθεια άλγεβρας μητρών, δηλαδή:

$$I = \frac{\mathbf{y}'\mathbf{W}\mathbf{y}}{\mathbf{y}'\mathbf{y}}$$

όπου \mathbf{y} είναι το $(n \times 1)$ διάνυσμα των τιμών της μεταβλητής y σε αποκλίσεις από τη μέση τιμή της και $\mathbf{W}\mathbf{y}$ το $(n \times 1)$ διάνυσμα της χωρικής υστέρησης.

1.5 Χωρικές στοχαστικές διαδικασίες

Στην ανάλυση χρονοσειρών θεωρείται ότι το εξεταζόμενο δείγμα είναι αποτέλεσμα από την πραγματοποίηση κάποιας στοχαστικής διαδικασίας (*stochastic process*). Κατ' αντιστοιχισμό, στα χωρικά δεδομένα το δείγμα προέρχεται από κάποια χωρική στοχαστική διαδικασία (*spatial stochastic process*). Όμως, οι χωρικές διαδικασίες είναι πιο σύνθετες από τις διαδικασίες των χρονοσειρών διότι η εξάρτηση στο χώρο ορίζεται προς πολλές κατευθύνσεις και ο δείκτης της διαδικασίας αποτελείται από γεωγραφικές συντεταγμένες. Οι επόμενοι ορισμοί προέρχονται από τους Schabenberger και Gotway (2005).

Με τον όρο χωρική στοχαστική διαδικασία εννοείται ένα διατεταγμένο σύνολο τυχαίων μεταβλητών που διατάσσονται σύμφωνα με τα στοιχεία του συνόλου $D \subset \mathbb{R}^d$ που περιλαμβάνει τις χωρικές συντεταγμένες $\mathbf{s} = [s_1, s_2, \dots, s_d]'$, δηλαδή $\{Z(\mathbf{s}) : \mathbf{s} \in D \subset \mathbb{R}^d\}$.

Όταν η διάσταση του d στο σύνολο των δεικτών είναι μεγαλύτερη του 1 τότε η στοχαστική διαδικασία ονομάζεται τυχαίο πεδίο (*random field*). Στην ανάλυση χωρικών δεδομένων συνήθως ισχύει ότι $d = 2$ διότι γίνεται αναφορά σε γεωγραφικές συντεταγμένες στο επίπεδο. Οι δείκτες στο σύνολο d μπορεί να είναι συνεχείς όταν αναφέρονται σε γεωγραφικές συντεταγμένες σημείων στο χώρο ή διακριτοί όταν αναφέρονται στις συντεταγμένες επάνω σε ένα ομαλό πλέγμα (*regular grid*) ή σε αύξοντες αριθμούς περιοχών (Arbia, 2006). Ως στοχαστική διαδικασία, η $Z(\mathbf{s})$ θα μπορούσε να ερμηνευθεί σαν η τιμή ενός χαρακτηριστικού Z στο σημείο του χώρου \mathbf{s} που προέρχεται από την εκτέλεση ενός τυχαίου πειράματος. Δηλαδή, η $Z(\mathbf{s})$ είναι στην πραγματικότητα μια τυχαία συνάρτηση.

Όπως για κάθε στοχαστική διαδικασία έτσι και στις χωρικές διαδικασίες ορίζεται η έννοια της στασιμότητας. Η στασιμότητα διακρίνεται στην αυστηρή στασιμότητα και στην ασθενή στασιμότητα. Ειδικότερα, ένα τυχαίο πεδίο $\{Z(\mathbf{s}) : \mathbf{s} \in D \subset \mathbb{R}^d\}$ θα ονομάζεται αυστηρά στάσιμο (*strict stationary*) όταν για τη χωρική κατανομή του ισχύει:

$$\begin{aligned} P(Z(\mathbf{s}_1) < z_1, Z(\mathbf{s}_2) < z_2, \dots, Z(\mathbf{s}_k) < z_k) = \\ P(Z(\mathbf{s}_1 + \mathbf{h}) < z_1, Z(\mathbf{s}_2 + \mathbf{h}) < z_2, \dots, Z(\mathbf{s}_k + \mathbf{h}) < z_k), \quad \forall k \text{ και } \forall \mathbf{h} \in D \end{aligned}$$

που σημαίνει ότι η χωρική κατανομή παραμένει ίδια σε κάθε μετατόπιση των χωρικών συντεταγμένων. Επειδή στο χώρο η σχέση εξάρτησης είναι προς πολλές κατευθύνσεις τυπικά για να είναι αυστηρά στάσιμο το πεδίο θα πρέπει να ισχύει και η υπόθεση της ισοτροπίας (*isotropy*). Η υπόθεση αυτή συνεπάγεται ότι η χωρική κατανομή δεν μεταβάλλεται όταν αλλάζει η κατεύθυνση όπως, παραδείγματος χάριν, όταν πραγματοποιείται περιστροφή των συντεταγμένων. Η αυστηρή στασιμότητα σπάνια συναντάται στις πρακτικές εφαρμογές και είναι δύσκολο να ελεγχθεί οπότε συνήθως θεωρείται η ασθενής στασιμότητα που αναφέρεται μόνο στις δύο πρώτες ροπές της κατανομής.

Το τυχαίο πεδίο θα ονομάζεται ασθενώς στάσιμο (*weak stationary*) ή στάσιμο β τάξης (*second order stationary*) όταν ισχύει:

$$\begin{aligned} E[Z(\mathbf{s})] &= \mu, \quad \forall \mathbf{s} \in D \\ \text{Var}[Z(\mathbf{s})] &= \sigma^2, \quad \forall \mathbf{s} \in D \\ \text{Cov}[Z(\mathbf{s}), Z(\mathbf{s} + \mathbf{h})] &= C(\mathbf{h}), \quad \forall \mathbf{s}, \mathbf{h} \in D \end{aligned}$$

Επομένως, εφόσον ισχύει η ασθενής στασιμότητα η χωρική διαδικασία θα έχει σταθερή μέση τιμή, σταθερή διακύμανση και συνδιακύμανση για κάθε σημείο στο χώρο που εξαρτάται μόνο από την απόστασή τους και όχι από τις γεωγραφικές συντεταγμένες καθ' αυτές. Αντίστοιχα, απαραίτητη είναι η υπόθεση της ισοτροπίας για τις ροπές που δεν θα πρέπει να εξαρτώνται από την κατεύθυνση στο χώρο.

Στη συνέχεια παρουσιάζονται συνοπτικά οι πιο βασικές χωρικές στοχαστικές διαδικασίες που συναντιούνται συχνότερα στη χωρική οικονομετρία και που θα χρησιμοποιηθούν στην ανάλυση των επόμενων κεφαλαίων. Οι διαδικασίες αυτές είναι η χωρική διαδικασία λευκού θόρυβου (*Spatial White Noise*), η χωρική αυτοπαλίνδρομη διαδικασία πρώτης τάξης (*Spatial Autoregressive Process–SAR(1)*) και η χωρική διαδικασία κινητού μέσου πρώτης τάξης (*Spatial Moving Average Process–SMA(1)*).

1.5.1 Η χωρική διαδικασία λευκού θορύβου

Η χωρική διαδικασία λευκού θορύβου (*Spatial White Noise*) αποτελεί επέκταση της γνωστής διαδικασίας λευκού θορύβου στις δύο διαστάσεις (Arbia, 2006). Ειδικότερα, ένα τυχαίο πεδίο $\{\varepsilon(\mathbf{s}) : \mathbf{s} \in D\}$ ονομάζεται χωρική διαδικασία λευκού θορύβου αν ισχύουν:

$$E[\varepsilon(\mathbf{s})] = 0, \quad \forall \mathbf{s} \in D$$

και

$$\text{Cov}[\varepsilon(\mathbf{s}_i), \varepsilon(\mathbf{s}_j)] = \begin{cases} \sigma^2 & \text{για } i = j \\ 0 & \text{για } i \neq j \end{cases}$$

δηλαδή, αν έχει μέση τιμή μηδέν, σταθερή διακύμανση και μηδενικές συνδιακυμάνσεις.

1.5.2 Η χωρική αυτοπαλίνδρομη διαδικασία πρώτης τάξης

Η χωρική αυτοπαλίνδρομη διαδικασία πρώτης τάξης (*Spatial Autoregressive Process—SAR(1)*) αποτελεί επέκταση της γνωστής από την ανάλυση χρονοσειρών αυτοπαλίνδρομης διαδικασίας AR(1) στις δύο διαστάσεις. Όπως δίνεται στον Arbia (2006), ένα τυχαίο πεδίο $\{Y(\mathbf{s}) : \mathbf{s} \in D\}$ θα ονομάζεται χωρική αυτοπαλίνδρομη διαδικασία πρώτης τάξης εάν παράγεται από το υπόδειγμα:

$$Y(\mathbf{s}_i) - \mu_i = \sum_{i \neq j} \rho_{ij} [Y(\mathbf{s}_j) - \mu_j] + \varepsilon(\mathbf{s}_i)$$

όπου $\rho_{ij} = \rho w_{ij}$ και $\varepsilon(\mathbf{s}_i)$ είναι μια χωρική διαδικασία λευκού θορύβου. Η διαδικασία αυτή αρχικά παρουσιάστηκε από τον Whittle (1954) και πολλές φορές συναντάται στη βιβλιογραφία και με την ονομασία χωρική ταυτόχρονη αυτοπαλίνδρομη διαδικασία (*Spatial Simultaneous Autoregressive Process*) διότι σε κάθε σημείο του χώρου πραγματοποιούνται ταυτόχρονα πολλές αυτοπαλινδρομήσεις. Συνήθως, στις εργασίες της χωρικής οικονομετρίας η διαδικασία αυτή γράφεται πιο απλά ως:

$$y_i = \rho \sum_{j=1}^n w_{ij} y_j + \varepsilon_i$$

όπου έχει γίνει η υπόθεση ότι η μέση τιμή ισούται με μηδέν ή ότι οι παρατηρήσεις εκφράζονται σε αποκλίσεις από το μέσο. Με τον τρόπο αυτό γραφής της διαδικασίας είναι φανερό ότι στη δεξιά πλευρά της ισότητας εμφανίζεται η χωρική υστέρηση της μεταβλητής και αποκαλύπτεται η ομοιότητα της διαδικασίας με την αντίστοιχη της ανάλυσης χρονοσειρών. Δηλαδή, η τιμή της μεταβλητής y σε μια περιοχή i σχηματίζεται ως το γινόμενο

ενός σταθμικού μέσου όρου των τιμών της στις γειτονικές περιοχές, όπως αυτές καθορίζονται από το κριτήριο γειτονίας που έχει εφαρμοστεί, επί την τιμή του αυτοπαλίνδρομου όρου συν το τυχαίο σφάλμα. Με τη βοήθεια της άλγεβρας μητρών (LeSage και Pace, 2009) η χωρική αυτοπαλίνδρομη διαδικασία γράφεται ως:

$$\mathbf{y} = \alpha \mathbf{1} + \rho \mathbf{W} \mathbf{y} + \boldsymbol{\varepsilon}$$

όπου \mathbf{y} είναι το $(n \times 1)$ διάνυσμα των παρατηρήσεων της διαδικασίας στα n σημεία του χώρου, α και $\mathbf{1}$ μια παράμετρος και ένα $(n \times 1)$ διάνυσμα από μονάδες που περιλαμβάνονται όταν η διαδικασία έχει μη μηδενική μέση τιμή, \mathbf{W} είναι η $(n \times n)$ μήτρα χωρικών σταθμίσεων, ρ είναι ο χωρικά αυτοπαλίνδρομος συντελεστής και $\boldsymbol{\varepsilon}$ ένα $(n \times 1)$ διάνυσμα με τιμές λευκού θορύβου. Αποδεικνύεται (LeSage και Pace, 2009) ότι η χωρική αυτοπαλίνδρομη διαδικασία πρώτης τάξης έχει μέση τιμή:

$$E[\mathbf{y}] = \frac{1}{1 - \rho} \mathbf{1} \alpha$$

και μήτρα διακυμάνσεων – συνδιακυμάνσεων που είναι:

$$\text{Cov}[(\mathbf{y} - \mu \mathbf{1}), (\mathbf{y} - \mu \mathbf{1})] = E[(\mathbf{y} - \mu \mathbf{1})(\mathbf{y} - \mu \mathbf{1})'] = \sigma^2 [(\mathbf{I} - \rho \mathbf{W})' (\mathbf{I} - \rho \mathbf{W})]^{-1}$$

Στο σημείο αυτό πρέπει να αναφερθεί ότι σε αντίθεση με την αντίστοιχη διαδικασία στην ανάλυση χρονοσειρών το εύρος τιμών του ρ δεν περιορίζεται αναγκαστικά στο διάστημα $(-1, 1)$. Συγκεκριμένα, για να είναι καλά ορισμένη η χωρική αυτοπαλίνδρομη διαδικασία πρώτης τάξης θα πρέπει η μήτρα $(\mathbf{I} - \rho \mathbf{W})$ να είναι μη ιδιάζων (*non - singular*). Ο Ord

(1975) έχει δείξει ότι για τις τιμές του ρ πρέπει να ισχύει $\frac{1}{\omega_{\min}} < \rho < \frac{1}{\omega_{\max}}$, όπου ω_{\min} και

ω_{\max} είναι η μικρότερη και η μεγαλύτερη ιδιοτιμή της μήτρας χωρικών σταθμίσεων \mathbf{W} εφόσον όλες οι ιδιοτιμές της είναι πραγματικές. Όμως, σε μια τυποποιημένη κατά γραμμή μήτρα χωρικών σταθμίσεων η μεγαλύτερη ιδιοτιμή της είναι πάντα +1 κάτι που εξασφαλίζει ότι $\rho < +1$ ενώ για το κάτω όριο δεν υπάρχει κάποια συγκεκριμένη τιμή που συνήθως είναι μικρότερο του -1 .

1.5.3 Η χωρική διαδικασία κινητού μέσου πρώτης τάξης

Ο Arbia (2006) ορίζει ένα τυχαίο πεδίο $\{Y(\mathbf{s}) : \mathbf{s} \in D\}$ ως χωρική διαδικασία κινητού μέσου πρώτης τάξης (*Spatial Moving Average Process*–SMA(1)) εάν παράγεται από το υπόδειγμα:

$$Y(\mathbf{s}_i) = \mu_i + \sum_{i \neq j} \lambda_{ij} \varepsilon(\mathbf{s}_j) + \varepsilon(\mathbf{s}_i)$$

όπου $\lambda_{ij} = \lambda w_{ij}$ και $\varepsilon(\mathbf{s}_i)$ είναι μια χωρική διαδικασία λευκού θορύβου. Η διαδικασία αυτή παρουσιάστηκε αρχικά από τον Haining (1978) και μπορεί να θεωρηθεί ως επέκταση στις δύο διαστάσεις της διαδικασίας κινητού μέσου MA(1) της ανάλυσης χρονοσειρών. Η διαδικασία αυτή μπορεί να γραφεί πιο απλά ως:

$$y_i = \lambda \sum_{j=1}^n w_{ij} \varepsilon_j + \varepsilon_i$$

Χρησιμοποιώντας μήτρες η χωρική διαδικασία κινητού μέσου πρώτης τάξης γράφεται ως:

$$\mathbf{y} = \lambda \mathbf{W} \boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}$$

όπου \mathbf{y} είναι το $(n \times 1)$ διάνυσμα των παρατηρήσεων, λ είναι ο συντελεστής χωρικού κινητού μέσου, \mathbf{W} η $(n \times n)$ μήτρα χωρικών σταθμίσεων και $\boldsymbol{\varepsilon}$ το $(n \times 1)$ διάνυσμα των τιμών του λευκού θορύβου. Η χωρική διαδικασία κινητού μέσου πρώτης τάξης έχει μέση τιμή:

$$E[\mathbf{y}] = 0$$

και μήτρα διακυμάνσεων – συνδιακυμάνσεων που είναι:

$$\text{Cov}[\mathbf{y}, \mathbf{y}] = E[\mathbf{y} \mathbf{y}'] = \sigma^2 [\mathbf{I} + \lambda (\mathbf{W} + \mathbf{W}') + \lambda^2 \mathbf{W} \mathbf{W}']$$

Για να είναι σωστά ορισμένη η χωρική διαδικασία κινητού μέσου πρώτης τάξης θα πρέπει η μήτρα $(\mathbf{I} + \lambda \mathbf{W})$ να είναι μη ιδιάζων. Αποδεικνύεται (Ord, 1975) ότι για να ισχύει αυτό θα

πρέπει να ικανοποιείται η συνθήκη $\frac{1}{\omega_{\min}} < -\lambda < \frac{1}{\omega_{\max}}$ όπου ω_{\min} και ω_{\max} είναι και πάλι η μικρότερη και η μεγαλύτερη ιδιοτιμή της μήτρας χωρικών σταθμίσεων \mathbf{W} .

1.6 Χωρικά οικονομετρικά υποδείγματα

Το βασικό χαρακτηριστικό των χωρικών δεδομένων είναι η ύπαρξη των χωρικών επιδράσεων, δηλαδή της χωρικής εξάρτησης και της χωρικής ετερογένειας. Οι επιδράσεις αυτές εάν αγνοηθούν από τον ερευνητή στα πλαίσια της κλασσικής οικονομετρικής ανάλυσης έχουν αποτέλεσμα τα αποτελέσματα που θα προκύψουν από την εκτίμηση των οικονομετρικών υποδειγμάτων να είναι μη αξιόπιστα. Για το λόγο αυτό έχουν αναπτυχθεί οικονομετρικά υποδείγματα που λαμβάνουν υπ' όψιν την παρουσία των χωρικών επιδράσεων στο δείγμα αντιμετωπίζοντας τις επιπτώσεις τους αλλά και προσφέροντας επιπλέον πληροφορίες στον ερευνητή. Στην ενότητα αυτή παρουσιάζονται συνοπτικά τα βασικότερα χωρικά οικονομετρικά υποδείγματα.

1.6.1 Το υπόδειγμα χωρικής υστέρησης

Το υπόδειγμα χωρικής υστέρησης (*Spatial Lag Model–SLM*) αντιμετωπίζει το πρόβλημα της εμφάνισης χωρικής αυτοσυσχέτισης στα σφάλματα ενός οικονομετρικού υποδείγματος που οφείλεται στην ύπαρξη χωρικών επιδράσεων της μορφής της χωρικής εξάρτησης περιλαμβάνοντας ως ανεξάρτητη μεταβλητή τη χωρική υστέρηση της εξαρτημένης μεταβλητής.⁴ Αυτό σημαίνει ότι το πρόβλημα αντιμετωπίζεται με τον ίδιο τρόπο όπως στην ανάλυση χρονοσειρών που προστίθεται η χρονική υστέρηση της εξαρτημένης μεταβλητής στις ανεξάρτητες μεταβλητές όταν υπάρχει χρονική αυτοσυσχέτιση. Η εκτίμηση αυτού του υποδείγματος γίνεται όταν η αιτία της χωρικής αυτοσυσχέτισης οφείλεται στην αλληλεπίδραση μεταξύ των διαφορετικών γεωγραφικών περιοχών που εκφράζεται μέσω διαδικασιών διάχυσης (*spillover*).

Με χρησιμοποίηση άλγεβρας μητρών το υπόδειγμα χωρικής υστέρησης ορίζεται ως:

$$\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

όπου \mathbf{y} είναι το $(n \times 1)$ διάνυσμα της εξαρτημένης μεταβλητής, $\mathbf{W}\mathbf{y}$ η χωρική υστέρηση της εξαρτημένης μεταβλητής που υπολογίζεται από την $(n \times n)$ μήτρα χωρικών σταθμίσεων \mathbf{W} , ρ ο συντελεστής χωρικής υστέρησης, \mathbf{X} η $(n \times (k+1))$ μήτρα των παρατηρήσεων των k ανεξάρτητων μεταβλητών περιλαμβάνοντας τον σταθερό όρο, $\boldsymbol{\beta}$ το $((k+1) \times 1)$ διάνυσμα

⁴ Μερικές φορές το υπόδειγμα χωρικής υστέρησης αναφέρεται και ως χωρικά αυτοπαλίνδρομο υπόδειγμα (*spatial autoregressive model*) ή ως μικτό παλίνδρομο – χωρικά αυτοπαλίνδρομο υπόδειγμα (*mixed regressive – spatial autoregressive model*) εννοώντας ότι πρόκειται για μια χωρικά αυτοπαλίνδρομη διαδικασία στην οποία έχουν προστεθεί ανεξάρτητες μεταβλητές.

των συντελεστών του υποδείγματος και $\boldsymbol{\varepsilon}$ το $(n \times 1)$ διάνυσμα των τυχαίων σφαλμάτων για τα οποία γίνεται η υπόθεση ότι $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$. Το υπόδειγμα αυτό ισοδύναμα μπορεί να γραφεί ως:

$$\mathbf{y} = (\mathbf{I} - \rho \mathbf{W})^{-1} \mathbf{X} \boldsymbol{\beta} + (\mathbf{I} - \rho \mathbf{W})^{-1} \boldsymbol{\varepsilon}$$

Από αυτόν τον τρόπο γραφής του υποδείγματος διαπιστώνεται ότι η ερμηνεία των συντελεστών των ανεξάρτητων μεταβλητών δεν πρέπει να γίνεται με τον τρόπο που ισχύει σε ένα απλό οικονομετρικό υπόδειγμα, δηλαδή ως μερικές παράγωγοι που εκφράζουν τη μεταβολή στις τιμές της εξαρτημένης μεταβλητής όταν μεταβάλλεται μία ανεξάρτητη μεταβλητή κατά μία μονάδα ενώ οι υπόλοιπες μεταβλητές του υποδείγματος παραμένουν σταθερές. Ειδικότερα, στο απλό οικονομετρικό υπόδειγμα η μεταβολή μίας ανεξάρτητης μεταβλητής σε μία συγκεκριμένη περιοχή επηρεάζει την τιμή της εξαρτημένης μεταβλητής αποκλειστικά σε αυτή την περιοχή. Αντίθετα, στο υπόδειγμα χωρικής υστέρησης μια μεταβολή στη τιμή κάποιας ανεξάρτητης μεταβλητής σε κάποια περιοχή επηρεάζει τις τιμές της εξαρτημένης μεταβλητής σε όλες τις άλλες περιοχές. Για να αποδειχθεί αυτό, οι LeSage και Pace (2009) βασίζόμενοι στην εργασία των Kim, Phipps και Anselin (2003) γράφουν το υπόδειγμα χωρικής υστέρησης για τις n γεωγραφικές περιοχές (παρατηρήσεις) και τις k ανεξάρτητες μεταβλητές στην εξής μορφή:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{r=1}^k \begin{bmatrix} S_r(W)_{11} & S_r(W)_{12} & \cdots & S_r(W)_{1n} \\ S_r(W)_{21} & S_r(W)_{22} & \cdots & S_r(W)_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ S_r(W)_{n1} & S_r(W)_{n2} & \cdots & S_r(W)_{nn} \end{bmatrix} \begin{bmatrix} x_{1r} \\ x_{2r} \\ \vdots \\ x_{nr} \end{bmatrix} + (\mathbf{I} - \rho \mathbf{W})^{-1} \boldsymbol{\varepsilon}$$

όπου με $S_r(W)_{ij}$ συμβολίζουν το i, j στοιχείο στη μήτρα $S_r(W) = (\mathbf{I} - \rho \mathbf{W})^{-1} \mathbf{I} \beta_r$ και αν με $V(W)_i$ συμβολιστεί η i γραμμή της μήτρας $(\mathbf{I} - \rho \mathbf{W})^{-1}$ τότε το υπόδειγμα χωρική υστέρησης για την i περιοχή γράφεται ως:

$$y_i = \sum_{r=1}^k [S_r(W)_{i1} x_{1r} + S_r(W)_{i2} x_{2r} + \dots + S_r(W)_{in} x_{nr}] + V(W)_i \boldsymbol{\varepsilon}$$

από τη σχέση αυτή φαίνεται ότι θα ισχύουν $\frac{\partial y_i}{\partial x_{jr}} = S_r(W)_{ij} \neq 0$ και $\frac{\partial y_i}{\partial x_{ir}} = S_r(W)_{ii} \neq \beta_r$.

Επειδή η ένταση των επιδράσεων διαφέρει από περιοχή σε περιοχή, οι LeSage και Pace (2009) προτείνουν τα παρακάτω τρία στατιστικά μέτρα για να μετριοούνται οι χωρικές επιδράσεις.⁵

1) Η μέση άμεση επίδραση

Η μέση άμεση επίδραση (*Average Direct Impact*) υπολογίζεται από τη σχέση:

$$\bar{M}(r)_{direct} = \frac{1}{n} tr[S_r(W)]$$

και εκφράζει την επίδραση των μεταβολών στην i παρατήρηση της r ανεξάρτητης μεταβλητής, x_{ir} στην i παρατήρηση της εξαρτημένης μεταβλητής y_i .

2) Η μέση συνολική επίδραση προς (από) μια παρατήρηση

Η μέση συνολική επίδραση προς μια παρατήρηση (*Average Total Impact to an Observation*) υπολογίζεται από τη σχέση:

$$\bar{M}(r)_{total} = \frac{1}{n} \mathbf{1}' S_r(W) \mathbf{1}$$

και εκφράζει τη συνολική επίδραση στην i παρατήρηση της εξαρτημένης μεταβλητής y_i που οφείλεται από τη μεταβολή της r ανεξάρτητης μεταβλητής κατά την ίδια ποσότητα σε όλες τις n περιοχές. Η μέση συνολική επίδραση προς μια παρατήρηση θα ισούται και με τη μέση συνολική επίδραση από μια παρατήρηση (*Average Total Impact from an Observation*) που εκφράζει τις συνολικές επιδράσεις σε όλες τις τιμές της y που οφείλονται στη μεταβολή της τιμής της j παρατήρησης στην r ανεξάρτητη μεταβλητή x_{jr} .

3) Η μέση έμμεση επίδραση

Η μέση έμμεση επίδραση (*Average Indirect Impact*) υπολογίζεται από τη σχέση:

$$\bar{M}(r)_{indirect} = \bar{M}(r)_{total} - \bar{M}(r)_{direct}$$

και εκφράζει την έμμεση επίδραση από τη μεταβολή της μεταβλητής x_r .

Η εκτίμηση του υποδείγματος χωρικής υστέρηση δεν μπορεί να γίνει με τη μέθοδο των ελαχίστων τετραγώνων διότι θα προκύψουν μη αμερόληπτοι εκτιμητές λόγω εμφάνισης προβλήματος ενδογένειας καθώς η χωρική υστέρηση της εξαρτημένης μεταβλητής συσχετίζεται με το τυχαίο σφάλμα. Για το λόγο αυτό, για την εκτίμηση του υποδείγματος χρησιμοποιείται η μέθοδος της μέγιστης πιθανοφάνειας. Η λογαριθμική συνάρτηση

⁵ Για λεπτομέρειες βλέπε Lesage και Pace (2009) σελίδες 34 – 39.

πιθανοφάνειας για το υπόδειγμα χωρικής υστέρησης σύμφωνα με τον Anselin (1988) ορίζεται ως:

$$\ln L(\boldsymbol{\beta}, \rho, \sigma^2 | \mathbf{y}, \mathbf{X}) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) + \ln |\mathbf{I} - \rho \mathbf{W}| - \frac{(\mathbf{y} - \rho \mathbf{W}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \rho \mathbf{W}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{2\sigma^2}$$

και πρέπει να μεγιστοποιηθεί ως προς τις παραμέτρους $\boldsymbol{\beta}$, ρ και σ^2 . Στη συνάρτηση πιθανοφάνειας υπάρχει και ο λογάριθμος της Ιακωβιανής ορίζουσας $|\mathbf{I} - \rho \mathbf{W}|$ η οποία θα πρέπει να είναι θετική. Επομένως, θα πρέπει να ισχύουν οι συνθήκες ομαλότητας που σύμφωνα με τον Ord (1975) είναι οι $1/\omega_{\min} < \rho < 1/\omega_{\max}$ όπως παρουσιάστηκαν για τη χωρική αυτοπαλίνδρομη διαδικασία.

Η παραγωγή ως προς το διάνυσμα $\boldsymbol{\beta}$ οδηγεί στους εκτιμητές:

$$\hat{\boldsymbol{\beta}}_{ML} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'(\mathbf{I} - \rho \mathbf{W})\mathbf{y}$$

που εξαρτώνται από την άγνωστη παράμετρο ρ και που ισοδύναμα γράφονται ως:

$$\hat{\boldsymbol{\beta}}_{ML} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' - \rho (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{W}\mathbf{y} = \hat{\boldsymbol{\beta}}_{olsy,X} - \rho \hat{\boldsymbol{\beta}}_{olsWy,X}$$

Δηλαδή, οι εκτιμητές μέγιστης πιθανοφάνειας διασπώνται σε δύο εκτιμητές ελαχίστων τετραγώνων του $\hat{\boldsymbol{\beta}}_{olsy,X}$ που προκύπτει από την παλινδρόμηση του υποδείγματος που έχει εξαρτημένη μεταβλητή την \mathbf{y} και ανεξάρτητες μεταβλητές τις \mathbf{X} και του εκτιμητή $\hat{\boldsymbol{\beta}}_{olsWy,X}$ που προκύπτει από την παλινδρόμηση του υποδείγματος με εξαρτημένη μεταβλητή τη χωρική υστέρησης $\mathbf{W}\mathbf{y}$ και ανεξάρτητες μεταβλητές τις \mathbf{X} . Εφόσον προκύψει με κάποιον τρόπο η εκτίμηση του ρ τότε οι εκτιμητές μέγιστης πιθανοφάνειας υπολογίζονται εύκολα. Για τα εκτιμηθέντα σφάλματα του υποδείγματος ισχύει:

$$\begin{aligned} \hat{\boldsymbol{\epsilon}}_{ML} &= \mathbf{y} - \rho \mathbf{W}\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{ML} \\ &= \mathbf{y} - \rho \mathbf{W}\mathbf{y} - \mathbf{X}(\hat{\boldsymbol{\beta}}_{olsy,X} - \rho \hat{\boldsymbol{\beta}}_{olsWy,X}) \\ &= \mathbf{y} - \rho \mathbf{W}\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{olsy,X} + \rho \mathbf{X}\hat{\boldsymbol{\beta}}_{olsWy,X} \\ &= \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{olsy,X} - \rho (\mathbf{W}\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{olsWy,X}) \\ &= \hat{\boldsymbol{\epsilon}}_{olsy,X} - \rho \hat{\boldsymbol{\epsilon}}_{olsWy,X} \end{aligned}$$

Επομένως και τα σφάλματα από την εκτίμηση με τη μέθοδο της μέγιστης πιθανοφάνειας μπορούν να διασπαστούν στα σφάλματα που προκύπτουν από τις εκτιμήσεις με τη μέθοδο των ελαχίστων τετραγώνων των δύο προαναφερθέντων υποδειγμάτων. Στη συνέχεια παραγωγίζοντας τη συνάρτηση πιθανοφάνειας ως προς τη διακύμανση σ^2 και αντικαθιστώντας τα εκτιμηθέντα σφάλματα προκύπτει:

$$\hat{\sigma}_{ML}^2 = \frac{1}{n} (\hat{\boldsymbol{\epsilon}}_{olsy,x} - \rho \hat{\boldsymbol{\epsilon}}_{olsWy,x})' (\hat{\boldsymbol{\epsilon}}_{olsy,x} - \rho \hat{\boldsymbol{\epsilon}}_{olsWy,x})$$

που σημαίνει ότι η διακύμανση από την εκτίμηση του υποδείγματος με τη μέθοδο της μέγιστης πιθανοφάνειας είναι συνάρτηση της άγνωστης παραμέτρου ρ . Κατά συνέπεια, διαπιστώνεται (Anselin, 1988) ότι η λογαριθμική συνάρτηση πιθανοφάνειας μπορεί να συγκεντρωθεί ως προς τις παραμέτρους $\boldsymbol{\beta}$ και σ^2 αντικαθιστώντας τις εκτιμήσεις τους και με τον τρόπο αυτό το πρόβλημα της ταυτόχρονης μεγιστοποίησης πολλών παραμέτρων μετατρέπεται σε πρόβλημα μεγιστοποίησης μίας άγνωστης παραμέτρου της ρ . Η λογαριθμική μορφή της συγκεντρωμένης συνάρτησης πιθανοφάνειας (*concentrated log – likelihood function*) είναι η:

$$\ln L(\rho) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} + \ln |\mathbf{I} - \rho \mathbf{W}| - \frac{n}{2} \ln \left[\frac{1}{n} (\hat{\boldsymbol{\epsilon}}_{olsy,x} - \rho \hat{\boldsymbol{\epsilon}}_{olsWy,x})' (\hat{\boldsymbol{\epsilon}}_{olsy,x} - \rho \hat{\boldsymbol{\epsilon}}_{olsWy,x}) \right]$$

η οποία είναι μη γραμμική συνάρτηση ως προς τη παράμετρο ρ . Η συνάρτηση μεγιστοποιείται με τη βοήθεια κάποιας αριθμητικής μεθόδου και η τιμή του ρ που βρίσκεται αντικαθίσταται στις παραπάνω σχέσεις και υπολογίζονται οι εκτιμητές μέγιστης πιθανοφάνειας των παραμέτρων $\boldsymbol{\beta}$ και σ^2 . Στη διαδικασία της μεγιστοποίησης διευκολύνει η διάσπαση της ορίζουσας στις ιδιοτιμές της όπως έχει προτείνει ο Ord (1975), δηλαδή ως:

$$\ln |\mathbf{I} - \rho \mathbf{W}| = \ln \prod_i (1 - \rho \omega_i) = \sum_i \ln (1 - \rho \omega_i)$$

όπου με ω_i συμβολίζονται οι ιδιοτιμές της μήτρας χωρικών σταθμίσεων \mathbf{W} . Η μήτρα διακυμάνσεων – συνδιακυμάνσεων των εκτιμητών του υποδείγματος που αναφέρεται στους Anselin και Bera (1998) υπολογίζεται αντιστρέφοντας τη μήτρα πληροφορίας όπως σε όλες τις διαδικασίες εκτίμησης με τη μέθοδο της μέγιστης πιθανοφάνειας.

1.6.2 Το υπόδειγμα χωρικού σφάλματος

Το δεύτερο πιο συνηθισμένο υπόδειγμα που χρησιμοποιείται όταν εντοπίζεται πρόβλημα χωρικής αυτοσυσχέτισης στα σφάλματα ενός οικονομετρικού υποδείγματος είναι το υπόδειγμα χωρικού σφάλματος (*Spatial Error Model–SEM*). Το υπόδειγμα αυτό εκτιμάται όταν ο ερευνητής θεωρεί ότι η παρουσία της αυτοσυσχέτισης δεν είναι ουσιαστική για την ερμηνεία της εξαρτημένης μεταβλητής αλλά προέρχεται εξαιτίας του είδους των δεδομένων και μπορεί να αντιμετωπιστεί με τον καθορισμό και την εκτίμηση μιας χωρικής διαδικασίας για τα σφάλματα που θα βοηθήσει στον προσδιορισμό των σωστών διακυμάνσεων των εκτιμητών, όπως συμβαίνει και στην επίλυση του προβλήματος της αυτοσυσχέτισης στα

οικονομετρικά υποδείγματα όταν χρησιμοποιούνται χρονοσειρές ως δείγμα. Το υπόδειγμα χωρικού σφάλματος με σφάλματα που προέρχονται από μια χωρική αυτοπαλίνδρομη διαδικασία πρώτης τάξης που είναι και η πιο συνηθισμένη περίπτωση ορίζεται ως:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \text{ με } \boldsymbol{\varepsilon} = \lambda \mathbf{W}\boldsymbol{\varepsilon} + \mathbf{u}$$

όπου \mathbf{y} είναι το διάνυσμα της εξαρτημένης μεταβλητής, \mathbf{X} η μήτρα των παρατηρήσεων των k ανεξάρτητων μεταβλητών, $\boldsymbol{\beta}$ το διάνυσμα των συντελεστών του υποδείγματος, όπως έχουν οριστεί στο υπόδειγμα χωρικής υστέρησης, $\mathbf{W}\boldsymbol{\varepsilon}$ είναι η χωρική υστέρηση των σφαλμάτων, λ ο συντελεστής χωρικής υστέρησης και για το διάνυσμα \mathbf{u} γίνεται η υπόθεση ότι $\mathbf{u} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$. Επιπρόσθετα, όπως και στο υπόδειγμα χωρικής υστέρησης για το συντελεστή λ ισχύουν οι συνθήκες ομαλότητας $1/\omega_{\min} < \lambda < 1/\omega_{\max}$. Ισοδύναμα το υπόδειγμα μπορεί να γραφεί ως:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \lambda \mathbf{W})^{-1} \mathbf{u}$$

Σε αντίθεση με το υπόδειγμα χωρικής υστέρησης στο υπόδειγμα αυτό η ερμηνεία των συντελεστών είναι η ίδια όπως σε ένα απλό οικονομικό υπόδειγμα. Η εκτίμηση του υποδείγματος χωρικού σφάλματος επιτυγχάνεται με τη βοήθεια της μεθόδου της μέγιστης πιθανοφάνειας η οποία στη λογαριθμική της μορφή (Anselin, 1988) ορίζεται ως:

$$\ln L(\boldsymbol{\beta}, \lambda, \sigma^2 | \mathbf{y}, \mathbf{X}) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) + \ln |\mathbf{I} - \lambda \mathbf{W}| - \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{I} - \lambda \mathbf{W})' (\mathbf{I} - \lambda \mathbf{W}) (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{2\sigma^2}$$

Όπως και στην εκτίμηση του υποδείγματος χωρικής υστέρησης η λογαριθμική συνάρτηση πιθανοφάνειας συγκεντρώνεται ως προς τις παραμέτρους $\boldsymbol{\beta}$ και σ^2 διευκολύνοντας τη διαδικασία. Ειδικότερα, παραγωγίζοντας ως προς το διάνυσμα $\boldsymbol{\beta}$ προκύπτουν οι εκτιμητές:

$$\hat{\boldsymbol{\beta}}_{ML} = \left[\mathbf{X}' (\mathbf{I} - \lambda \mathbf{W})' (\mathbf{I} - \lambda \mathbf{W}) \mathbf{X} \right]^{-1} \mathbf{X}' (\mathbf{I} - \lambda \mathbf{W})' (\mathbf{I} - \lambda \mathbf{W}) \mathbf{y} = \hat{\boldsymbol{\beta}}_{GLS}$$

οι οποίοι ισούνται (Anselin και Bera, 1998) με τους εκτιμητές $\hat{\boldsymbol{\beta}}_{GLS}$ που προκύπτουν από την εφαρμογή της γενικευμένης μεθόδου των ελαχίστων τετραγώνων (*Generalized Least Squares*), δηλαδή από την παλινδρόμηση με εξαρτημένη μεταβλητή την $(\mathbf{I} - \lambda \mathbf{W})\mathbf{y}$ και ανεξάρτητες τις $(\mathbf{I} - \lambda \mathbf{W})\mathbf{X}$ προϋποθέτοντας ότι η τιμή της παραμέτρου λ είναι γνωστή. Η παραγωγή ως προς τη διακύμανση σ^2 και η αντικατάσταση των εκτιμηθέντων σφαλμάτων οδηγεί στην εξής εκτίμηση της διακύμανσης:

$$\hat{\sigma}_{ML}^2 = \frac{1}{n} \hat{\boldsymbol{\varepsilon}}'_{GLS} (\mathbf{I} - \lambda \mathbf{W})' (\mathbf{I} - \lambda \mathbf{W}) \hat{\boldsymbol{\varepsilon}}_{GLS}$$

Με αντικατάσταση των εκτιμήσεων στη συνάρτηση πιθανοφάνειας προκύπτει η συγκεντρωμένη συνάρτηση πιθανοφάνειας (*concentrated log – likelihood function*), που είναι:

$$\ln L(\lambda) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} + \ln |\mathbf{I} - \lambda \mathbf{W}| - \frac{n}{2} \ln \left[\frac{1}{n} \hat{\boldsymbol{\varepsilon}}'_{GLS} (\mathbf{I} - \lambda \mathbf{W})' (\mathbf{I} - \lambda \mathbf{W}) \hat{\boldsymbol{\varepsilon}}_{GLS} \right]$$

η οποία είναι μη γραμμική συνάρτηση ως προς την παράμετρο λ και μπορεί να μεγιστοποιηθεί με την εφαρμογή κάποιας αριθμητικής μεθόδου. Όμως, η διαδικασία είναι πιο περίπλοκη σε σχέση με το υπόδειγμα χωρικής υστέρησης διότι στη συγκεντρωμένη συνάρτηση πιθανοφάνειας του υποδείγματος χωρικού σφάλματος τα εκτιμηθέντα σφάλματα εξαρτώνται έμμεσα από την άγνωστη παράμετρο λ καθώς η εκτίμηση $\hat{\boldsymbol{\beta}}_{GLS}$ απαιτεί τη γνώση της. Για το λόγο αυτό, είναι αναγκαία μια επαναληπτική διαδικασία εκτίμησης που και οι δύο παράμετροι να προσδιορίζονται ταυτόχρονα υπό συνθήκη η μία της άλλης. Ειδικότερα, ο Anselin (1988) προτείνει αρχικά την εκτίμηση με τη μέθοδο των ελαχίστων τετραγώνων του απλού υποδείγματος και τη χρησιμοποίηση των εκτιμηθέντων σφαλμάτων στη συγκεντρωμένη συνάρτηση πιθανοφάνειας ως αρχικά σφάλματα για την εύρεση της τιμής της παραμέτρου λ που τη μεγιστοποιεί.⁶ Στη συνέχεια, η τιμή αυτή χρησιμοποιείται για την εύρεση των εκτιμητών $\hat{\boldsymbol{\beta}}_{GLS}$ και τον υπολογισμό νέων σφαλμάτων που αντικαθίστανται στη συγκεντρωμένη συνάρτηση πιθανοφάνειας και οδηγούν σε καινούρια εκτίμηση του λ . Η διαδικασία αυτή επαναλαμβάνεται μέχρι να επέλθει σύγκλιση στα αποτελέσματα σύμφωνα με κάποιο κριτήριο. Η τελική τιμή του λ και τα αντίστοιχα σφάλματα αντικαθίστανται στο τύπο της διακύμανσης για την εκτίμησή της. Αφού βρεθούν όλες οι εκτιμήσεις είναι εφικτό να υπολογιστεί και η μήτρα διακυμάνσεων – συνδιακυμάνσεων των εκτιμητών του υποδείγματος που παρουσιάζεται στους Anselin και Bera (1998). Όπως και στο υπόδειγμα χωρικής υστέρησης η διάσπαση της ορίζουσας που βρίσκεται στη συνάρτηση πιθανοφάνειας με τη τεχνική του Ord (1975) διευκολύνει τη διαδικασία.

⁶ Ο Ord (1975) περιγράφει μια διαδικασία παρόμοια με τη μέθοδο των Cochrane και Orcutt της ανάλυσης χρονοσειρών στην οποία η τιμή του λ εκτιμάται άμεσα στο SAR(1) υπόδειγμα των σφαλμάτων με τη μέθοδο της μέγιστης πιθανοφάνειας αποφεύγοντας τη μεγιστοποίηση της συγκεντρωμένης πιθανοφάνειας αμφισβητώντας όμως ταυτόχρονα την εγκυρότητά της στη χωρική ανάλυση.

1.6.3 Υποδείγματα χωρικής ετερογένειας

Η παρουσία της χωρικής ετερογένειας σε μια οικονομετρική ανάλυση έχει ως συνέπεια η σχέση μεταξύ των μεταβλητών να μην είναι σταθερή σε όλη την εξεταζόμενη γεωγραφική περιοχή κάτι που οδηγεί στην εμφάνιση του προβλήματος της ετεροσκεδαστικότητας στα εκτιμηθέντα σφάλματα ενός οικονομετρικού υποδείγματος. Σε σχέση με τη χωρική εξάρτηση η χωρική ετερογένεια, ανάλογα με τον τύπο της, μπορεί να αντιμετωπιστεί αποτελεσματικά και μέσα από τα μη χωρικά οικονομετρικά υποδείγματα εφαρμόζοντας κάποια από τις γνωστές μεθόδους διόρθωσης του προβλήματος. Ειδικότερα, η χωρική ετερογένεια θα μπορούσε να χωριστεί σε δύο τύπους, δηλαδή στη διακριτή χωρική ετερογένεια (*discrete spatial heterogeneity*) και στη συνεχή χωρική ετερογένεια (*continuous spatial heterogeneity*).

Διακριτή χωρική ετερογένεια σημαίνει ότι οι παράμετροι του υποδείγματος σε κάποιες συγκεκριμένες γεωγραφικές περιοχές έχουν διαφορετικές τιμές από ότι στις άλλες περιοχές. Παραδείγματος χάριν, οι τιμές των παραμέτρων μπορεί να διαφοροποιούνται ανάλογα με το εάν μια περιοχή είναι αστική ή αγροτική ή εάν ανήκει στο βόριο ή στο νότιο μέρος μιας χώρας. Στις περιπτώσεις αυτές το πρόβλημα αντιμετωπίζεται με την εισαγωγή στο υπόδειγμα ψευδομεταβλητών (*dummy variables*) που καθορίζουν τη γεωγραφική θέση κάθε παρατήρησης και που οδηγούν στην εκτίμηση διαφορετικών σταθερών όρων ή και συντελεστών των ανεξάρτητων μεταβλητών.

Η συνεχής χωρική ετερογένεια είναι πιο σύνθετη στο να αντιμετωπιστεί και να ερμηνευθεί και η χωρική οικονομετρία προτείνει ειδική μεθοδολογία. Ο τύπος αυτός ετερογένειας συνεπάγεται ότι οι συντελεστές του υποδείγματος διαφοροποιούνται σε όλες τις περιοχές και προσδιορίζονται για κάθε γεωγραφικό σημείο ως αποτέλεσμα εφαρμογής κάποιας συνάρτησης των γεωγραφικών συντεταγμένων. Από τις πρώτες τεχνικές που χρησιμοποιήθηκαν για την αντιμετώπιση της χωρικής ετερογένειας αυτής της μορφής είναι η μέθοδος της χωρικής επέκτασης (*spatial expansion method*) που έχει προταθεί από τον Casetti (1972 και 1986). Η μέθοδος αυτή βοηθάει στην εκτίμηση υποδειγμάτων με μη σταθερούς συντελεστές. Συγκεκριμένα, κάθε παράμετρος του υποδείγματος θεωρείται συνάρτηση άλλων βοηθητικών μεταβλητών που τις περισσότερες φορές είναι οι γεωγραφικές συντεταγμένες. Με τον τρόπο αυτό οι παράμετροι του απλού υποδείγματος που ο Casetti το ονομάζει αρχικό (*initial model*) επεκτείνονται μέσω μιας συνήθως γραμμικής συνάρτησης των συντεταγμένων και προκύπτει ένα καινούριο υπόδειγμα που ονομάζεται τελικό (*terminal*

model). Για να γίνει κατανοητή η μέθοδος της χωρικής επέκτασης θεωρείται ένα απλό γραμμικό υπόδειγμα με μία ανεξάρτητη μεταβλητή, δηλαδή:

$$y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i$$

και υποτίθεται ότι η παράμετρος β_1 δεν σταθερή αλλά γραμμική συνάρτηση των γεωγραφικών συντεταγμένων z_1 και z_2 . Η παράμετρος β_1 επεκτείνεται και γίνεται:

$$\beta_1 = b_0 + b_1 z_{1i} + b_2 z_{2i}$$

Με αντικατάσταση στο αρχικό υπόδειγμα προκύπτει το τελικό υπόδειγμα που είναι:

$$\begin{aligned} y_i &= \beta_0 + (b_0 + b_1 z_{1i} + b_2 z_{2i}) x_{1i} + \varepsilon_i \\ &= \beta_0 + b_0 x_{1i} + b_1 z_{1i} x_{1i} + b_2 z_{2i} x_{1i} + \varepsilon_i \end{aligned}$$

Το υπόδειγμα αυτό εκτιμάται με τη μέθοδο των ελαχίστων τετραγώνων και οι εκτιμηθείσες τιμές των συντελεστών αντικαθίστανται στην εξίσωση επέκτασης για να παραχθεί για κάθε ζεύγος των γεωγραφικών συντεταγμένων η τιμή του β_1 .

Παρά την υιοθέτηση της μεθόδου της χωρικής επέκτασης σε πολλές εμπειρικές εργασίες τα αποτελέσματα από την εκτίμηση του τελικού υποδείγματος πρέπει να ερμηνεύονται με προσοχή διότι αρκετές φορές δεν ικανοποιούνται οι προϋποθέσεις της μεθόδου των ελαχίστων τετραγώνων που χρησιμοποιείται στην εκτίμησή του. Ένα πρόβλημα που εμφανίζεται στην περίπτωση αυτή είναι ότι οι μεταβλητές που εμπλέκονται στο τελικό υπόδειγμα παρουσιάζουν υψηλό κίνδυνο πολυσυγγραμμικότητας εξαιτίας της ύπαρξης των μεταβλητών αλληλοεπίδρασης. Οι Casetti και Jones (1987) αντιμετωπίζουν την εμφάνιση πολυσυγγραμμικότητας αντικαθιστώντας τις μεταβλητές στην εξίσωση επέκτασης με τις κύριες συνιστώσες τους (*principal components*). Επιπρόσθετα, η υπόθεση της ακριβούς γραμμική σχέσης στην εξίσωση επέκτασης είναι πολύ περιοριστική καθώς στην πράξη θα υπάρχει πάντα κάποιο σφάλμα που θα αντιπροσωπεύεται με κάποιο στοχαστικό όρο, δηλαδή:

$$\beta_1 = b_0 + b_1 z_{1i} + b_2 z_{2i} + u_i$$

με $u_i \sim iid N(0, \sigma_u^2)$. Ο Anselin (1988) δείχνει ότι όταν στην εξίσωση επέκτασης δεν υπάρχει ακριβής σχέση και εμπλέκεται όρος σφάλματος τότε πάντοτε στο τελικό υπόδειγμα θα εμφανίζεται πρόβλημα ετεροσκεδαστικότητας στα σφάλματα που θα πρέπει να διορθωθεί για να γίνει ο σωστός προσδιορισμός των διακυμάνσεων των εκτιμητών.

Εξέλιξη της μεθόδου χωρικής επέκτασης αποτελεί η τεχνική της γεωγραφικά σταθμισμένης παλινδρόμησης (*Geographically Weighted Regression – GWR*) που παρουσιάστηκε από τους Fotheringham, Brunson και Charlton, (2002). Ουσιαστικά

πρόκειται για μια μη παραμετρική μέθοδο παλινδρόμησης που παρέχει διαφορετικούς συντελεστές σε κάθε γεωγραφικό σημείο εκτιμώντας χωριστά υποδείγματα παλινδρόμησης με υποσύνολα του δείγματος που περιλαμβάνουν τις παρατηρήσεις που βρίσκονται κοντά στην εξεταζόμενη περιοχή με διαφορετικούς βαθμούς στάθμισης, ανάλογα με τη γεωγραφική τους απόσταση, που καθορίζονται από την εφαρμογή μιας συνάρτησης στάθμισης γνωστής ως πυρήνας (*kernel*). Η γεωγραφικά σταθμισμένη παλινδρόμηση εξετάζεται αναλυτικά στο τρίτο κεφάλαιο.

1.7 Ανακεφαλαίωση

Στα πλαίσια του κεφαλαίου αυτού έγινε προσπάθεια να παρουσιαστούν οι βασικές έννοιες και τα υποδείγματα που αφορούν τον κλάδο της χωρική οικονομετρίας και να εξηγηθούν οι διαφορές της από την κλασσική οικονομετρία και τον ευρύτερο επιστημονικό κλάδο της χωρικής ανάλυσης. Σκοπός της ανάλυσης ήταν να τονίσει τη χρησιμότητα και την αναγκαιότητά της θέτοντας τις θεωρητικές βάσεις για την κατανόηση των επόμενων κεφαλαίων και όχι να εξαντλήσει όλα τα θέματα της χωρικής οικονομετρίας. Ειδικότερα, αναλύθηκε η ιδιαιτερότητα των χωρικών δεδομένων που οδηγεί στην εμφάνιση των χωρικών επιδράσεων. Οι χωρικές επιδράσεις, δηλαδή η χωρική εξάρτηση και η χωρική ετερογένεια δεν μπορούν να αγνοηθούν από τον ερευνητή κατά την πραγματοποίηση μιας οικονομετρικής ανάλυσης με δεδομένα που συλλέγονται από διαφορετικές γεωγραφικές περιοχές και απαιτούν ειδικό χειρισμό και εξειδικευμένα υποδείγματα.

Στη συνέχεια του κεφαλαίου αναφέρθηκαν οι τρόποι με τους οποίους ορίζεται η μήτρα των χωρικών σταθμίσεων και η εφαρμογή της για τον ορισμό της χωρικής γειτονίας και κατά συνέπεια της χωρικής εξάρτησης. Η μήτρα αυτή είναι απαραίτητη για τον έλεγχο της ύπαρξης χωρικής αυτοσυσχέτισης με το συντελεστή I του Moran και για τον ορισμό των υποδειγμάτων που περιλαμβάνουν χωρική εξάρτηση. Επιπρόσθετα, παρουσιάστηκαν οι χωρικές διαδικασίες ως τυχαία πεδία, δηλαδή διαδικασίες δύο διαστάσεων, και περιγράφηκαν οι βασικές χωρικές διαδικασίες από τις οποίες θεωρείται ότι σχηματίζονται τα χωρικά δεδομένα. Στην πραγματικότητα, οι διαδικασίες αυτές αποτελούν επέκταση στο χώρο των γνωστών διαδικασιών από την ανάλυση χρονοσειρών. Στο τελευταίο μέρος του κεφαλαίου αναλύθηκαν τα δύο βασικά χωρικά οικονομετρικά υποδείγματα που χρησιμοποιούνται στην αντιμετώπιση του προβλήματος της χωρικής αυτοσυσχέτισης, δηλαδή του υποδείγματος

χωρικής υστέρησης και του υποδείγματος χωρικού σφάλματος και περιγράφηκε η μεθοδολογία που συνήθως χρησιμοποιείται για την εκτίμησή τους. Τα δύο αυτά υποδείγματα είναι απαραίτητα για την ανάλυση των αποτελεσμάτων του δευτέρου κεφαλαίου. Τέλος, δείχτηκαν κάποιες μεθοδολογίες που εφαρμόζονται στην κατασκευή υποδειγμάτων που αντιμετωπίζουν τη χωρική ετερογένεια θέμα το οποίο είναι αντικείμενο του τρίτου κεφαλαίου.

Η ανάλυση που έγινε ήταν συνοπτική και αφορούσε μόνο τα απαραίτητα θέματα που θα συναντηθούν στα επόμενα κεφάλαια. Περισσότερες λεπτομέρειες για θέματα χωρικής οικονομετρίας και τα υποδείγματα που εφαρμόζονται μπορούν να βρεθούν στους Anselin (1988), Arbia (2006) και LeSage και Pace (2009).

ΚΕΦΑΛΑΙΟ 2

ΧΩΡΙΚΗ ΑΥΤΟΣΥΣΧΕΤΙΣΗ ΣΤΟ ΓΡΑΜΜΙΚΟ ΥΠΟΔΕΙΓΜΑ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

2.1 Εισαγωγή

Η εκτίμηση υποδειγμάτων γραμμικής παλινδρόμησης με χωρικά δεδομένα, δηλαδή με δεδομένα που προέρχονται από διαφορετικές γεωγραφικές περιοχές, πολλές φορές υποκρύπτει τον κίνδυνο εμφάνισης του προβλήματος της χωρικής αυτοσυσχέτισης στα σφάλματα. Πάρα πολλοί ερευνητές αγνοούν τον κίνδυνο αυτό και θεωρούν ότι λόγω των διαστρωματικών δεδομένων η ετεροσκεδαστικότητα των σφαλμάτων είναι ένα πιθανό πρόβλημα που μπορεί να προκύψει και ότι η αυτοσυσχέτιση εμφανίζεται μόνο όταν το δείγμα αποτελείται από χρονοσειρές. Η χωρική αυτοσυσχέτιση (*spatial autocorrelation*) είναι αποτέλεσμα της χωρικής εξάρτησης (*spatial dependence*), η οποία μαζί με τη χωρική ετερογένεια (*spatial heterogeneity*) αποτελούν τις χωρικές επιδράσεις (*spatial effects*) που οφείλονται στο είδος των δεδομένων. Η ύπαρξη χωρικής αυτοσυσχέτισης, δηλαδή συσχέτισης των παρατηρήσεων σύμφωνα με τις γεωγραφικές τους θέσεις, καθιστά τα αποτελέσματα από την εκτίμηση ενός οικονομετρικού υποδείγματος αναξιόπιστα. Κατά συνέπεια, ο ερευνητής θα πρέπει να είναι σε θέση να εντοπίσει και να αντιμετωπίσει το πρόβλημα αυτό, εκτιμώντας χωρικά οικονομετρικά υποδείγματα που περιλαμβάνουν χωρικές επιδράσεις.

Στο κεφάλαιο αυτό παρουσιάζονται οι πιο συνηθισμένες αιτίες που προκαλούν την εμφάνιση του προβλήματος της χωρικής αυτοσυσχέτισης στα σφάλματα ενός γραμμικού οικονομετρικού υποδείγματος παλινδρόμησης και οι αντίστοιχες συνέπειες στη στατιστική συμπερασματολογία. Στη συνέχεια, εξετάζονται οι πιο διαδομένοι έλεγχοι που χρησιμοποιούνται στην ανίχνευση του προβλήματος και στην επιλογή του κατάλληλου χωρικού υποδείγματος. Με τη βοήθεια προσομοίωσης μελετάται η αξιοπιστία αυτών των ελέγχων για τον εντοπισμό της αιτίας της χωρικής αυτοσυσχέτισης και την εκλογή του σωστού χωρικού υποδείγματος. Τα αποτελέσματα της προσομοίωσης επιβεβαιώνουν

παλιότερα γνωστά στη βιβλιογραφία της χωρικής οικονομετρίας συμπεράσματα για μικρά και μέτρια δείγματα και προσθέτουν περαιτέρω πληροφορίες για τη συμπεριφορά αυτών των ελέγχων στα μεγάλα δείγματα. Επιπρόσθετα, παρουσιάζονται οι έλεγχοι με τους οποίους διερευνάται η ύπαρξη χωρικής αυτοσυσχέτισης στο υπόδειγμα χωρικής υστέρησης της εξαρτημένης μεταβλητής και εξετάζεται η συμπεριφορά τους μέσω προσομοίωσης.

Μια άλλη αιτία που μπορεί να προκαλέσει χωρική αυτοσυσχέτιση στα σφάλματα ενός γραμμικού υποδείγματος είναι η γνωστή από την ανάλυση χρονοσειρών υποφαινομενική παλινδρόμηση (*spurious regression*). Με τη βοήθεια προσομοίωσης μελετάται το πρόβλημα αυτό όταν οι μεταβλητές του υποδείγματος αποτελούν στάσιμες χωρικές αυτοπαλινδρομες διαδικασίες πρώτης τάξης (*Spatial Autoregressive Process–SAR(1)*) και συγκρίνεται με ανάλογα γνωστά αποτελέσματα από την ανάλυση χρονοσειρών. Τα αποτελέσματα της προσομοίωσης δείχνουν ότι υπάρχουν ομοιότητες στην εμφάνιση του φαινομένου στα χωρικά δεδομένα με τα δεδομένα των χρονοσειρών. Ωστόσο, εάν ο ερευνητής ακολουθήσει τις υποδείξεις από την εφαρμογή των ελέγχων χωρικής εξάρτησης και εκτιμήσει κάποιο χωρικό οικονομετρικό υπόδειγμα δεν εμφανίζεται το πρόβλημα της υποφαινομενικής παλινδρόμησης.

2.2 Οι συνέπειες της χωρικής αυτοσυσχέτισης στα σφάλματα του γραμμικού υποδείγματος παλινδρόμησης

Στην οικονομετρική ανάλυση η μέθοδος της ανάλυσης παλινδρόμησης έχει σκοπό τη διερεύνηση της σχέσης ανάμεσα σε μια εξαρτημένη μεταβλητή y και ένα σύνολο ανεξάρτητων μεταβλητών x_1, x_2, \dots, x_k . Συνήθως, θεωρείται ότι η σχέση αυτή είναι γραμμική, οπότε ένα υπόδειγμα που θα μπορούσε να χρησιμοποιηθεί για αυτό το σκοπό είναι το ακόλουθο:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \varepsilon_i \quad \text{για } i = 1, 2, \dots, n$$

το οποίο με τη βοήθεια της άλγεβρας μητρών γράφεται ως εξής:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2.1)$$

όπου \mathbf{y} είναι το $(n \times 1)$ διάνυσμα των παρατηρήσεων της εξαρτημένης μεταβλητής, \mathbf{X} η $(n \times (k+1))$ μήτρα των παρατηρήσεων των k ανεξάρτητων μεταβλητών έχοντας στην πρώτη στήλη μονάδες για την εκτίμηση του σταθερού όρου, $\boldsymbol{\beta}$ το $((k+1) \times 1)$ διάνυσμα των

συντελεστών του υποδείγματος και ε το $(n \times 1)$ διάνυσμα των τυχαίων σφαλμάτων που περιλαμβάνουν όλους τους παράγοντες που επηρεάζουν την εξαρτημένη μεταβλητή και δεν έχουν συμπεριληφθεί στις ανεξάρτητες μεταβλητές του υποδείγματος. Το υπόδειγμα της σχέσης (2.1) δεν λαμβάνει υπ' όψιν του την παρουσία χωρικών επιδράσεων και για το λόγο αυτό θα ονομάζεται απλό οικονομετρικό υπόδειγμα για να διακρίνεται από τα χωρικά υποδείγματα.

Επιπρόσθετα, για την εκτίμηση του παραπάνω υποδείγματος με τη μέθοδο των ελαχίστων τετραγώνων (OLS) χρησιμοποιούνται οι ακόλουθες υποθέσεις:

- 1) $E[\varepsilon] = \mathbf{0}$, δηλαδή τα σφάλματα έχουν μέσο όρο μηδέν.
- 2) $E[\varepsilon\varepsilon'] = \sigma^2\mathbf{I}$, δηλαδή τα σφάλματα έχουν σταθερή διακύμανση και είναι μεταξύ τους ανεξάρτητα, όπου με \mathbf{I} συμβολίζεται η μοναδιαία μήτρα.
- 3) $E[\mathbf{X}'\varepsilon] = \mathbf{0}$, δηλαδή τα σφάλματα δεν συσχετίζονται με τις ανεξάρτητες μεταβλητές.
- 4) $r[\mathbf{X}] = k \Rightarrow |\mathbf{X}'\mathbf{X}| \neq 0$, δηλαδή η μήτρα \mathbf{X} είναι πλήρους τάξεως, που σημαίνει ότι οι ανεξάρτητες μεταβλητές δεν συσχετίζονται γραμμικά μεταξύ τους.
- 5) $\varepsilon \sim N(\mathbf{0}, \sigma^2\mathbf{I})$, δηλαδή τα σφάλματα ακολουθούν την πολυμεταβλητή κανονική κατανομή με μέσο όρο μηδέν και σταθερή διακύμανση.

Όταν δεν ισχύει η υπόθεση της σταθερής διακύμανσης των σφαλμάτων εμφανίζεται το πρόβλημα της ετεροσκεδαστικότητας, όταν παραβιάζεται η υπόθεση της ανεξαρτησίας των σφαλμάτων εμφανίζεται το πρόβλημα της αυτοσυσχέτισης και όταν οι ανεξάρτητες μεταβλητές συσχετίζονται γραμμικά εμφανίζεται το πρόβλημα της πολυσυγγραμμικότητας. Εφόσον τηρούνται όλες οι προηγούμενες υποθέσεις, οι εκτιμητές των συντελεστών του υποδείγματος (2.1) που προκύπτουν από την εφαρμογή της μεθόδου των ελαχίστων τετραγώνων είναι οι $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ που σύμφωνα με το θεώρημα Gauss-Markov ονομάζονται BLUE (Best Linear Unbiased Estimators) διότι είναι γραμμικοί ως προς τις τιμές της εξαρτημένης μεταβλητής, αμερόληπτοι και άριστοι, δηλαδή έχουν τη μικρότερη διακύμανση από όλους τους γραμμικούς και αμερόληπτους εκτιμητές των συντελεστών του υποδείγματος. Η διακύμανση των εκτιμητών σε αυτήν την περίπτωση ορίζεται ως $\text{Var}[\hat{\boldsymbol{\beta}}] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$. Όταν οι τιμές των σφαλμάτων του υποδείγματος (2.1) δεν είναι ανεξάρτητες αλλά παρουσιάζουν κάποια μορφή χωρικής αυτοσυσχέτισης, τότε αποδεικνύεται

ότι η μέθοδος των ελαχίστων τετραγώνων θα οδηγήσει σε εκτιμητές μη αποδοτικούς και ανάλογα με την αιτία δημιουργίας του προβλήματος μπορεί να είναι αμερόληπτοι ή μεροληπτικοί.

Μια αιτία εμφάνισης του φαινομένου της χωρικής αυτοσυσχέτισης στα σφάλματα ενός γραμμικού υποδείγματος μπορεί να είναι η ύπαρξη κάποιου μη μετρήσιμου παράγοντα που επηρεάζει την εξαρτημένη μεταβλητή ή κάποιου σφάλματος μέτρησης στις ανεξάρτητες μεταβλητές. Στην περίπτωση αυτή το πρόβλημα δεν θεωρείται ουσιαστικό για την εξειδίκευση του υποδείγματος και την ερμηνεία της εξαρτημένης μεταβλητής και ο ερευνητής επιδιώκει να το αντιμετωπίσει με τη βοήθεια καταλλήλων οικονομετρικών τεχνικών οι οποίες θα το εξαλείψουν και θα προσδιορίσουν τις σωστές διακυμάνσεις των εκτιμητών όπως συμβαίνει στην ανάλυση χρονοσειρών. Αυτό επιτυγχάνεται με την εκτίμηση του υπόδειγματος χωρικού σφάλματος (*Spatial Error Model–SEM*) που εξειδικεύει και εκτιμά ταυτόχρονα τη χωρική διαδικασία από την οποία σχηματίζονται τα σφάλματα.

Για να εξεταστούν οι συνέπειες της χωρικής αυτοσυσχέτισης που οφείλονται στους λόγους που αναφέρθηκαν στην προηγούμενη παράγραφο θα θεωρηθεί ότι τα σφάλματα σχηματίζονται σύμφωνα με μια στάσιμη χωρική αυτοπαλίνδρομη διαδικασία πρώτης τάξης (*Spatial Autoregressive Process–SAR(1)*), δηλαδή ως εξής:

$$\boldsymbol{\varepsilon} = \lambda \mathbf{W}\boldsymbol{\varepsilon} + \mathbf{u} \quad (2.2)$$

όπου $\mathbf{u} \sim N(0, \sigma^2 \mathbf{I})$, λ είναι ο χωρικά αυτοπαλίνδρομος συντελεστής για τον οποίο γίνεται η υπόθεση ότι $1/\omega_{\min} < \lambda < 1/\omega_{\max}$ και \mathbf{W} είναι η $(n \times n)$ μήτρα χωρικών σταθμίσεων με ω_{\min} και ω_{\max} τη μικρότερη και μεγαλύτερη ιδιοτιμή της. Η σχέση (2.2) ισοδύναμα γράφεται ως:

$$\boldsymbol{\varepsilon} = (\mathbf{I} - \lambda \mathbf{W})^{-1} \mathbf{u} \quad (2.3)$$

Υπολογίζοντας τη μέση τιμή των σφαλμάτων σύμφωνα με τη σχέση (2.3), προκύπτει:

$$E[\boldsymbol{\varepsilon}] = E\left[(\mathbf{I} - \lambda \mathbf{W})^{-1} \mathbf{u}\right] = (\mathbf{I} - \lambda \mathbf{W})^{-1} E[\mathbf{u}] = 0$$

Κατά συνέπεια, τα σφάλματα από την εφαρμογή της μεθόδου των ελαχίστων τετραγώνων στο υπόδειγμα (2.1) θα εξακολουθούν να έχουν μέση τιμή μηδέν.

Η μήτρα διακυμάνσεων – συνδιακυμάνσεων των σφαλμάτων θα είναι:

$$\begin{aligned} E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] &= E\left[\left((\mathbf{I} - \lambda \mathbf{W})^{-1} \mathbf{u}\right)\left((\mathbf{I} - \lambda \mathbf{W})^{-1} \mathbf{u}\right)'\right] = E\left[(\mathbf{I} - \lambda \mathbf{W})^{-1} \mathbf{u}\mathbf{u}'\left((\mathbf{I} - \lambda \mathbf{W})^{-1}\right)'\right] \\ &= (\mathbf{I} - \lambda \mathbf{W})^{-1} E[\mathbf{u}\mathbf{u}']\left((\mathbf{I} - \lambda \mathbf{W})^{-1}\right)' = \sigma^2 (\mathbf{I} - \lambda \mathbf{W})^{-1} \left((\mathbf{I} - \lambda \mathbf{W})^{-1}\right)' \neq \sigma^2 \mathbf{I} \end{aligned}$$

και επειδή ισχύει ότι:

$$(\mathbf{I} - \lambda \mathbf{W})^{-1} = \mathbf{I} + \lambda \mathbf{W} + \lambda^2 \mathbf{W}^2 + \lambda^3 \mathbf{W}^3 + \dots$$

η μήτρα διακυμάνσεων – συνδιακυμάνσεων των τιμών του τυχαίου σφάλματος γράφεται αναλυτικά ως:

$$E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] = \sigma^2 \left[\mathbf{I} + \lambda (\mathbf{W} + \mathbf{W}') + \lambda^2 (\mathbf{W}^2 + \mathbf{W}\mathbf{W}' + \mathbf{W}'^2) + \dots \right]$$

Επομένως, διαπιστώνεται ότι εμφανίζεται πρόβλημα αυτοσυσχέτισης στα σφάλματα τα οποία συσχετίζονται σύμφωνα με τη γεωγραφική τους θέση, όπως αυτή καθορίζεται από τη μήτρα \mathbf{W} , κατά τέτοιον τρόπο ώστε κάθε περιοχή να συσχετίζεται με όλες τις άλλες περιοχές αλλά οι περιοχές που βρίσκονται χωρικά πιο κοντά μεταξύ τους να παρουσιάζουν πιο έντονη συσχέτιση από τις πιο απομακρυσμένες. Ταυτόχρονα, παρατηρείται ότι τα στοιχεία στην κύρια διαγώνιο της μήτρας διακυμάνσεων – συνδιακυμάνσεων δεν είναι σταθερά που σημαίνει ότι εμφανίζεται και το πρόβλημα της ετεροσκεδαστικότητας κάτι που περιπλέκει περισσότερο την ανάλυση. Αυτό έρχεται σε άμεση αντίθεση με την περίπτωση της παλινδρόμησης όπου το δείγμα αποτελείται από χρονοσειρές στην οποία όταν τα σφάλματα σχηματίζονται σύμφωνα με μια αυτοπαλινδρομη διαδικασία εμφανίζεται μόνο το πρόβλημα της αυτοσυσχέτισης και όχι και της ετεροσκεδαστικότητας.

Για τους εκτιμητές των συντελεστών του υποδείγματος προκύπτει ότι ισχύει:

$$\begin{aligned} E[\hat{\boldsymbol{\beta}}] &= E\left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}\right] = E\left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})\right] \\ &= E\left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\varepsilon}\right] \\ &= E\left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'(\mathbf{I} - \lambda \mathbf{W})^{-1} \mathbf{u}\right] \\ &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{X}E[\boldsymbol{\beta}] + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'(\mathbf{I} - \lambda \mathbf{W})^{-1} E[\mathbf{u}] = \boldsymbol{\beta} + \mathbf{0} = \boldsymbol{\beta} \end{aligned}$$

και έτσι, οι εκτιμητές από την εφαρμογή της μεθόδου των ελαχίστων τετραγώνων είναι αμερόληπτοι.

Η μήτρα διακυμάνσεων – συνδιακυμάνσεων των εκτιμητών του υποδείγματος θα είναι:

$$\begin{aligned} Var[\hat{\boldsymbol{\beta}}] &= E\left[\left(\hat{\boldsymbol{\beta}} - E[\hat{\boldsymbol{\beta}}]\right)\left(\hat{\boldsymbol{\beta}} - E[\hat{\boldsymbol{\beta}}]\right)'\right] \\ &= E\left[\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)'\right] \\ &= E\left[\left(\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\varepsilon} - \boldsymbol{\beta}\right)\left(\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\varepsilon} - \boldsymbol{\beta}\right)'\right] \end{aligned}$$

$$\begin{aligned}
&= E \left[\left((\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\varepsilon} \right) \left((\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\varepsilon} \right)' \right] \\
&= E \left[\left((\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\varepsilon} \right) \left(\boldsymbol{\varepsilon}'\mathbf{X} \left((\mathbf{X}'\mathbf{X})^{-1} \right)' \right) \right] \\
&= E \left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \right] \\
&= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}']\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\sigma^2(\mathbf{I}-\lambda\mathbf{W})^{-1} \left((\mathbf{I}-\lambda\mathbf{W})^{-1} \right)' \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}
\end{aligned}$$

Κατά συνέπεια, διαπιστώνεται ότι $\text{Var}[\hat{\boldsymbol{\beta}}] \neq \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ που σημαίνει ότι οι εκτιμητές που προκύπτουν από την εφαρμογή της μεθόδου των ελαχίστων τετραγώνων παύουν να είναι αποδοτικοί, καθώς η μήτρα διακυμάνσεων – συνδιακυμάνσεων των εκτιμητών λαμβάνει μια αρκετά σύνθετη μορφή που καθιστά μη αξιόπιστη τη στατιστική συμπερασματολογία και την εφαρμογή των στατιστικών ελέγχων.

Μια άλλη στοχαστική διαδικασία που μπορεί να θεωρηθεί ότι ακολουθούν τα σφάλματα του υποδείγματος (2.1) είναι η χωρική διαδικασία κινητού μέσου πρώτης τάξης (*Spatial Moving Average Process*–SMA(1)), η οποία ορίζεται ως:

$$\boldsymbol{\varepsilon} = \lambda\mathbf{W}\mathbf{u} + \mathbf{u}$$

ή ισοδύναμα ως:

$$\boldsymbol{\varepsilon} = (\mathbf{I} + \lambda\mathbf{W})\mathbf{u} \quad (2.4)$$

όπου για το λ υποτίθεται ότι $1/\omega_{\min} < -\lambda < 1/\omega_{\max}$.

Εύκολα μπορεί να αποδειχθεί ότι και στην περίπτωση του χωρικού κινητού μέσου τα σφάλματα του απλού οικονομετρικού υποδείγματος θα έχουν μέση τιμή μηδέν αφού σύμφωνα με τη σχέση (2.4) θα ισχύει:

$$E[\boldsymbol{\varepsilon}] = E[(\mathbf{I} + \lambda\mathbf{W})\mathbf{u}] = (\mathbf{I} + \lambda\mathbf{W})E[\mathbf{u}] = \mathbf{0}$$

Η μήτρα διακυμάνσεων – συνδιακυμάνσεων των σφαλμάτων θα είναι:

$$\begin{aligned}
E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] &= E \left[\left((\mathbf{I} + \lambda\mathbf{W})\mathbf{u} \right) \left((\mathbf{I} + \lambda\mathbf{W})\mathbf{u} \right)' \right] = E \left[(\mathbf{I} + \lambda\mathbf{W})\mathbf{u}\mathbf{u}'(\mathbf{I} + \lambda\mathbf{W}') \right] \\
&= E[\mathbf{u}\mathbf{u}'](\mathbf{I} + \lambda\mathbf{W})(\mathbf{I} + \lambda\mathbf{W}') = \sigma^2 \left(\mathbf{I} + \lambda(\mathbf{W} + \mathbf{W}') + \lambda^2\mathbf{W}\mathbf{W}' \right) \neq \sigma^2\mathbf{I}
\end{aligned}$$

και κατά συνέπεια, όπως και η χωρική αυτοπαλίνδρομη διαδικασία, που εξετάστηκε προηγουμένως, έτσι και η χωρική διαδικασία κινητού μέσου θα επιφέρει το πρόβλημα της αυτοσυσχέτισης και της ετεροσκεδαστικότητας στα σφάλματα του υποδείγματος (2.1). Η

ουσιαστική διαφορά ανάμεσα στις δύο διαδικασίες είναι ότι η χωρική αυτοσυσχέτιση που προκαλείται από τη διαδικασία κινητού μέσου είναι πιο ασθενής, αφού η αλληλεπίδραση μεταξύ των περιοχών εμφανίζεται μόνο μεταξύ των πρώτων και δεύτερων γειτόνων σε αντίθεση με την αυτοπαλίνδρομη διαδικασία που όλες οι γεωγραφικές περιοχές αλληλοεξαρτώνται. Ανάλογα αποδεικνύεται ότι οι εκτιμητές του υποδείγματος θα είναι αμερόληπτοι αλλά όχι αποδοτικοί καθώς η διακύμανσή τους θα είναι ίση με:

$$\text{Var}[\hat{\boldsymbol{\beta}}] = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\sigma^2 (\mathbf{I} + \lambda\mathbf{W})(\mathbf{I} + \lambda\mathbf{W}')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

Όπως αναφέρουν οι Anselin και Bera (1998) κάθε διαδικασία που εμπλέκει χωρική εξάρτηση στα σφάλματα θα οδηγεί στην εμφάνιση των προβλημάτων της αυτοσυσχέτισης και της ετεροσκεδαστικότητας. Μοναδική εξαίρεση που δεν θα προκληθεί ετεροσκεδαστικότητα αποτελεί η περίπτωση στην οποία τα στοιχεία της μήτρας \mathbf{W} έχουν οριστεί με τέτοιο τρόπο ώστε όλες οι παρατηρήσεις να έχουν τον ίδιο αριθμό γειτόνων και ίδιες σταθμίσεις.

Ένα άλλο χαρακτηριστικό που προκαλεί την εμφάνιση του προβλήματος της αυτοσυσχέτισης στα σφάλματα του γραμμικού υποδείγματος είναι η ύπαρξη χωρικής αυτοσυσχέτισης στις τιμές της εξαρτημένης μεταβλητής. Στην περίπτωση αυτή, η παρουσία της αυτοσυσχέτισης έχει ουσιαστική σημασία καθώς συνεισφέρει στην ερμηνευτική ικανότητα του υποδείγματος, δηλαδή οι τιμές της εξαρτημένης μεταβλητής σε κάποια γεωγραφική περιοχή επηρεάζονται με άμεσο τρόπο από τις τιμές της εξαρτημένης μεταβλητής στις γειτονικές της περιοχές προκαλώντας μια διαδικασία διάχυσης (*Spillover effect*) των χωρικών επιδράσεων σε όλες τις περιοχές. Αυτός ο τύπος αυτοσυσχέτισης είναι πιο περίπλοκος να αναλυθεί και εάν αγνοηθεί από τον ερευνητή έχει σοβαρές επιπτώσεις στην ερμηνεία και αξιοπιστία των αποτελεσμάτων που προκύπτουν από την εκτίμηση του απλού οικονομετρικού υποδείγματος. Η αντιμετώπιση αυτού του είδους αυτοσυσχέτισης γίνεται με την εκτίμηση του υποδείγματος χωρικής υστέρησης (*Spatial Lag Model-SLM*) που περιλαμβάνει τη χωρική υστέρηση της εξαρτημένης μεταβλητής στις ερμηνευτικές μεταβλητές.

Έστω ότι η πραγματική σχέση ανάμεσα στην εξαρτημένη μεταβλητή και τις ανεξάρτητες μεταβλητές είναι της μορφής:

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2.5)$$

όπου $\mathbf{W}\mathbf{y}$ είναι η χωρική υστέρηση της εξαρτημένης μεταβλητής και ρ ο συντελεστής χωρικής υστέρησης για τον οποίο ισχύει ότι $1/\omega_{\min} < \rho < 1/\omega_{\max}$. Στη σχέση του υποδείγματος

(2.5), που είναι το υπόδειγμα χωρικής υστέρησης, η εξαρτημένη μεταβλητή y σε κάθε περιοχή επηρεάζεται από τις τιμές της στις γειτονικές της περιοχές. Εάν υποθεθεί ότι ο ερευνητής αγνοεί την παρουσία της χωρικής υστέρησης στην εξαρτημένη μεταβλητή και εκτιμήσει με τη μέθοδο των ελαχίστων τετραγώνων το απλό οικονομετρικό υπόδειγμα που ορίστηκε στη σχέση (2.1), τότε η αναμενόμενη τιμή του διανύσματος των εκτιμητών είναι:

$$\begin{aligned} E[\hat{\boldsymbol{\beta}}] &= E\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}\right] = E\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})\right] \\ &= E\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\rho\mathbf{W}\mathbf{y} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\right] \\ &= E[\boldsymbol{\beta}] + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y}E[\rho] + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\boldsymbol{\varepsilon}] \\ &= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y}\rho = \boldsymbol{\beta} + \mathbf{b}_{\mathbf{w}_y, \mathbf{x}}\rho \end{aligned}$$

όπου με $\mathbf{b}_{\mathbf{w}_y, \mathbf{x}}$ συμβολίζεται το διάνυσμα των εκτιμητών από την εκτίμηση με τη μέθοδο των ελαχίστων τετραγώνων του γραμμικού υποδείγματος που έχει ως εξαρτημένη μεταβλητή τη χωρική υστέρηση της εξαρτημένης μεταβλητής και ανεξάρτητες μεταβλητές τις υπόλοιπες ανεξάρτητες μεταβλητές του υποδείγματος (2.5). Επομένως, είναι φανερό ότι οι εκτιμητές που προκύπτουν είναι μη συνεπείς και μεροληπτικοί κάτι που συνεπάγεται ότι και η διακύμανσή τους δεν θα προσδιοριστεί σωστά. Ειδικότερα, εάν θεωρηθεί ότι η εξαρτημένη και οι ανεξάρτητες μεταβλητές αποτελούνται από θετικές τιμές κάτι που συνήθως ισχύει για τα οικονομικά δεδομένα και ο συντελεστής ρ είναι θετικός οι τιμές των εκτιμητών που προκύπτουν θα είναι μεγαλύτερες από τις πραγματικές. Επιπρόσθετα, η παράλειψη της χωρικής υστέρησης της εξαρτημένης μεταβλητής από τις ερμηνευτικές μεταβλητές του υποδείγματος θα οδηγήσει στην εμφάνιση του φαινομένου της χωρικής αυτοσυσχέτισης στα σφάλματα του απλού οικονομετρικού υποδείγματος για τον ίδιο λόγο που εμφανίζεται η χρονική αυτοσυσχέτιση στα σφάλματα ενός οικονομετρικού υποδείγματος που εκτιμάται με χρονοσειρές όταν παραλείπεται η χρονική υστέρηση της εξαρτημένης μεταβλητής εφόσον αυτή συνεισφέρει στην ερμηνεία του υποδείγματος. Αυτό φαίνεται εύκολα καθώς η παράλειψη της χωρικής υστέρησης συνεπάγεται την ενσωμάτωσή της στα σφάλματα του υποδείγματος που θα ορίζονται ως $\mathbf{v} = \rho\mathbf{W}\mathbf{y} + \boldsymbol{\varepsilon}$, οπότε η μήτρα διακυμάνσεων – συνδιακυμάνσεων των σφαλμάτων θα είναι:

$$\begin{aligned} E[\mathbf{v}\mathbf{v}'] &= E\left[(\rho\mathbf{W}\mathbf{y} + \boldsymbol{\varepsilon})(\rho\mathbf{W}\mathbf{y} + \boldsymbol{\varepsilon})'\right] = E\left[(\rho\mathbf{W}\mathbf{y} + \boldsymbol{\varepsilon})(\rho(\mathbf{W}\mathbf{y})' + \boldsymbol{\varepsilon}')\right] \\ &= E\left[\rho^2(\mathbf{W}\mathbf{y})(\mathbf{W}\mathbf{y})' + \rho(\mathbf{W}\mathbf{y})\boldsymbol{\varepsilon}' + \rho\boldsymbol{\varepsilon}(\mathbf{W}\mathbf{y})' + \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\right] \neq \sigma^2\mathbf{I} \end{aligned}$$

στην οποία οι τιμές των στοιχείων εκτός της κυρίας διαγωνίου δεν είναι μηδέν.

Συνοψίζοντας, η χωρική αυτοσυσχέτιση στα σφάλματα του απλού οικονομετρικού υποδείγματος είναι ένα πρόβλημα που μπορεί, αν αγνοηθεί, να δημιουργήσει σοβαρά προβλήματα στην οικονομετρική ανάλυση. Για το λόγο αυτό, ο ερευνητής θα πρέπει να είναι σε θέση να την εντοπίσει και αφού διερευνήσει την αιτία εμφάνισής της να την αντιμετωπίσει με τη βοήθεια του κατάλληλου χωρικού οικονομετρικού υποδείγματος.

2.3 Έλεγχοι για την ύπαρξη χωρικής αυτοσυσχέτισης στα σφάλματα του γραμμικού υποδείγματος παλινδρόμησης

Στην ενότητα αυτή παρουσιάζονται οι πιο σημαντικοί έλεγχοι που συνήθως χρησιμοποιούνται στη χωρική οικονομετρία για τον έλεγχο της ύπαρξης χωρικής αυτοσυσχέτισης στα σφάλματα ενός απλού οικονομετρικού υποδείγματος και τον εντοπισμό της αιτίας δημιουργίας του προβλήματος. Πιο συγκεκριμένα, περιγράφονται η διαδικασία εφαρμογής του ελέγχου με τη στατιστική I του Moran, που αναφέρθηκε στο Κεφάλαιο 1, για τα σφάλματα ενός γραμμικού υποδείγματος παλινδρόμησης και μια κατηγορία διαγνωστικών ελέγχων που συμβάλλουν στον εντοπισμό της γενεσιουργού αιτίας του προβλήματος και στην εξειδίκευση του σωστού χωρικού οικονομετρικού υποδείγματος.

2.3.1 Ο έλεγχος I του Moran για σφάλματα παλινδρόμησης

Η πιο γνωστή διαδικασία έλεγχου για την ύπαρξη χωρικής αυτοσυσχέτισης στα σφάλματα ενός γραμμικού υποδείγματος παλινδρόμησης είναι με την εφαρμογή της στατιστικής I του Moran (1950) όπως έχει προταθεί από τους Cliff και Ord (1972) για την περίπτωση καταλοίπων παλινδρόμησης. Η στατιστική I του Moran ορίζεται ως:

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\hat{\boldsymbol{\varepsilon}}' \mathbf{W} \hat{\boldsymbol{\varepsilon}}}{\hat{\boldsymbol{\varepsilon}}' \hat{\boldsymbol{\varepsilon}}}$$

όπου n είναι το μέγεθος του δείγματος. $\hat{\boldsymbol{\varepsilon}}$ το $(n \times 1)$ διάνυσμα των καταλοίπων που έχουν προκύψει από την εκτίμηση του υποδείγματος με τη μέθοδο των ελαχίστων τετραγώνων, \mathbf{W} είναι η $(n \times n)$ μήτρα χωρικών σταθμίσεων και $\sum_{i=1}^n \sum_{j=1}^n w_{ij}$ ένας παράγοντας κανονικοποίησης που ισούται με το άθροισμα όλων των χωρικών σταθμίσεων. Επειδή για μια τυποποιημένη

κατά γραμμή μήτρα χωρικών σταθμίσεων ισχύει $\sum_{i=1}^n \sum_{j=1}^n w_{ij} = n$ η στατιστική I του Moran

γράφεται στη μορφή:

$$I = \frac{\hat{\boldsymbol{\varepsilon}}' \mathbf{W} \hat{\boldsymbol{\varepsilon}}}{\hat{\boldsymbol{\varepsilon}}' \hat{\boldsymbol{\varepsilon}}}$$

Οι Cliff και Ord (1972) απέδειξαν την ασυμπτωτική κατανομή της στατιστικής I όταν η εξεταζόμενη μεταβλητή αποτελείται από κατάλοιπα παλινδρόμησης. Ειδικότερα, υπό τη μηδενική υπόθεση της απουσίας χωρικής αυτοσχέτισης και υποθέτοντας κανονικότητα η στατιστική I του Moran έχει μέση τιμή και διακύμανση που ορίζονται αντίστοιχα ως:

$$E(I) = \frac{tr[\mathbf{M}\mathbf{W}]}{n - (k + 1)}$$

και

$$Var(I) = \frac{tr[\mathbf{M}\mathbf{W}\mathbf{M}\mathbf{W}'] + tr[\mathbf{M}\mathbf{W}]^2 + (tr[\mathbf{M}\mathbf{W}])^2}{(n - k - 1)(n - k + 1)} - [E(I)]^2,$$

όπου $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ και k είναι ο αριθμός των ανεξάρτητων μεταβλητών του υποδείγματος. Κατά συνέπεια, εφόσον το δείγμα είναι αρκετά μεγάλο, η μετασχηματισμένη μεταβλητή:

$$Z = \frac{I - E(I)}{\sqrt{Var(I)}}$$

ακολουθεί την τυποποιημένη κανονική κατανομή και μπορεί να χρησιμοποιηθεί για τη διεξαγωγή του ελέγχου της στατιστικής σημαντικότητας του I . Η εφαρμογή του ελέγχου που βασίζεται στην κατασκευή, με διαδικασία μεταθέσεων, της εμπειρικής κατανομή της I υπό τη μηδενική υπόθεση ή που βασίζεται στην υπόθεση της τυχαιοποίησης οδηγεί σε αναξιόπιστα αποτελέσματα στην περίπτωση των καταλοίπων παλινδρόμησης.

Αξίζει να αναφερθεί ότι ενώ ως μηδενική υπόθεση του ελέγχου ορίζεται η μη ύπαρξη χωρικής αυτοσχέτισης δεν υπάρχει κάποια συγκεκριμένη μορφή για την εναλλακτική υπόθεση και απλά υποτίθεται η παρουσία κάποιου τύπου χωρικής αυτοσχέτισης. Στην πράξη, έχει αποδειχθεί ότι ο έλεγχος αυτός έχει ισχύ για αρκετές μορφές χωρικής αυτοσχέτισης με πιο συνηθισμένες τη χωρική αυτοπαλίνδρομη διαδικασία και τη χωρική διαδικασία κινητού μέσου. Επιπρόσθετα, η γνωστή από την ανάλυση χρονοσειρών στατιστική των Durbin και Watson αποτελεί ειδική περίπτωση της στατιστικής I όταν η

μήτρα χωρικών σταθμίσεων οριστεί με ειδικό τρόπο.¹ Ωστόσο, η εφαρμογή της στατιστικής I του Moran για χωρικά δεδομένα παρουσιάζει πρόβλημα καθορισμού κριτικών τιμών, κάτι το οποίο δεν ισχύει με το στατιστικό έλεγχο των Durbin και Watson. Το πρόβλημα αυτό επικεντρώνεται στη δυσκολία των μαθηματικών υπολογισμών που απαιτούνται και στο ότι οι προκύπτουσες κριτικές τιμές, με τη μεθοδολογία που έχουν αναπτύξει οι Tiefelsdorf και Boots (1995), εξαρτώνται από τη μήτρα χωρικών σταθμίσεων η οποία διαφέρει σε κάθε χωρική ανάλυση. Για το λόγο αυτό, ο έλεγχος εφαρμόζεται με κριτικές τιμές από την τυπική κανονική κατανομή, όπως έχουν δείξει οι Cliff και Ord (1972).

2.3.2 Οι έλεγχοι χωρικής εξάρτησης

Ο έλεγχος I του Moran αποτελεί σημαντικό διαγνωστικό εργαλείο για τον εντοπισμό της ύπαρξης χωρικής αυτοσυσχέτισης στα σφάλματα ενός γραμμικού υποδείγματος αλλά δεν παρέχει καμία πληροφορία για τον τύπο της αυτοσυσχέτισης και την αιτία που την προκαλεί. Όπως έχει αναφερθεί, η αυτοσυσχέτιση στα σφάλματα μπορεί να είναι αποτέλεσμα χωρικής αυτοσυσχέτισης στην εξαρτημένη μεταβλητή αλλά και κάποιου μη μετρήσιμου παράγοντα που επηρεάζει την εξαρτημένη μεταβλητή. Και στις δύο περιπτώσεις ο ερευνητής θα πρέπει να εκτιμήσει κάποιο υπόδειγμα που περιλαμβάνει χωρικές επιδράσεις. Τα δύο πιο συχνά χρησιμοποιούμενα χωρικά υποδείγματα είναι το υπόδειγμα χωρικής υστέρησης (*Spatial Lag Model*) και το υπόδειγμα χωρικού σφάλματος (*Spatial Error Model*). Στην ενότητα αυτή παρουσιάζονται μια κατηγορία στατιστικών ελέγχων που συνεισφέρουν στον εντοπισμό της αιτίας της εμφάνισης χωρικής αυτοσυσχέτισης και στην εξειδίκευση του σωστού χωρικού οικονομετρικού υποδείγματος. Οι έλεγχοι αυτοί είναι γνωστοί ως έλεγχοι χωρικής εξάρτησης (*Spatial Dependence Tests*) και ανήκουν στην κατηγορία των ελέγχων σκορ ή πολλαπλασιαστή Lagrange (*Score test* ή *Lagrange Multiplier test*). Το βασικό τους πλεονέκτημα είναι ότι ο υπολογισμός των στατιστικών τους βασίζεται αποκλειστικά στα κατάλοιπα που προκύπτουν από την εκτίμηση του απλού οικονομετρικού υποδείγματος, με τη μέθοδο των ελαχίστων τετραγώνων, χωρίς να απαιτείται η εκτίμηση του αντίστοιχου χωρικού οικονομετρικού υποδείγματος. Αξίζει να αναφερθεί ότι σε αντίθεση με τους περισσότερους ελέγχους πολλαπλασιαστή Lagrange που εφαρμόζονται στην οικονομετρία, η διαδικασία εφαρμογής των ελέγχων χωρικής εξάρτησης δεν περιλαμβάνει το στάδιο υπολογισμού του συντελεστή προσδιορισμού από κάποια βοηθητική παλινδρόμηση. Η

¹ Για λεπτομέρειες βλέπε Anselin και Bera (1998) σελίδα 266.

περιγραφή των ελέγχων βασίζεται στο γενικό χωρικό οικονομετρικό υπόδειγμα του Anselin (1988), γνωστό και με την ονομασία SAC (LeSage και Pace, 2009).²

Ειδικότερα, το γενικό χωρικό οικονομετρικό υπόδειγμα, το οποίο παρουσιάζεται λεπτομερώς στην εργασία των Anselin, Bera, Florax, και Yoon (1996), αποτελείται από ένα ταυτόχρονο σύστημα δύο εξισώσεων που μπορούν να ορίσουν το απλό οικονομετρικό υπόδειγμα, το υπόδειγμα χωρικής υστέρησης (*Spatial Lag Model – SLM*) και το υπόδειγμα χωρικού σφάλματος (*Spatial Error Model – SEM*) και έχει ως εξής:

$$\begin{aligned} \mathbf{y} &= \rho \mathbf{W}_1 \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &= \lambda \mathbf{W}_2 \boldsymbol{\varepsilon} + \mathbf{u} \\ \mathbf{u} &\sim N(\mathbf{0}, \sigma^2 \mathbf{I}) \end{aligned} \quad (2.6)$$

όπου \mathbf{y} είναι το $(n \times 1)$ διάνυσμα της εξαρτημένης μεταβλητής, \mathbf{X} η $(n \times (k+1))$ μήτρα των k ανεξάρτητων μεταβλητών περιλαμβάνοντας μία στήλη με μονάδες για την εκτίμηση του σταθερού όρου, $\boldsymbol{\beta}$ το $((k+1) \times 1)$ διάνυσμα των παραμέτρων, $\boldsymbol{\varepsilon}$ και \mathbf{u} τα $(n \times 1)$ διανύσματα των σφαλμάτων και ρ και λ οι συντελεστές χωρικών υστερήσεων της εξαρτημένης μεταβλητής και των σφαλμάτων. Οι $(n \times n)$ μήτρες \mathbf{W}_1 και \mathbf{W}_2 περιέχουν τις χωρικές σταθμίσεις για τον ορισμό των χωρικών υστερήσεων στις δύο εξισώσεις του υποδείγματος. Για λόγους απλοποίησης θεωρείται ότι $\mathbf{W}_1 = \mathbf{W}_2 = \mathbf{W}$ και ότι στη μήτρα \mathbf{X} υπάρχει μία τουλάχιστον ανεξάρτητη μεταβλητή εκτός του σταθερού όρου ώστε να μπορεί το υπόδειγμα (2.6) να ταυτοποιηθεί διότι διαφορετικά μπορεί να εμφανιστούν προβλήματα προσδιορισμού των παραμέτρων ρ και λ . Είναι φανερό ότι εάν στο υπόδειγμα (2.6) τεθούν $\rho = \lambda = 0$ προκύπτει το απλό οικονομετρικό υπόδειγμα, εάν τεθεί $\lambda = 0$ προκύπτει το υπόδειγμα χωρικής υστέρησης και εάν τεθεί $\rho = 0$ το υπόδειγμα χωρικού σφάλματος.

Ο πρώτος έλεγχος αυτής της κατηγορίας εξετάζει την αναγκαιότητα εκτίμησης του υποδείγματος χωρικού σφάλματος σε σχέση με το απλό υπόδειγμα και παρουσιάστηκε αρχικά από τον Burridge (1980). Δηλαδή, υποθέτοντας ότι $\rho = 0$ η μηδενική και η εναλλακτική υπόθεση ορίζονται αντίστοιχα ως:

$$\begin{aligned} H_0 : \lambda = 0 &\quad \Rightarrow \text{Απλό Οικονομετρικό Υπόδειγμα} \\ H_1 : \lambda \neq 0 &\quad \Rightarrow \text{Υπόδειγμα Χωρικού Σφάλματος} \end{aligned}$$

Για να εκτελεστεί ο έλεγχος εκτιμάται με τη μέθοδο των ελαχίστων τετραγώνων το απλό οικονομετρικό υπόδειγμα και υπολογίζεται η στατιστική συνάρτηση:

² Ο Elhorst (2014) στη σελίδα 6 αναφέρει ότι δεν υπάρχει κάποια συγκεκριμένη ερμηνεία του συμβολισμού SAC. Προφανώς, ο συμβολισμός αναφέρεται στα αρχικά των ονομάτων LeSage και Pace (2009).

$$\text{LM-ERR} = \frac{\left[\frac{\hat{\boldsymbol{\varepsilon}}' \mathbf{W} \hat{\boldsymbol{\varepsilon}}}{S^2} \right]^2}{T} \quad (2.7)$$

η οποία υπό τη μηδενική υπόθεση ακολουθεί ασυμπτωτικά την κατανομή χ^2 με ένα βαθμό ελευθερίας. Στη σχέση (2.7) με $\hat{\boldsymbol{\varepsilon}}$ συμβολίζεται το διάνυσμα των καταλοίπων από την εκτίμηση, $S^2 = \hat{\boldsymbol{\varepsilon}}' \hat{\boldsymbol{\varepsilon}} / n$ είναι η εκτιμηθείσα διακύμανση της παλινδρόμησης και $T = \text{tr}[\mathbf{W}^2 + \mathbf{W}'\mathbf{W}]$. Αποδεικνύεται ότι η στατιστική LM-ERR είναι η ίδια και για την περίπτωση που η εναλλακτική υπόθεση του ελέγχου είναι μια χωρική διαδικασία κινητού μέσου της μορφής $\boldsymbol{\varepsilon} = \lambda \mathbf{W}\mathbf{u} + \mathbf{u}$ κάτι που συνεπάγεται ότι πρακτικά είναι αδύνατη η διάκριση μεταξύ των δύο τύπων αυτοσυσχέτισης. Επιπρόσθετα, η στατιστική LM-ERR μπορεί ισοδύναμα να γραφεί ως:

$$\text{LM-ERR} = \frac{\left[n \frac{\hat{\boldsymbol{\varepsilon}}' \mathbf{W} \hat{\boldsymbol{\varepsilon}}}{\hat{\boldsymbol{\varepsilon}}' \hat{\boldsymbol{\varepsilon}}} \right]^2}{T}$$

από όπου προκύπτει ότι η στατιστική αυτή στην πραγματικότητα είναι το τετράγωνο της στατιστικής I του Moran διαιρεμένο με τον παράγοντα T/n^2 κάτι που εξηγεί και την ισχύ της για διαφορετικές μορφές χωρικής αυτοσυσχέτισης.

Ο έλεγχος πολλαπλασιαστή Lagrange για την παρουσία στο υπόδειγμα χωρικής υστέρησης στην εξαρτημένη μεταβλητή προτάθηκε από τον Anselin (1988b) και υποθέτοντας ότι $\lambda = 0$ έχει ως μηδενική και εναλλακτική υπόθεση:

$$\begin{aligned} H_0 : \rho = 0 & \Rightarrow \text{Απλό Οικονομετρικό Υπόδειγμα} \\ H_1 : \rho \neq 0 & \Rightarrow \text{Υπόδειγμα Χωρικής Υστέρησης} \end{aligned}$$

και πραγματοποιείται με βάση την ακόλουθη στατιστική:

$$\text{LM-LAG} = \frac{\left[\frac{\hat{\boldsymbol{\varepsilon}}' \mathbf{W} \mathbf{y}}{S^2} \right]^2}{T + \left[\frac{(\mathbf{W} \mathbf{X} \hat{\boldsymbol{\beta}})' \mathbf{M} (\mathbf{W} \mathbf{X} \hat{\boldsymbol{\beta}})}{S^2} \right]} \quad (2.8)$$

όπου $\hat{\boldsymbol{\beta}}$ είναι το διάνυσμα με τις εκτιμηθείσες τιμές των παραμέτρων του απλού υποδείματος και $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'$. Η στατιστική LM-LAG έχει την ίδια ασυμπτωτική κατανομή με την LM-ERR, δηλαδή ακολουθεί την κατανομή χ^2 με ένα βαθμό ελευθερίας.

Οι στατιστικές LM-ERR και LM-LAG δεν είναι έγκυρες όταν αντίστοιχα $\rho \neq 0$ και $\lambda \neq 0$. Οι Anselin, Bera, Florax, και Yoon (1996) ανέπτυξαν ανθεκτικές (*robust*) μορφές των ελέγχων που διορθώνουν για την ύπαρξη τοπικής λανθασμένης εξειδίκευσης (*local misspecification*) και μπορούν να χρησιμοποιηθούν για την επιλογή του σωστού χωρικού οικονομετρικού υποδείγματος όταν υπάρχει ασάφεια από την εφαρμογή των απλών ελέγχων, σημειώνοντας ότι ως τοπική λανθασμένη εξειδίκευση εννοείται ότι οι τιμές των λ και ρ δεν είναι μηδέν αλλά πολύ κοντά στο μηδέν. Ωστόσο, πρέπει να αναφερθεί πως και οι ανθεκτικές μορφές των ελέγχων αντιμετωπίζουν προβλήματα όταν οι τιμές των λ και ρ είναι πολύ μακριά από το μηδέν. Ο ανθεκτικός έλεγχος της υπόθεσης $H_0 : \lambda = 0$ που λαμβάνει υπ' όψιν την παρουσία χωρικής υστέρησης της εξαρτημένης μεταβλητής πραγματοποιείται με τη στατιστική:

$$\text{LM-EL} = \frac{\left[\frac{\hat{\boldsymbol{\varepsilon}}' \mathbf{W} \hat{\boldsymbol{\varepsilon}}}{S^2} - T \frac{S^2}{S^2 T + (\mathbf{W} \mathbf{X} \hat{\boldsymbol{\beta}})' \mathbf{M}(\mathbf{W} \mathbf{X} \hat{\boldsymbol{\beta}})} \frac{\hat{\boldsymbol{\varepsilon}}' \mathbf{W} \mathbf{y}}{S^2} \right]^2}{T - T^2 \left[\frac{S^2}{S^2 T + (\mathbf{W} \mathbf{X} \hat{\boldsymbol{\beta}})' \mathbf{M}(\mathbf{W} \mathbf{X} \hat{\boldsymbol{\beta}})} \right]} \quad (2.9)$$

και για τον έλεγχο της υπόθεσης $H_0 : \rho = 0$ που λαμβάνει υπ' όψιν την παρουσία χωρικής αυτοσυσχέτισης στα σφάλματα με τη στατιστική:

$$\text{LM-LE} = \frac{\left[\frac{\hat{\boldsymbol{\varepsilon}}' \mathbf{W} \mathbf{y}}{S^2} - \frac{\hat{\boldsymbol{\varepsilon}}' \mathbf{W} \hat{\boldsymbol{\varepsilon}}}{S^2} \right]^2}{\frac{1}{S^2} \left[(\mathbf{W} \mathbf{X} \hat{\boldsymbol{\beta}})' \mathbf{M}(\mathbf{W} \mathbf{X} \hat{\boldsymbol{\beta}}) + S^2 T \right] - T} \quad (2.10)$$

Οι στατιστικές αυτές όπως και οι μη ανθεκτικές εκδόσεις τους ακολουθούν την κατανομή χ^2 με ένα βαθμό ελευθερίας.

Ένας άλλος τρόπος διενέργειας των ελέγχων με τον οποίο δεν απαιτείται η εξασφάλιση ότι η μία από τις δύο χωρικές παραμέτρους λαμβάνει την τιμή μηδέν είναι να πραγματοποιηθεί από κοινού έλεγχος ότι υπάρχει χωρική υστέρηση στην εξαρτημένη μεταβλητή και ταυτόχρονα τα σφάλματα ακολουθούν μια χωρική αυτοπαλίνδρομη διαδικασία ή μια διαδικασία κινητού μέσου. Ο έλεγχος αυτός που παρουσιάστηκε από τον Anselin (1988b) ορίζεται ως:

$$H_0 : \lambda = \rho = 0$$

$$H_1 : \text{Ένα τουλάχιστον από τα } \lambda, \rho \neq 0$$

και γίνεται με τη στατιστική:

$$\text{LM-SARMA} = \frac{\left[\frac{\hat{\boldsymbol{\varepsilon}}' \mathbf{W} \hat{\boldsymbol{\varepsilon}}}{S^2} \right]^2}{T} + \frac{\left[\frac{\hat{\boldsymbol{\varepsilon}}' \mathbf{W} \mathbf{y}}{S^2} - \frac{\hat{\boldsymbol{\varepsilon}}' \mathbf{W} \hat{\boldsymbol{\varepsilon}}}{S^2} \right]^2}{\frac{1}{S^2} \left[(\mathbf{W} \mathbf{X} \hat{\boldsymbol{\beta}})' \mathbf{M} (\mathbf{W} \mathbf{X} \hat{\boldsymbol{\beta}}) + S^2 T \right]} - T \quad (2.11)$$

η οποία υπολογίζεται από την εκτίμηση του απλού υποδείγματος και υπό τη μηδενική υπόθεση ακολουθεί την κατανομή χ^2 με δύο βαθμούς ελευθερίας. Από τη σχέση (2.11) διαπιστώνεται ότι $\text{LM-SARMA} = \text{LM-ERR} + \text{LM-LE}$. Ανάλογα, μετά από μερικές αλγεβρικές πράξεις, προκύπτει ότι $\text{LM-SARMA} = \text{LM-LAG} + \text{LM-EL}$. Κατά συνέπεια, η στατιστική LM-SARMA αποτελείται από το άθροισμα της στατιστικής του απλού ελέγχου μίας χωρικής παραμέτρου με τη στατιστική του ανθεκτικού ελέγχου της άλλης παραμέτρου.

Ο από κοινού έλεγχος για την ύπαρξη των δύο τύπων χωρικής εξάρτησης έχει δύο βασικά μειονεκτήματα. Το πρώτο είναι ότι όταν απορρίπτεται η μηδενική υπόθεση ο ερευνητής δεν έχει καμία ένδειξη για το χωρικό οικονομετρικό υπόδειγμα που θα πρέπει να εκτιμήσει και το δεύτερο είναι ότι ο έλεγχος έχει απώλεια σε ισχύ όταν στην πραγματικότητα υπό τη μηδενική υπόθεση εμφανίζεται μόνο η μία μορφή χωρικής εξάρτησης.

Οι έλεγχοι χωρικής εξάρτησης που παρουσιάστηκαν μερικές φορές οδηγούν σε αντιφατικά συμπεράσματα ως προς την επιλογή του σωστού χωρικού οικονομετρικού υποδείγματος. Στις περιπτώσεις που υπάρχει ασάφεια από τα αποτελέσματα των ελέγχων οι Anselin και Florax (1995) και οι Anselin, Bera, Florax, και Yoon (1996) προτείνουν την ακόλουθη στρατηγική εντοπισμού του σωστού υποδείγματος. Αρχικά διεξάγονται οι έλεγχοι με τις στατιστικές LM-ERR και LM-LAG. Εάν κανένας από τους δύο ελέγχους δεν απορρίπτει τη μηδενική του υπόθεση τότε πρέπει να εξεταστούν τα κατάλοιπα του απλού οικονομετρικού υποδείγματος για την πιθανότητα ύπαρξης άλλων προβλημάτων που μπορεί να οδήγησαν στην εσφαλμένη απόρριψη του ελέγχου χωρικής αυτοσυσχέτισης με τη στατιστική I του Moran όπως η ετεροσκεδαστικότητα και η έλλειψη κανονικότητας. Εάν στον έναν από τους δύο ελέγχους απορρίπτεται η μηδενική υπόθεση, τότε επιλέγεται και εκτιμάται το αντίστοιχο χωρικό υπόδειγμα. Στην περίπτωση που απορρίπτεται η μηδενική υπόθεση και στους δύο απλούς ελέγχους διεξάγονται οι έλεγχοι με τις στατιστικές LM-EL και LM-LE και επιλέγεται το χωρικό υπόδειγμα του ελέγχου που απορρίπτεται η μηδενική του υπόθεση. Εάν και στους δύο

ανθεκτικούς ελέγχους απορρίπτεται η μηδενική υπόθεση τότε επιλέγεται και εκτιμάται το υπόδειγμα που υποδεικνύει η ανθεκτική στατιστική με τη μεγαλύτερη τιμή.

2.4 Το μέγεθος και η ισχύς των ελέγχων χωρικής εξάρτησης

Οι έλεγχοι χωρικής εξάρτησης έχουν γίνει γενικά αποδεκτοί στη βιβλιογραφία σαν ένα σύνολο διαγνωστικών εργαλείων που βοηθάνε στον εντοπισμό του προβλήματος της χωρικής αυτοσυσχέτισης στα σφάλματα του απλού οικονομετρικού υποδείγματος και στην επίλυσή του μέσω της εκτίμησης του κατάλληλου χωρικού υποδείγματος, κάτι που φανερώνεται από τη χρησιμοποίησή τους σε μεγάλο αριθμό εμπειρικών εφαρμογών που έχουν δημοσιευτεί τα τελευταία χρόνια. Όμως, όπως έχει αναφερθεί, αρκετές φορές στην πράξη η εφαρμογή τους προκαλεί σύγχυση καθώς δεν γίνεται πάντα ξεκάθαρο το χωρικό υπόδειγμα που θα πρέπει να εκτιμηθεί. Η αποδοχή των ελέγχων αυτών από τους ερευνητές και η προτεινόμενη διαδικασία επιλογής του χωρικού οικονομετρικού υποδείγματος βασίζονται σε μεγάλο βαθμό στα αποτελέσματα των προσομοιώσεων που περιγράφονται στις εργασίες των Anselin και Rey (1991), Anselin και Florax (1995), καθώς και στους Anselin, Bera, Florax και Yoon (1996). Στην ενότητα αυτή εξετάζεται με τη βοήθεια προσομοίωσης η συμπεριφορά και η επίδοση των ελέγχων χωρικής εξάρτησης για διαφορετικά μεγέθη δείγματος και μεγαλύτερο αριθμό επαναλήψεων σε σχέση με τις προηγούμενες εργασίες. Στόχος είναι να διερευνηθεί περαιτέρω η στατιστική ισχύς των ελέγχων αυτών και η δυνατότητά τους να υποδεικνύουν το σωστό χωρικό υπόδειγμα.

Για λόγους σύγκρισης με τις προαναφερθείσες εργασίες διατηρείται ο ίδιος σχεδιασμός στην προσομοίωση που θεωρεί υπό τη μηδενική υπόθεση της απουσίας χωρικής αυτοσυσχέτισης το απλό οικονομετρικό υπόδειγμα $y = X\beta + \varepsilon$, όπου X είναι μία $(n \times 3)$ μήτρα που αποτελείται από δύο ανεξάρτητες μεταβλητές και μία στήλη με μονάδες για την εκτίμηση του σταθερού όρου και β το (3×1) διάνυσμα των συντελεστών που θέτονται και οι τρεις ίσοι με τη μονάδα. Οι δύο ανεξάρτητες μεταβλητές προέρχονται από ομοιόμορφη $U(0, 10)$ κατανομή και παρήχθησαν μόνο μία φορά παραμένοντας σταθερές σε όλες τις επαναλήψεις. Το $(n \times 1)$ διάνυσμα των τυχαίων σφαλμάτων ε προέρχεται από $N(0, I)$ κατανομή και προστίθεται στο διάνυσμα $X\beta$ για να παραχθεί το $(n \times 1)$ διάνυσμα της εξαρτημένης μεταβλητής y .

Η χωρική εξάρτηση ορίστηκε με μία ($n \times n$) τυποποιημένη κατά γραμμή μήτρα χωρικών σταθμίσεων \mathbf{W} που κατασκευάστηκε με βάση τα κριτήρια του πύργου (*rook contiguity*) και τη βασίλισσας (*queen contiguity*) για ένα ομαλό τετραγωνικό πλέγμα διαστάσεων 7x7, 10x10, 20x20 και 30x30 παρέχοντας δείγματα 49, 100, 400 και 900 παρατηρήσεων αντίστοιχα. Επιπρόσθετα, για κάθε κριτήριο χωρικής γειννίας εξετάστηκε και η εφαρμογή διόρθωσης τόρου (*torus correction*) ώστε να εξαλειφθεί η μεροληψία που μπορεί να προκληθεί από τις ακραίες χωρικές μονάδες και όλες να έχουν τον ίδιο αριθμό γειτόνων. Τα χωρικά οικονομετρικά υποδείγματα της εναλλακτικής υπόθεσης είναι το υπόδειγμα χωρικής υστέρησης (*Spatial Lag Model-SLM*) και το υπόδειγμα χωρικού σφάλματος (*Spatial Error Model-SEM*).

Το υπόδειγμα χωρικής υστέρησης παράγεται λύνοντας τη σχέση (2.5) ως προς \mathbf{y} , δηλαδή ουσιαστικά πολλαπλασιάζοντας τη δεξιά πλευρά του απλού οικονομετρικού υποδείγματος με το χωρικό πολλαπλασιαστή $(\mathbf{I} - \rho\mathbf{W})^{-1}$ οπότε προκύπτει:

$$\mathbf{y} = (\mathbf{I} - \rho\mathbf{W})^{-1} \mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \rho\mathbf{W})^{-1} \boldsymbol{\varepsilon}$$

Για το υπόδειγμα χωρικού σφάλματος θεωρούνται οι περιπτώσεις που τα σφάλματα ακολουθούν μια στάσιμη χωρική αυτοπαλίνδρομη διαδικασία πρώτης τάξης, (SAR(1)), και μια χωρική διαδικασία κινητού μέσου πρώτης τάξης, (SMA(1)). Σε κάθε περίπτωση το διάνυσμα της εξαρτημένης μεταβλητής παράγεται προσθέτοντας τη σχέση (2.3) και τη σχέση (2.4) στο απλό οικονομετρικό υπόδειγμα, δηλαδή αντίστοιχα ως:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \lambda\mathbf{W})^{-1} \mathbf{u}$$

και

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + (\mathbf{I} + \lambda\mathbf{W})\mathbf{u}$$

όπου το ($n \times 1$) διάνυσμα \mathbf{u} προέρχεται από $N(0, \mathbf{I})$ κατανομή. Οι συντελεστές ρ και λ των χωρικών υποδειγμάτων θεωρήθηκε ότι λαμβάνουν τις τιμές 0,0, 0,2, 0,5, 0,8 και 0,9.

Η διαδικασία της προσομοίωσης πραγματοποιήθηκε στο στατιστικό πρόγραμμα R με τη βοήθεια του πακέτου SPDEP που έχει αναπτυχθεί από τον Bivand (2015). Εκτελέστηκαν 10.000 επαναλήψεις και για κάθε χωρικό υπόδειγμα εκτιμήθηκε με τη μέθοδο των ελαχίστων τετραγώνων το απλό οικονομετρικό υπόδειγμα και κατεγράφησαν η μέση τιμή της στατιστικής I του Moran (1950) για την ύπαρξη χωρικής αυτοσυσχέτισης στα σφάλματα και το ποσοστό απόρριψης της μηδενικής υπόθεσης του σχετικού ελέγχου για την απουσία χωρικής αυτοσυσχέτισης με την προσέγγιση των Cliff και Ord (1972). Επιπρόσθετα,

εξετάστηκαν τα ποσοστά απόρριψης των LM ελέγχων χωρικής εξάρτησης που χρησιμοποιούνται για τον εντοπισμό της αιτίας της αυτοσυσχέτισης και την επιλογή των χωρικών οικονομετρικών υποδειγμάτων θεωρώντας επίπεδο σημαντικότητας 5%.

Πίνακας 2.1

Μέσες τιμές I και ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% των ελέγχων χωρικής εξάρτησης για το απλό οικονομετρικό υπόδειγμα για όλα τα μεγέθη δείγματος

| N | | Τόρος Όχι | | Τόρος Ναι | |
|----------------------|----------|------------|-----------|------------|-----------|
| | | Σταθμίσεις | | Σταθμίσεις | |
| | | Πύργος | Βασίλισσα | Πύργος | Βασίλισσα |
| 49 | Μέσο I | -0,030 | -0,026 | -0,027 | -0,025 |
| Ποσοστά Απόρριψης | I | 4,83 | 5,31 | 5,05 | 5,08 |
| | LM-LAG | 5,91 | 5,56 | 5,82 | 5,36 |
| | LM-LE | 6,28 | 5,86 | 6,23 | 5,87 |
| | LM-ERR | 4,78 | 3,86 | 4,70 | 3,96 |
| | LM-EL | 5,13 | 4,36 | 4,94 | 4,13 |
| | LM-SARMA | 5,91 | 4,59 | 5,58 | 4,59 |
| 100 | Μέσο I | -0,010 | -0,010 | -0,010 | -0,010 |
| Ποσοστά Απόρριψης | I | 4,79 | 4,77 | 5,01 | 4,80 |
| | LM-LAG | 4,95 | 5,12 | 5,18 | 4,94 |
| | LM-LE | 5,39 | 5,37 | 5,12 | 5,59 |
| | LM-ERR | 4,59 | 4,24 | 4,68 | 4,06 |
| | LM-EL | 4,81 | 4,51 | 5,00 | 4,34 |
| | LM-SARMA | 4,96 | 4,51 | 5,09 | 4,74 |
| 400 | Μέσο I | -0,003 | -0,002 | -0,002 | -0,002 |
| Ποσοστά Απόρριψης | I | 4,84 | 5,08 | 4,93 | 4,83 |
| | LM-LAG | 4,99 | 4,82 | 4,91 | 4,68 |
| | LM-LE | 5,01 | 4,91 | 5,19 | 4,81 |
| | LM-ERR | 4,75 | 4,97 | 4,87 | 4,77 |
| | LM-EL | 5,14 | 5,03 | 5,11 | 4,82 |
| | LM-SARMA | 4,85 | 4,92 | 4,95 | 4,56 |
| 900 | Μέσο I | -0,001 | -0,001 | -0,001 | -0,001 |
| Ποσοστά Απόρριψης | I | 4,97 | 5,05 | 4,92 | 4,81 |
| | LM-LAG | 4,95 | 4,78 | 5,08 | 4,55 |
| | LM-LE | 4,94 | 4,97 | 4,87 | 4,93 |
| | LM-ERR | 4,89 | 5,01 | 4,96 | 4,72 |
| | LM-EL | 5,04 | 5,02 | 4,74 | 4,88 |
| | LM-SARMA | 4,82 | 5,01 | 4,84 | 4,81 |

Ο Πίνακας 2.1 περιλαμβάνει τις μέσες τιμές της στατιστικής I του Moran και τα ποσοστά απόρριψης (μέγεθος – σφάλμα τύπου I) των ελέγχων χωρικής εξάρτησης που προέκυψαν από την προσομοίωση όταν δεν εμφανίζεται καμία μορφή χωρικής εξάρτησης στο υπόδειγμα ($\rho=0$ και $\lambda=0$), δηλαδή όταν ισχύει το απλό οικονομετρικό υπόδειγμα. Όπως φαίνεται από αυτόν τον πίνακα, οι μέσες τιμές της στατιστικής I είναι όλες πολύ κοντά στο μηδέν και

συγκεκριμένα πολύ κοντά στις θεωρητικές αναμενόμενες τιμές της στατιστικής, όπως έχουν δείξει οι Cliff και Ord (1972). Οι θεωρητικές αναμενόμενες τιμές της στατιστικής I έχουν υπολογιστεί για λόγους σύγκρισης σύμφωνα με τη σχέση $tr[\mathbf{MW}]/(N-3)$ και παρουσιάζονται στον Πίνακα 2.2 για όλα τα μεγέθη δείγματος και μήτρες χωρικών σταθμίσεων. Επισημαίνεται ότι αυτές οι θεωρητικές τιμές εξαρτώνται μόνο από τη μορφή της μήτρας χωρικής στάθμισης που ορίζεται εξωγενώς και τις ανεξάρτητες μεταβλητές που έχουν παραμείνει σταθερές σε όλες τις επαναλήψεις.

Πίνακας 2.2
Θεωρητικές αναμενόμενες τιμές της στατιστικής I για το απλό οικονομετρικό υπόδειγμα για όλα τα μεγέθη δείγματος και μήτρες χωρικών σταθμίσεων

| N | Τόρος Όχι | | Τόρος Ναι | |
|------------|-------------|-------------|-------------|-------------|
| | Σταθμίσεις | | Σταθμίσεις | |
| | Πύργος | Βασίλισσα | Πύργος | Βασίλισσα |
| 49 | -0,03000057 | -0,02625530 | -0,02762416 | -0,02518227 |
| 100 | -0,01006390 | -0,00961985 | -0,01019006 | -0,00987551 |
| 400 | -0,00233129 | -0,00233487 | -0,00231588 | -0,00231720 |
| 900 | -0,00111152 | -0,00110318 | -0,00111544 | -0,00111298 |

Τα ποσοστά απόρριψης της στατιστικής I , που συνοψίζονται στον Πίνακα 2.1, είναι όλα κοντά στο επίπεδο σημαντικότητας του 5% κάτι που φανερώνει την απουσία αυτοσυσχέτισης. Ανάλογα αποτελέσματα εμφανίζονται και για τους πέντε ελέγχους πολλαπλασιαστή Lagrange για τον εντοπισμό της αιτίας εμφάνισης της χωρικής αυτοσυσχέτισης για τους οποίους τα ποσοστά απόρριψης είναι κοντά στο 5%. Εξάιρεση υπάρχει μόνο στα μικρά δείγματα των 49 και 100 παρατηρήσεων στα οποία τα ποσοστά απόρριψης των στατιστικών LM-LAG, LM-LE και LM-SARMA είναι λίγο μεγαλύτερα από το 5%, ενώ των στατιστικών LM-ERR και LM-EL είναι λίγο μικρότερα. Συμπερασματικά, η συμπεριφορά των ελέγχων, όπως εμφανίζεται με το μέγεθός τους όταν ισχύει το απλό οικονομετρικό υπόδειγμα και απουσιάζει το πρόβλημα της χωρικής αυτοσυσχέτισης, κρίνεται ότι είναι ικανοποιητική.

Τα αποτελέσματα της προσομοίωσης για την περίπτωση που υπάρχει χωρική υστέρηση στην εξαρτημένη μεταβλητή αλλά έχει αγνοηθεί από τον ερευνητή εκτιμώντας το απλό οικονομετρικό υπόδειγμα για όλες τις τιμές του συντελεστή ρ και για όλα τα μεγέθη δείγματος παρουσιάζονται στον Πίνακα 2.3 χωρίς διόρθωση τόρος και στον Πίνακα 2.4 με διόρθωση τόρος. Οι πίνακες αυτοί περιλαμβάνουν για τη δυνατότητα συγκρίσεων και τις

περιπτώσεις απουσίας χωρικής εξάρτησης που παρουσιάστηκαν στον Πίνακα 2.1. Αναμφισβήτητα, καθώς αυξάνεται η τιμή του ρ η μέση τιμή της στατιστική I αυξάνεται κάτι που συμβαίνει και με τα ποσοστά απόρριψης του αντίστοιχου ελέγχου δηλώνοντας την έντονη παρουσία χωρικής αυτοσυσχέτισης. Τα ποσοστά αυτά για μεγάλα μεγέθη δείγματος προσεγγίζουν το 100% ακόμη και για χαμηλές τιμές του ρ ανεξάρτητα από τη μήτρα χωρικών σταθμίσεων που χρησιμοποιείται και την εφαρμογή διόρθωσης τόρος. Κατά συνέπεια, διαπιστώνεται η ικανότητα της στατιστικής I να εντοπίζει το πρόβλημα στα κατάλοιπα από τη εκτίμηση του απλού γραμμικού υποδείγματος παλινδρόμησης.

Οι συμπεριφορές της στατιστικής LM-LAG για την παρουσία χωρικής υστέρησης στην εξαρτημένη μεταβλητή και της ανθεκτικής της εκδοχής LM-LE κρίνονται εξαιρετικές, καθώς απορρίπτουν τη μηδενική υπόθεση της ύπαρξης χωρικής υστέρησης στην εξαρτημένη μεταβλητή σε ποσοστό σχεδόν 100% ακόμη και στα μικρά δείγματα και για τιμές του ρ ίσες με 0,2. Τα ποσοστά απόρριψης της στατιστικής LM-LAG είναι μεγαλύτερα από τα αντίστοιχα της στατιστικής I , ιδιαίτερα στα μικρά δείγματα, κάτι που φανερώνει ότι η στατιστική LM-LAG έχει περισσότερη ικανότητα να εντοπίζει το πρόβλημα. Ταυτόχρονα, διαπιστώνεται ότι η στατιστική LM-ERR για τον έλεγχο του υποδείγματος χωρικού σφάλματος έχει σημαντική στατιστική ισχύ, ιδιότητα που δεν θα έπρεπε να συμβαίνει, με ποσοστά απόρριψης που ειδικά σε μεγάλα δείγματα προσεγγίζουν το 100% ακόμη και για μικρά ρ , προκαλώντας στον ερευνητή σύγχυση ως προς τον εντοπισμό της αιτίας της αυτοσυσχέτισης. Το αποτέλεσμα αυτό δεν πρέπει να προκαλεί εντύπωση αν ληφθεί υπ' όψιν η άμεση σχέση της στατιστικής LM-ERR με τη στατιστική I του Moran.

Αξιοσημείωτο είναι το γεγονός ότι η ανθεκτική της εκδοχή, δηλαδή η στατιστική LM-EL, παρουσιάζει ιδιόζουσα συμπεριφορά και ειδικότερα, όταν η χωρική εξάρτηση ορίζεται με το κριτήριο του πύργου το ποσοστό απόρριψης της στατιστικής LM-EL για μικρές τιμές του ρ είναι κοντά στο 5% αλλά για μεγάλες τιμές μειώνεται και τελικά μηδενίζεται.³ Το πρόβλημα όμως αυτό στην πράξη δεν προκαλεί δυσκολία στην επιλογή του υποδείγματος χωρικής υστέρησης.

³ Η ιδιότροπη συμπεριφορά της στατιστικής LM-EL με το κριτήριο του πύργου είχε αναφερθεί και στις εργασίες των Anselin και Florax (1995) καθώς και Anselin, Bera, Florax, και Yoon (1996). Πρέπει όμως να σημειωθεί ότι στις εργασίες αυτές το μεγαλύτερο δείγμα που χρησιμοποιήθηκε αφορά 127 παρατηρήσεις και από όλες τις περιπτώσεις μόνο δείγμα 81 παρατηρήσεων ορίστηκε με βάση ένα ομαλό τετραγωνικό πλέγμα. Κατά συνέπεια, η συμπεριφορά της LM-EL σε πολύ μεγάλα δείγματα που με το κριτήριο της βασίλισσας προσεγγίζει ποσοστά απόρριψης 100% προκαλώντας ασάφεια δεν είχε εντοπιστεί.

Πίνακας 2.3
Μέσες τιμές I και ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% των
ελέγχων χωρικής εξάρτησης για όταν ισχύει το υπόδειγμα χωρικής υστέρησης
για όλες τις τιμές του ρ και μεγέθη δείγματος
Χωρίς διόρθωση τόρος

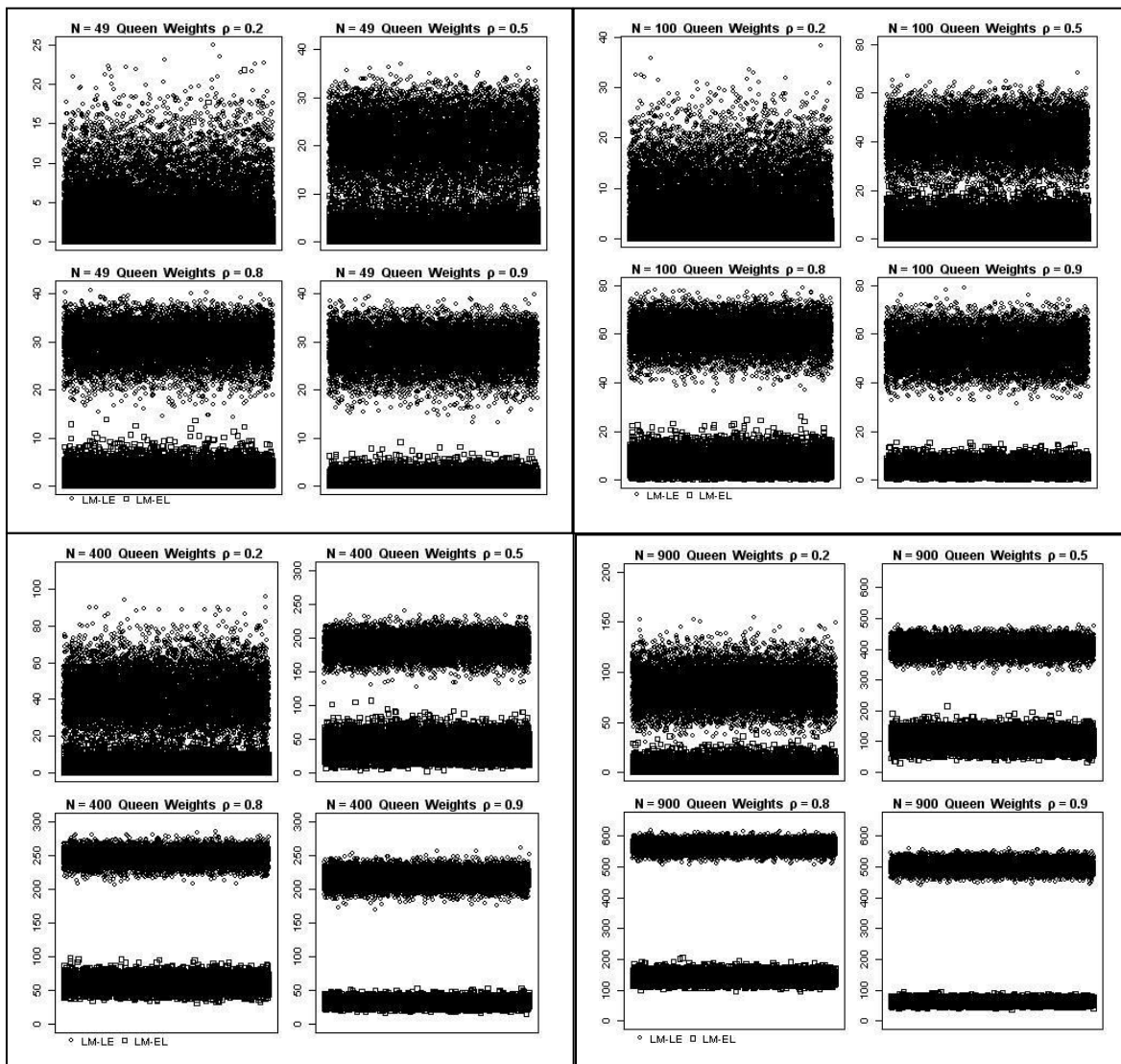
| N | Σταθμίσεις | Πύργος | | | | | Βασίλισσα | | | | |
|----------------------|------------|--------|-------|-------|-------|-------|-----------|-------|-------|-------|-------|
| | ρ | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 |
| 49 | Μέσο I | -0,030 | 0,050 | 0,147 | 0,354 | 0,480 | -0,026 | 0,056 | 0,275 | 0,535 | 0,627 |
| Ποσοστά Απόρριψης | I | 4,83 | 11,64 | 37,20 | 96,98 | 100 | 5,31 | 22,67 | 92,50 | 100 | 100 |
| | LM-LAG | 5,91 | 80,33 | 100 | 100 | 100 | 5,56 | 57,91 | 100 | 100 | 100 |
| | LM-LE | 6,28 | 78,77 | 100 | 100 | 100 | 5,86 | 55,39 | 99,98 | 100 | 100 |
| | LM-ERR | 4,78 | 6,50 | 23,71 | 92,73 | 99,80 | 3,86 | 13,45 | 85,69 | 100 | 100 |
| | LM-EL | 5,13 | 8,74 | 16,86 | 5,18 | 0,8 | 4,36 | 5,09 | 13,30 | 14,09 | 2,94 |
| | LM-SARMA | 5,91 | 70,75 | 100 | 100 | 100 | 4,59 | 48,52 | 99,96 | 100 | 100 |
| 100 | Μέσο I | -0,010 | 0,088 | 0,249 | 0,493 | 0,634 | -0,01 | 0,079 | 0,315 | 0,591 | 0,71 |
| Ποσοστά Απόρριψης | I | 4,79 | 26,55 | 94,98 | 100 | 100 | 4,77 | 39,05 | 99,91 | 100 | 100 |
| | LM-LAG | 4,95 | 98,71 | 100 | 100 | 100 | 5,12 | 85,98 | 100 | 100 | 100 |
| | LM-LE | 5,39 | 97,76 | 100 | 100 | 100 | 5,37 | 82,04 | 100 | 100 | 100 |
| | LM-ERR | 4,59 | 21,39 | 92,86 | 100 | 100 | 4,24 | 31,67 | 99,81 | 100 | 100 |
| | LM-EL | 4,81 | 5,23 | 3,70 | 0,28 | 0,03 | 4,51 | 6,46 | 49,43 | 80,28 | 56,8 |
| | LM-SARMA | 4,96 | 96,91 | 100 | 100 | 100 | 4,51 | 80 | 100 | 100 | 100 |
| 400 | Μέσο I | -0,003 | 0,100 | 0,292 | 0,576 | 0,719 | -0,002 | 0,093 | 0,353 | 0,655 | 0,774 |
| Ποσοστά Απόρριψης | I | 4,84 | 79,73 | 100 | 100 | 100 | 5,08 | 91,56 | 100 | 100 | 100 |
| | LM-LAG | 4,99 | 100 | 100 | 100 | 100 | 4,82 | 100 | 100 | 100 | 100 |
| | LM-LE | 5,01 | 100 | 100 | 100 | 100 | 4,91 | 100 | 100 | 100 | 100 |
| | LM-ERR | 4,75 | 77,50 | 100 | 100 | 100 | 4,97 | 90,28 | 100 | 100 | 100 |
| | LM-EL | 5,14 | 5,29 | 1,65 | 0,03 | 0,00 | 5,03 | 23,92 | 99,98 | 100 | 100 |
| | LM-SARMA | 4,85 | 100 | 100 | 100 | 100 | 4,92 | 100 | 100 | 100 | 100 |
| 900 | Μέσο I | -0,001 | 0,105 | 0,306 | 0,591 | 0,728 | -0,001 | 0,094 | 0,368 | 0,677 | 0,789 |
| Ποσοστά Απόρριψης | I | 4,97 | 99,12 | 100 | 100 | 100 | 5,05 | 99,75 | 100 | 100 | 100 |
| | LM-LAG | 4,95 | 100 | 100 | 100 | 100 | 4,78 | 100 | 100 | 100 | 100 |
| | LM-LE | 4,94 | 100 | 100 | 100 | 100 | 4,97 | 100 | 100 | 100 | 100 |
| | LM-ERR | 4,89 | 98,95 | 100 | 100 | 100 | 5,01 | 99,7 | 100 | 100 | 100 |
| | LM-EL | 5,04 | 5,30 | 8,75 | 2,09 | 0,00 | 5,02 | 48,68 | 100 | 100 | 100 |
| | LM-SARMA | 4,82 | 100 | 100 | 100 | 100 | 5,01 | 100 | 100 | 100 | 100 |

Αντίθετα, σύγχυση προκαλείται από τη στατιστική LM-EL όταν χρησιμοποιείται το κριτήριο της βασίλισσας σε μεγάλα δείγματα και για τιμές του ρ μεγαλύτερες του 0,5, καθώς το ποσοστό απόρριψης είναι 100% κάτι που επιφέρει ασάφεια για την επιλογή μεταξύ του υποδείγματος χωρικής υστέρησης και του υποδείγματος χωρικού σφάλματος. Στις περιπτώσεις αυτές έχει ενδιαφέρον να μελετηθούν οι τιμές των στατιστικών LM-LE και LM-EL που προέκυψαν από την προσομοίωση. Στο Διάγραμμα 2.1 απεικονίζονται οι αριθμητικές τιμές των δύο ανθεκτικών στατιστικών για τα τέσσερα διαφορετικά δείγματα και για ρ ίσα με 0,2, 0,5, 0,8 και 0,9.

Πίνακας 2.4
Μέσες τιμές I και ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% των
ελέγχων χωρικής εξάρτησης για όταν ισχύει το υπόδειγμα χωρικής υστέρησης
για όλες τις τιμές του ρ και μεγέθη δείγματος
Με διόρθωσης τóρος

| N | Σταθμίσεις | Πύργος | | | | | Βασίλισσα | | | | |
|----------------------|------------|--------|-------|-------|-------|-------|-----------|-------|-------|-------|-------|
| | ρ | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 |
| 49 | Μέσο I | -0,027 | 0,046 | 0,146 | 0,300 | 0,374 | -0,025 | 0,035 | 0,189 | 0,346 | 0,394 |
| Ποσοστά Απόρριψης | I | 5,05 | 11,75 | 43,12 | 97,28 | 99,92 | 5,08 | 18,95 | 86,70 | 100 | 100 |
| | LM-LAG | 5,82 | 77,37 | 100 | 100 | 100 | 5,36 | 50,91 | 99,90 | 100 | 100 |
| | LM-LE | 6,23 | 76,71 | 100 | 100 | 100 | 5,87 | 49,71 | 99,88 | 100 | 100 |
| | LM-ERR | 4,70 | 6,46 | 27,85 | 92,23 | 99,51 | 3,96 | 9,19 | 73,51 | 99,93 | 100 |
| | LM-EL | 4,94 | 6,73 | 9,02 | 3,63 | 1,28 | 4,13 | 4,05 | 6,43 | 5,64 | 2,08 |
| | LM-SARMA | 5,58 | 67,57 | 100 | 100 | 100 | 4,59 | 41,33 | 99,75 | 100 | 100 |
| 100 | Μέσο I | -0,010 | 0,080 | 0,243 | 0,480 | 0,582 | -0,01 | 0,064 | 0,279 | 0,525 | 0,601 |
| Ποσοστά Απόρριψης | I | 5,01 | 25,94 | 95,02 | 100 | 100 | 4,80 | 35,68 | 99,78 | 100 | 100 |
| | LM-LAG | 5,18 | 97,47 | 100 | 100 | 100 | 4,94 | 80,51 | 100 | 100 | 100 |
| | LM-LE | 5,12 | 96,13 | 100 | 100 | 100 | 5,59 | 76,78 | 100 | 100 | 100 |
| | LM-ERR | 4,68 | 20,45 | 93,01 | 100 | 100 | 4,06 | 27,46 | 99,50 | 100 | 100 |
| | LM-EL | 5,00 | 5,33 | 2,65 | 0,02 | 0,0 | 4,34 | 6,52 | 58,51 | 97,09 | 89,06 |
| | LM-SARMA | 5,09 | 94,82 | 100 | 100 | 100 | 4,74 | 73,29 | 100 | 100 | 100 |
| 400 | Μέσο I | -0,002 | 0,092 | 0,267 | 0,538 | 0,674 | -0,002 | 0,080 | 0,314 | 0,608 | 0,721 |
| Ποσοστά Απόρριψης | I | 4,93 | 75,17 | 100 | 100 | 100 | 4,83 | 86,62 | 100 | 100 | 100 |
| | LM-LAG | 4,91 | 100 | 100 | 100 | 100 | 4,68 | 100 | 100 | 100 | 100 |
| | LM-LE | 5,19 | 100 | 100 | 100 | 100 | 4,81 | 99,99 | 100 | 100 | 100 |
| | LM-ERR | 4,87 | 72,89 | 100 | 100 | 100 | 4,77 | 84,53 | 100 | 100 | 100 |
| | LM-EL | 5,11 | 5,97 | 2,26 | 0,79 | 0,0 | 4,82 | 19,3 | 99,9 | 100 | 100 |
| | LM-SARMA | 4,95 | 100 | 100 | 100 | 100 | 4,56 | 100 | 100 | 100 | 100 |
| 900 | Μέσο I | -0,001 | 0,101 | 0,292 | 0,562 | 0,693 | -0,001 | 0,087 | 0,341 | 0,635 | 0,746 |
| Ποσοστά Απόρριψης | I | 4,92 | 98,83 | 100 | 100 | 100 | 4,81 | 99,64 | 100 | 100 | 100 |
| | LM-LAG | 5,08 | 100 | 100 | 100 | 100 | 4,55 | 100 | 100 | 100 | 100 |
| | LM-LE | 4,87 | 100 | 100 | 100 | 100 | 4,93 | 100 | 100 | 100 | 100 |
| | LM-ERR | 4,96 | 98,65 | 100 | 100 | 100 | 4,72 | 99,59 | 100 | 100 | 100 |
| | LM-EL | 4,74 | 5,41 | 8,99 | 6,11 | 0,00 | 4,88 | 44,45 | 100 | 100 | 100 |
| | LM-SARMA | 4,84 | 100 | 100 | 100 | 100 | 4,81 | 100 | 100 | 100 | 100 |

Η παρατήρηση του Διαγράμματος 2.1 αποκαλύπτει ότι για τιμές του ρ μεγαλύτερες ή ίσες του 0,5 οι τιμές της στατιστικής LM-LE είναι πάντα μεγαλύτερες από τις τιμές της στατιστικής LM-EL. Επιπλέον, καθώς αυξάνεται το μέγεθος του δείγματος η απόσταση μεταξύ των τιμών των δύο στατιστικών αυξάνεται σημαντικά. Επομένως, εάν η τιμή της στατιστικής LM-LE είναι μεγαλύτερη από τη τιμή της LM-EL, στην περίπτωση που και οι δύο απλές στατιστικές απορρίπτουν τις μηδενικές τους υποθέσεις και συμβαίνει το ίδιο και με τις ανθεκτικές τους μορφές, τότε υπάρχουν σαφείς ενδείξεις για την επιλογή του υποδείγματος χωρικής υστέρησης που είναι το σωστό υπόδειγμα. Το κριτήριο αυτό έχει χρησιμοποιηθεί αρκετές φορές σε εμπειρικές εφαρμογές.



Διάγραμμα 2.1

Προσομοιωμένες τιμές των στατιστικών LM-LE και LM-EL για το υπόδειγμα χωρικής υστέρησης με το κριτήριο της βασίλισσας για όλα τα μεγέθη δείγματος και τιμές του ρ Χωρίς διόρθωση τóρος

Ο έλεγχος με τη στατιστική LM-SARMA αποδίδει ποσοστά απόρριψης που προσεγγίζουν το 100% για όλα τα μεγέθη δείγματος και τιμές του ρ . Το αποτέλεσμα αυτό είναι αναμενόμενο διότι η στατιστική LM-SARMA έχει στατιστική ισχύ και για τα δύο χωρικά οικονομετρικά υποδείγματα. Εντοπίζει, δηλαδή, το πρόβλημα της χωρικής αυτοσυσχέτισης, αλλά δεν μπορεί να φανερώσει την αιτία του. Τα προηγούμενα συμπεράσματα για τους ελέγχους χωρικής εξάρτησης ισχύουν ανεξάρτητα από τη χρησιμοποίηση ή την απουσία διόρθωσης τóρος.

Οι Πίνακες 2.5 και 2.6 περιέχουν τα αποτελέσματα της προσομοίωσης, χωρίς διόρθωση τύπος και με διόρθωση τύπος, αντίστοιχα, για την περίπτωση που το πρόβλημα της χωρικής αυτοσυσχέτισης οφείλεται στην ύπαρξη του υποδείγματος χωρικού σφάλματος και συγκεκριμένα με σφάλματα που ακολουθούν μια στάσιμη χωρική αυτοπαλίνδρομη διαδικασία πρώτης τάξης, (SAR(1)), για όλες τις τιμές του συντελεστή λ και μεγέθη δείγματος. Όπως φαίνεται στον Πίνακα 2.5 η μέση τιμή της στατιστικής I αυξάνεται καθώς αυξάνεται η τιμή του λ ως συνέπεια της ύπαρξης χωρικής αυτοσυσχέτισης. Τα ποσοστά απόρριψης της στατιστικής I είναι μεγαλύτερα από τα αντίστοιχα ποσοστά όλων των υπολοίπων στατιστικών στα δείγματα των 49 και 100 παρατηρήσεων και για τιμές του λ μέχρι 0,5. Τα ποσοστά απόρριψης της στατιστικής LM-ERR είναι τα δεύτερα πιο μεγάλα, πολύ κοντά στα ποσοστά απόρριψης της στατιστικής I . Στα δείγματα των 400 και 900 παρατηρήσεων και για τιμές του λ μεγαλύτερες ή ίσες του 0,5 οι στατιστικές I και LM-ERR ουσιαστικά έχουν τα ίδια ποσοστά απόρριψης του 100%. Το συμπέρασμα αυτό έχει επισημανθεί και στην εργασία των Anselin και Florax (1995) για τα δικά τους δείγματα.

Η στατιστική LM-LAG παρουσιάζει και αυτή αξιολογικά ποσοστά απόρριψης ιδιαίτερα για υψηλές τιμές του λ σε μεγάλα δείγματα προκαλώντας σύγχυση ως προς τον εντοπισμό της σωστής αιτίας της χωρικής αυτοσυσχέτισης κάτι, που όπως προηγουμένως έχει αναφερθεί, ισχύει και με τη στατιστική LM-ERR όταν η αιτία της αυτοσυσχέτισης είναι η χωρική υστέρηση της εξαρτημένης μεταβλητής. Στην περίπτωση αυτή ο ερευνητής χρησιμοποιεί τα αποτελέσματα των ανθεκτικών στατιστικών. Ειδικότερα, από τον Πίνακα 2.5 φαίνεται ότι η ανθεκτική στατιστική LM-EL έχει ποσοστά απόρριψης που προσεγγίζουν το 100% ακόμη και σε μετρίου μεγέθους δείγματα και για τιμές του λ μεγαλύτερες ή ίσες του 0,5. Αντίθετα, τα ποσοστά απόρριψης της ανθεκτικής στατιστικής LM-LE είναι χαμηλά, αν και ξεπερνούν το επίπεδο σημαντικότητας του 5%, στα μικρά δείγματα και μέτρια στα δείγματα των 400 και 900 παρατηρήσεων. Λίγο πιο αυξημένα ποσοστά απόρριψης εντοπίζονται όταν έχει χρησιμοποιηθεί το κριτήριο της βασίλισσας.

Πίνακας 2.5
Μέσες τιμές I και ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% των
ελέγχων χωρικής εξάρτησης για όταν ισχύει το υπόδειγμα χωρικού σφάλματος με
σφάλματα που ακολουθούν χωρικές αυτοπαλίνδρομες διαδικασίες
για όλες τις τιμές του λ και μεγέθη δείγματος
Χωρίς διόρθωση τόρος

| N | Σταθμίσεις | Πύργος | | | | | Βασίλισσα | | | | |
|----------------------|------------|--------|-------|-------|-------|-------|-----------|-------|-------|-------|-------|
| | λ | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 |
| 49 | Μέσο I | -0,030 | 0,076 | 0,252 | 0,502 | 0,627 | -0,026 | 0,034 | 0,165 | 0,399 | 0,526 |
| Ποσοστά Απόρριψης | I | 4,83 | 17,66 | 72,34 | 99,11 | 99,91 | 5,31 | 16,38 | 61,27 | 96,30 | 99,12 |
| | LM-LAG | 5,91 | 6,22 | 16,68 | 64,45 | 86,82 | 5,56 | 7,65 | 20,29 | 60,83 | 80,17 |
| | LM-LE | 6,28 | 5,67 | 5,46 | 5,44 | 4,60 | 5,86 | 7,58 | 11,47 | 17,69 | 19,78 |
| | LM-ERR | 4,78 | 10,42 | 61,28 | 98,23 | 99,75 | 3,86 | 9,43 | 48,82 | 93,35 | 98,39 |
| | LM-EL | 5,13 | 9,49 | 54,77 | 95,78 | 99,07 | 4,36 | 9,09 | 44,51 | 90,09 | 97,09 |
| | LM-SARMA | 5,91 | 8,47 | 52,72 | 96,98 | 99,51 | 4,59 | 9,89 | 46,38 | 92,21 | 98,07 |
| 100 | Μέσο I | -0,010 | 0,098 | 0,281 | 0,548 | 0,686 | -0,010 | 0,053 | 0,191 | 0,452 | 0,605 |
| Ποσοστά Απόρριψης | I | 4,79 | 30,50 | 96,29 | 100 | 100 | 4,77 | 24,22 | 87,64 | 99,98 | 100 |
| | LM-LAG | 4,95 | 7,44 | 32,22 | 89,12 | 98,72 | 5,12 | 8,73 | 32,63 | 85,84 | 97,42 |
| | LM-LE | 5,39 | 5,35 | 6,61 | 10,40 | 11,84 | 5,37 | 7,29 | 12,56 | 20,98 | 24,43 |
| | LM-ERR | 4,59 | 25,27 | 94,84 | 100 | 100 | 4,24 | 18,97 | 83,60 | 99,93 | 100 |
| | LM-EL | 4,81 | 23,07 | 91,9 | 100 | 100 | 4,51 | 17,58 | 78,93 | 99,74 | 99,99 |
| | LM-SARMA | 4,96 | 19,88 | 91,49 | 99,99 | 100 | 4,51 | 17,35 | 79,94 | 99,87 | 100 |
| 400 | Μέσο I | -0,003 | 0,104 | 0,286 | 0,562 | 0,712 | -0,002 | 0,059 | 0,196 | 0,469 | 0,643 |
| Ποσοστά Απόρριψης | I | 4,84 | 83,23 | 100 | 100 | 100 | 5,08 | 61,76 | 100 | 100 | 100 |
| | LM-LAG | 4,99 | 15,18 | 76,82 | 99,98 | 100 | 4,82 | 12,44 | 65,02 | 99,87 | 100 |
| | LM-LE | 5,01 | 5,24 | 7,67 | 14,63 | 19,19 | 4,91 | 6,80 | 12,06 | 25,80 | 34,54 |
| | LM-ERR | 4,75 | 81,18 | 100 | 100 | 100 | 4,97 | 58,33 | 100 | 100 | 100 |
| | LM-EL | 5,14 | 76,61 | 100 | 100 | 100 | 5,03 | 54,57 | 99,98 | 100 | 100 |
| | LM-SARMA | 4,85 | 72,40 | 100 | 100 | 100 | 4,92 | 50,36 | 100 | 100 | 100 |
| 900 | Μέσο I | -0,001 | 0,104 | 0,285 | 0,562 | 0,715 | -0,001 | 0,059 | 0,194 | 0,467 | 0,647 |
| Ποσοστά Απόρριψης | I | 4,97 | 99,12 | 100 | 100 | 100 | 5,05 | 90,65 | 100 | 100 | 100 |
| | LM-LAG | 4,95 | 31,29 | 98,36 | 100 | 100 | 4,78 | 22,52 | 92,91 | 100 | 100 |
| | LM-LE | 4,94 | 5,17 | 7,58 | 15,55 | 21,51 | 4,97 | 7,15 | 13,64 | 28,15 | 36,32 |
| | LM-ERR | 4,89 | 99,01 | 100 | 100 | 100 | 5,01 | 89,52 | 100 | 100 | 100 |
| | LM-EL | 5,04 | 98,19 | 100 | 100 | 100 | 5,02 | 86,14 | 100 | 100 | 100 |
| | LM-SARMA | 4,82 | 97,72 | 100 | 100 | 100 | 5,01 | 83,93 | 100 | 100 | 100 |

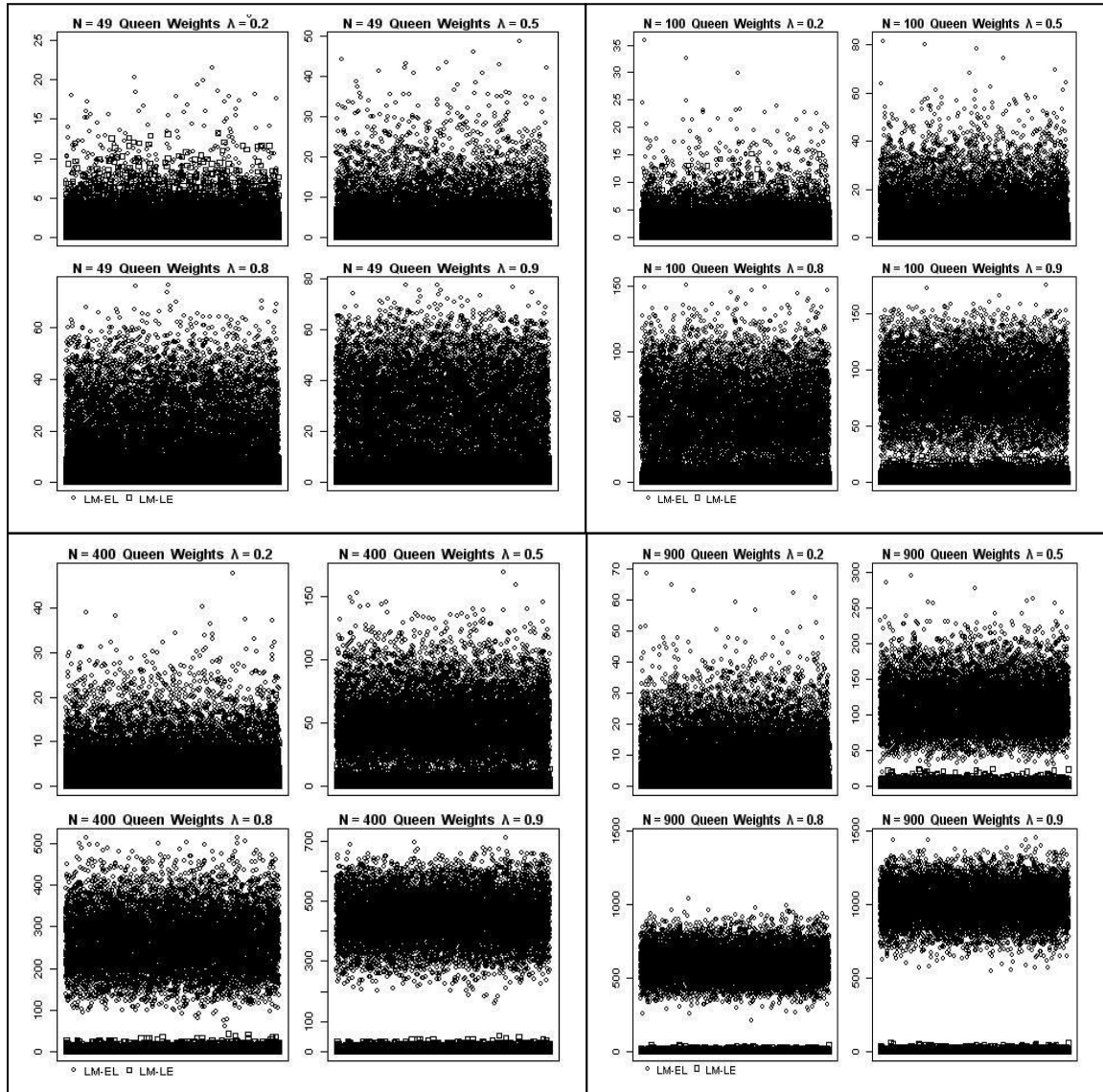
Όπως και στο υπόδειγμα χωρικής υστέρησης, έτσι και στην περίπτωση του υποδείγματος χωρικού σφάλματος όταν τα αποτελέσματα των ελέγχων των δύο ανθεκτικών στατιστικών είναι αντικρουόμενα υπάρχει μεγάλη πιθανότητα το χωρικό υπόδειγμα που υποδεικνύει η ανθεκτική στατιστική με τη μεγαλύτερη αριθμητική τιμή να είναι το σωστό. Αυτό εξετάζεται για τα αποτελέσματα της προσομοίωσης που προέκυψαν από την εφαρμογή της μήτρας χωρικών σταθμίσεων που κατασκευάστηκε σύμφωνα με το κριτήριο της βασίλισσας καθώς εκεί εμφανίζεται πιο σοβαρό το πρόβλημα. Οι τιμές των δύο ανθεκτικών στατιστικών για όλα τα δείγματα και τιμές του λ ίσες με 0,2, 0,5, 0,8 και 0,9 απεικονίζονται στο Διάγραμμα 2.2.

Πίνακας 2.6
Μέσες τιμές I και ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% των
ελέγχων χωρικής εξάρτησης για όταν ισχύει το υπόδειγμα χωρικού σφάλματος με
σφάλματα που ακολουθούν χωρικές αυτοπαλίνδρομες διαδικασίες
για όλες τις τιμές του λ και μεγέθη δείγματος
Με διόρθωσης τόρος

| N | Σταθμίσεις | Πύργος | | | | | Βασίλισσα | | | | |
|----------------------|------------|--------|-------|-------|-------|-------|-----------|-------|-------|-------|-------|
| | λ | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 |
| 49 | Μέσο I | -0,027 | 0,058 | 0,197 | 0,375 | 0,452 | -0,025 | 0,017 | 0,101 | 0,228 | 0,285 |
| Ποσοστά Απόρριψης | I | 5,05 | 15,08 | 63,17 | 97,05 | 99,21 | 5,08 | 13,63 | 47,86 | 86,85 | 94,10 |
| | LM-LAG | 5,82 | 6,18 | 13,68 | 43,20 | 61,08 | 5,36 | 7,06 | 14,48 | 36,04 | 47,34 |
| | LM-LE | 6,23 | 5,89 | 6,51 | 9,38 | 10,78 | 5,87 | 7,30 | 10,83 | 16,78 | 19,47 |
| | LM-ERR | 4,70 | 7,99 | 49,62 | 93,61 | 98,32 | 3,96 | 6,40 | 32,13 | 77,97 | 88,75 |
| | LM-EL | 4,94 | 7,82 | 43,85 | 89,33 | 96,21 | 4,13 | 6,7 | 28,97 | 73,07 | 85,08 |
| | LM-SARMA | 5,58 | 7,23 | 41,73 | 90,82 | 97,55 | 4,59 | 7,74 | 31,00 | 76,16 | 87,59 |
| 100 | Μέσο I | -0,010 | 0,084 | 0,242 | 0,466 | 0,576 | -0,010 | 0,040 | 0,146 | 0,335 | 0,437 |
| Ποσοστά Απόρριψης | I | 5,01 | 27,89 | 93,78 | 100 | 100 | 4,80 | 20,94 | 79,28 | 99,60 | 99,97 |
| | LM-LAG | 5,18 | 7,22 | 28,17 | 79,83 | 94,27 | 4,94 | 8,32 | 26,61 | 70,56 | 87,12 |
| | LM-LE | 5,12 | 5,37 | 7,20 | 13,18 | 16,45 | 5,59 | 7,46 | 12,98 | 23,88 | 28,66 |
| | LM-ERR | 4,68 | 22,17 | 91,36 | 99,99 | 100 | 4,06 | 15,34 | 72,94 | 99,36 | 99,94 |
| | LM-EL | 5,00 | 20,18 | 87,3 | 99,95 | 100 | 4,34 | 14,73 | 68,28 | 98,94 | 99,84 |
| | LM-SARMA | 5,09 | 17,27 | 86,37 | 99,97 | 100 | 4,74 | 14,58 | 69,15 | 99,05 | 99,92 |
| 400 | Μέσο I | -0,002 | 0,097 | 0,268 | 0,526 | 0,668 | -0,002 | 0,053 | 0,174 | 0,416 | 0,571 |
| Ποσοστά Απόρριψης | I | 4,93 | 80,79 | 100 | 100 | 100 | 4,83 | 58,16 | 99,97 | 100 | 100 |
| | LM-LAG | 4,91 | 14,05 | 74,41 | 99,96 | 100 | 4,68 | 12,37 | 63,04 | 99,69 | 100 |
| | LM-LE | 5,19 | 5,16 | 7,09 | 12,90 | 17,03 | 4,81 | 6,42 | 11,77 | 23,91 | 31,03 |
| | LM-ERR | 4,87 | 78,58 | 100 | 100 | 100 | 4,77 | 54,14 | 99,95 | 100 | 100 |
| | LM-EL | 5,11 | 73,73 | 100 | 100 | 100 | 4,82 | 50,12 | 99,88 | 100 | 100 |
| | LM-SARMA | 4,95 | 69,06 | 100 | 100 | 100 | 4,56 | 46,28 | 99,87 | 100 | 100 |
| 900 | Μέσο I | -0,001 | 0,100 | 0,273 | 0,538 | 0,686 | -0,001 | 0,055 | 0,179 | 0,432 | 0,600 |
| Ποσοστά Απόρριψης | I | 4,92 | 98,97 | 100 | 100 | 100 | 4,81 | 88,71 | 100 | 100 | 100 |
| | LM-LAG | 5,08 | 29,64 | 97,76 | 100 | 100 | 4,55 | 21,23 | 91,01 | 100 | 100 |
| | LM-LE | 4,87 | 5,15 | 7,76 | 15,10 | 20,19 | 4,93 | 7,00 | 13,04 | 26,48 | 33,79 |
| | LM-ERR | 4,96 | 98,76 | 100 | 100 | 100 | 4,72 | 87,35 | 100 | 100 | 100 |
| | LM-EL | 4,74 | 97,87 | 100 | 100 | 100 | 4,88 | 83,55 | 100 | 100 | 100 |
| | LM-SARMA | 4,84 | 97,16 | 100 | 100 | 100 | 4,81 | 81,4 | 100 | 100 | 100 |

Από το Διάγραμμα 2.2 διαπιστώνεται ότι οι τιμές που προέκυψαν από την προσομοίωση για τη στατιστική LM-EL είναι μεγαλύτερες από τις τιμές της στατιστικής LM-LE. Το χαρακτηριστικό αυτό φαίνεται πιο έντονα στα δείγματα των 400 και 900 παρατηρήσεων και για τιμές των λ ίσες με 0,8 και 0,9. Πρέπει όμως να αναφερθεί ότι, παρόλο που δεν διακρίνεται καθαρά στο διάγραμμα, η ιδιότητα αυτή ισχύει για τις περισσότερες περιπτώσεις της προσομοίωσης και στα μικρά δείγματα και για μικρές τιμές του λ . Εμφανίζονται μόνο μερικές μεμονωμένες εξαιρέσεις. Επομένως, και στο υπόδειγμα χωρικού σφάλματος, όταν η

τιμή της στατιστικής LM-EL είναι μεγαλύτερη από την τιμή της στατιστικής LM-LE υπάρχουν ισχυρές ενδείξεις που οδηγούν στην εκτίμηση του σωστού χωρικού υποδείγματος.



Διάγραμμα 2.2

Προσομοιωμένες τιμές των στατιστικών LM-EL και LM-LE για το υπόδειγμα χωρικού σφάλματος με σφάλματα που ακολουθούν χωρικές αυτοπαλίνδρομες διαδικασίες με το κριτήριο της βασίλισσας για όλα τα μεγέθη δείγματος και τιμές του λ
Χωρίς διόρθωση τóρος

Πίνακας 2.7

Μέσες τιμές I και ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% των ελέγχων χωρικής εξάρτησης για όταν ισχύει το υπόδειγμα χωρικού σφάλματος με σφάλματα που ακολουθούν χωρικές διαδικασίες κινητού μέσου για όλες τις τιμές του λ και μεγέθη δείγματος Χωρίς διόρθωση τόρος

| N | Σταθμίσεις | Πύργος | | | | | Βασίλισσα | | | | |
|-------------------|------------|--------|-------|-------|-------|-------|-----------|-------|-------|-------|-------|
| | λ | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 |
| 49 | Μέσο I | -0,030 | 0,074 | 0,218 | 0,332 | 0,362 | -0,026 | 0,029 | 0,117 | 0,203 | 0,230 |
| Ποσοστά Απόρριψης | I | 4,83 | 16,64 | 64,42 | 93,62 | 96,86 | 5,31 | 14,59 | 45,06 | 74,43 | 81,14 |
| | LM-LAG | 5,91 | 6,05 | 11,06 | 21,89 | 26,17 | 5,56 | 7,10 | 13,49 | 23,66 | 27,31 |
| | LM-LE | 6,28 | 5,54 | 4,63 | 4,47 | 4,52 | 5,86 | 7,37 | 9,61 | 12,04 | 13,00 |
| | LM-ERR | 4,78 | 9,50 | 50,10 | 86,96 | 92,43 | 3,86 | 8,19 | 32,18 | 62,18 | 71,05 |
| | LM-EL | 5,13 | 8,91 | 44,04 | 81,14 | 87,20 | 4,36 | 7,93 | 28,97 | 56,84 | 65,56 |
| | LM-SARMA | 5,91 | 7,84 | 40,00 | 79,41 | 86,44 | 4,59 | 8,97 | 29,68 | 58,10 | 66,78 |
| 100 | Μέσο I | -0,010 | 0,096 | 0,241 | 0,355 | 0,385 | -0,010 | 0,047 | 0,137 | 0,226 | 0,254 |
| Ποσοστά Απόρριψης | I | 4,79 | 29,07 | 93,31 | 99,95 | 99,98 | 4,77 | 21,53 | 70,83 | 95,86 | 98,04 |
| | LM-LAG | 4,95 | 6,93 | 21,31 | 41,73 | 48,74 | 5,12 | 8,04 | 20,28 | 38,12 | 44,00 |
| | LM-LE | 5,39 | 5,18 | 5,26 | 5,72 | 6,03 | 5,37 | 6,98 | 10,42 | 13,02 | 14,03 |
| | LM-ERR | 4,59 | 23,97 | 90,53 | 99,93 | 99,98 | 4,24 | 16,35 | 64,00 | 94,09 | 97,07 |
| | LM-EL | 4,81 | 21,93 | 86,11 | 99,59 | 99,90 | 4,51 | 15,29 | 58,97 | 91,11 | 95,40 |
| | LM-SARMA | 4,96 | 18,36 | 83,48 | 99,60 | 99,93 | 4,51 | 14,84 | 58,68 | 91,14 | 95,68 |
| 400 | Μέσο I | -0,003 | 0,101 | 0,243 | 0,356 | 0,386 | -0,002 | 0,053 | 0,141 | 0,227 | 0,254 |
| Ποσοστά Απόρριψης | I | 4,84 | 82,00 | 100 | 100 | 100 | 5,08 | 55,12 | 99,63 | 100 | 100 |
| | LM-LAG | 4,99 | 14,24 | 59,20 | 90,20 | 94,78 | 4,82 | 11,10 | 40,79 | 73,54 | 81,02 |
| | LM-LE | 5,01 | 4,87 | 5,56 | 6,45 | 6,90 | 4,91 | 6,57 | 9,33 | 12,14 | 13,15 |
| | LM-ERR | 4,75 | 79,84 | 100 | 100 | 100 | 4,97 | 51,40 | 99,57 | 100 | 100 |
| | LM-EL | 5,14 | 75,13 | 100 | 100 | 100 | 5,03 | 48,04 | 99,30 | 100 | 100 |
| | LM-SARMA | 4,85 | 70,22 | 100 | 100 | 100 | 4,92 | 43,37 | 99,08 | 100 | 100 |
| 900 | Μέσο I | -0,001 | 0,102 | 0,242 | 0,354 | 0,384 | -0,001 | 0,054 | 0,139 | 0,224 | 0,251 |
| Ποσοστά Απόρριψης | I | 4,97 | 98,99 | 100 | 100 | 100 | 5,05 | 85,81 | 100 | 100 | 100 |
| | LM-LAG | 4,95 | 29,24 | 92,39 | 99,88 | 99,99 | 4,78 | 19,45 | 73,69 | 96,98 | 98,9 |
| | LM-LE | 4,94 | 4,76 | 5,35 | 6,43 | 6,86 | 4,97 | 6,84 | 9,90 | 13,40 | 14,41 |
| | LM-ERR | 4,89 | 98,79 | 100 | 100 | 100 | 5,01 | 84,13 | 100 | 100 | 100 |
| | LM-EL | 5,04 | 97,94 | 100 | 100 | 100 | 5,02 | 79,81 | 100 | 100 | 100 |
| | LM-SARMA | 4,82 | 97,36 | 100 | 100 | 100 | 5,01 | 76,99 | 100 | 100 | 100 |

Η προσομοίωση επαναλήφθηκε θεωρώντας ότι τα σφάλματα στο υπόδειγμα χωρικού σφάλματος αυτοσυσχετίζονται ακολουθώντας μια χωρική διαδικασία κινητού μέσου πρώτης τάξης (SMA(1)). Τα αποτελέσματα που προέκυψαν χωρίς διόρθωση τόρος συνοψίζονται στον Πίνακα 2.7, ενώ ο Πίνακας 2.8 περιλαμβάνει τα αποτελέσματα με διόρθωση τόρος. Το βασικό συμπέρασμα που εξάγεται από την μελέτη του Πίνακα 2.7 είναι ότι τα ποσοστά απόρριψης όλων των εξεταζόμενων ελέγχων είναι φανερά μικρότερα σε σύγκριση με την περίπτωση των χωρικά αυτοπαλίνδρομων σφαλμάτων ιδιαίτερα στα δείγματα των 49 και 100 παρατηρήσεων.

Πίνακας 2.8

Μέσες τιμές I και ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% των ελέγχων χωρικής εξάρτησης για όταν ισχύει το υπόδειγμα χωρικού σφάλματος με σφάλματα που ακολουθούν χωρικές διαδικασίες κινητού μέσου για όλες τις τιμές του λ και μεγέθη δείγματος Με διόρθωσης τόρος

| N | Σταθμίσεις | Πύργος | | | | | Βασίλισσα | | | | |
|-------------------|------------|--------|-------|-------|-------|-------|-----------|-------|-------|-------|-------|
| | λ | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 |
| 49 | Μέσο I | -0,027 | 0,058 | 0,177 | 0,274 | 0,301 | -0,025 | 0,014 | 0,076 | 0,138 | 0,158 |
| Ποσοστά Απόρριψης | I | 5,05 | 14,25 | 56,41 | 89,28 | 94,01 | 5,08 | 12,52 | 36,12 | 64,25 | 72,13 |
| | LM-LAG | 5,82 | 5,98 | 10,07 | 18,31 | 21,60 | 5,36 | 6,85 | 10,84 | 18,27 | 20,73 |
| | LM-LE | 6,23 | 5,73 | 5,35 | 5,70 | 5,87 | 5,87 | 7,13 | 9,36 | 11,85 | 12,64 |
| | LM-ERR | 4,70 | 7,43 | 40,87 | 79,76 | 87,09 | 3,96 | 5,85 | 21,60 | 47,81 | 56,59 |
| | LM-EL | 4,94 | 7,37 | 35,77 | 73,55 | 81,54 | 4,13 | 6,15 | 19,78 | 42,86 | 51,72 |
| | LM-SARMA | 5,58 | 6,75 | 32,01 | 70,32 | 79,71 | 4,59 | 7,25 | 21,04 | 44,48 | 53,10 |
| 100 | Μέσο I | -0,010 | 0,083 | 0,211 | 0,315 | 0,343 | -0,010 | 0,036 | 0,108 | 0,180 | 0,203 |
| Ποσοστά Απόρριψης | I | 5,01 | 26,73 | 89,69 | 99,86 | 99,96 | 4,80 | 18,81 | 62,62 | 92,08 | 95,64 |
| | LM-LAG | 5,18 | 6,79 | 19,28 | 38,39 | 44,30 | 4,94 | 7,80 | 17,30 | 31,70 | 36,96 |
| | LM-LE | 5,12 | 5,22 | 5,47 | 5,85 | 6,27 | 5,59 | 7,15 | 10,14 | 13,35 | 14,38 |
| | LM-ERR | 4,68 | 20,98 | 85,67 | 99,68 | 99,95 | 4,06 | 13,49 | 54,16 | 88,18 | 93,60 |
| | LM-EL | 5,00 | 19,17 | 80,02 | 99,19 | 99,81 | 4,34 | 13,01 | 50,23 | 84,05 | 90,54 |
| | LM-SARMA | 5,09 | 16,19 | 77,49 | 99,06 | 99,73 | 4,74 | 12,92 | 49,16 | 83,99 | 90,64 |
| 400 | Μέσο I | -0,002 | 0,095 | 0,229 | 0,337 | 0,366 | -0,002 | 0,048 | 0,127 | 0,205 | 0,230 |
| Ποσοστά Απόρριψης | I | 4,93 | 79,38 | 100 | 100 | 100 | 4,83 | 51,37 | 99,31 | 100 | 100 |
| | LM-LAG | 4,91 | 13,33 | 56,85 | 88,86 | 93,86 | 4,68 | 11,00 | 40,22 | 72,93 | 80,14 |
| | LM-LE | 5,19 | 4,88 | 5,10 | 5,94 | 6,13 | 4,81 | 6,11 | 9,06 | 11,85 | 12,97 |
| | LM-ERR | 4,87 | 76,92 | 100 | 100 | 100 | 4,77 | 47,49 | 99,12 | 100 | 100 |
| | LM-EL | 5,11 | 72,31 | 100 | 100 | 100 | 4,82 | 43,90 | 98,72 | 100 | 100 |
| | LM-SARMA | 4,95 | 67,02 | 100 | 100 | 100 | 4,56 | 39,98 | 98,41 | 100 | 100 |
| 900 | Μέσο I | -0,001 | 0,097 | 0,233 | 0,341 | 0,371 | -0,001 | 0,050 | 0,130 | 0,209 | 0,235 |
| Ποσοστά Απόρριψης | I | 4,92 | 98,79 | 100 | 100 | 100 | 4,81 | 83,50 | 100 | 100 | 100 |
| | LM-LAG | 5,08 | 27,71 | 90,85 | 99,81 | 99,97 | 4,55 | 18,52 | 70,88 | 96,03 | 98,42 |
| | LM-LE | 4,87 | 4,81 | 5,38 | 6,36 | 6,84 | 4,93 | 6,66 | 9,87 | 13,08 | 14,19 |
| | LM-ERR | 4,96 | 98,61 | 100 | 100 | 100 | 4,72 | 81,58 | 100 | 100 | 100 |
| | LM-EL | 4,74 | 97,62 | 100 | 100 | 100 | 4,88 | 76,86 | 100 | 100 | 100 |
| | LM-SARMA | 4,84 | 96,83 | 100 | 100 | 100 | 4,81 | 73,95 | 99,99 | 100 | 100 |

Στα μικρά δείγματα και για τιμές του λ μέχρι το 0,5 οι έλεγχοι χωρικής εξάρτησης αντιμετωπίζουν δυσκολίες στον εντοπισμό του προβλήματος της χωρικής αυτοσυσχέτισης, συμπέρασμα που έχουν καταλήξει στην εργασία τους και οι Anselin και Florax (1995). Το χαρακτηριστικό αυτό εντοπίζεται ιδιαίτερα στο κριτήριο της βασίλισσας. Αυτό ίσως θα μπορούσε να αποδοθεί στο γεγονός πως οι συνέπειες της χωρικής αυτοσυσχέτισης, όπως εμφανίζεται μέσα από ένα υπόδειγμα κινητού μέσου, είναι πιο ήπιες κάτι που δικαιολογεί και τις μέσες τιμές της στατιστικής I του Moran που είναι περίπου οι μισές σε μέγεθος από αυτές που εμφανίζονται στον Πίνακα 2.5. Αντίθετα, στα μεγάλα δείγματα των 400 και 900

παρατηρήσεων οι έλεγχοι έχουν πιο μεγάλα ποσοστά απόρριψης. Μάλιστα, η στατιστική LM-LAG έχει υψηλά ποσοστά απόρριψης που στα δείγματα των 900 παρατηρήσεων τείνουν στο 100%. Υπάρχει όμως μια τάση τα ποσοστά της να είναι πιο μικρά από τα ποσοστά απόρριψης της LM-ERR η οποία στα μεγάλα δείγματα έχει στατιστική ισχύ 100% εντοπίζοντας το πρόβλημα της χωρικής αυτοσυσχέτισης στα σφάλματα του υποδείγματος κάτι που επίσης ισχύει και για τη στατιστική LM-SARMA.

Συγκρίνοντας τις δύο ανθεκτικές στατιστικές διαπιστώνεται ότι τα ποσοστά απόρριψης της LM-EL οδηγούν στην επιλογή του υποδείγματος χωρικού σφάλματος, καθώς στα μεγάλα δείγματα και για και για τιμές του λ μεγαλύτερες ή ίσες του 0,5 τα ποσοστά απόρριψης είναι 100% σε σχέση με τα ποσοστά απόρριψης της LM-LE που κυμαίνονται κοντά στο επίπεδο σημαντικότητας του 5%. Μικρή εξαίρεση υπάρχει, όπως είχε παρατηρηθεί και στα χωρικά αυτοπαλίνδρομα σφάλματα, όταν χρησιμοποιείται το κριτήριο της βασίλισσας όπου τα ποσοστά απόρριψης της LM-LE είναι λίγο πιο αυξημένα όταν το λ λαμβάνει μεγάλες τιμές φτάνοντας σε κάποιες περιπτώσεις κοντά στο 14%. Όμως, τα ποσοστά αυτά είναι αισθητά πιο μικρά συγκρίνοντάς τα με τα αντίστοιχα στα χωρικά αυτοπαλίνδρομα σφάλματα. Τα αποτελέσματα αυτά γενικά ισχύουν και όταν έχει χρησιμοποιηθεί η διόρθωση τόρος χωρίς ιδιαίτερες διαφοροποιήσεις.

Από την ανάλυση των αποτελεσμάτων της προσομοίωσης για τα δύο διαφορετικά χωρικά υποδείγματα γίνεται φανερό ότι ενώ οι εξεταζόμενοι έλεγχοι χωρικής εξάρτησης προσφέρουν σημαντική βοήθεια στον εντοπισμό της ύπαρξης του προβλήματος της χωρικής αυτοσυσχέτισης στα σφάλματα ενός γραμμικού οικονομετρικού υποδείγματος η ερμηνεία τους ως προς την ανίχνευση της αιτίας που τη προκαλεί θα πρέπει να γίνεται με προσοχή από τον ερευνητή. Ειδικότερα, διαπιστώθηκε ότι η συμπεριφορά των ελέγχων αυτών ως προς την ικανότητά τους να διακρίνουν την αιτία του προβλήματος εξαρτάται από το μέγεθος του δείγματος και σε κάποιες περιπτώσεις από τον ορισμό της χωρικής γειννίας μέσω της μήτρας των σταθμίσεων καθώς μπορεί να οδηγήσει σε τελείως διαφορετικά αποτελέσματα.

ΑΣυνέπεια στα αποτελέσματα προέκυψε για τις ανθεκτικές στατιστικές LM-LE και LM-EL όταν η αιτία της αυτοσυσχέτισης είναι η ύπαρξη χωρικής υστέρησης στην εξαρτημένη μεταβλητή που έχει παραληφθεί από το υπόδειγμα όπου τα ποσοστά απόρριψης των ελέγχων με τις δύο στατιστικές παρουσιάζουν δραματικές διαφοροποιήσεις μεταξύ των κριτηρίων του πύργου και της βασίλισσας. Επομένως, ο ερευνητής θα πρέπει να εξετάζει τη συνέπεια των αποτελεσμάτων από τους ελέγχους ορίζοντας και χρησιμοποιώντας

διαφορετικά κριτήρια χωρικής γειτνίασης πριν καταλήξει στο τελικό του συμπέρασμα. Η τακτική αυτή πρέπει να εφαρμόζεται σε κάθε χωρική οικονομετρική ανάλυση λαμβάνοντας μάλιστα υπ' όψιν ότι στις προσομοιώσεις χρησιμοποιήθηκαν για τον ορισμό των χωρικών μονάδων ομαλά πλέγματα, κάτι που σπάνια ισχύει στην πράξη όπου οι χωρικές μονάδες αποτελούν μέρους ενός πραγματικού γεωγραφικού χάρτη δυσκολεύοντας την κατάσταση ακόμη περισσότερο.

Η χρησιμοποίηση μεγαλύτερων δειγμάτων δεν βελτιώνει απαραίτητα τα αποτελέσματα αλλά αυξάνει τα ποσοστά απόρριψης των ελέγχων προκαλώντας μερικές φορές ασάφεια ως προς τον εντοπισμό της αιτίας της αυτοσυσχέτισης. Το συμπέρασμα αυτό δεν είχε επισημανθεί στις παλιότερες γνωστές εργασίες διότι οι προσομοιώσεις τους χρησιμοποιούσαν μικρότερα δείγματα. Στις περιπτώσεις που δεν μπορεί να εντοπιστεί με σαφήνεια η αιτία του προβλήματος η σύγκριση των τιμών των στατιστικών LM-LE και LM-EL μπορεί να οδηγήσει στο σωστό υπόδειγμα. Η στρατηγική αυτή έχει αναφερθεί και σε παλιότερες εργασίες χωρίς να έχει αρκετά τεκμηριωθεί και χρησιμοποιείται σε πολλές εμπειρικές εφαρμογές.

2.5 Έλεγχος για χωρική αυτοσυσχέτιση στο υπόδειγμα χωρικής υστέρησης

Το υπόδειγμα χωρικής υστέρησης (*Spatial Lag Model-SLM*) αποτελεί ένα από τα πιο διαδομένα χωρικά οικονομετρικά υποδείγματα που χρησιμοποιείται για την αντιμετώπιση της χωρικής αυτοσυσχέτισης στα σφάλματα ενός απλού οικονομετρικού υποδείγματος όταν η πηγή του προβλήματος οφείλεται στην ύπαρξή της στις τιμές της εξαρτημένης μεταβλητής. Το υπόδειγμα αυτό επιλύει το πρόβλημα περιλαμβάνοντας στις ανεξάρτητες μεταβλητές του υποδείγματος τη χωρική υστέρηση της εξαρτημένης μεταβλητής με παρόμοιο τρόπο όπως περιλαμβάνεται σε ένα υπόδειγμα παλινδρόμησης με χρονοσειρές ή χρονική υστέρηση της εξαρτημένης μεταβλητής για την αντιμετώπιση της αυτοσυσχέτισης στα σφάλματα. Κατά συνέπεια, η εκτίμηση ενός υποδείγματος χωρικής υστέρησης με τη μέθοδο της μέγιστης πιθανοφάνειας θα πρέπει να οδηγεί σε σφάλματα ασυσχέτιστα εφόσον το υπόδειγμα έχει ταυτοποιηθεί σωστά. Στην ενότητα αυτή παρουσιάζονται οι βασικοί έλεγχοι με τους οποίους ελέγχεται η ύπαρξη χωρικής αυτοσυσχέτισης στα σφάλματα του υποδείγματος χωρικής υστέρησης και με την πραγματοποίηση προσομοίωσης διερευνάται η συμπεριφορά τους.

2.5.1 Η στατιστική $LM_{\lambda\rho}$

Ο πιο απλός τρόπος για να ερευνηθεί η ύπαρξη χωρικής αυτοσυσχέτισης στα σφάλματα του υποδείγματος χωρικής υστέρησης είναι με τη στατιστική I του Moran. Όμως, στην περίπτωση αυτή η ασυμπτωτική κατανομή της στατιστικής που απέδειξαν οι Cliff και Ord (1972) δεν μπορεί να χρησιμοποιηθεί διότι ισχύει για σφάλματα απλού γραμμικού υποδείγματος που εκτιμήθηκε με την μέθοδο των ελαχίστων τετραγώνων. Ωστόσο, θα μπορούσε να πραγματοποιηθεί ο έλεγχος υπολογίζοντας τη διακύμανση της στατιστικής σύμφωνα με την προσέγγιση της τυχαιοποίησης (*randomization*), όπως για οποιαδήποτε μεταβλητή, χωρίς τεκμηριωμένα αποτελέσματα για την εγκυρότητα αυτής της πρακτικής. Επιπρόσθετα, η παρουσία της χωρικής υστέρησης της εξαρτημένης μεταβλητής θέτει σε αμφισβήτηση την καταλληλότητα της στατιστικής I του Moran για τη διεξαγωγή του ελέγχου λόγω εμφάνισης του προβλήματος της ενδογένειας όπως συμβαίνει και με τη στατιστική των Durbin και Watson στην ανάλυση χρονοσειρών όταν υπάρχει σε κάποιο οικονομετρικό υπόδειγμα η χρονική υστέρηση της εξαρτημένης μεταβλητής.

Ο Anselin (1988b) βασιζόμενος στη γενικότερη μεθοδολογία των ελέγχων χωρικής εξάρτησης που παρουσιάστηκαν αναλυτικά σε προηγούμενη ενότητα κατασκεύασε έλεγχο πολλαπλασιαστή Lagrange για την ύπαρξη χωρικής αυτοσυσχέτισης στο υπόδειγμα χωρικής υστέρησης. Ο έλεγχος αυτός θεωρεί το υπόδειγμα χωρικής υστέρησης:

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

με

$$\boldsymbol{\varepsilon} = \lambda\mathbf{W}\boldsymbol{\varepsilon} + \mathbf{u} \text{ και } \mathbf{u} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$$

για το οποίο οι συντελεστές, οι μήτρες και τα διανύσματα έχουν οριστεί αναλυτικά πιο επάνω και ορίζει ως μηδενική και εναλλακτική υπόθεση, αντίστοιχα τις:

$$H_0 : \lambda = 0$$

$$H_1 : \lambda \neq 0$$

και πραγματοποιείται με την ακόλουθη στατιστική:

$$LM_{\lambda\rho} = \frac{\left[\frac{\hat{\boldsymbol{\varepsilon}}'\mathbf{W}\hat{\boldsymbol{\varepsilon}}}{S_{ML}^2} \right]^2}{T - T_A^2 \text{Var}(\hat{\rho})} \quad (2.12)$$

όπου $\hat{\boldsymbol{\varepsilon}}$ είναι το $(n \times 1)$ διάνυσμα των καταλοίπων που έχουν προκύψει από την εκτίμηση του υποδείγματος χωρικής υστέρησης με τη μέθοδο της μέγιστης πιθανοφάνειας, $S_{ML}^2 = \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}/n$,

$T = tr[\mathbf{W}^2 + \mathbf{W}'\mathbf{W}]$, $T_A = tr[\mathbf{W}^2\mathbf{A}^{-1} + \mathbf{W}'\mathbf{W}\mathbf{A}^{-1}]$ με $\mathbf{A} = [\mathbf{I} - \hat{\rho}\mathbf{W}]$ και $\text{Var}(\hat{\rho})$ η

εκτιμηθείσα διακύμανση της παραμέτρου ρ . Η στατιστική $LM_{\lambda\rho}$ υπό τη μηδενική υπόθεση ακολουθεί την κατανομή χ^2 με ένα βαθμό ελευθερίας. Εφόσον $LM_{\lambda\rho} \sim \chi^2_1$ για την τετραγωνική της ρίζα θα ισχύει:

$$\frac{\frac{\hat{e}'\mathbf{W}\hat{e}}{S_{ML}^2}}{\sqrt{T - T_A^2 \text{Var}(\hat{\rho})}} \sim N(0, 1) \quad (2.13)$$

Αλλά στη σχέση (2.13) ο αριθμητής μπορεί να γραφεί ως:

$$\frac{\hat{e}'\mathbf{W}\hat{e}}{S_{ML}^2} = \frac{\hat{e}'\mathbf{W}\hat{e}}{\frac{\hat{e}'\hat{e}}{n}} = n \frac{\hat{e}'\mathbf{W}\hat{e}}{\hat{e}'\hat{e}} = nI$$

οπότε

$$\sqrt{LM_{\lambda\rho}} = \frac{n}{\sqrt{T - T_A^2 \text{Var}(\hat{\rho})}} I$$

Κατά συνέπεια, η στατιστική $LM_{\lambda\rho}$ στην πραγματικότητα αποτελείται από το γινόμενο μίας στατιστικής I του Moran που έχει υπολογιστεί με τα κατάλοιπα από την εκτίμηση με τη μέθοδο της μέγιστης πιθανοφάνειας του υποδείγματος χωρικής υστέρησης επί κάποιον παράγοντα διόρθωσης. Όπως αναφέρουν οι Anselin, Bera, Florax, και Yoon (1996) η στατιστική αυτή θα μπορούσε να χαρακτηριστεί ως η χωρική εκδοχή της γνωστής από την ανάλυση χρονοσειρών στατιστικής h του Durbin και ελέγχει για την ύπαρξη χωρικής αυτοσυσχέτισης στα σφάλματα ενός υποδείγματος λαμβάνοντας υπ' όψιν την παρουσία της χωρικής υστέρησης της εξαρτημένης μεταβλητής. Η διαφορά της σε σχέση με τους υπόλοιπους ελέγχους χωρικής εξάρτησης της κατηγορίας πολλαπλασιαστική Lagrange είναι ότι για τον υπολογισμό της απαιτείται η εκτίμηση με τη μέθοδο της μέγιστης πιθανοφάνειας του χωρικού υποδείγματος της χωρικής υστέρησης και δεν αρκεί η εκτίμηση μόνο του απλού οικονομετρικού υποδείγματος.

2.5.2 Η ισχύς των ελέγχων χωρικής αυτοσυσχέτισης στο υπόδειγμα χωρικής υστέρησης

Η αξιοπιστία των στατιστικών I του Moran και της $LM_{\lambda\rho}$ για τον εντοπισμό πιθανής χωρικής αυτοσυσχέτισης στα σφάλματα του υποδείγματος χωρικής υστέρησης μελετώνται με τη βοήθεια προσομοίωσης. Η προσομοίωση αυτή πραγματοποιήθηκε επεκτείνοντας τη διαδικασία που παρουσιάστηκε αναλυτικά στην ενότητα 2.4. Πιο συγκεκριμένα, θεωρήθηκε ότι τα σφάλματα του υποδείγματος χωρικής υστέρησης δεν είναι ανεξάρτητα αλλά παρουσιάζουν χωρική αυτοσυσχέτιση σύμφωνα με μια στάσιμη χωρική αυτοπαλίνδρομη

διαδικασία πρώτης τάξης, (SAR(1)). Το υπόδειγμα χωρικής υστέρησης με χωρική αυτοσυσχέτιση στα σφάλματα παράγεται ως εξής:

$$\mathbf{y} = (\mathbf{I} - \rho \mathbf{W})^{-1} [\mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \lambda \mathbf{W})^{-1} \mathbf{u}]$$

κατόπιν μερικών απλών αλγεβρικών πράξεων στη σχέση (2.6) και θεωρώντας ότι $\mathbf{W}_1 = \mathbf{W}_2 = \mathbf{W}$. Χρησιμοποιήθηκαν πάλι δύο ανεξάρτητες μεταβλητές προερχόμενες από $U(0, 10)$ που διατηρήθηκαν σταθερές σε όλες τις επαναλήψεις και τρεις συντελεστές ίσοι με τη μονάδα. Διατηρήθηκαν τα ίδια μεγέθη δειγμάτων και μήτρες χωρικών σταθμίσεων. Τα σφάλματα \mathbf{u} παρείχθησαν από $N(0, \mathbf{I})$ κατανομή και εκτελέστηκαν 10.000 επαναλήψεις με τους συντελεστές ρ και λ να παίρνουν τις τιμές 0,0, 0,2, 0,5, 0,8 και 0,9. Το υπόδειγμα χωρικής υστέρησης εκτιμήθηκε με τη μέθοδο της μέγιστης πιθανοφάνειας και καταγράφηκαν οι μέσες τιμές της στατιστικής I του Moran που προέκυψαν για τα σφάλματα, τα ποσοστά απόρριψης της ίδιας στατιστικής για τη μηδενική υπόθεση της απουσίας χωρικής αυτοσυσχέτισης με τη διακύμανσής της να υπολογίζεται θεωρώντας την υπόθεση της τυχαιοποίησης και τα ποσοστά απόρριψης του ίδιου ελέγχου με τη στατιστική $LM_{\lambda\rho}$.

Οι μέσες τιμές της στατιστικής I του Moran που προέκυψαν από την προσομοίωση για όλα τα δείγματα και τις μήτρες χωρικών σταθμίσεων συνοψίζονται στον Πίνακα 2.9 χωρίς διόρθωση τόρος και στον Πίνακα 2.10 με διόρθωση τόρος. Όπως διαπιστώνεται, για δεδομένο μέγεθος δείγματος η μέση τιμή της I αυξάνεται καθώς η τιμή του συντελεστή λ του αυτοπαλίνδρομου χωρικού υποδείγματος των σφαλμάτων αυξάνεται υποδεικνύοντας έντονα την ύπαρξη του προβλήματος της χωρικής αυτοσυσχέτισης στα εκτιμηθέντα σφάλματα. Όμως, για κάθε συγκεκριμένο αριθμό χωρικών μονάδων η τιμή της στατιστικής I δεν επηρεάζεται από το συντελεστή χωρικής υστέρησης της εξαρτημένης μεταβλητής και παραμένει στο ίδιο περίπου επίπεδο ανεξάρτητα των τιμών του ρ . Αντίθετα, η αύξηση του μεγέθους του δείγματος οδηγεί σε μικρή αύξηση στη μέση τιμή της I . Η συμπεριφορά της στατιστικής που περιγράφηκε ισχύει ανεξάρτητα της μήτρας χωρικών σταθμίσεων με τη μοναδική διαφορά που εντοπίζεται να είναι ότι με το κριτήριο της βασίλισσας έχουν προκύψει λίγο πιο μικρές μέσες τιμές κάτι που συμβαίνει και όταν εφαρμόζεται διόρθωση τόρος.

Πίνακας 2.9

Μέσες τιμές I για τα σφάλματα του υποδείγματος χωρικής υστέρησης με χωρικά αυτοπαλίνδρομα σφάλματα για όλες τις τιμές των ρ και λ και μεγέθη δείγματος Χωρίς διόρθωση τόρος

| Σταθμίσεις | | Πύργος | | | | | Βασίλισσα | | | | |
|------------|---------------------------|--------|-------|-------|-------|-------|-----------|-------|-------|-------|-------|
| N | $\rho \backslash \lambda$ | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 |
| 49 | 0,0 | -0,024 | 0,067 | 0,209 | 0,348 | 0,359 | -0,029 | 0,022 | 0,124 | 0,264 | 0,307 |
| | 0,2 | -0,025 | 0,065 | 0,206 | 0,337 | 0,344 | -0,030 | 0,021 | 0,122 | 0,260 | 0,304 |
| | 0,5 | -0,028 | 0,063 | 0,201 | 0,325 | 0,336 | -0,031 | 0,019 | 0,119 | 0,256 | 0,302 |
| | 0,8 | -0,030 | 0,061 | 0,198 | 0,332 | 0,378 | -0,032 | 0,018 | 0,118 | 0,261 | 0,327 |
| | 0,9 | -0,028 | 0,062 | 0,201 | 0,363 | 0,449 | -0,030 | 0,021 | 0,123 | 0,285 | 0,379 |
| 100 | 0,0 | -0,008 | 0,086 | 0,236 | 0,389 | 0,398 | -0,012 | 0,042 | 0,151 | 0,304 | 0,348 |
| | 0,2 | -0,009 | 0,085 | 0,233 | 0,379 | 0,386 | -0,012 | 0,041 | 0,149 | 0,299 | 0,341 |
| | 0,5 | -0,010 | 0,084 | 0,230 | 0,369 | 0,381 | -0,013 | 0,041 | 0,147 | 0,290 | 0,332 |
| | 0,8 | -0,011 | 0,084 | 0,231 | 0,376 | 0,417 | -0,013 | 0,041 | 0,147 | 0,290 | 0,347 |
| | 0,9 | -0,010 | 0,086 | 0,236 | 0,401 | 0,476 | -0,012 | 0,042 | 0,151 | 0,309 | 0,396 |
| 400 | 0,0 | -0,002 | 0,093 | 0,247 | 0,416 | 0,431 | -0,003 | 0,052 | 0,166 | 0,341 | 0,394 |
| | 0,2 | -0,002 | 0,092 | 0,245 | 0,406 | 0,418 | -0,003 | 0,051 | 0,164 | 0,334 | 0,384 |
| | 0,5 | -0,003 | 0,092 | 0,244 | 0,400 | 0,416 | -0,003 | 0,051 | 0,163 | 0,326 | 0,375 |
| | 0,8 | -0,003 | 0,094 | 0,248 | 0,413 | 0,455 | -0,003 | 0,052 | 0,165 | 0,331 | 0,389 |
| | 0,9 | -0,003 | 0,095 | 0,254 | 0,435 | 0,501 | -0,003 | 0,053 | 0,169 | 0,345 | 0,422 |
| 900 | 0,0 | -0,001 | 0,092 | 0,243 | 0,408 | 0,423 | -0,001 | 0,052 | 0,162 | 0,333 | 0,386 |
| | 0,2 | -0,001 | 0,092 | 0,242 | 0,400 | 0,412 | -0,001 | 0,051 | 0,161 | 0,328 | 0,381 |
| | 0,5 | -0,001 | 0,092 | 0,241 | 0,396 | 0,413 | -0,001 | 0,052 | 0,161 | 0,325 | 0,377 |
| | 0,8 | -0,001 | 0,094 | 0,247 | 0,412 | 0,455 | -0,002 | 0,053 | 0,165 | 0,332 | 0,392 |
| | 0,9 | -0,001 | 0,096 | 0,254 | 0,433 | 0,498 | -0,002 | 0,054 | 0,169 | 0,345 | 0,421 |

Πίνακας 2.10

Μέσες τιμές I για τα σφάλματα του υποδείγματος χωρικής υστέρησης με χωρικά αυτοπαλίνδρομα σφάλματα για όλες τις τιμές των ρ και λ και μεγέθη δείγματος Με διόρθωση τόρος

| Σταθμίσεις | | Πύργος | | | | | Βασίλισσα | | | | |
|------------|---------------------------|--------|-------|-------|-------|-------|-----------|-------|-------|-------|-------|
| N | $\rho \backslash \lambda$ | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 |
| 49 | 0,0 | -0,023 | 0,051 | 0,168 | 0,292 | 0,328 | -0,028 | 0,009 | 0,078 | 0,173 | 0,210 |
| | 0,2 | -0,025 | 0,050 | 0,167 | 0,289 | 0,324 | -0,028 | 0,008 | 0,077 | 0,172 | 0,210 |
| | 0,5 | -0,026 | 0,049 | 0,165 | 0,288 | 0,324 | -0,029 | 0,008 | 0,077 | 0,173 | 0,211 |
| | 0,8 | -0,027 | 0,049 | 0,167 | 0,299 | 0,348 | -0,027 | 0,010 | 0,080 | 0,182 | 0,226 |
| | 0,9 | -0,025 | 0,052 | 0,174 | 0,322 | 0,387 | -0,023 | 0,015 | 0,088 | 0,199 | 0,250 |
| 100 | 0,0 | -0,008 | 0,074 | 0,204 | 0,346 | 0,377 | -0,012 | 0,031 | 0,116 | 0,241 | 0,290 |
| | 0,2 | -0,009 | 0,073 | 0,202 | 0,339 | 0,368 | -0,012 | 0,030 | 0,115 | 0,240 | 0,288 |
| | 0,5 | -0,010 | 0,072 | 0,201 | 0,337 | 0,368 | -0,013 | 0,030 | 0,115 | 0,238 | 0,287 |
| | 0,8 | -0,011 | 0,073 | 0,204 | 0,348 | 0,397 | -0,012 | 0,031 | 0,116 | 0,243 | 0,300 |
| | 0,9 | -0,010 | 0,074 | 0,208 | 0,367 | 0,444 | -0,011 | 0,032 | 0,120 | 0,259 | 0,335 |
| 400 | 0,0 | -0,002 | 0,087 | 0,231 | 0,393 | 0,418 | -0,003 | 0,046 | 0,146 | 0,300 | 0,354 |
| | 0,2 | -0,002 | 0,086 | 0,229 | 0,383 | 0,403 | -0,003 | 0,045 | 0,144 | 0,294 | 0,344 |
| | 0,5 | -0,003 | 0,086 | 0,227 | 0,374 | 0,394 | -0,003 | 0,045 | 0,143 | 0,286 | 0,333 |
| | 0,8 | -0,003 | 0,087 | 0,230 | 0,382 | 0,422 | -0,003 | 0,046 | 0,144 | 0,287 | 0,340 |
| | 0,9 | -0,003 | 0,089 | 0,235 | 0,399 | 0,466 | -0,003 | 0,046 | 0,147 | 0,296 | 0,371 |
| 900 | 0,0 | -0,001 | 0,088 | 0,234 | 0,397 | 0,420 | -0,001 | 0,048 | 0,150 | 0,311 | 0,368 |
| | 0,2 | -0,001 | 0,088 | 0,232 | 0,389 | 0,409 | -0,001 | 0,048 | 0,150 | 0,306 | 0,360 |
| | 0,5 | -0,001 | 0,089 | 0,232 | 0,384 | 0,406 | -0,001 | 0,048 | 0,149 | 0,301 | 0,352 |
| | 0,8 | -0,001 | 0,090 | 0,237 | 0,395 | 0,436 | -0,001 | 0,049 | 0,151 | 0,303 | 0,359 |
| | 0,9 | -0,001 | 0,092 | 0,242 | 0,412 | 0,476 | -0,001 | 0,049 | 0,154 | 0,313 | 0,385 |

Τα αντίστοιχα ποσοστά απόρριψης της στατιστικής I του Moran για τον έλεγχο της απουσίας χωρικής αυτοσυσχέτισης υποθέτοντας τυχαιοποίηση δίνονται στους Πίνακες 2.11 και 2.12 χωρίς τόρος και με τόρος αντίστοιχα. Το χαρακτηριστικό που πρέπει να αναφερθεί είναι ότι σε όλες τις περιπτώσεις όταν ο συντελεστής λ ισούται με το μηδέν, δηλαδή τα σφάλματα δεν αυτοσυσχετίζονται, τα ποσοστά απόρριψης (το μέγεθος του ελέγχου) είναι μικρότερα από το επίπεδο σημαντικότητας του 5%. Αυτό δηλώνει ότι η στατιστική I δεν ακολουθεί ακριβώς την τυπική κανονική κατανομή όταν η εξεταζόμενη για την παρουσία χωρικής αυτοσυσχέτισης μεταβλητή είναι τα κατάλοιπα από την εκτίμηση ενός υποδείγματος χωρικής υστέρησης. Στις άλλες περιπτώσεις τα ποσοστά απόρριψης της I παρουσιάζουν παρόμοια συμπεριφορά με τις μέσες τιμές της. Δηλαδή, με δεδομένο δείγμα, το ποσοστό απόρριψης του ελέγχου αυξάνεται με την αύξηση της τιμής του συντελεστή λ . Τα ποσοστά απόρριψης δεν φαίνεται να επηρεάζονται από τη μεταβολή της τιμής του ρ αλλά μόνο από τον αριθμό των παρατηρήσεων παραμένοντας περίπου τα ίδια για δεδομένο λ ενώ αυξάνουν με την αύξηση του δείγματος.

Η μέγιστη ισχύς του ελέγχου επιτυγχάνεται στα δείγματα των 400 και 900 παρατηρήσεων και για τιμές του λ μεγαλύτερες ή ίσες του 0,5 με ποσοστά απόρριψης 100%. Η χρησιμοποίηση του κριτηρίου της βασίλισσας για τον ορισμό της μήτρας χωρικών σταθμίσεων αλλά και η εφαρμογή διόρθωσης τόρος δεν αλλάζουν σημαντικά τα αποτελέσματα και η μοναδική διαφορά που εντοπίζεται είναι τα ελαφρά μικρότερα ποσοστά απόρριψης στα μικρά δείγματα και στις μικρές τιμές του λ .

Τα ποσοστά απόρριψης του ελέγχου για την απουσία χωρικής αυτοσυσχέτισης στα σφάλματα του υποδείγματος χωρικής υστέρησης που προέκυψαν από την προσομοίωση με την εφαρμογή της στατιστικής $LM_{\lambda\rho}$ περιλαμβάνονται στον Πίνακα 2.13 χωρίς διόρθωση τόρος και στον Πίνακα 2.14 με διόρθωση τόρος. Όπως φαίνεται στους πίνακες αυτούς, όταν ο συντελεστής λ ισούται με το μηδέν για όλα τα μεγέθη δείγματος το ποσοστό απόρριψης, δηλαδή το μέγεθος του ελέγχου, είναι κοντά στο επίπεδο σημαντικότητας του 5% κάτι που ισχύει ανεξάρτητα από τη κριτήριο γειτνίασης και την απουσία ή παρουσία διόρθωσης τόρος. Όπως και με τα ποσοστά απόρριψης της στατιστικής I , τα ποσοστά απόρριψης της στατιστικής $LM_{\lambda\rho}$ επηρεάζονται αποκλειστικά από τις τιμές του λ και αυξάνονται καθώς η χωρική αυτοσυσχέτιση στα σφάλματα γίνεται πιο ισχυρή. Το αποτέλεσμα αυτό είναι σε κάποιο βαθμό αναμενόμενο λαμβάνοντας υπ' όψιν ότι στην πραγματικότητα η εξεταζόμενη στατιστική αποτελεί μια διορθωμένη στατιστική I του Moran.

Πίνακας 2.11

Ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% της στατιστικής I για τα σφάλματα του υποδείγματος χωρικής υστέρησης με χωρικά αυτοπαλίνδρομα σφάλματα για όλες τις τιμές των ρ και λ και μεγέθη δείγματος Χωρίς διόρθωση τόρος

| Σταθμίσεις | | Πύργος | | | | | Βασίλισσα | | | | |
|------------|--------------------------|--------|-------|-------|-------|-------|-----------|-------|-------|-------|-------|
| N | $\rho \setminus \lambda$ | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 |
| 49 | 0,0 | 3,66 | 12,08 | 58,39 | 96,05 | 98,61 | 3,33 | 10,93 | 46,24 | 87,78 | 94,89 |
| | 0,2 | 3,71 | 11,77 | 57,19 | 95,10 | 97,44 | 3,39 | 10,54 | 45,29 | 86,56 | 94,09 |
| | 0,5 | 3,69 | 11,31 | 55,49 | 92,99 | 95,60 | 3,36 | 10,06 | 44,30 | 85,25 | 92,90 |
| | 0,8 | 3,69 | 10,68 | 54,18 | 92,46 | 97,07 | 3,32 | 9,78 | 43,76 | 85,07 | 93,56 |
| | 0,9 | 3,62 | 10,88 | 55,33 | 95,04 | 98,84 | 3,08 | 10,14 | 45,16 | 88,28 | 96,22 |
| 100 | 0,0 | 3,91 | 24,25 | 91,64 | 99,95 | 99,99 | 3,64 | 18,12 | 77,09 | 99,57 | 99,95 |
| | 0,2 | 3,91 | 24,01 | 91,21 | 99,94 | 99,97 | 3,58 | 17,97 | 76,50 | 99,46 | 99,93 |
| | 0,5 | 3,88 | 23,69 | 90,47 | 99,92 | 99,98 | 3,53 | 17,68 | 75,67 | 99,31 | 99,86 |
| | 0,8 | 4,14 | 23,84 | 90,58 | 99,91 | 99,98 | 3,57 | 17,77 | 75,57 | 99,21 | 99,86 |
| | 0,9 | 4,12 | 24,52 | 91,68 | 99,97 | 100 | 3,73 | 18,44 | 76,96 | 99,43 | 99,95 |
| 400 | 0,0 | 4,11 | 75,83 | 100 | 100 | 100 | 3,99 | 53,11 | 99,95 | 100 | 100 |
| | 0,2 | 4,11 | 75,67 | 100 | 100 | 100 | 3,87 | 52,63 | 99,94 | 100 | 100 |
| | 0,5 | 4,14 | 75,50 | 100 | 100 | 100 | 3,87 | 52,16 | 99,95 | 100 | 100 |
| | 0,8 | 4,18 | 76,74 | 100 | 100 | 100 | 4,11 | 53,23 | 99,96 | 100 | 100 |
| | 0,9 | 4,49 | 78,19 | 100 | 100 | 100 | 4,24 | 54,50 | 99,97 | 100 | 100 |
| 900 | 0,0 | 3,86 | 97,76 | 100 | 100 | 100 | 3,89 | 83,95 | 100 | 100 | 100 |
| | 0,2 | 3,81 | 97,72 | 100 | 100 | 100 | 3,87 | 83,79 | 100 | 100 | 100 |
| | 0,5 | 3,85 | 97,80 | 100 | 100 | 100 | 3,83 | 84,04 | 100 | 100 | 100 |
| | 0,8 | 4,12 | 97,99 | 100 | 100 | 100 | 4,07 | 85,34 | 100 | 100 | 100 |
| | 0,9 | 4,32 | 98,29 | 100 | 100 | 100 | 4,28 | 86,40 | 100 | 100 | 100 |

Πίνακας 2.12

Ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% της στατιστικής I για τα σφάλματα του υποδείγματος χωρικής υστέρησης με χωρικά αυτοπαλίνδρομα σφάλματα για όλες τις τιμές των ρ και λ και μεγέθη δείγματος Με διόρθωση τόρος

| Σταθμίσεις | | Πύργος | | | | | Βασίλισσα | | | | |
|------------|--------------------------|--------|-------|-------|-------|-------|-----------|-------|-------|-------|-------|
| N | $\rho \setminus \lambda$ | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 |
| 49 | 0,0 | 3,81 | 10,44 | 50,02 | 90,67 | 96,55 | 3,81 | 9,04 | 34,89 | 75,10 | 85,02 |
| | 0,2 | 3,84 | 10,16 | 49,54 | 89,91 | 96,05 | 3,87 | 8,90 | 34,60 | 74,71 | 84,77 |
| | 0,5 | 3,94 | 10,09 | 48,82 | 89,16 | 95,49 | 3,88 | 8,87 | 34,59 | 74,38 | 84,82 |
| | 0,8 | 3,96 | 10,16 | 49,55 | 90,07 | 96,29 | 3,84 | 9,08 | 36,08 | 76,79 | 86,82 |
| | 0,9 | 3,91 | 10,76 | 52,07 | 92,77 | 97,85 | 3,93 | 10,09 | 39,61 | 81,26 | 90,29 |
| 100 | 0,0 | 3,95 | 21,74 | 87,58 | 99,89 | 99,99 | 3,56 | 15,76 | 67,01 | 97,78 | 99,50 |
| | 0,2 | 3,91 | 21,41 | 87,00 | 99,85 | 99,98 | 3,56 | 15,47 | 66,69 | 97,69 | 99,42 |
| | 0,5 | 4,04 | 21,06 | 86,57 | 99,82 | 99,96 | 3,57 | 15,44 | 66,57 | 97,75 | 99,33 |
| | 0,8 | 4,13 | 21,22 | 87,13 | 99,85 | 99,98 | 3,62 | 15,77 | 67,17 | 98,07 | 99,48 |
| | 0,9 | 4,18 | 21,90 | 88,41 | 99,92 | 100 | 3,68 | 16,58 | 69,19 | 98,69 | 99,83 |
| 400 | 0,0 | 4,12 | 72,97 | 100 | 100 | 100 | 3,73 | 48,59 | 99,81 | 100 | 100 |
| | 0,2 | 4,12 | 72,55 | 100 | 100 | 100 | 3,75 | 47,99 | 99,80 | 100 | 100 |
| | 0,5 | 4,13 | 72,35 | 100 | 100 | 100 | 3,81 | 47,18 | 99,81 | 100 | 100 |
| | 0,8 | 4,17 | 73,22 | 100 | 100 | 100 | 3,88 | 47,96 | 99,78 | 100 | 100 |
| | 0,9 | 4,19 | 74,51 | 100 | 100 | 100 | 4,04 | 49,36 | 99,84 | 100 | 100 |
| 900 | 0,0 | 3,73 | 97,48 | 100 | 100 | 100 | 3,73 | 81,22 | 100 | 100 | 100 |
| | 0,2 | 3,72 | 97,45 | 100 | 100 | 100 | 3,69 | 81,17 | 100 | 100 | 100 |
| | 0,5 | 3,63 | 97,53 | 100 | 100 | 100 | 3,73 | 81,28 | 100 | 100 | 100 |
| | 0,8 | 3,94 | 97,77 | 100 | 100 | 100 | 3,85 | 82,23 | 100 | 100 | 100 |
| | 0,9 | 4,12 | 98,00 | 100 | 100 | 100 | 3,97 | 83,46 | 100 | 100 | 100 |

Πίνακας 2.13

Ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% της στατιστικής $LM_{\lambda\rho}$ για τα σφάλματα του υποδείγματος χωρικής υστέρησης με χωρικά αυτοπαλίνδρομα σφάλματα για όλες τις τιμές των ρ και λ και μεγέθη δείγματος Χωρίς διόρθωση τόρος

| Σταθμίσεις | | Πύργος | | | | | Βασίλισσα | | | | |
|------------|---------------------------|--------|-------|-------|-------|-------|-----------|-------|-------|-------|-------|
| N | $\rho \backslash \lambda$ | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 |
| 49 | 0,0 | 5,03 | 9,69 | 54,21 | 95,32 | 98,44 | 4,68 | 8,09 | 39,38 | 84,19 | 93,41 |
| | 0,2 | 5,06 | 9,51 | 52,82 | 94,26 | 97,49 | 4,74 | 7,86 | 38,62 | 83,36 | 92,52 |
| | 0,5 | 5,11 | 9,30 | 51,03 | 92,00 | 95,41 | 4,89 | 7,68 | 37,85 | 81,95 | 91,17 |
| | 0,8 | 5,36 | 8,94 | 50,10 | 91,49 | 96,72 | 5,01 | 7,83 | 38,02 | 82,28 | 92,19 |
| | 0,9 | 5,39 | 9,37 | 52,41 | 94,22 | 98,62 | 5,04 | 8,31 | 40,51 | 86,08 | 95,43 |
| 100 | 0,0 | 4,85 | 22,97 | 91,14 | 99,94 | 99,99 | 4,56 | 16,11 | 74,44 | 99,46 | 99,92 |
| | 0,2 | 4,90 | 22,77 | 90,67 | 99,94 | 99,97 | 4,55 | 16,01 | 73,98 | 99,32 | 99,90 |
| | 0,5 | 4,99 | 22,38 | 89,87 | 99,91 | 99,98 | 4,66 | 15,87 | 73,14 | 99,15 | 99,81 |
| | 0,8 | 5,19 | 22,20 | 89,75 | 99,90 | 99,98 | 4,81 | 16,01 | 73,25 | 99,05 | 99,83 |
| | 0,9 | 5,28 | 22,61 | 90,90 | 99,97 | 100 | 4,90 | 16,53 | 74,61 | 99,29 | 99,92 |
| 400 | 0,0 | 5,19 | 76,63 | 100 | 100 | 100 | 5,09 | 52,63 | 99,95 | 100 | 100 |
| | 0,2 | 5,17 | 76,44 | 100 | 100 | 100 | 5,11 | 52,22 | 99,94 | 100 | 100 |
| | 0,5 | 5,12 | 76,16 | 100 | 100 | 100 | 5,05 | 51,77 | 99,95 | 100 | 100 |
| | 0,8 | 5,09 | 76,79 | 100 | 100 | 100 | 5,03 | 52,43 | 99,96 | 100 | 100 |
| | 0,9 | 5,25 | 77,72 | 100 | 100 | 100 | 5,10 | 53,22 | 99,97 | 100 | 100 |
| 900 | 0,0 | 4,98 | 98,13 | 100 | 100 | 100 | 5,02 | 85,00 | 100 | 100 | 100 |
| | 0,2 | 5,14 | 98,07 | 100 | 100 | 100 | 5,04 | 84,77 | 100 | 100 | 100 |
| | 0,5 | 5,10 | 98,05 | 100 | 100 | 100 | 4,96 | 84,83 | 100 | 100 | 100 |
| | 0,8 | 4,94 | 98,12 | 100 | 100 | 100 | 4,96 | 85,66 | 100 | 100 | 100 |
| | 0,9 | 4,98 | 98,33 | 100 | 100 | 100 | 5,03 | 86,33 | 100 | 100 | 100 |

Πίνακας 2.14

Ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% της στατιστικής $LM_{\lambda\rho}$ για τα σφάλματα του υποδείγματος χωρικής υστέρησης με χωρικά αυτοπαλίνδρομα σφάλματα για όλες τις τιμές των ρ και λ και μεγέθη δείγματος Με διόρθωση τόρος

| Σταθμίσεις | | Πύργος | | | | | Βασίλισσα | | | | |
|------------|---------------------------|--------|-------|-------|-------|-------|-----------|-------|-------|-------|-------|
| N | $\rho \backslash \lambda$ | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 |
| 49 | 0,0 | 4,89 | 7,73 | 43,63 | 88,02 | 95,32 | 4,33 | 5,95 | 25,78 | 67,14 | 79,92 |
| | 0,2 | 4,91 | 7,62 | 42,94 | 87,21 | 94,74 | 4,44 | 5,94 | 25,70 | 66,94 | 79,73 |
| | 0,5 | 4,96 | 7,45 | 42,37 | 86,37 | 93,95 | 4,61 | 6,13 | 26,02 | 67,34 | 79,82 |
| | 0,8 | 5,16 | 7,73 | 43,43 | 87,70 | 95,07 | 4,95 | 6,70 | 28,63 | 71,20 | 83,44 |
| | 0,9 | 5,20 | 8,34 | 46,62 | 90,94 | 97,14 | 4,87 | 7,76 | 32,44 | 76,83 | 87,63 |
| 100 | 0,0 | 4,99 | 20,11 | 86,23 | 99,87 | 99,99 | 4,27 | 13,09 | 62,97 | 97,24 | 99,32 |
| | 0,2 | 4,97 | 19,77 | 85,72 | 99,84 | 99,98 | 4,38 | 13,05 | 62,70 | 97,12 | 99,18 |
| | 0,5 | 5,12 | 19,42 | 85,08 | 99,73 | 99,96 | 4,47 | 12,86 | 62,53 | 97,08 | 99,16 |
| | 0,8 | 5,27 | 19,42 | 85,47 | 99,78 | 99,98 | 4,70 | 13,22 | 63,37 | 97,56 | 99,41 |
| | 0,9 | 5,20 | 19,88 | 86,56 | 99,92 | 100 | 4,78 | 14,13 | 65,73 | 98,44 | 99,78 |
| 400 | 0,0 | 5,11 | 73,79 | 100 | 100 | 100 | 4,83 | 48,14 | 99,81 | 100 | 100 |
| | 0,2 | 5,14 | 73,39 | 100 | 100 | 100 | 4,94 | 47,57 | 99,80 | 100 | 100 |
| | 0,5 | 5,08 | 73,05 | 100 | 100 | 100 | 4,98 | 46,85 | 99,80 | 100 | 100 |
| | 0,8 | 5,18 | 73,51 | 100 | 100 | 100 | 5,02 | 47,40 | 99,77 | 100 | 100 |
| | 0,9 | 5,06 | 74,31 | 100 | 100 | 100 | 4,99 | 48,46 | 99,81 | 100 | 100 |
| 900 | 0,0 | 4,84 | 97,87 | 100 | 100 | 100 | 4,86 | 82,17 | 100 | 100 | 100 |
| | 0,2 | 4,86 | 97,83 | 100 | 100 | 100 | 4,90 | 82,24 | 100 | 100 | 100 |
| | 0,5 | 4,76 | 97,81 | 100 | 100 | 100 | 4,74 | 82,09 | 100 | 100 | 100 |
| | 0,8 | 4,80 | 97,90 | 100 | 100 | 100 | 4,73 | 82,73 | 100 | 100 | 100 |
| | 0,9 | 4,86 | 98,02 | 100 | 100 | 100 | 4,79 | 83,52 | 100 | 100 | 100 |

Η αύξηση του δείγματος οδηγεί σε αύξηση στην ισχύ της στατιστικής $LM_{\lambda\rho}$ όπως προκύπτει από τα μεγαλύτερα ποσοστά απόρριψης. Η προηγούμενη συμπεριφορά εμφανίζεται και όταν εφαρμόζεται το κριτήριο της βασίλισσας αλλά με μικρότερα ποσοστά απόρριψης. Η εφαρμογή της διόρθωσης τóρος οδηγεί στα ίδια συμπεράσματα με τη διαφορά ότι επιτυγχάνονται πιο μικρά ποσοστά απόρριψης σε κάθε τύπο χωρικής γειννίασης σε σχέση με τα αντίστοιχα αποτελέσματα όταν απουσιάζει η διόρθωση. Μία εξαίρεση στα παραπάνω συμπεράσματα που παρατηρείται όταν εφαρμόζεται η διόρθωση τóρος είναι ότι για δείγμα 49 παρατηρήσεων και με τη χρησιμοποίηση του κριτηρίου της βασίλισσας το ποσοστό απόρριψης αυξάνεται όταν αυξάνονται οι τιμές της παραμέτρου χωρικής υστέρησης της εξαρτημένης μεταβλητής ρ .

Γενικά, ο έλεγχος με τη στατιστική $LM_{\lambda\rho}$ επιτυγχάνει τη μέγιστη ισχύ του στα δείγματα των 400 και 900 παρατηρήσεων και για τιμές του λ ίσες με 0,5, 0,8 και 0,9. Στα μικρά δείγματα και ιδιαίτερα για μικρές τιμές του λ η στατιστική δεν έχει ικανοποιητική ισχύ. Η έλλειψη ικανοποιητικής ισχύς έχει αναφερθεί και στην εργασία των Anselin, Bera, Florax, και Yoon (1996) χωρίς όμως να συνοδεύεται από αποτελέσματα κάποιας προσομοίωσης. Συγκρίνοντας την ισχύ της στατιστικής $LM_{\lambda\rho}$ με την ισχύ της στατιστικής I διαπιστώνεται ότι στα δείγματα των 49 και 100 παρατηρήσεων η στατιστική I είναι πιο ισχυρή στον εντοπισμό του προβλήματος της χωρικής αυτοσυσχέτισης όπως αποδεικνύεται από τα μεγαλύτερα ποσοστά απόρριψης που επιτυγχάνει και περιλαμβάνονται στους Πίνακες 2.11 και 2.12. Αντίθετα, στα δείγματα των 400 και 900 παρατηρήσεων η ισχύς των δύο στατιστικών είναι ουσιαστικά η ίδια.

Το τελικό συμπέρασμα από την ανάλυση των αποτελεσμάτων της προσομοίωσης είναι ότι η στατιστική I του Moran όταν χρησιμοποιείται στον έλεγχο της ύπαρξης χωρικής αυτοσυσχέτισης στα σφάλματα ενός υποδείγματος χωρικής υστέρησης στην εξαρτημένη μεταβλητή και υπολογίζοντας για την πραγματοποίηση του ελέγχου τη διακύμανσή της με την προσέγγιση της τυχαιοποίησης παρουσιάζει σημαντική στατιστική ισχύ στον εντοπισμό του προβλήματος. Μάλιστα, στα μικρά δείγματα η ισχύς της είναι ανώτερη από την ισχύ της στατιστικής $LM_{\lambda\rho}$. Στα μεγαλύτερα δείγματα οι δύο στατιστικές έχουν παρόμοια ισχύ. Κατά συνέπεια, συνιστάται ο ερευνητής να χρησιμοποιεί και τις δύο στατιστικές για την πραγματοποίηση του συγκεκριμένου ελέγχου συγκρίνοντας και επαληθεύοντας τα αποτελέσματά τους.

2.6 Χωρική αυτοσυσχέτιση στα σφάλματα του γραμμικού υποδείγματος παλινδρόμησης λόγω ύπαρξης υποφαινομενικής σχέσης

Στις προηγούμενες ενότητες εξετάστηκαν οι συνέπειες που επέρχονται στη στατιστική συμπερασματολογία ενός οικονομετρικού υποδείγματος λόγω της παρουσίας χωρικής αυτοσυσχέτισης στα σφάλματα και μελετήθηκε η εφαρμογή των ελέγχων που χρησιμοποιούνται στον εντοπισμό του προβλήματος και στην εξειδίκευση του σωστού χωρικού οικονομετρικού υποδείγματος. Το βασικό χαρακτηριστικό στην ανάλυση ήταν ότι στο υπόδειγμα υπάρχει πραγματική σχέση ανάμεσα στην εξαρτημένη και τις ανεξάρτητες μεταβλητές και η εμφάνιση της χωρικής αυτοσυσχέτισης στα σφάλματα οφείλεται στην παρουσία των χωρικών επιδράσεων εξαιτίας της φύσης των στατιστικών δεδομένων. Στην ενότητα αυτή εξετάζεται ένα άλλο πρόβλημα που μπορεί να εμφανιστεί ταυτόχρονα με την ύπαρξη της χωρικής αυτοσυσχέτισης στα σφάλματα, δηλαδή η γνωστή από την ανάλυση χρονοσειρών υποφαινομενική παλινδρόμηση (*spurious regression*). Αφού παρουσιαστεί συνοπτικά η έννοια της υποφαινομενικής συμπεριφοράς διερευνάται με τη βοήθεια προσομοίωσης η εμφάνιση του προβλήματος στην περίπτωση που οι εμπλεκόμενες στην παλινδρόμηση μεταβλητές είναι στάσιμες χωρικές αυτοπαλίνδρομες διαδικασίες πρώτης τάξης (Spatial Autoregressive Process–SAR(1)) και μελετάται η συνεισφορά των ελέγχων χωρικής εξάρτησης στην αντιμετώπισή του μέσω της εκτίμησης κάποιου χωρικού οικονομετρικού υποδείγματος.

2.6.1 Η έννοια της υποφαινομενικής συμπεριφοράς

Η έννοια της υποφαινομενικής συμπεριφοράς (*spurious behavior*) είναι γνωστή από την ανάλυση χρονοσειρών ως υποφαινομενική παλινδρόμηση (*spurious regression*). Το πρόβλημα εμφανίζεται όταν δύο ή περισσότερες ανεξάρτητες χρονοσειρές εντοπίζονται να συσχετίζονται σύμφωνα με τη στατιστική συμπερασματολογία της μεθόδου των ελαχίστων τετραγώνων. Δηλαδή, οι έλεγχοι στατιστικής σημαντικότητας των συντελεστών ενός υποδείγματος απορρίπτουν τη μηδενική υπόθεση ότι κάποιος συντελεστής ισούται με το μηδέν ενώ στην πραγματικότητα δεν υπάρχει πραγματική σχέση ανάμεσα στην εξαρτημένη και στην εξεταζόμενη ανεξάρτητη μεταβλητή. Όταν συμβαίνει αυτό, η στατιστική σημαντικότητα μεταξύ των δύο μεταβλητών είναι υποφαινομενική και ο ερευνητής οδηγείται σε λανθασμένα συμπεράσματα. Το φαινόμενο αυτό περιγράφηκε αρχικά από τον Yule (1926) μέσω του συντελεστή γραμμικής συσχέτισης ως υποφαινομενική συσχέτιση (*spurious*

correlation) αλλά έλαβε ιδιαίτερη προσοχή μετά τη δημοσίευση της εργασίας των Granger και Newbold (1974), οι οποίοι το ονόμασαν υποφαινομενική παλινδρόμηση (*spurious regression*). Πιο συγκεκριμένα, οι Granger και Newbold (1974) έδειξαν χρησιμοποιώντας προσομοίωση ότι όταν δύο ανεξάρτητοι τυχαίοι περίπατοι χωρίς σταθερό όρο (*drift – free random walks*) εμπλακούν σε μία απλή γραμμική παλινδρόμηση η t – στατιστική για τον έλεγχο της σημαντικότητας της ανεξάρτητης μεταβλητής απορρίπτεται σε ποσοστό 76% των επαναλήψεων και τα σφάλματα από την εκτίμηση του υποδείγματος παρουσιάζουν υψηλή αυτοσυσχέτιση. Από τότε, αρκετοί ερευνητές ασχολήθηκαν με το φαινόμενο της υποφαινομενικής συμπεριφοράς παρέχοντας μαθηματικές αποδείξεις, ενδείξεις αυτοσυσχέτισης στα σφάλματα και εμπειρικές μεθόδους διόρθωσης του προβλήματος. Ειδικότερα, οι Newbold και Davies (1978) διερεύνησαν την ύπαρξη αυτοσυσχέτισης στα σφάλματα ενός υποδείγματος πολλαπλής παλινδρόμησης όταν οι μεταβλητές αποτελούνται από μη στάσιμες διαδικασίες κινητών μέσων εντοπίζοντας και πάλι την παρουσία της. Ο Phillips (1986) απέδειξε μαθηματικά τα αποτελέσματα της προσομοίωσης των Granger και Newbold (1974) αναπτύσσοντας την κατάλληλη ασυμπτωτική θεωρία και αποδεικνύοντας ότι στην περίπτωση της υποφαινομενικής παλινδρόμησης η συνηθισμένη t – στατιστική που χρησιμοποιείται για τον έλεγχο της στατιστικής σημαντικότητας της ανεξάρτητης μεταβλητής δεν έχει οριακή κατανομή. Ο Marmol (1995) επέκτεινε την εργασία του Phillips (1986) για ολοκληρωμένες διαδικασίες υψηλότερης τάξης και απέδειξε ότι η στατιστική των Durbin και Watson για τον έλεγχο της ύπαρξης αυτοσυσχέτισης στα σφάλματα ενός υποδείγματος παλινδρόμησης θα συγκλίνει κατά πιθανότητα στο μηδέν όταν υπάρχει το πρόβλημα της υποφαινομενικής παλινδρόμησης. Επομένως, όταν εμφανίζεται το φαινόμενο θα πρέπει να αναμένονται χαμηλές τιμές της στατιστικής των Durbin και Watson κάτι που είχε επισημανθεί και στην εργασία των Granger και Newbold (1974). Οι Granger, Hyung, και Jeon (2001) χρησιμοποιώντας προσομοίωση βρήκαν υποφαινομενικά αποτελέσματα ακόμη και όταν οι μεταβλητές του υποδείγματος είναι στάσιμες αυτοπαλίνδρομες διαδικασίες πρώτης τάξης, AR(1). Ο Agiakloglou (2009) έδειξε ότι η αυτοσυσχέτιση στα σφάλματα ενός γραμμικού υποδείγματος στην περίπτωση δύο στάσιμων AR(1) διαδικασιών μπορεί να εμφανιστεί όχι μόνο μέσω των πρώτων ροπών αλλά και μέσω των δευτέρων, όπως ορίζονται από ένα αυτοπαλίνδρομο υπό συνθήκη ετεροοσκεδαστικότητας υπόδειγμα πρώτης τάξης, ARCH(1). Οι Agiakloglou και Tsimpanos (2012) εξέτασαν το φαινόμενο της υποφαινομενικής συσχέτισης για στάσιμες AR(1) διαδικασίες και έδειξαν ότι το πρόβλημα

αντιμετωπίζεται εάν χρησιμοποιηθεί στο έλεγχο της στατιστικής σημαντικότητας του συντελεστή γραμμικής συσχέτισης η πραγματική διακύμανση του συντελεστή, όπως δίνεται για δύο ανεξάρτητες AR(1) διαδικασίες από τον Bartlett (1935).

Η υποφαινομενική συμπεριφορά μπορεί να εμφανιστεί και όταν χρησιμοποιούνται χωρικά δεδομένα τα οποία έχουν ομοιότητες με τις χρονοσειρές. Η διαφορά μεταξύ χωρικών δεδομένων και χρονοσειρών υφίσταται στον τρόπο σχηματισμού των τιμών της εξεταζόμενης μεταβλητής. Ειδικότερα, στα χωρικά δεδομένα οι τιμές μιας μεταβλητής συσχετίζονται σύμφωνα με τις γεωγραφικές τους θέσεις δημιουργώντας συμπλέγματα (*clusters*) και όχι σύμφωνα με τη χρονική τους σειρά όπως συμβαίνει στις χρονοσειρές. Δηλαδή, η σχέση εξάρτησης είναι προς πολλές κατευθύνσεις και όχι μόνο προς δύο όπως στις χρονοσειρές. Κατά συνέπεια, όταν αναλύονται χωρικά δεδομένα οι συμβατικές οικονομετρικές τεχνικές οδηγούν σε αναξιόπιστα αποτελέσματα και πρέπει να χρησιμοποιούνται εξειδικευμένα υποδείγματα που ενσωματώνουν χωρικές επιδράσεις (Anselin, 1988).

Ο Fingleton (1999) εισήγαγε τις έννοιες των χωρικών μοναδιαίων ριζών (*spatial unit roots*), της χωρικής συν-ολοκλήρωσης (*spatial co-integration*) και της χωρικής υποφαινομενικής παλινδρόμησης (*spurious spatial regression*). Συγκεκριμένα, έδειξε ότι η παλινδρόμηση μεταξύ δύο ανεξάρτητων αυτοσυσχετιζόμενων ή με ύπαρξη μοναδιαίας ρίζας χωρικών διαδικασιών οδηγεί σε υποφαινομενικά αποτελέσματα (*spurious results*) παρόμοια με αυτά που εμφανίζονται με χρονοσειρές. Επιπρόσθετα, διαπίστωσε ότι τα σφάλματα από αυτή την παλινδρόμηση παρουσιάζουν υψηλή χωρική αυτοσυσχέτιση όπως δηλώνεται από τις πολύ υψηλές τιμές της στατιστικής I του Moran κατά κάποιο τρόπο όπως εμφανίζεται το πρόβλημα στις χρονοσειρές με τις πολύ χαμηλές τιμές της στατιστικής των Durbin και Watson. Οι Mur και Trivez (2003) επέκτειναν την εργασία του Fingleton (1999) εισάγοντας την έννοια της χωρικής αιτιοκρατικής τάσης (*spatial deterministic trend*) με την προσθήκη αιτιοκρατικών σταθερών όρων (*deterministic intercepts*) στις χωρικές διαδικασίες και βρήκαν υποφαινομενικά αποτελέσματα. Επίσης, διαπίστωσαν ότι οι έλεγχοι χωρικής εξάρτησης πολλαπλασιαστή Lagrange οδηγούν σε παραπλανητικά αποτελέσματα.

2.6.2 Υποφαινομενική συμπεριφορά με στάσιμες χωρικές αυτοπαλίνδρομες διαδικασίες πρώτης τάξης

Η διερεύνηση της ύπαρξης χωρικής γραμμικής σχέσης μεταξύ δύο μεταβλητών απαιτεί την εκτίμηση ενός υποδείγματος της μορφής της σχέσης (2.1) με μία ανεξάρτητη μεταβλητή το οποίο για λόγους διευκόλυνσης της παρουσίασης ξανά ορίζεται, δηλαδή:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2.14)$$

όπου \mathbf{y} είναι το $(n \times 1)$ διάνυσμα των τιμών της εξαρτημένης μεταβλητής, \mathbf{X} είναι η $(n \times 2)$ μήτρα της ανεξάρτητης μεταβλητής περιλαμβάνοντας μία στήλη με μονάδες για την εκτίμηση του σταθερού όρου, $\boldsymbol{\beta}$ είναι το (2×1) διάνυσμα των δύο παραμέτρων β_0 και β_1 για τον σταθερό όρο και την ανεξάρτητη μεταβλητή, αντίστοιχα, $\boldsymbol{\varepsilon}$ είναι το $(n \times 1)$ διάνυσμα των τυχαίων σφαλμάτων για το οποία ισχύει ότι $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$ όπου \mathbf{I} είναι ο μοναδιαίος πίνακας και σ^2 η διακύμανση. Εάν υπάρχει γραμμική σχέση μεταξύ των δύο μεταβλητών που διερευνώνται τότε η εκτίμηση του υποδείγματος της σχέσης (2.14) θα πρέπει να οδηγήσει σε σημαντικά στατιστικά αποτελέσματα και ο έλεγχος της μηδενικής υπόθεσης ότι ο συντελεστής β_1 ισούται με το μηδέν θα πρέπει να απορριφθεί.

Για να εξεταστεί με την εφαρμογή προσομοίωσης η εμφάνιση του φαινομένου της υποφαινομενικής παλινδρόμησης με χωρικές μεταβλητές, το πρόβλημα της χωρικής αυτοσυσχέτισης στα εκτιμηθέντα σφάλματα και η συμπεριφορά των ελέγχων χωρικής εξάρτησης θεωρούνται δύο ανεξάρτητες στάσιμες χωρικές αυτοπαλίνδρομες διαδικασίες πρώτης τάξης, SAR(1), \mathbf{y} και \mathbf{x} οι οποίες παράγονται αντίστοιχα ως $\mathbf{y} = (\mathbf{I} - \phi \mathbf{W})^{-1} \boldsymbol{\varepsilon}_y$ και $\mathbf{x} = (\mathbf{I} - \phi \mathbf{W})^{-1} \boldsymbol{\varepsilon}_x$, όπου τα σφάλματα $\boldsymbol{\varepsilon}_y$ και $\boldsymbol{\varepsilon}_x$ είναι χωρικές διαδικασίες λευκού θορύβου $N(0, \mathbf{I})$ και ο αυτοπαλίνδρομος συντελεστής ϕ λαμβάνει τις τιμές 0,0, 0,2, 0,5, 0,8 και 0,9. Η $(n \times n)$ μήτρα χωρικών σταθμίσεων \mathbf{W} ορίστηκε όπως και στις προηγούμενες προσομοιώσεις στο κεφάλαιο αυτό, δηλαδή σύμφωνα με τα κριτήρια του πύργου (*rook contiguity*) και τη βασίλισσας (*queen contiguity*), χωρίς διόρθωση τόρος και με διόρθωση τόρος, επάνω σε ένα ομαλό τετραγωνικό πλέγμα διαστάσεων 7x7, 10x10, 20x20 και 30x30 δίνοντας δείγματα 49, 100, 400 και 900 παρατηρήσεων. Εκτελέστηκαν 10.000 επαναλήψεις.

Ο Πίνακας 2.15 παρουσιάζει τα αποτελέσματα της προσομοίωσης από την εκτίμηση του απλού οικονομετρικού υποδείγματος της σχέσης (2.14) για δύο ανεξάρτητες στάσιμες SAR(1) διαδικασίες που παρείχθησαν σύμφωνα με το μηχανισμό που παρουσιάστηκε στην προηγούμενη παράγραφο. Το πιο ενδιαφέρον αποτέλεσμα που διαπιστώνεται από τον πίνακα αυτό είναι ότι η υποφαινομενική συμπεριφορά έχει πολλές ομοιότητες με την υποφαινομενική συμπεριφορά που εμφανίζεται στην ανάλυση χρονοσειρών για δύο στάσιμες AR(1) διαδικασίες όπως έχει υποδειχθεί στις εργασίες των Granger, Hyung, και Jeon (2001), Agiakloglou (2009) και Agiakloglou και Tsimpanos (2012) με τη μοναδική διαφορά το ελαφρά μικρότερο ποσοστό απόρριψης της μηδενικής υπόθεσης $\beta_1 = 0$ για κάθε δεδομένη

Πίνακας 2.15

Ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% της μηδενικής υπόθεσης $\beta_1 = 0$ για το απλό οικονομετρικό υπόδειγμα (2.14) για δύο ανεξάρτητες στάσιμες SAR(1) διαδικασίες με μέσες τιμές στατιστικής I και ποσοστά απόρριψης των ελέγχων χωρικής εξάρτησης για όλες τις τιμές του φ και μεγέθη δείγματος Χωρίς διόρθωση τόρος

| N | Σταθμίσεις | Πύργος | | | | | Βασίλισσα | | | | |
|----------------------|------------|--------|-------|-------|-------|-------|-----------|-------|-------|-------|-------|
| | φ | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 |
| 49 | Μέσο I | -0,021 | 0,085 | 0,258 | 0,489 | 0,596 | -0,020 | 0,040 | 0,166 | 0,371 | 0,474 |
| Ποσοστά Απόρριψης | β_1 | 5,48 | 6,03 | 10,04 | 27,25 | 42,10 | 5,48 | 5,79 | 9,13 | 27,63 | 42,10 |
| | I | 4,98 | 17,84 | 74,01 | 99,06 | 99,92 | 5,19 | 16,71 | 61,97 | 95,51 | 98,71 |
| | LM-LAG | 4,57 | 11,89 | 64,85 | 98,46 | 99,83 | 3,81 | 10,40 | 51,19 | 92,86 | 97,96 |
| | LM-LE | 5,19 | 5,94 | 8,92 | 17,20 | 22,72 | 5,61 | 7,84 | 13,68 | 25,58 | 30,96 |
| | LM-ERR | 4,60 | 11,67 | 63,23 | 98,00 | 99,68 | 3,87 | 10,02 | 49,02 | 91,35 | 97,06 |
| | LM-EL | 5,26 | 5,54 | 5,99 | 4,72 | 3,18 | 5,56 | 7,39 | 10,48 | 7,85 | 4,77 |
| | LM-SARMA | 4,75 | 10,02 | 56,15 | 97,11 | 99,49 | 4,55 | 10,92 | 47,60 | 90,71 | 97,09 |
| 100 | Μέσο I | -0,010 | 0,097 | 0,277 | 0,534 | 0,662 | -0,011 | 0,052 | 0,186 | 0,429 | 0,566 |
| Ποσοστά Απόρριψης | β_1 | 5,28 | 5,70 | 10,29 | 29,71 | 46,91 | 5,28 | 5,58 | 9,53 | 31,00 | 49,01 |
| | I | 4,87 | 30,85 | 96,39 | 100 | 100 | 5,19 | 24,50 | 87,25 | 99,94 | 100 |
| | LM-LAG | 4,77 | 25,13 | 94,81 | 100 | 100 | 4,35 | 18,50 | 83,12 | 99,92 | 100 |
| | LM-LE | 5,34 | 5,79 | 9,07 | 19,11 | 25,89 | 4,88 | 7,29 | 14,53 | 28,18 | 36,00 |
| | LM-ERR | 4,70 | 24,79 | 94,48 | 100 | 100 | 4,44 | 18,21 | 82,40 | 99,89 | 100 |
| | LM-EL | 5,45 | 5,62 | 5,62 | 3,98 | 2,21 | 4,90 | 6,89 | 10,20 | 6,15 | 3,89 |
| | LM-SARMA | 4,96 | 19,76 | 91,16 | 100 | 100 | 4,54 | 17,08 | 78,48 | 99,82 | 100 |
| 400 | Μέσο I | -0,002 | 0,104 | 0,287 | 0,560 | 0,708 | -0,002 | 0,059 | 0,195 | 0,465 | 0,634 |
| Ποσοστά Απόρριψης | β_1 | 4,93 | 5,32 | 9,72 | 29,85 | 48,33 | 4,93 | 5,20 | 8,72 | 31,74 | 51,96 |
| | I | 4,80 | 83,61 | 100 | 100 | 100 | 4,95 | 62,42 | 99,99 | 100 | 100 |
| | LM-LAG | 4,94 | 81,49 | 100 | 100 | 100 | 4,88 | 58,69 | 99,98 | 100 | 100 |
| | LM-LE | 4,94 | 5,67 | 9,44 | 20,45 | 28,49 | 5,16 | 7,68 | 14,85 | 29,99 | 38,28 |
| | LM-ERR | 4,92 | 81,35 | 100 | 100 | 100 | 4,81 | 58,47 | 99,98 | 100 | 100 |
| | LM-EL | 4,88 | 5,24 | 5,80 | 3,58 | 1,38 | 5,12 | 7,17 | 10,38 | 5,01 | 9,89 |
| | LM-SARMA | 4,86 | 73,10 | 100 | 100 | 100 | 4,99 | 51,16 | 99,95 | 100 | 100 |
| 900 | Μέσο I | -0,001 | 0,104 | 0,285 | 0,560 | 0,713 | -0,001 | 0,059 | 0,194 | 0,465 | 0,642 |
| Ποσοστά Απόρριψης | β_1 | 5,17 | 5,25 | 9,04 | 29,25 | 48,03 | 5,17 | 5,22 | 8,50 | 31,14 | 52,54 |
| | I | 4,86 | 99,26 | 100 | 100 | 100 | 5,04 | 90,85 | 100 | 100 | 100 |
| | LM-LAG | 4,82 | 99,12 | 100 | 100 | 100 | 5,05 | 89,38 | 100 | 100 | 100 |
| | LM-LE | 5,15 | 5,69 | 9,06 | 20,23 | 28,75 | 5,26 | 7,55 | 14,30 | 30,14 | 39,27 |
| | LM-ERR | 4,81 | 99,12 | 100 | 100 | 100 | 5,03 | 89,39 | 100 | 100 | 100 |
| | LM-EL | 5,18 | 5,38 | 5,73 | 3,63 | 1,11 | 5,31 | 7,20 | 10,36 | 4,96 | 11,80 |
| | LM-SARMA | 5,04 | 97,70 | 100 | 100 | 100 | 5,08 | 84,20 | 100 | 100 | 100 |

τιμή της αυτοπαλίνδρομης παραμέτρου. Επιπρόσθετα, όπως στην περίπτωση που χρησιμοποιούνται χρονοσειρές, το ποσοστό απόρριψης δεν φαίνεται να επηρεάζεται ιδιαίτερα από το μέγεθος του δείγματος, κάτι που έχει υποδειχθεί και από τον Fingleton (1999). Παραδείγματος χάριν, όπως παρουσιάζεται στον Πίνακα 2.15, το ποσοστό απόρριψης σε δείγμα 100 παρατηρήσεων και τιμές των αυτοπαλίνδρομων παραμέτρων στις SAR(1) διαδικασίες ίσες με 0,9 είναι 46,91% με το κριτήριο του πύργου και 49,01% με το κριτήριο της βασίλισσας, ενώ σύμφωνα με τους Agiakloglou και Tsimpanos (2012) το αντίστοιχο ποσοστό απόρριψης του ελέγχου σημαντικότητας του συντελεστή γραμμικής συσχέτισης για χρονοσειρές 100 παρατηρήσεων και ίδια τιμή της αυτοπαλίνδρομης παραμέτρου σε δύο AR(1) διαδικασίες είναι 50,27%. Κατ' αναλογία, για δείγμα 900 παρατηρήσεων και τις ίδιες τιμές των αυτοπαλίνδρομων παραμέτρων στις SAR(1) διαδικασίες τα ποσοστά απόρριψης είναι με το κριτήριο του πύργου 48,03% και με το κριτήριο της βασίλισσας 52,54% ενώ στις χρονοσειρές σε δείγμα 1000 παρατηρήσεων το ποσοστό απόρριψης είναι 52,01%.

Ένα άλλο αξιοσημείωτο αποτέλεσμα που περιγράφεται στον Πίνακα 2.15 είναι ότι η παλινδρόμηση των δύο ανεξάρτητων SAR(1) διαδικασιών οδηγεί στην εμφάνιση του προβλήματος της χωρικής αυτοσυσχέτισης στα σφάλματα του υποδείγματος κάτι που έχει εντοπιστεί και από τον Agiakloglou (2013) στην υποφαινομενική παλινδρόμηση με χρονοσειρές. Πραγματικά, οι μέσες τιμές της στατιστικής I του Moran για τα εκτιμηθέντα σφάλματα αυξάνουν καθώς αυξάνεται η τιμή της χωρικής αυτοπαλίνδρομης παραμέτρου, για δεδομένο αριθμό χωρικών μονάδων, παρέχοντας ισχυρές ενδείξεις ύπαρξης χωρικής αυτοσυσχέτισης. Όμως, αξίζει να αναφερθεί ότι μολονότι οι μέσες τιμές της στατιστικής I του Moran δεν φαίνονται να επηρεάζονται σημαντικά από το μέγεθος του δείγματος η εφαρμογή του ανάλογου ελέγχου για την ύπαρξη χωρικής αυτοσυσχέτισης, όπως έχει προταθεί από τους Cliff και Ord (1972), δείχνει ισχυρότερες ενδείξεις ύπαρξης χωρικής αυτοσυσχέτισης αφού τα ποσοστά απόρριψής του αυξάνονται με την αύξηση του δείγματος, ακόμη και για μικρές τιμές της αυτοπαλίνδρομης παραμέτρου.

Τα αποτελέσματα της προσομοίωσης που προέκυψαν με διόρθωσης τόρος στις μήτρες χωρικών σταθμίσεων συνοψίζονται στον Πίνακα 2.16. Από τον πίνακα αυτόν διαπιστώνεται ότι δεν υπάρχουν σημαντικές διαφορές σε σύγκριση με την απουσία τόρος. Μικρή εξαίρεση αποτελεί ότι στα μικρά δείγματα και για δεδομένες τιμές της αυτοπαλίνδρομης παραμέτρου τα ποσοστά απόρριψης του ελέγχου στατιστικής σημαντικότητας του συντελεστή β_1 είναι λίγο πιο μικρά.

Πίνακας 2.16

**Ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% της μηδενικής υπόθεσης $\beta_1 = 0$ για το απλό οικονομετρικό υπόδειγμα (2.14) για δύο ανεξάρτητες στάσιμες SAR(1) διαδικασίες με μέσες τιμές στατιστικής I και ποσοστά απόρριψης των ελέγχων χωρικής εξάρτησης για όλες τις τιμές του φ και μεγέθη δείγματος
Με διόρθωση τόρος**

| N | Σταθμίσεις | Πύργος | | | | | Βασίλισσα | | | | |
|----------------------|------------|--------|-------|-------|-------|-------|-----------|-------|-------|-------|-------|
| | φ | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 |
| 49 | Μέσο I | -0,021 | 0,065 | 0,204 | 0,377 | 0,451 | -0,021 | 0,022 | 0,105 | 0,227 | 0,281 |
| Ποσοστά Απόρριψης | β_1 | 5,48 | 5,93 | 8,96 | 19,91 | 27,59 | 5,48 | 5,67 | 7,78 | 15,63 | 21,22 |
| | I | 4,68 | 15,27 | 64,46 | 97,06 | 99,21 | 5,27 | 13,9 | 48,92 | 86,75 | 93,71 |
| | LM-LAG | 4,43 | 9,07 | 51,84 | 93,66 | 98,10 | 3,98 | 7,23 | 35,63 | 79,28 | 89,23 |
| | LM-LE | 5,44 | 5,64 | 5,87 | 5,32 | 4,19 | 5,29 | 6,96 | 11,45 | 17,65 | 20,05 |
| | LM-ERR | 4,44 | 9,43 | 53,68 | 94,73 | 98,51 | 3,90 | 7,10 | 34,17 | 77,27 | 87,62 |
| | LM-EL | 5,40 | 5,94 | 8,17 | 13,24 | 15,46 | 5,38 | 6,94 | 9,49 | 9,28 | 7,80 |
| | LM-SARMA | 4,73 | 8,22 | 44,82 | 91,37 | 97,26 | 4,84 | 8,13 | 32,63 | 75,65 | 86,59 |
| 100 | Μέσο I | -0,010 | 0,084 | 0,240 | 0,457 | 0,562 | -0,010 | 0,039 | 0,143 | 0,322 | 0,417 |
| Ποσοστά Απόρριψης | β_1 | 5,28 | 5,71 | 9,42 | 25,01 | 37,70 | 5,28 | 5,43 | 8,06 | 22,62 | 34,68 |
| | I | 4,94 | 27,75 | 94,01 | 100 | 100 | 4,89 | 20,88 | 79,53 | 99,50 | 99,91 |
| | LM-LAG | 4,66 | 21,48 | 91,23 | 100 | 100 | 4,20 | 15,02 | 73,01 | 99,21 | 99,90 |
| | LM-LE | 5,75 | 5,45 | 5,94 | 4,61 | 2,97 | 5,02 | 7,09 | 13,17 | 24,13 | 29,11 |
| | LM-ERR | 4,65 | 21,77 | 91,92 | 100 | 100 | 4,18 | 14,77 | 72,19 | 99,10 | 99,86 |
| | LM-EL | 5,64 | 5,69 | 8,82 | 17,35 | 22,06 | 4,89 | 6,76 | 10,18 | 8,13 | 4,92 |
| | LM-SARMA | 4,86 | 17,40 | 86,36 | 99,98 | 100 | 4,77 | 14,41 | 68,07 | 98,78 | 99,80 |
| 400 | Μέσο I | -0,002 | 0,098 | 0,268 | 0,524 | 0,663 | -0,002 | 0,053 | 0,174 | 0,411 | 0,562 |
| Ποσοστά Απόρριψης | β_1 | 4,93 | 5,29 | 9,34 | 28,13 | 45,29 | 4,93 | 5,25 | 8,51 | 28,93 | 47,46 |
| | I | 4,95 | 81,07 | 100 | 100 | 100 | 5,19 | 58,72 | 99,96 | 100 | 100 |
| | LM-LAG | 4,90 | 78,41 | 100 | 100 | 100 | 4,86 | 54,27 | 99,95 | 100 | 100 |
| | LM-LE | 4,61 | 5,17 | 6,12 | 4,16 | 2,04 | 4,97 | 7,34 | 14,72 | 30,27 | 38,84 |
| | LM-ERR | 4,88 | 78,69 | 100 | 100 | 100 | 4,79 | 54,15 | 99,95 | 100 | 100 |
| | LM-EL | 4,57 | 5,50 | 9,16 | 20,76 | 29,41 | 4,97 | 7,12 | 10,71 | 6,64 | 7,26 |
| | LM-SARMA | 4,79 | 69,34 | 100 | 100 | 100 | 4,76 | 46,82 | 99,87 | 100 | 100 |
| 900 | Μέσο I | -0,001 | 0,100 | 0,273 | 0,536 | 0,684 | -0,001 | 0,055 | 0,179 | 0,429 | 0,595 |
| Ποσοστά Απόρριψης | β_1 | 5,17 | 5,33 | 8,97 | 28,03 | 46,54 | 5,17 | 5,25 | 8,33 | 29,00 | 49,91 |
| | I | 4,78 | 99,12 | 100 | 100 | 100 | 4,77 | 88,60 | 100 | 100 | 100 |
| | LM-LAG | 4,73 | 98,90 | 100 | 100 | 100 | 4,81 | 87,29 | 100 | 100 | 100 |
| | LM-LE | 5,06 | 5,08 | 5,99 | 4,13 | 1,69 | 5,10 | 7,45 | 14,08 | 29,65 | 39,95 |
| | LM-ERR | 4,69 | 98,92 | 100 | 100 | 100 | 4,78 | 87,26 | 100 | 100 | 100 |
| | LM-EL | 5,13 | 5,58 | 9,15 | 20,55 | 29 | 5,09 | 7,10 | 10,57 | 6,10 | 9,84 |
| | LM-SARMA | 4,97 | 97,50 | 100 | 100 | 100 | 5,23 | 81,16 | 100 | 100 | 100 |

Όταν ο έλεγχος με τη στατιστικής I του Moran απορρίπτει τη μηδενική υπόθεση της απουσίας χωρικής αυτοσυσχέτισης, ο αναλυτής θα πρέπει να χρησιμοποιήσει τους ελέγχους χωρικής εξάρτησης για να εντοπίσει την αιτία του προβλήματος και να προχωρήσει στην εκτίμηση του κατάλληλου χωρικού οικονομετρικού υποδείγματος, δηλαδή του υποδείγματος χωρικής υστέρησης (*Spatial Lag Model – SLM*) που ορίζεται ως:

$$\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2.15)$$

ή του υποδείγματος χωρικού σφάλματος (*Spatial Error Model – SEM*) που ορίζεται ως:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \text{ με } \boldsymbol{\varepsilon} = \lambda \mathbf{W}\boldsymbol{\varepsilon} + \mathbf{u} \quad (2.16)$$

όπου \mathbf{u} είναι ένα $(n \times 1)$ διάνυσμα σφαλμάτων και οι μήτρες $\mathbf{W}\mathbf{y}$ και $\mathbf{W}\boldsymbol{\varepsilon}$ είναι οι μήτρες χωρικών υστερήσεων που περιλαμβάνονται στα υποδείγματα και ουσιαστικά αποτελούνται από τους σταθμισμένους μέσους όρους των τιμών των μεταβλητών στις γειτονικές περιοχές. Τα υποδείγματα αυτά είναι παρόμοια με αυτά που εφαρμόζονται στην ανάλυση χρονοσειρών με τη διαφορά ότι εδώ η εξάρτηση είναι προς πολλές κατευθύνσεις.

Ο έλεγχος με τη στατιστική LM-LAG καθορίζει εάν πρέπει στο υπόδειγμα να περιληφθεί η χωρική υστέρηση της εξαρτημένης μεταβλητής ενώ ο έλεγχος με τη στατιστική LM-ERR καθορίζει εάν πρέπει να εκτιμηθεί το υπόδειγμα χωρικού σφάλματος. Επιπρόσθετα, υπάρχουν οι ανθεκτικές (*robust*) μορφές των στατιστικών, δηλαδή η στατιστική LM-LE που ελέγχει για την ύπαρξη της χωρικής υστέρησης της εξαρτημένης μεταβλητής στο υπόδειγμα λαμβάνοντας υπ' όψιν την παρουσία τοπικής λανθασμένης εξειδίκευσης (*local misspecification*) χωρικής αυτοσυσχέτισης στα σφάλματα και η στατιστική LM-EL που ελέγχει την ύπαρξη χωρικής αυτοσυσχέτισης στα σφάλματα λαμβάνοντας υπ' όψιν την παρουσία τοπικής λανθασμένης εξειδίκευσης (*local misspecification*) χωρικής υστέρησης στην εξαρτημένη μεταβλητή καθώς και η στατιστική LM-SARMA που ελέγχει και για τις δύο αιτίες χωρικής εξάρτησης. Υπό τη μηδενική υπόθεση της απουσίας χωρικής εξάρτησης οι τέσσερις πρώτες στατιστικές ακολουθούν την κατανομή χ^2 με ένα βαθμό ελευθερίας και η πέμπτη στατιστική, δηλαδή, η LM-SARMA την κατανομή χ^2 με δύο βαθμούς ελευθερίας.

Ο Πίνακας 2.15 περιλαμβάνει επίσης τα ποσοστά απόρριψης της μηδενικής υπόθεσης εξειδίκευσης του απλού οικονομετρικού υποδείγματος (2.14) από την εφαρμογή των πέντε ελέγχων χωρικής εξάρτησης για επίπεδο σημαντικότητας 5%. Όπως διαπιστώνεται, ο έλεγχος με τη στατιστική LM-LAG απορρίπτει πολύ συχνά τη μηδενική υπόθεση προτείνοντας ότι πρέπει να εκτιμηθεί το υπόδειγμα χωρικής υστέρησης που ορίζεται στη σχέση (2.15). Επιπλέον, τα ποσοστά απόρριψης αυξάνονται καθώς αυξάνεται η τιμή της χωρικής

αυτοπαλίνδρομης παραμέτρου ϕ και το μέγεθος του δείγματος. Κατά συνέπεια, το επόμενο στάδιο είναι η εκτίμηση του υποδείγματος χωρικής υστέρησης της σχέσης (2.15). Ωστόσο, η εκτίμηση με τη μέθοδο της μέγιστης πιθανοφάνειας του υποδείγματος χωρικής υστέρησης αποκαλύπτει πολύ ενδιαφέροντα αποτελέσματα παρόμοια με αυτά που έχουν εντοπιστεί από τον Agiakloglou (2013) στην ανάλυση χρονοσειρών. Πιο συγκεκριμένα, το φαινόμενο της υποφαινομενικής συμπεριφοράς φαίνεται ότι αντιμετωπίζεται, καθώς τα εμπειρικά επίπεδα σημαντικότητας για τον έλεγχο της μηδενικής υπόθεσης ότι ο συντελεστής β_1 στο υπόδειγμα χωρικής υστέρησης ισούται με το μηδέν είναι όλα κοντά στα ονοματικά επίπεδα σημαντικότητας του 5%, όπως δείχνεται στον Πίνακα 2.17 χωρίς τόρος και στον Πίνακα 2.18 με διόρθωση τόρος, για όλες τις τιμές της αυτοπαλίνδρομης παραμέτρου ϕ και όλα τα μεγέθη δείγματος. Επιπρόσθετα, το πρόβλημα της χωρικής αυτοσυσχέτισης στα εκτιμηθέντα σφάλματα δεν εμφανίζεται σε αυτή την περίπτωση καθώς τα εμπειρικά επίπεδα απόρριψης της χωρικής στατιστικής h του Durbin, LM_{lp} , των Anselin, Bera, Florax, και Yoon (1996) για τον έλεγχο της απουσίας χωρικής αυτοσυσχέτισης στο υπόδειγμα χωρικής υστέρησης λαμβάνοντας υπ' όψιν την παρουσία της χωρικής υστέρησης της εξαρτημένης μεταβλητής είναι όλα κοντά στα επίπεδα σημαντικότητας του 5%, όπως παρατηρείται στους ίδιους πίνακες.

Παρόμοια συμπεριφορά με τη στατιστική LM-LAG εντοπίζεται και για τη στατιστική LM-ERR, όπως φαίνεται στους Πίνακες 2.15 και 2.16. Δηλαδή, ο έλεγχος με τη στατιστική LM-ERR υποδεικνύει την εκτίμηση του υποδείγματος χωρικού σφάλματος (2.16) για το οποίο η υποφαινομενική συμπεριφορά εξαφανίζεται αφού, όπως περιγράφεται στους Πίνακες 2.17 και 2.18, τα ποσοστά απόρριψης του ελέγχου της υπόθεσης ότι ο συντελεστής β_1 στο υπόδειγμα χωρικού σφάλματος είναι μηδέν είναι όλα κοντά στο 5%. Αξίζει να αναφερθεί ότι για όλα τα χωρικά υποδείγματα έχουν υπολογιστεί και οι τιμές της στατιστικής I του Moran για την οποία οι μέσες τιμές δεν παρουσιάζονται καθώς σε όλες τις περιπτώσεις είναι κοντά στο μηδέν.

Ένα αλλοπρόσαλλο αποτέλεσμα που πρέπει να αναφερθεί είναι ότι οι ανθεκτικοί έλεγχοι χωρικής εξάρτησης δεν συμπεριφέρονται όπως οι απλοί έλεγχοι κάτι που δεν μπορεί εύκολα να εξηγηθεί, αν και στη βιβλιογραφία υπάρχουν δημοσιευμένες εργασίες, όπως των Mur και Trivez (2003) και των Acosta, de Matos και Fava (2006) που δείχνουν ότι οι συγκεκριμένοι έλεγχοι έχουν ιδιότροπη συμπεριφορά σε ειδικές περιπτώσεις και δεν οδηγούν στην επιλογή του σωστού υποδείγματος. Υπενθυμίζεται ότι και στις προσομοιώσεις της ενότητας 2.4 είχε

Πίνακας 2.17

Ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% της μηδενικής υπόθεσης $\beta_1 = 0$ για τα χωρικά υποδείγματα (2.15) και (2.16) για δύο ανεξάρτητες στάσιμες SAR(1) διαδικασίες και ποσοστά απόρριψης στατιστικής $LM_{\lambda\rho}$ για τον έλεγχο της απουσίας χωρικής αυτοσυσχέτισης στο υπόδειγμα (2.15)

Χωρίς διόρθωση τόρος

| φ | Πύργος | | | | | Βασίλισσα | | | | |
|---|--|------|------|------|------|-----------|------|------|------|-------|
| | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 |
| Υπόδειγμα Χωρικής Υστέρησης (2.15) – SLM | | | | | | | | | | |
| N | Ποσοστά απόρριψης συντελεστή β_1 | | | | | | | | | |
| 49 | 6,28 | 6,46 | 6,85 | 8,56 | 10,6 | 6,27 | 6,46 | 7,12 | 9,31 | 11,97 |
| 100 | 5,66 | 5,68 | 5,94 | 6,69 | 7,86 | 5,68 | 5,70 | 6,02 | 6,95 | 8,68 |
| 400 | 5,01 | 4,84 | 4,91 | 5,61 | 5,77 | 5,00 | 4,90 | 4,95 | 5,39 | 5,90 |
| 900 | 5,16 | 4,92 | 4,80 | 5,06 | 5,12 | 5,15 | 4,87 | 4,86 | 5,15 | 5,29 |
| N | Ποσοστά απόρριψης στατιστικής $LM_{\lambda\rho}$ | | | | | | | | | |
| 49 | 5,04 | 4,84 | 4,42 | 4,52 | 5,01 | 5,19 | 5,38 | 5,15 | 6,29 | 7,33 |
| 100 | 5,01 | 4,49 | 4,53 | 4,62 | 5,10 | 5,29 | 4,82 | 4,83 | 5,73 | 6,42 |
| 400 | 5,20 | 4,78 | 4,61 | 4,88 | 5,05 | 5,11 | 4,78 | 4,41 | 5,31 | 5,72 |
| 900 | 5,15 | 5,14 | 4,84 | 4,66 | 4,79 | 5,42 | 5,16 | 5,06 | 5,17 | 5,28 |
| Υπόδειγμα Χωρικού Σφάλματος (2.16) – SEM | | | | | | | | | | |
| N | Ποσοστά απόρριψης συντελεστή β_1 | | | | | | | | | |
| 49 | 7,17 | 7,13 | 7,31 | 7,81 | 8,16 | 7,47 | 7,49 | 7,72 | 8,11 | 8,43 |
| 100 | 6,02 | 6,07 | 6,09 | 6,08 | 6,14 | 5,99 | 6,07 | 6,07 | 6,09 | 6,29 |
| 400 | 5,05 | 4,99 | 5,04 | 5,00 | 4,99 | 5,04 | 5,02 | 4,97 | 5,02 | 5,05 |
| 900 | 5,13 | 5,15 | 5,19 | 5,20 | 5,22 | 5,18 | 5,19 | 5,23 | 5,21 | 5,21 |

Πίνακας 2.18

Ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% της μηδενικής υπόθεσης $\beta_1 = 0$ για τα χωρικά υποδείγματα (2.15) και (2.16) για δύο ανεξάρτητες στάσιμες SAR(1) διαδικασίες και ποσοστά απόρριψης στατιστικής $LM_{\lambda\rho}$ για τον έλεγχο της απουσίας χωρικής αυτοσυσχέτισης στο υπόδειγμα (2.15)

Με διόρθωση τόρος

| φ | Πύργος | | | | | Βασίλισσα | | | | |
|---|--|------|------|------|------|-----------|------|------|------|------|
| | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 |
| Υπόδειγμα Χωρικής Υστέρησης (2.15) – SLM | | | | | | | | | | |
| N | Ποσοστά απόρριψης συντελεστή β_1 | | | | | | | | | |
| 49 | 6,26 | 6,39 | 6,79 | 7,96 | 9,35 | 6,12 | 6,28 | 7,02 | 8,23 | 9,36 |
| 100 | 5,66 | 5,79 | 6,01 | 6,83 | 7,97 | 5,67 | 5,74 | 5,93 | 6,87 | 8,15 |
| 400 | 5,00 | 4,92 | 4,88 | 5,37 | 5,88 | 5,00 | 4,93 | 4,91 | 5,55 | 6,00 |
| 900 | 5,17 | 4,92 | 4,79 | 4,90 | 5,17 | 5,17 | 4,96 | 4,85 | 5,15 | 5,36 |
| N | Ποσοστά απόρριψης στατιστικής $LM_{\lambda\rho}$ | | | | | | | | | |
| 49 | 4,83 | 4,81 | 4,08 | 4,09 | 4,84 | 5,20 | 5,08 | 4,74 | 4,78 | 6,30 |
| 100 | 5,08 | 4,66 | 4,55 | 4,72 | 5,07 | 4,90 | 4,96 | 4,88 | 5,41 | 6,25 |
| 400 | 5,02 | 4,88 | 4,79 | 4,57 | 4,87 | 5,05 | 4,82 | 4,80 | 5,42 | 5,90 |
| 900 | 5,21 | 5,17 | 4,98 | 4,73 | 4,66 | 5,19 | 5,05 | 5,05 | 4,99 | 5,36 |
| Υπόδειγμα Χωρικού Σφάλματος (2.16) – SEM | | | | | | | | | | |
| N | Ποσοστά απόρριψης συντελεστή β_1 | | | | | | | | | |
| 49 | 7,11 | 7,03 | 7,15 | 7,38 | 7,60 | 7,05 | 7,16 | 7,34 | 7,62 | 7,90 |
| 100 | 6,14 | 6,22 | 6,23 | 6,27 | 6,34 | 6,15 | 6,16 | 6,19 | 6,30 | 6,22 |
| 400 | 5,07 | 5,04 | 5,04 | 5,07 | 5,04 | 4,99 | 5,00 | 5,05 | 5,01 | 5,06 |
| 900 | 5,12 | 5,12 | 5,17 | 5,16 | 5,17 | 5,14 | 5,12 | 5,16 | 5,19 | 5,13 |

παρατηρηθεί παράξενη συμπεριφορά για τη στατιστική LM-EL με το κριτήριο του πύργου. Ειδικότερα, όπως φαίνεται στους Πίνακες 2.15 και 2.16, η ανθεκτική στατιστική LM-LE έχει παρόμοια συμπεριφορά με τη στατιστική LM-LAG αλλά με μικρότερα ποσοστά απόρριψης. Αντίθετα, η ανθεκτική στατιστική LM-EL έχει ακριβώς την αντίθετη συμπεριφορά με τη στατιστική LM-ERR στο ότι δεν επιλέγει ποτέ το υπόδειγμα χωρικού σφάλματος. Τέλος, η στατιστική LM-SARMA, όπως οι δύο μη ανθεκτικές στατιστικές, υποδεικνύει πάντοτε την εκτίμηση ενός χωρικού οικονομετρικού υποδείγματος.

Συνοψίζοντας τα αποτελέσματα αυτής της ενότητας, η εμφάνιση υποφαινομενικών στατιστικών σχέσεων μεταξύ ανεξάρτητων μεταβλητών στα πλαίσια ενός υποδείγματος γραμμικής παλινδρόμησης, μπορεί να προκύψει εξίσου και με χωρικά δεδομένα όπως συμβαίνει στην ανάλυση χρονοσειρών. Η συμπεριφορά του φαινομένου αυτού παρουσιάστηκε για δύο ανεξάρτητες στάσιμες χωρικές αυτοπαλίνδρομες διαδικασίες πρώτης τάξης, SAR(1). Τα ποσοστά απόρριψης του ελέγχου της στατιστικής σημαντικότητας του συντελεστή της ανεξάρτητης μεταβλητής που προέκυψαν είναι πολύ κοντά στα αντίστοιχα των χρονοσειρών για δύο στάσιμες αυτοπαλίνδρομες διαδικασίες πρώτης τάξης, AR(1). Ταυτόχρονα, η στατιστική I του Moran για τα εκτιμηθέντα σφάλματα του υποδείγματος έχει υψηλές τιμές εκφράζοντας την ύπαρξη χωρικής αυτοσυσχέτισης κατ' ανάλογο τρόπο με τις χαμηλές τιμές της στατιστικής των Durbin και Watson στις χρονοσειρές.

Ο εντοπισμός του προβλήματος της χωρικής αυτοσυσχέτισης στα σφάλματα υποχρεώνει τον ερευνητή να εξετάσει τη στατιστική εγκυρότητα του εκτιμηθέντος υποδείγματος και με τη βοήθεια των ελέγχων χωρικής εξάρτησης να εντοπίσει την αιτία του προβλήματος. Όμως, το πιο σημαντικό αποτέλεσμα της διεξαχθείσας προσομοίωσης είναι ότι με την εφαρμογή των μη ανθεκτικών ελέγχων χωρικής εξάρτησης η υποφαινομενική σχέση δεν αποκαλύπτεται. Οι έλεγχοι αυτοί υποδεικνύουν ότι η εξειδίκευση ενός απλού οικονομετρικού υποδείγματος δεν είναι η σωστή και ότι πρέπει να εκτιμηθεί είτε το υπόδειγμα χωρικής υστέρησης είτε το υπόδειγμα χωρικού σφάλματος. Εκτιμώντας αυτά τα υποδείγματα δεν εμφανίζεται υποφαινομενική συμπεριφορά ούτε χωρική αυτοσυσχέτιση στα σφάλματα. Κάτι αντίστοιχο παρατηρείται και στην υποφαινομενική παλινδρόμηση με δείγματα χρονοσειρές όπου, για δύο στάσιμες αυτοπαλίνδρομες διαδικασίες πρώτης τάξης, ο Agiakloglou (2013) έχει δείξει ότι η εκτίμηση του υποδείγματος με τη χρονική υστέρηση της εξαρτημένης μεταβλητής στις ερμηνευτικές μεταβλητές διορθώνει το πρόβλημα.

Η εφαρμογή των ανθεκτικών ελέγχων υποδεικνύει την επιλογή του απλού οικονομετρικού υποδείγματος στο οποίο εμφανίζεται υποφαινομενική συμπεριφορά και χωρική αυτοσυσχέτιση στα σφάλματα. Η ανακολουθία στα αποτελέσματα μεταξύ μη ανθεκτικών και ανθεκτικών ελέγχων θα μπορούσε να αποτελέσει για τον ερευνητή ένα κριτήριο για τον εντοπισμό του προβλήματος της υποφαινομενικής παλινδρόμησης.

2.7 Ανακεφαλαίωση

Στο κεφάλαιο αυτό παρουσιάστηκαν οι επιπτώσεις στο υπόδειγμα γραμμικής παλινδρόμησης που μπορούν να προκύψουν από την αγνόηση του πρώτου τύπου των χωρικών επιδράσεων, δηλαδή της χωρικής εξάρτησης. Οι επιπτώσεις αυτές εμφανίζονται ως χωρική αυτοσυσχέτιση στα σφάλματα του εκτιμηθέντος υποδείγματος και ανάλογα με την αιτία πρόκλησής τους επηρεάζουν την αποτελεσματικότητα των εκτιμητών ελαχίστων τετραγώνων άλλα και την αμεροληψία τους όταν ο λόγος που οδηγεί στο πρόβλημα είναι η ύπαρξη χωρικής υστέρησης στην εξαρτημένη μεταβλητή που έχει παραληφθεί από το υπόδειγμα.

Η επίλυση του προβλήματος επιτυγχάνεται με την εφαρμογή των ελέγχων χωρικής εξάρτησης που βοηθούν στον εντοπισμό της αιτίας εμφάνισης της χωρικής αυτοσυσχέτισης και στην εξειδίκευση του σωστού χωρικού οικονομετρικού υποδείγματος. Η προσομοίωση που πραγματοποιήθηκε επιβεβαίωσε παλιότερα, γνωστά στη βιβλιογραφία, συμπεράσματα για την αποτελεσματικότητα αυτών των ελέγχων σε μικρά και μέτρια δείγματα και πρόσθεσε καινούριες πληροφορίες για τη συμπεριφορά τους σε μεγάλα δείγματα. Επιπρόσθετα, έδειξε ότι αρκετές φορές οι έλεγχοι μπορεί να οδηγήσουν σε ασάφειες ως προς την επιλογή του σωστού χωρικού υποδείγματος που μπορεί να οφείλονται στον ορισμό της μήτρας χωρικής γειτνίασης αλλά και στο μέγεθος του δείγματος. Σε όλες τις περιπτώσεις ο ερευνητής θα πρέπει να είναι ιδιαίτερα προσεχτικός ως προς το χωρικό υπόδειγμα που τελικά θα εκτιμήσει και να ελέγχει τη συνέπεια των ελέγχων ορίζοντας διαφορετικά κριτήρια γειτνίασης. Η σύγκριση των τιμών των στατιστικών LM-LE και LM-EL προσφέρει σημαντική βοήθεια στην επιλογή του σωστού υποδείγματος.

Στη συνέχεια, εξετάστηκαν, με τη βοήθεια προσομοίωσης, η ικανότητα της στατιστικής I του Moran και της στατιστικής LM_{lp} , να ανιχνεύουν την παρουσία χωρικής αυτοσυσχέτισης στα σφάλματα του υποδείγματος χωρικής υστέρησης διαπιστώνοντας ότι, παρόλο που η

συμπεριφορά της δεύτερης στατιστικής είναι πιο αποδεκτή σύμφωνα με τη στατιστική θεωρία, στα μεγάλα δείγματα οι δύο στατιστικές έχουν παρόμοια στατιστική ισχύ ενώ στα μικρότερα δείγματα η στατιστική I είναι ανώτερη. Ως συνέπεια, μπορούν να χρησιμοποιούνται και οι δύο στατιστικές για σύγκριση και επαλήθευση των αποτελεσμάτων.

Στην τελευταία ενότητα του κεφαλαίου παρουσιάστηκε το φαινόμενο της υποφαινομενικής συμπεριφοράς με χωρικά δεδομένα ως μια πρόσθετη αιτία που οδηγεί στην εμφάνιση χωρικής αυτοσυσχέτισης στα σφάλματα ενός γραμμικού υποδείγματος. Η προσομοίωση που πραγματοποιήθηκε, για δύο ανεξάρτητες στάσιμες χωρικές αυτοπαλίνδρομες διαδικασίες πρώτης τάξης, αποκάλυψε τις ομοιότητες του φαινομένου με την ανάλυση των χρονοσειρών και κατέληξε στο συμπέρασμα ότι η εφαρμογή των μη ανθεκτικών ελέγχων χωρικής εξάρτησης οδηγεί στην εκτίμηση χωρικών οικονομετρικών υποδειγμάτων με τα οποία τόσο η υποφαινομενική συμπεριφορά όσο και η χωρική αυτοσυσχέτιση στα σφάλματα αντιμετωπίζονται.

ΚΕΦΑΛΑΙΟ 3

ΧΩΡΙΚΗ ΕΤΕΡΟΓΕΝΕΙΑ ΣΤΟ ΓΡΑΜΜΙΚΟ ΥΠΟΔΕΙΓΜΑ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

3.1 Εισαγωγή

Η χωρική ετερογένεια (*spatial heterogeneity*) αποτελεί τη δεύτερη μορφή χωρικών επιδράσεων που μαζί με τη χωρική εξάρτηση (*spatial dependence*) μπορούν να εμφανιστούν, λόγω της φύσης των δεδομένων που χρησιμοποιούνται, σε κάποια οικονομετρική ανάλυση επηρεάζοντας την αξιοπιστία των αποτελεσμάτων. Με τον όρο χωρική ετερογένεια εννοείται η ύπαρξη αστάθειας στη σχέση μεταξύ των μεταβλητών ενός υποδείγματος στην περιοχή μελέτης ενός φαινομένου που συνήθως εμφανίζεται ως ετεροσκεδαστικότητα στα σφάλματα του υποδείγματος (Anselin, 1988).

Γενικά, δύο βασικές αιτίες θα μπορούσε να θεωρηθεί ότι οδηγούν στην εμφάνιση χωρικής ετερογένειας σε ένα υπόδειγμα. Η πρώτη αιτία είναι η ύπαρξη κάποιου σφάλματος μέτρησης στις μεταβλητές του υποδείγματος ή η παράλειψη κάποιας στατιστικά σημαντικής μεταβλητής. Στην περίπτωση αυτή, το πρόβλημα δεν θεωρείται ουσιαστικό καθώς αντιμετωπίζεται στα πλαίσια της κλασικής οικονομετρίας με τη βοήθεια κάποιας μεθόδου διόρθωσης της μη σταθερής διακύμανσης των σφαλμάτων. Η δεύτερη αιτία είναι η ύπαρξη διαφορετικής σχέσης μεταξύ εξαρτημένης και ανεξάρτητων μεταβλητών από περιοχή σε περιοχή που εκφράζεται με μεταβλητούς συντελεστές η ακόμη και με διαφορετική συναρτησιακή μορφή. Η αιτία αυτή που είναι πιο ουσιαστική μπορεί να οφείλεται στην παρουσία διακριτής χωρικής ετερογένειας (*discrete spatial heterogeneity*) ή σε συνεχή χωρική ετερογένεια (*continuous spatial heterogeneity*). Διακριτή χωρική ετερογένεια εμφανίζεται όταν η γεωγραφική περιοχή μελέτης ενός φαινομένου διαχωρίζεται σε ομάδες περιοχών με διαφορετικούς συντελεστές αλλά που παραμένουν σταθεροί μέσα σε κάθε ομάδα. Αυτός ο τύπος ετερογένειας μπορεί να αντιμετωπιστεί και με μη χωρικά υποδείγματα με τη βοήθεια ψευδομεταβλητών ή εξειδικεύοντας υποδείγματα χωρικού καθεστώτος (*spatial regimes models*) και χρησιμοποιώντας τεχνικές εκτίμησης όπως η παλινδρόμηση εναλλαγής

(*switching regression*). Περισσότερο δύσκολο να αντιμετωπιστεί είναι η συνεχής χωρική ετερογένεια καθώς σε αυτή την περίπτωση οι συντελεστές του υποδείγματος διαφοροποιούνται με συνεχή τρόπο σε όλες τις περιοχές ως αποτέλεσμα κάποιας συνεχούς συνάρτησης. Η επίλυση του προβλήματος χωρικής ετερογένειας αυτής της μορφής γίνεται με την εκτίμηση τοπικών υποδειγμάτων (*local models*) μέσω εξειδικευμένων τεχνικών εκτίμησης με πιο συχνά χρησιμοποιούμενη τη γεωγραφικά σταθμισμένη παλινδρόμηση (*Geographically Weighted Regression*) που έχει προταθεί από τους Fotheringham, Brunson και Charlton (2002).

Τα τοπικά υποδείγματα (Fotheringham και Brunson, 1999) θεωρούν ότι οι σχέσεις μεταξύ των μεταβλητών δεν είναι σταθερές αλλά μεταβάλλονται από περιοχή σε περιοχή.¹ Ουσιαστικά, αποτελούν επέκταση μιας ευρύτερης κατηγορίας στατιστικών μέτρων που ονομάζονται τοπικές στατιστικές συναρτήσεις (*local statistics*) και που μελετάνε ένα χαρακτηριστικό σε μια συγκεκριμένη περιοχή του χώρου περιλαμβάνοντας στον υπολογισμό τους τη σχέση που υφίσταται με τις γειτονικές περιοχές. Βασική ιδιότητα των τοπικών στατιστικών αλλά και των τοπικών υποδειγμάτων, ως επέκτασή τους, είναι η πολυμεταβλητή διάστασή τους με την έννοια ότι αποδίδουν διαφορετικές τιμές σε κάθε μία από τις n περιοχές του δείγματος και η ερμηνεία τους αποσκοπεί στον εντοπισμό των χωρικών διαφοροποιήσεων του εξεταζόμενου φαινομένου μέσω της χαρτογράφησης των τιμών τους με ένα γεωγραφικό σύστημα πληροφοριών.

Στο κεφάλαιο αυτό παρουσιάζεται η μεθοδολογία της γεωγραφικά σταθμισμένης παλινδρόμησης και οι πιο διαδομένοι έλεγχοι που χρησιμοποιούνται για την αξιολόγηση των αποτελεσμάτων της εκτίμησης τοπικών υποδειγμάτων. Με τη βοήθεια προσομοίωσης εξετάζονται η εφαρμογή της τεχνικής και η συμπεριφορά των ελέγχων στον εντοπισμό χωρικών προτύπων για μια ποικιλία πραγματικών υποδειγμάτων. Τα αποτελέσματα έδειξαν ότι, παρά την καλή εφαρμογή και τα πλεονεκτήματα της μεθόδου, υπάρχουν περιπτώσεις που μπορεί να οδηγήσει σε παραπλανητικά αποτελέσματα. Επιπρόσθετα, εισάγεται η έννοια της τοπικής υποφαινομενικής παλινδρόμησης (*local spurious regression*) και μέσω προσομοίωσης διερευνάται η συμπεριφορά του φαινομένου για δύο ανεξάρτητες στάσιμες χωρικές αυτοπαλίνδρομες διαδικασίες πρώτης τάξης, SAR(1), για τα υποδείγματα που εκτιμώνται με τη γεωγραφικά σταθμισμένη παλινδρόμηση.

¹ Η μη σταθερή σχέση μεταξύ των μεταβλητών αναφέρεται και ως χωρική μη στασιμότητα (*spatial non-stationarity*). Στο κεφάλαιο αυτό αποφεύγεται η χρήση αυτού του όρου για να μη γίνεται σύγχυση με την έννοια της στασιμότητας των χωρικών στοχαστικών διαδικασιών.

3.2 Η γεωγραφικά σταθμισμένη παλινδρόμηση

Έστω το υπόδειγμα της πολλαπλής γραμμικής παλινδρόμησης:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \varepsilon_i \quad (3.1)$$

όπου $\varepsilon_i \sim iid N(0, \sigma^2)$ και το $i = 1, 2, \dots, n$ αναφέρεται στα γεωγραφικά σημεία στα οποία έχει συλλεχθεί η κάθε παρατήρηση. Το υπόδειγμα (3.1) υποθέτει ότι η σχέση εξάρτησης μεταξύ της εξαρτημένης μεταβλητής και των ανεξάρτητων μεταβλητών παραμένει σταθερή σε όλη την εξεταζόμενη γεωγραφική περιοχή και για το λόγο αυτό, πολλές φορές στη χωρική ανάλυση ονομάζεται ολικό υπόδειγμα (*global model*). Εάν θεωρηθεί ότι η σχέση εξάρτησης δεν είναι σταθερή σε όλες τις γεωγραφικές περιοχές, το υπόδειγμα (3.1) μπορεί να γραφεί ως εξής:

$$y_i = \beta_{i0} + \beta_{i1} x_{1i} + \beta_{i2} x_{2i} + \dots + \beta_{ik} x_{ki} + \varepsilon_i \quad (3.2)$$

όπου οι συντελεστές λαμβάνουν διαφορετικές τιμές από περιοχή σε περιοχή εκφράζοντας διαφορετική κατεύθυνση αλλά και ένταση στη σχέση των ανεξάρτητων μεταβλητών με την εξαρτημένη μεταβλητή. Το υπόδειγμα (3.2) ονομάζεται τοπικό υπόδειγμα (*local model*) και εκτιμάται χωριστά για κάθε γεωγραφικό σημείο στο χώρο.

Σε αντίθεση με το υπόδειγμα (3.1) που μπορεί να εκτιμηθεί εύκολα με τη μέθοδο των ελαχίστων τετραγώνων, η εκτίμηση του υποδείματος (3.2) παρουσιάζει δυσκολίες και απαιτεί την εφαρμογή εξειδικευμένων μεθόδων εκτίμησης. Συγκεκριμένα, η δυσκολία που εμφανίζεται είναι ότι για κάθε γεωγραφικό σημείο υπάρχουν περισσότερες παράμετροι που πρέπει να εκτιμηθούν από τις διαθέσιμες παρατηρήσεις. Οι Brunson, Fotheringham και Charlton (1996) και Fotheringham, Brunson και Charlton (2002) χρησιμοποιώντας τις αρχές της μη-παραμετρικής τοπικά σταθμισμένης παλινδρόμησης (*non-parametric locally weighted regression*) και της παλινδρόμησης πυρήνος (*kernel regression*), που έχουν παρουσιαστεί από τους Cleveland (1979) και Cleveland και Devlin (1988), ανέπτυξαν τη μέθοδο της γεωγραφικά σταθμισμένης παλινδρόμησης (*Geographically Weighted Regression – GWR*), δηλαδή μια μη-παραμετρική μέθοδο παλινδρόμησης που επιτρέπει την εκτίμηση υποδειγμάτων με παραμέτρους που δεν παραμένουν σταθερές στο χώρο αλλά μεταβάλλονται. Ουσιαστικά, αποτελεί επέκταση της κλασικής μεθόδου παλινδρόμησης που όμως παρέχει τοπικές εκτιμήσεις των παραμέτρων του υποδείματος σε κάθε γεωγραφικό σημείο.

Η βασική ιδέα της μεθόδου της γεωγραφικά σταθμισμένης παλινδρόμησης είναι η εκτίμηση του υποδείματος (3.2) σε κάθε σημείο εφαρμόζοντας τη σταθμική μέθοδο των

ελαχίστων τετραγώνων (*Weighted Least Squares – WLS*) χρησιμοποιώντας ένα υποσύνολο του δείγματος με παρατηρήσεις που βρίσκονται στη γειτονική γεωγραφική περιοχή, υποθέτοντας ότι οι περιοχές που είναι γειτονικές μεταξύ τους έχουν ομοιότητες. Με τον τρόπο αυτό προκύπτουν εκτιμήσεις για τις παραμέτρους στο i σημείο που είναι μεροληπτικές αλλά με μικρό βαθμό μεροληψίας και με τυπικά σφάλματα που εξαρτώνται από το μέγεθος του δείγματος που χρησιμοποιήθηκε. Όσο μεγαλύτερο είναι το μέγεθος του δείγματος τόσο μικρότερα θα είναι τα σφάλματα των εκτιμήσεων αλλά και μεγαλύτερη η μεροληψία τους και αντίστροφα.

Χρησιμοποιώντας άλγεβρα μητρών το υπόδειγμα της γεωγραφικά σταθμισμένης παλινδρόμησης γράφεται ως εξής:

$$\mathbf{y} = (\boldsymbol{\beta} \otimes \mathbf{X})\mathbf{1} + \boldsymbol{\varepsilon} \quad (3.3)$$

όπου \mathbf{y} είναι το $(n \times 1)$ διάνυσμα των τιμών της εξαρτημένης μεταβλητής, $\boldsymbol{\beta}$ και \mathbf{X} μήτρες διαστάσεων $(n \times (k+1))$, $\mathbf{1}$ είναι ένα $((k+1) \times 1)$ διάνυσμα που περιέχει μονάδες και $\boldsymbol{\varepsilon}$ είναι το $(n \times 1)$ διάνυσμα των σφαλμάτων. Η μήτρα $\boldsymbol{\beta}$ αποτελείται από n γραμμές (όσες οι διαφορετικές περιοχές του δείγματος) που περιλαμβάνουν τους τοπικούς συντελεστές για κάθε γεωγραφικό σημείο και έχει τη μορφή:

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_{10} & \beta_{11} & \beta_{12} & \dots & \beta_{1k} \\ \beta_{20} & \beta_{21} & \beta_{22} & \dots & \beta_{2k} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \beta_{i0} & \beta_{i1} & \beta_{i2} & \dots & \beta_{ik} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \beta_{n0} & \beta_{n1} & \beta_{n2} & \dots & \beta_{nk} \end{bmatrix}$$

Κατά συνέπεια, οι εκτιμήσεις των συντελεστών του υποδείγματος (3.2) που προκύπτουν από την εφαρμογή της μεθόδου της γεωγραφικά σταθμισμένης παλινδρόμησης για το i γεωγραφικό σημείο θα είναι:

$$\hat{\boldsymbol{\beta}}_i = (\mathbf{X}' \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}' \mathbf{W}_i \mathbf{y}$$

με $\mathbf{W}_i = \text{diag}[w_{i1}, w_{i2}, \dots, w_{im}]$ να ορίζεται μια διαγώνια μήτρα που περιέχει τις σταθμίσεις των παρατηρήσεων στο i σημείο. Η μέθοδος της γεωγραφικά σταθμισμένης παλινδρόμησης (GWR) παρουσιάζει ομοιότητες με τη σταθμική μέθοδο των ελαχίστων τετραγώνων (WLS) με τη διαφορά ότι οι σταθμίσεις δεν παραμένουν σταθερές αλλά αναπροσαρμόζονται κάθε φορά που γίνεται η εκτίμηση σε ένα γεωγραφικό σημείο. Ειδικότερα, οι σταθμίσεις

υπολογίζονται με τη χρησιμοποίηση μίας συνάρτησης πυρήνα (*kernel function*) για κάθε σημείο i κατά τη διαδικασία εκτίμησης, έτσι ώστε οι παρατηρήσεις που βρίσκονται πιο κοντά στο γεωγραφικό σημείο i να έχουν μεγαλύτερη επίδραση στην εκτίμηση των συντελεστών από αυτές που είναι πιο απομακρυσμένες.

Οι δύο πιο διαδεδομένες συναρτήσεις πυρήνα, που παρουσιάζονται αναλυτικά στους Fotheringham, Brunson και Charlton (2002), και που εφαρμόζονται στον υπολογισμό των σταθμίσεων, είναι η Γκαουσιανή (*Gaussian weight function*) και η διτετράγωνη (*bi-squared weight function*). Οι σταθμίσεις με τη Γκαουσιανή συνάρτηση υπολογίζονται από τη σχέση:

$$w_{ij} = \exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{b}\right)^2\right) \quad (3.4)$$

και με τη διτετράγωνη συνάρτηση από τη σχέση:

$$w_{ij} = \begin{cases} \left[1 - \left(\frac{d_{ij}}{b}\right)^2\right]^2, & \text{εάν } d_{ij} < b \\ 0, & \text{διαφορετικά} \end{cases} \quad (3.5)$$

Στις δύο προηγούμενες συναρτήσεις σταθμίσεων με w_{ij} συμβολίζεται η στάθμιση της j παρατήρησης στο σημείο i , d_{ij} είναι η απόσταση μεταξύ των σημείων i και j και b το εύρος (*bandwidth*), μια θετική παράμετρος που καθορίζει τη χωρική απόσταση και κατά συνέπεια τον αριθμό γειτονικών παρατηρήσεων που εμπλέκονται σε κάθε τοπική εκτίμηση. Με άλλα λόγια, το εύρος χρησιμοποιείται για να ορίζονται οι σταθμίσεις με τέτοιο τρόπο ώστε η επίδραση (χωρική αυτοσυσχέτιση) να μειώνεται για τις παρατηρήσεις που είναι πιο απομακρυσμένες από το σημείο της εκτίμησης και να προκύπτουν σταθμίσεις αντίστροφες της απόστασης των δύο περιοχών.

Οι συναρτήσεις πυρήνα στις σχέσεις (3.4) και (3.5) ονομάζονται σταθερού πυρήνα διότι χρησιμοποιούν το ίδιο χωρικό εύρος για όλες τις εκτιμήσεις και είναι κατάλληλες στις περιπτώσεις που οι παρατηρήσεις κατανέμονται ομοιόμορφα στο γεωγραφικό χώρο. Όταν όμως οι παρατηρήσεις δεν είναι ομοιόμορφα κατανεμημένες αλλά υπάρχουν περιοχές με πυκνότερα ή αραιότερα γεωγραφικά σημεία δεδομένων, τότε είναι αναγκαία η χρησιμοποίηση προσαρμοσμένου πυρήνα (*adaptive kernel*) ώστε να μεταβάλλεται το εύρος και να εξασφαλίζεται ότι κάθε τοπικό υπόδειγμα θα εκτιμάται με τον ίδιο αριθμό παρατηρήσεων. Μια συνάρτηση αυτής της μορφής είναι η εξής:

$$w_{ij} = \begin{cases} \left[1 - \left(\frac{d_{ij}}{d_{ik}} \right)^2 \right]^2, & \text{εάν η } j \text{ είναι ένας από τους } k \text{ κοντινούς γείτονες της } i \\ 0, & \text{διαφορετικά} \end{cases} \quad (3.6)$$

όπου d_{ik} είναι η απόσταση του i σημείου μέχρι τον k κοντινό του γείτονα.

Το άριστο εύρος προκύπτει από μια διαδικασία διασταυρωτικής επικύρωσης (*cross validation*) με την οποία ελαχιστοποιείται, με τη βοήθεια κάποιας αριθμητικής μεθόδου, το ακόλουθο άθροισμα:

$$CV = \sum_{i=1}^n [y_i - \hat{y}_{\neq i}(b)]^2 \quad (3.7)$$

όπου με $\hat{y}_{\neq i}(b)$ συμβολίζεται η εκτίμηση της y_i με τις παρατηρήσεις για το σημείο i να παραλείπονται από τη διαδικασία της εκτίμησης. Εναλλακτικά, το άριστο εύρος μπορεί να είναι αποτέλεσμα της ελαχιστοποίησης ενός πληροφοριακού κριτηρίου που παρέχει κάποια ισορροπία ανάμεσα στη καλή προσαρμογή και τους βαθμούς ελευθερίας του υποδείγματος. Το πιο συνηθισμένο πληροφοριακό κριτήριο στις εμπειρικές εφαρμογές εκτίμησης τοπικών υποδειγμάτων είναι το διορθωμένο κριτήριο του Akaike που έχει προταθεί από τους Hurvich, Simonoff και Tsai (1998) το οποίο για την περίπτωση της γεωγραφικά σταθμισμένης παλινδρόμησης γράφεται:

$$AIC_c = 2n \ln(\hat{\sigma}) + n \ln(2\pi) + n \left[\frac{(n + tr(\mathbf{S}_1))}{(n - 2 - tr(\mathbf{S}_1))} \right] \quad (3.8)$$

με $\hat{\sigma}$ την εκτιμηθείσα τυπική απόκλιση των σφαλμάτων.

Στη σχέση (3.8) με \mathbf{S}_1 συμβολίζεται ο πίνακας προβολής (*hat* ή *projection matrix*) του υποδείγματος της γεωγραφικά σταθμισμένης παλινδρόμησης που αντιστοιχίζει τις εκτιμηθείσες τιμές \hat{y} με τις πραγματικές τιμές y ικανοποιώντας την εξίσωση $\hat{y} = \mathbf{S}_1 \mathbf{y}$.² Η i γραμμή του πίνακα προβολής θα ισούται με $\mathbf{X}_i (\mathbf{X}' \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}' \mathbf{W}_i$ ενώ τα αποτελέσματα των αλγεβρικών πράξεων $2tr[\mathbf{S}_1] - tr[\mathbf{S}_1' \mathbf{S}_1]$ και $n - (2tr[\mathbf{S}_1] - tr[\mathbf{S}_1' \mathbf{S}_1])$ ονομάζονται, αντίστοιχα, αποτελεσματικός αριθμός παραμέτρων (*effective number of parameters*) και αποτελεσματικοί βαθμοί ελευθερίας (*effective degrees of freedom*), τις περισσότερες φορές είναι μη ακέραια νούμερα και μπορούν να θεωρηθούν οι ανάλογες ποσότητες του αριθμού

² Για λεπτομέρειες βλέπε Fotheringham, Brunson και Charlton (2002 σελίδα 55).

των παραμέτρων και των βαθμών ελευθερίας ενός γραμμικού υποδείγματος παλινδρόμησης που εκτιμάται με τη κλασική μέθοδο των ελαχίστων τετραγώνων.³

Η εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης οδηγεί σε ένα σύνολο τοπικών εκτιμήσεων για τους συντελεστές του υποδείγματος σε όλα τα γεωγραφικά σημεία με αντίστοιχα τοπικά τυπικά σφάλματα και συντελεστές προσδιορισμού. Ειδικότερα, η κατά προσέγγιση διακύμανση των εκτιμηθέντων συντελεστών στο i σημείο είναι:

$$Var(\hat{\beta}_i) = \mathbf{C}_i \mathbf{C}_i' \hat{\sigma}^2 \quad (3.9)$$

όπου $\mathbf{C}_i = (\mathbf{X}'\mathbf{W}_i\mathbf{X})^{-1} \mathbf{X}'\mathbf{W}_i$ και $\hat{\sigma}^2$ είναι η εκτιμηθείσα διακύμανση των σφαλμάτων που προκύπτει από την κανονικοποίηση του αθροίσματος των τετραγώνων των καταλοίπων που επιτυγχάνεται με τη διαίρεσή του με τους αποτελεσματικούς βαθμούς ελευθερίας, δηλαδή:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - 2tr(\mathbf{S}_1) + tr(\mathbf{S}_1^T \mathbf{S}_1)} \quad (3.10)$$

Αντικαθιστώντας την εκτιμηθείσα διακύμανση της σχέσης (3.10) στη σχέση (3.9) υπολογίζονται τα τοπικά τυπικά σφάλματα των εκτιμητών ως εξής:

$$Se(\hat{\beta}_i) = \sqrt{Var(\hat{\beta}_i)} = \sqrt{\mathbf{C}_i \mathbf{C}_i' \hat{\sigma}^2}$$

Οι τοπικοί συντελεστές διαιρούνται με τα τυπικά τους σφάλματα και παράγονται τοπικές ψευδό- t στατιστικές (*local pseudo t-values*) που μπορούν να χρησιμοποιηθούν για τον ανεπίσημο έλεγχο της στατιστικής τους σημαντικότητας (Fotheringham, Brunson και Charlton, 2002 και Wheeler και Paez, 2010).

Εκτός από τοπικές ψευδό- t στατιστικές μπορούν να υπολογιστούν τοπικοί συντελεστές προσδιορισμού που εκφράζουν την ερμηνευτική ικανότητα κάθε τοπικού υποδείγματος.⁴ Ο τοπικός συντελεστής προσδιορισμού για το i γεωγραφικό σημείο ορίζεται ως εξής:

$$R_i^2 = 1 - \frac{\sum_{j=1}^n w_{ij} (y_j - \hat{y}_j)^2}{\sum_{j=1}^n w_{ij} (y_j - \bar{y}_{weighted})^2}$$

³ Σε ένα γραμμικό υπόδειγμα παλινδρόμησης που εκτιμάται με τη μέθοδο των ελαχίστων τετραγώνων ο πίνακας προβολής ορίζεται ως $\mathbf{S}_0 = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'$ και ισχύουν πάντα οι ισοδυναμίες $2tr[\mathbf{S}_0] - tr[\mathbf{S}_0' \mathbf{S}_0] = k + 1$ (ο αριθμός των εκτιμηθέντων παραμέτρων) και $n - (2tr[\mathbf{S}_0] - tr[\mathbf{S}_0' \mathbf{S}_0]) = n - (k + 1)$ (οι βαθμοί ελευθερίας).

⁴ Το εύρος τιμών ενός τοπικού συντελεστή προσδιορισμού δεν περιορίζεται υποχρεωτικά στο διάστημα $[0, 1]$.

όπου w_{ij} είναι η στάθμιση της j παρατήρησης στο σημείο i και $\bar{y}_{weighted}$ ο τοπικά σταθμισμένος μέσος όρος της εξαρτημένης μεταβλητής. Τα αποτελέσματα από την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης μπορούν να χαρτογραφηθούν και να οδηγήσουν στην εξαγωγή συμπερασμάτων σε σχέση με την κατεύθυνση και το μέγεθος της εξάρτησης μεταξύ των μεταβλητών βοηθώντας τον ερευνητή στον εντοπισμό των χωρικών διαφοροποιήσεων στην περιοχή μελέτης ενός φαινομένου.

Η συνολική ερμηνευτική ικανότητα ενός υποδείγματος γεωγραφικά σταθμισμένης παλινδρόμησης μπορεί να αξιολογηθεί με τον υπολογισμό ενός ψευδό συντελεστή προσδιορισμού (*pseudo* ή *quasi-coefficient of determination*) που συγκρίνει την εκτιμηθείσα με την πραγματική μεταβλητότητα της εξαρτημένης μεταβλητής και που ορίζεται ως εξής:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Εφόσον ο ψευδό συντελεστής προσδιορισμού έχει μεγαλύτερη τιμή από το συντελεστή προσδιορισμού του ολικού υποδείγματος, που έχει εκτιμηθεί με τη μέθοδο των ελαχίστων τετραγώνων, ο ερευνητής έχει ενδείξεις υπέρ του υποδείγματος της γεωγραφικά σταθμισμένης παλινδρόμησης.

3.3 Έλεγχοι βελτίωσης από την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης

Η συνεισφορά της γεωγραφικά σταθμισμένης παλινδρόμησης στη βελτίωση της ερμηνευτικής ικανότητας ενός υποδείγματος σε σχέση με την εκτίμησή του μέσω της κλασσικής μεθόδου των ελαχίστων τετραγώνων μπορεί να αξιολογηθεί και με τη βοήθεια στατιστικών ελέγχων. Οι έλεγχοι αυτοί, που βασίζονται στη φιλοσοφία των ελέγχων ανάλυσης διακύμανσης και του πηλίκου δύο διακυμάνσεων, συγκρίνουν τα αθροίσματα των τετραγώνων των εκτιμηθέντων σφαλμάτων που προκύπτουν από τις δύο μεθόδους εκτίμησης. Ο υπολογισμός των στατιστικών που θα παρουσιαστούν σε αυτή την ενότητα και των αντίστοιχων βαθμών ελευθερίας για τη διεξαγωγή των ελέγχων διευκολύνεται εκφράζοντας τα αθροίσματα των τετραγώνων των εκτιμηθέντων σφαλμάτων με τη βοήθεια των πινάκων προβολής.

Ειδικότερα, τόσο για τη μέθοδο των ελαχίστων τετραγώνων όσο και για τη γεωγραφικά σταθμισμένη παλινδρόμηση ορίζεται ο $(n \times n)$ πίνακας προβολής (*hat* ή *projection matrix*) \mathbf{S} που αντιστοιχίζει τις εκτιμηθείσες τιμές \hat{y} με τις πραγματικές τιμές y μέσω της σχέσης $\hat{y} = \mathbf{S}y$. Τα εκτιμηθέντα σφάλματα ορίζονται ως εξής:

$$\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{S}\mathbf{y} = (\mathbf{I} - \mathbf{S})\mathbf{y}$$

και κατά συνέπεια, το άθροισμα τετραγώνων των εκτιμηθέντων σφαλμάτων θα είναι:

$$\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} = \mathbf{y}'(\mathbf{I} - \mathbf{S})'(\mathbf{I} - \mathbf{S})\mathbf{y} = \mathbf{y}'\mathbf{R}\mathbf{y}$$

όπου $\mathbf{R} = (\mathbf{I} - \mathbf{S})'(\mathbf{I} - \mathbf{S})$. Επομένως, εάν με \mathbf{S}_0 και \mathbf{S}_1 συμβολιστούν αντίστοιχα, οι πίνακες προβολής της μεθόδου των ελαχίστων τετραγώνων και της γεωγραφικά σταθμισμένης παλινδρόμησης, το άθροισμα τετραγώνων των εκτιμηθέντων σφαλμάτων θα είναι

$$\text{RSS}_{\text{OLS}} = \mathbf{y}'\mathbf{R}_0\mathbf{y} \quad (3.11)$$

και

$$\text{RSS}_{\text{GWR}} = \mathbf{y}'\mathbf{R}_1\mathbf{y} \quad (3.12)$$

με $\mathbf{R}_0 = (\mathbf{I} - \mathbf{S}_0)'(\mathbf{I} - \mathbf{S}_0)$ και $\mathbf{R}_1 = (\mathbf{I} - \mathbf{S}_1)'(\mathbf{I} - \mathbf{S}_1)$. Αξίζει να αναφερθεί ότι το ίχνος του πίνακα \mathbf{R}_0 ισούται με τους βαθμούς ελευθερίας που προκύπτουν από την εφαρμογή της μεθόδου των ελαχίστων τετραγώνων για ένα γραμμικό υπόδειγμα καθώς ισχύει η σχέση:

$$\text{tr}[\mathbf{R}_0] = \text{tr}\left[(\mathbf{I} - \mathbf{S}_0)'(\mathbf{I} - \mathbf{S}_0)\right] = \text{tr}[\mathbf{I} - \mathbf{S}_0 - \mathbf{S}_0' + \mathbf{S}_0'\mathbf{S}_0] = \text{tr}[\mathbf{I} - \mathbf{S}_0] = n - (k + 1) \quad (3.13)$$

διότι ο πίνακας \mathbf{S}_0 είναι εκθετικά αναλλοίωτος (*idempotent*) και έχει τις ιδιότητες:

$$\text{tr}[\mathbf{S}_0] = \text{tr}[\mathbf{S}_0'] = \text{tr}[\mathbf{S}_0'\mathbf{S}_0]$$

Επειδή για τον πίνακα \mathbf{S}_1 διατηρείται η ιδιότητα $\text{tr}[\mathbf{S}_1] = \text{tr}[\mathbf{S}_1']$ αλλά όχι και η ιδιότητα $\text{tr}[\mathbf{S}_1] = \text{tr}[\mathbf{S}_1'\mathbf{S}_1]$ προκύπτει ότι το ίχνος του πίνακα \mathbf{R}_1 της γεωγραφικά σταθμισμένης παλινδρόμησης θα είναι:

$$\text{tr}[\mathbf{R}_1] = \text{tr}\left[(\mathbf{I} - \mathbf{S}_1)'(\mathbf{I} - \mathbf{S}_1)\right] = \text{tr}[\mathbf{I} - \mathbf{S}_1 - \mathbf{S}_1' + \mathbf{S}_1'\mathbf{S}_1] = n - (2\text{tr}[\mathbf{S}_1] - \text{tr}[\mathbf{S}_1'\mathbf{S}_1]) \quad (3.14)$$

που ισούται με τους αποτελεσματικούς βαθμούς ελευθερίας (*effective degrees of freedom*) από την εκτίμηση των τοπικών υποδειγμάτων.

Οι Brunson, Fotheringham και Charlton (1999) ελέγχουν τη σταθερότητα των συντελεστών του υποδείγματος σε όλη τη γεωγραφική περιοχή με τρόπο παρόμοιο με τη

διαδικασία σύγκρισης δύο υποδειγμάτων παλινδρόμησης όταν σε κάποιο προστίθενται περισσότερες ανεξάρτητες μεταβλητές, δηλαδή συγκρίνοντας το μέσο άθροισμα τετραγώνων των καταλοίπων της μεθόδου των ελαχίστων τετραγώνων και της γεωγραφικά σταθμισμένης παλινδρόμησης με τη βοήθεια ενός πίνακα ανάλυσης διακύμανσης. Η μηδενική και η εναλλακτική υπόθεση του ελέγχου ορίζονται αντίστοιχα ως:

$$H_0 : \beta_{ik} = \beta_k \text{ για κάθε } k$$

$$H_1 : \text{Ένα τουλάχιστον από τα } \beta_{ik} \neq \beta_k$$

και πραγματοποιείται με βάση την ακόλουθη στατιστική:

$$\text{ANOVA-BCM} = \frac{\frac{\text{RSS}_{\text{OLS}} - \text{RSS}_{\text{GWR}}}{\text{tr}[\mathbf{R}_0 - \mathbf{R}_1]}}{\frac{\text{RSS}_{\text{GWR}}}{\text{tr}[\mathbf{R}_1]}}$$

η οποία λόγω των σχέσεων (3.11), (3.12), (3.13) και (3.14) γράφεται ισοδύναμα ως:

$$\text{ANOVA-BCM} = \frac{\frac{\mathbf{y}'\mathbf{R}_0\mathbf{y} - \mathbf{y}'\mathbf{R}_1\mathbf{y}}{2\text{tr}[\mathbf{S}_1] - \text{tr}[\mathbf{S}'_1\mathbf{S}_1] - (k+1)}}{\frac{\mathbf{y}'\mathbf{R}_1\mathbf{y}}{n - (2\text{tr}[\mathbf{S}_1] - \text{tr}[\mathbf{S}'_1\mathbf{S}_1])}} \quad (3.15)$$

Εφόσον τα σφάλματα του υποδείγματος είναι μεταξύ τους ανεξάρτητα και ακολουθούν κανονική κατανομή με σταθερή διακύμανση, η στατιστική ANOVA-BCM ακολουθεί κατά

προσέγγιση την κατανομή F με $\frac{(\text{tr}[\mathbf{R}_0 - \mathbf{R}_1])^2}{\text{tr}[(\mathbf{R}_0 - \mathbf{R}_1)^2]}$ και $\frac{(\text{tr}[\mathbf{R}_1])^2}{\text{tr}[\mathbf{R}_1^2]}$ βαθμούς ελευθερίας,

αντίστοιχα, στον αριθμητή και στον παρανομαστή. Εάν απορριφθεί η μηδενική υπόθεση του ελέγχου τότε μπορεί να θεωρηθεί ότι η γεωγραφικά σταθμισμένη παλινδρόμηση βελτιώνει την ερμηνευτική ικανότητα του υποδείγματος.

Οι Fotheringham, Brunson και Charlton (2002) αναφέρουν ένα διαφορετικό τρόπο εφαρμογής του ελέγχου με τη στατιστική (3.15) που περιλαμβάνεται στο λογισμικό εκτίμησης υποδειγμάτων γεωγραφικής παλινδρόμησης που έχει αναπτυχθεί από τους ίδιους τους συγγραφείς. Συγκεκριμένα, ο έλεγχος εκτελείται θεωρώντας ότι η στατιστική της σχέσης (3.15), που για λόγους διάκρισης θα συμβολίζεται στην περίπτωση αυτή ως ANOVA-FBC,

ακολουθεί την κατανομή F με βαθμούς ελευθερίας $2tr[\mathbf{S}_1] - tr[\mathbf{S}'_1\mathbf{S}_1] - (k+1)$ στον αριθμητή και $n - (2tr[\mathbf{S}_1] - tr[\mathbf{S}'_1\mathbf{S}_1])$ στον παρανομαστή.⁵

Οι Leung, Mei και Zhang (2000a) προτείνουν για τον έλεγχο της συνεισφοράς της γεωγραφικά σταθμισμένης παλινδρόμησης στην ερμηνευτική ικανότητα ενός υποδείγματος μια διαφορετική εκδοχή της στατιστικής ANOVA-BCM της σχέσης (3.15) που τη συμβολίζουν F_2 και που ορίζεται ως εξής:

$$F_2 = \frac{\frac{RSS_{OLS} - RSS_{GWR}}{tr[\mathbf{R}_0 - \mathbf{R}_1]}}{\frac{RSS_{OLS}}{tr[\mathbf{R}_0]}} = \frac{\frac{\mathbf{y}'\mathbf{R}_0\mathbf{y} - \mathbf{y}'\mathbf{R}_1\mathbf{y}}{2tr[\mathbf{S}_1] - tr[\mathbf{S}'_1\mathbf{S}_1] - (k+1)}}{\frac{\mathbf{y}'\mathbf{R}_0\mathbf{y}}{n - (k+1)}} \quad (3.16)$$

Η στατιστική F_2 ακολουθεί κατά προσέγγιση την κατανομή F με $\frac{(tr[\mathbf{R}_0 - \mathbf{R}_1])^2}{tr[(\mathbf{R}_0 - \mathbf{R}_1)^2]}$ βαθμούς ελευθερίας στον αριθμητή και $tr[\mathbf{R}_0]$ βαθμούς ελευθερίας στον παρανομαστή.

Ένας άλλος τρόπος με το οποίο μπορούν να συγκριθούν οι επιδόσεις της μεθόδου των ελαχίστων τετραγώνων και της γεωγραφικά σταθμισμένης παλινδρόμησης στην εκτίμηση ενός υποδείγματος είναι με την εφαρμογή ελέγχων πηλίκου F όπως συνηθίζεται όταν μελετάται η ισότητα δύο διακυμάνσεων. Με βάση αυτή τη μεθοδολογία, οι Fotheringham, Brunson και Charlton (2002) προτείνουν ως στατιστική για τη διεξαγωγή του ελέγχου το πηλίκο του αθροίσματος τετραγώνων των καταλοίπων από την εκτίμηση του ολικού υποδείγματος προς το άθροισμα τετραγώνων των καταλοίπων από την εκτίμηση με τη μέθοδο της γεωγραφικά σταθμισμένης παλινδρόμησης, δηλαδή η στατιστική ορίζεται ως εξής:

$$F\text{-RATIO} = \frac{RSS_{OLS}}{RSS_{GWR}} = \frac{\mathbf{y}'\mathbf{R}_0\mathbf{y}}{\mathbf{y}'\mathbf{R}_1\mathbf{y}} \quad (3.17)$$

και στην πραγματικότητα ο έλεγχος αντιστοιχεί στην ακόλουθη μηδενική και εναλλακτική υπόθεση:

⁵ Αυτός ο τρόπος εφαρμογή του ελέγχου περιλαμβάνεται στο λογισμικό GWR 4.0. Στην παλιότερη έκδοση GWR 3.0 του ίδιου λογισμικού που περιγράφεται αναλυτικά στους Fotheringham, Brunson και Charlton (2002) θεωρείται η προσέγγιση $tr[\mathbf{S}_1] \square tr[\mathbf{S}'_1\mathbf{S}_1]$ με συνέπεια τη μικρή διαφοροποίηση στα δύο πηλικά στη σχέση (3.15) και στους βαθμούς ελευθερίας. Επειδή τα αποτελέσματα των προσομοιώσεων που αναφέρονται στις επόμενες ενότητες δεν αλλάζουν ουσιαστικά με τη χρησιμοποίηση της παραπάνω προσέγγισης, αυτή η επιπρόσθετη παραλλαγή στον υπολογισμό της στατιστικής δεν παρουσιάζεται.

$$H_0 : \frac{\sigma_{OLS}^2}{\sigma_{GWR}^2} = 1 \Rightarrow \text{Δεν υπάρχει διαφορά μεταξύ } OLS \text{ και } GWR$$

$$H_1 : \frac{\sigma_{OLS}^2}{\sigma_{GWR}^2} > 1 \Rightarrow \text{Η } GWR \text{ είναι καλύτερη}$$

όπου με σ_{OLS}^2 και σ_{GWR}^2 συμβολίζονται οι διακυμάνσεις των σφαλμάτων των δύο μεθόδων εκτίμησης.⁶ Σύμφωνα με τους Fotheringham, Brunson και Charlton (2002) το πηλίκο αυτό ακολουθεί την κατανομή F με βαθμούς ελευθερίας στον αριθμητή τους βαθμούς ελευθερίας $n - (k + 1)$ της κλασσικής μεθόδου των ελαχίστων τετραγώνων και στον παρανομαστή τους αποτελεσματικούς βαθμούς ελευθερίας $n - (2tr[\mathbf{S}_1] - tr[\mathbf{S}'_1\mathbf{S}_1])$ από την εκτίμηση με τη γεωγραφικά σταθμισμένη παλινδρόμηση. Μια υψηλή τιμή της στατιστικής F -RATIO οδηγεί στην απόρριψη της μηδενικής υπόθεσης και στο συμπέρασμα ότι η γεωγραφικά σταθμισμένη παλινδρόμηση εκφράζει καλύτερα το υπόδειγμα.

Περισσότερο τεκμηριωμένα αποτελέσματα ως προς την εγκυρότητα των ελέγχων πηλίκου F για τη σύγκριση του ολικού υποδείγματος με τα τοπικά υποδείγματα υπάρχουν στην εργασία των Leung, Mei και Zhang (2000a) οι οποίοι αποδεικνύουν ότι το πηλίκο $RSS_{GWR}/tr[\mathbf{R}_1]$ αποτελεί αμερόληπτη εκτίμηση της διακύμανσης σ_{GWR}^2 των σφαλμάτων της γεωγραφικά σταθμισμένης παλινδρόμησης. Επιπρόσθετα, οι ίδιοι συγγραφείς δείχνουν ότι το

πηλίκο $\frac{tr[\mathbf{R}_1]}{tr[\mathbf{R}_1^2]} \frac{RSS_{GWR}}{\sigma_{GWR}^2}$ ακολουθεί κατά προσέγγιση την κατανομή χ^2 με

$(tr[\mathbf{R}_1])^2/tr[\mathbf{R}_1^2]$ βαθμούς ελευθερίας. Αντίστοιχα, είναι γνωστό από τη μεθοδολογία της ανάλυσης παλινδρόμησης ότι το πηλίκο RSS_{OLS}/σ_{OLS}^2 ακολουθεί ακριβώς την κατανομή χ^2 με $n - (k + 1)$ βαθμούς ελευθερίας. Επομένως, όπως αναφέρουν οι Leung, Mei και Zhang (2000a), όταν η γεωγραφικά σταθμισμένη παλινδρόμηση δεν βελτιώνει σημαντικά τα αποτελέσματα σε σχέση με την απλή παλινδρόμηση, ο λόγος RSS_{GWR}/RSS_{OLS} θα πρέπει να έχει τιμές κοντά στη μονάδα ενώ σε διαφορετική περίπτωση, όταν υπάρχει πραγματική βελτίωση, ο λόγος αυτός θα πρέπει να λαμβάνει πολύ χαμηλές τιμές. Κατά συνέπεια, στη λογική των ελέγχων ισότητας δύο διακυμάνσεων, θα μπορούσαν να οριστούν η ακόλουθη μηδενική και εναλλακτική υπόθεση:

⁶ Ο έλεγχος αυτός αναφέρεται στους Fotheringham, Brunson και Charlton (2002) στη σελίδα 92 χωρίς περαιτέρω διευκρινήσεις και μαθηματική απόδειξη.

$$H_0 : \frac{\sigma_{GWR}^2}{\sigma_{OLS}^2} = 1 \Rightarrow \text{Δεν υπάρχει διαφορά μεταξύ } OLS \text{ και } GWR$$

$$H_1 : \frac{\sigma_{GWR}^2}{\sigma_{OLS}^2} < 1 \Rightarrow \text{Η } GWR \text{ είναι καλύτερη}$$

και να πραγματοποιηθεί ο έλεγχος με τη στατιστική F_1 που είναι πιο αποδεκτή σε σύγκριση με τη στατιστική F -RATIO στη σχέση (3.17) και που ορίζεται ως εξής:

$$F_1 = \frac{\frac{RSS_{GWR}}{tr[\mathbf{R}_1]}}{\frac{RSS_{OLS}}{n-(k+1)}} = \frac{\frac{\mathbf{y}'\mathbf{R}_1\mathbf{y}}{tr[\mathbf{R}_1]}}{\frac{\mathbf{y}'\mathbf{R}_0\mathbf{y}}{n-(k+1)}} \quad (3.18)$$

Η στατιστική F_1 θα ακολουθεί κατά προσέγγιση την κατανομή F με $\frac{(tr[\mathbf{R}_1])^2}{tr[\mathbf{R}_1^2]}$ βαθμούς ελευθερίας στον αριθμητή και $n-(k+1)$ βαθμούς ελευθερίας στον παρανομαστή. Για δεδομένο επίπεδο σημαντικότητας α εάν ισχύει ότι $F_1 < F_{1-\alpha} \left(\frac{(tr[\mathbf{R}_1])^2}{tr[\mathbf{R}_1^2]}, n-(k+1) \right)$, όπου με $F_{1-\alpha} \left(\frac{(tr[\mathbf{R}_1])^2}{tr[\mathbf{R}_1^2]}, n-(k+1) \right)$ συμβολίζεται το επάνω $100(1-\alpha)$ ποσοστιαίο σημείο της F κατανομής, απορρίπτεται η μηδενική υπόθεση και ερευνητής μπορεί να θεωρήσει ότι η γεωγραφικά σταθμισμένη παλινδρόμηση βελτιώνει τα αποτελέσματα.

3.4 Έλεγχος για χωρική αυτοσυσχέτιση στα σφάλματα του υποδείγματος της γεωγραφικά σταθμισμένης παλινδρόμησης

Η εκτίμηση ενός υποδείγματος με τη μέθοδο της γεωγραφικά σταθμισμένης παλινδρόμησης αντιμετωπίζει το πρόβλημα της χωρικής ετερογένειας και οδηγεί τον ερευνητή στην αξιολόγηση των τοπικών σχέσεων μεταξύ των εξεταζόμενων μεταβλητών. Ωστόσο, η εγκυρότητα των αποτελεσμάτων επηρεάζεται από την παρουσία χωρικής αυτοσυσχέτισης στα εκτιμηθέντα σφάλματα όπως γενικότερα συμβαίνει και στη μέθοδο των ελαχίστων τετραγώνων, καθώς στις βασικές της αρχές βασίζεται και η γεωγραφικά σταθμισμένη παλινδρόμηση. Τα σφάλματα από την εκτίμηση ενός ολικού υποδείγματος με τη κλασική μέθοδο των ελαχίστων τετραγώνων μπορούν να εξεταστούν για την ύπαρξη χωρικής αυτοσυσχέτισης εφαρμόζοντας τον έλεγχο με τη στατιστική I του Moran, όπως έχει προταθεί

από τους Cliff και Ord (1972) για τα κατάλοιπα παλινδρόμησης, και με τους ελέγχους χωρικής εξάρτησης των Anselin, Bera, Florax, και Yoon (1996). Σε αντίθεση με τη μέθοδο των ελαχίστων τετραγώνων, που οι προηγούμενοι έλεγχοι χρησιμοποιούνται ευρέως για τον εντοπισμό του προβλήματος της χωρικής αυτοσυσχέτισης στα εκτιμηθέντα σφάλματα, στη γεωγραφικά σταθμισμένη παλινδρόμηση δεν υπάρχει κάποιο καθορισμένο μεθοδολογικό πλαίσιο ελέγχων για την εμφάνιση χωρικής αυτοσυσχέτισης στα σφάλματα. Συνήθως, οι συγγραφείς, στις περισσότερες εμπειρικές εφαρμογές, ελέγχουν το ενδεχόμενο χωρικής αυτοσυσχέτισης υπολογίζοντας τη στατιστική I του Moran και πραγματοποιούν τον έλεγχο με την προσέγγιση της τυχαιοποίησης (*randomization*) ή της κανονικότητας (*normalization*), ή χρησιμοποιούν τις τοπικές στατιστικές I του Moran (*local Moran's I*) που έχουν παρουσιαστεί από τον Anselin (1995). Η πρακτική αυτή, αν και πολύ διαδεδομένη, δεν έχει αποδειχθεί ότι αποδίδει αξιόπιστα αποτελέσματα για τα σφάλματα στη γεωγραφικά σταθμισμένη παλινδρόμηση.

Οι Leung, Mei και Zhang (2000b) ανέπτυξαν μια στατιστική διαδικασία για τον έλεγχο της ύπαρξης χωρικής αυτοσυσχέτισης στα κατάλοιπα από την εκτίμηση ενός υποδείγματος με τη τεχνική της γεωγραφικά σταθμισμένης παλινδρόμησης που βασίζεται στην προσέγγιση τιμών πιθανότητας p (*p-values*) για τη στατιστική I του Moran παράγοντας ταυτόχρονα την ακριβή κατανομή της υπό τη μηδενική υπόθεση της απουσίας χωρικής αυτοσυσχέτισης. Ειδικότερα, οι Leung, Mei και Zhang (2000b) ορίζουν για τα σφάλματα την ακόλουθη μηδενική και εναλλακτική υπόθεση:

$$H_0 : E(\mathbf{\epsilon}\mathbf{\epsilon}') = \sigma^2\mathbf{I}$$

$$H_1 : E(\mathbf{\epsilon}\mathbf{\epsilon}') \neq \sigma^2\mathbf{I}$$

δηλαδή, υπό τη μηδενική υπόθεση τα σφάλματα δεν έχουν χωρική αυτοσυσχέτιση ενώ υπό την εναλλακτική υπόθεση υπάρχει θετική ή αρνητική αυτοσυσχέτιση. Ο έλεγχος πραγματοποιείται με τη γνωστή στατιστική I του Moran που ορίζεται ως εξής:

$$I = \frac{\hat{\mathbf{\epsilon}}'\mathbf{W}\hat{\mathbf{\epsilon}}}{\hat{\mathbf{\epsilon}}'\hat{\mathbf{\epsilon}}} \quad (3.19)$$

όπου $\hat{\mathbf{\epsilon}}$ είναι το $(n \times 1)$ διάνυσμα των εκτιμηθέντων σφαλμάτων που έχουν προκύψει από την εκτίμηση του υποδείγματος με τη μέθοδο της γεωγραφικά σταθμισμένης παλινδρόμησης και \mathbf{W} είναι μια $(n \times n)$ μήτρα χωρικών σταθμίσεων.⁷

⁷ Επισημαίνεται ότι τα στοιχεία της μήτρας \mathbf{W} δεν έχουν καμία σχέση με τις τιμές w_{ij} που προκύπτουν από τις συναρτήσεις στάθμισης κατά τη διαδικασία της γεωγραφικά σταθμισμένης παλινδρόμησης.

Μια υψηλή θετική τιμή της στατιστικής I υποστηρίζει την εναλλακτική υπόθεση της ύπαρξης θετικής χωρικής αυτοσυσχέτισης ενώ μια υψηλή αρνητική τιμή την εναλλακτική υπόθεση της αρνητικής χωρικής αυτοσυσχέτισης. Κατά συνέπεια, οι αντίστοιχες τιμές πιθανότητας p θα πρέπει να ορίζονται ως $p = P(I \geq r)$ και $p = P(I \leq r)$ με r να συμβολίζεται η παρατηρηθείσα τιμή της στατιστικής I . Για έλεγχο με δίπλευρη εναλλακτική υπόθεση, οι τιμές πιθανότητας p θα είναι $2P(I \geq r)$ όταν $P(I \geq r) \leq 1/2$ ή $2[1 - P(I \geq r)]$ όταν $P(I \geq r) \geq 1/2$. Ο υπολογισμός των προαναφερθέντων τιμών πιθανότητας p απαιτεί τη γνώση της κατανομής της στατιστικής I όταν ισχύει η μηδενική υπόθεση της απουσίας χωρικής αυτοσυσχέτισης. Για να παραγάγουν την κατανομή της I , οι Leung, Mei και Zhang (2000b) κάνουν τη βασική υπόθεση ότι τα σφάλματα έχουν μέσο όρο μηδέν και ακολουθούν κανονική κατανομή, δηλαδή ότι ισχύει $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$. Άμεση συνέπεια της υπόθεσης αυτής είναι ότι θα πρέπει να ισχύει και η υπόθεση $E[\hat{y}_i] = E[y_i] \forall i$.⁸ Επομένως, τα εκτιμηθέντα σφάλματα θα πρέπει να έχουν μέσο όρο μηδέν διότι προκύπτει:

$$E[\hat{\boldsymbol{\varepsilon}}] = E[\mathbf{y} - \hat{\mathbf{y}}] = \mathbf{0} \quad (3.20)$$

και επειδή ισχύει ότι:

$$\hat{\boldsymbol{\varepsilon}} = \mathbf{N}\mathbf{y} \quad (3.21)$$

όπου $\mathbf{N} = (\mathbf{I} - \mathbf{S}_1)$, τα εκτιμηθέντα σφάλματα με βάση τις σχέσεις (3.20) και (3.21) γράφονται ως εξής:

$$\hat{\boldsymbol{\varepsilon}} = \hat{\boldsymbol{\varepsilon}} - E[\hat{\boldsymbol{\varepsilon}}] = \mathbf{N}\mathbf{y} - E[\mathbf{N}\mathbf{y}] = \mathbf{N}(\mathbf{y} - E[\mathbf{y}]) = \mathbf{N}\boldsymbol{\varepsilon} \quad (3.22)$$

Αντικαθιστώντας τη σχέση (3.22) στη σχέση (3.19) η στατιστική I του Moran παίρνει τη μορφή:

$$I = \frac{\boldsymbol{\varepsilon}'\mathbf{N}'\mathbf{W}\mathbf{N}\boldsymbol{\varepsilon}}{\boldsymbol{\varepsilon}'\mathbf{N}'\mathbf{N}\boldsymbol{\varepsilon}} \quad (3.23)$$

Οι Leung, Mei και Zhang (2000b) χρησιμοποιούν τη σχέση (3.23) και δείχνουν τον τρόπο με τον οποίο μπορούν να υπολογιστούν οι τιμές πιθανότητας p αφού παραχθεί η ακριβής κατανομή της στατιστικής I όταν ισχύει η μηδενική υπόθεση επεκτείνοντας τη μεθοδολογία

⁸ Σε αντίθεση με τη μέθοδο των ελαχίστων τετραγώνων, τα κατάλοιπα από τη γεωγραφικά σταθμισμένη παλινδρόμηση δεν έχουν μέσο όρο μηδέν και κατά συνέπεια και η εκτιμηθείσα τιμή της y στο σημείο i δεν θα είναι αμερόληπτη εκτίμηση της αναμενόμενης τιμής της. Ωστόσο, εάν η μέθοδος έχει εντοπίσει σωστά το μέγιστο εύρος, τότε η μεροληψία σε κάθε σημείο θα είναι πάρα πολύ μικρή και ως επακόλουθο το άθροισμα των καταλοίπων θα είναι πολύ κοντά στο μηδέν.

των Tiefelsdorf και Boots (1995) για τα εκτιμηθέντα σφάλματα της μεθόδου των ελαχίστων τετραγώνων στα εκτιμηθέντα σφάλματα της γεωγραφικά σταθμισμένης παλινδρόμησης, με τη βοήθεια της εργασίας του Imhof (1961).⁹ Ουσιαστικά, σύμφωνα με τη σχέση (3.23) και μετά από απλές αλγεβρικές πράξεις προκύπτει ότι η πιθανότητα που πρέπει να υπολογιστεί είναι της μορφής:

$$P(I \leq r) = P(\boldsymbol{\varepsilon}'\mathbf{N}'(\mathbf{W} - r\mathbf{I})\mathbf{N}\boldsymbol{\varepsilon} \leq 0) \quad (3.24)$$

Επειδή η ακριβής κατανομή της στατιστικής I υπό τη μηδενική υπόθεση είναι εξαιρετικά δύσκολο να προκύψει στην πράξη εξαιτίας των μαθηματικών υπολογισμών που απαιτούνται, οι Leung, Mei και Zhang (2000b) προτείνουν την προσέγγισή της από μια κατανομή χ^2 εφαρμόζοντας την τεχνική της προσέγγισης των τριών πρώτων ροπών (*the three-moment χ^2 approximation*) όπως έχει παρουσιαστεί από τον Imhof (1961) για την περίπτωση τετραγωνικών μορφών κανονικών τυχαίων μεταβλητών. Συγκεκριμένα, η τεχνική αυτή προσεγγίζει την κατανομή μίας τετραγωνικής μορφής (*quadratic form*) κανονικών μεταβλητών με μια γραμμική συνάρτηση μίας χ^2 μεταβλητής, με τους κατάλληλους βαθμούς ελευθερίας, ώστε οι τρεις πρώτες ροπές των δύο κατανομών να συμπίπτουν. Η προσέγγιση αυτή μπορεί να πραγματοποιηθεί για την κατανομή της στατιστικής I διότι η σχέση (3.23) είναι ένα πηλίκο δύο τετραγωνικών μορφών ανεξάρτητων κανονικών τυχαίων μεταβλητών.

Στην πραγματικότητα, η προσέγγιση αρκεί να γίνει για το γινόμενο:

$$Q = \boldsymbol{\varepsilon}'\mathbf{N}'(\mathbf{W} - r\mathbf{I})\mathbf{N}\boldsymbol{\varepsilon} \quad (3.25)$$

διότι αποτελεί τετραγωνική μορφή αφού η μήτρα $\mathbf{N}'(\mathbf{W} - r\mathbf{I})\mathbf{N}$ είναι συμμετρική με πραγματικά στοιχεία. Όπως δείχνουν οι Leung, Mei και Zhang (2000b), οι ροπές της Q στη σχέση (3.25) μπορούν να προκύψουν υπολογίζοντας τα ίχνη των δυνάμεων του γινομένου $\mathbf{N}'(\mathbf{W} - r\mathbf{I})\mathbf{N}$. Κατά συνέπεια, οι τρεις πρώτες ροπές της Q θα είναι:

$$E[Q] = tr[\mathbf{N}'(\mathbf{W} - r\mathbf{I})\mathbf{N}] \quad (3.26)$$

$$Var[Q] = 2tr[\mathbf{N}'(\mathbf{W} - r\mathbf{I})\mathbf{N}]^2 \quad (3.27)$$

$$E[Q - E[Q]]^3 = 8tr[\mathbf{N}'(\mathbf{W} - r\mathbf{I})\mathbf{N}]^3 \quad (3.28)$$

και οι βαθμοί ελευθερίας h της κατανομής χ^2 θα υπολογίζονται από τη σχέση:

⁹ Για λεπτομέρειες βλέπε Leung, Mei και Zhang (2000b) σελίδες 877-878.

$$h = \frac{8(\text{Var}[Q])^3}{\left(\text{E}[Q - \text{E}[Q]]\right)^2} = \frac{\left(\text{tr}[\mathbf{N}'(\mathbf{W} - r\mathbf{I})\mathbf{N}]^2\right)^3}{\left(\text{tr}[\mathbf{N}'(\mathbf{W} - r\mathbf{I})\mathbf{N}]^3\right)^2} \quad (3.29)$$

Ο προσδιορισμός της τελικής τιμής πιθανότητας p εξαρτάται από το πρόσημο της τρίτης ροπής που υπολογίζεται στη σχέση (3.28). Συγκεκριμένα, οι Leung, Mei και Zhang (2000b) αποδεικνύουν ότι στην περίπτωση που η τρίτη ροπή είναι θετική, δηλαδή ισχύει $\text{E}[Q - \text{E}[Q]]^3 > 0$, η τιμή πιθανότητας p υπολογίζεται ως εξής:

$$P(I \leq r) = P(Q \leq 0) \approx P\left\{\chi_h^2 \leq h - \frac{\sqrt{2h} \text{E}[Q]}{\sqrt{\text{Var}[Q]}}\right\}$$

και όταν η τρίτη ροπή είναι αρνητική, δηλαδή ισχύει $\text{E}[Q - \text{E}[Q]]^3 < 0$, η τιμή πιθανότητας p είναι:

$$P(I \leq r) \approx 1 - P\left\{\chi_h^2 \leq h + \frac{\sqrt{2h} \text{E}[Q]}{\sqrt{\text{Var}[Q]}}\right\}$$

Οι δύο προηγούμενες πιθανότητες υπολογίζονται με αντικατάσταση των αποτελεσμάτων των σχέσεων (3.26) – (3.29) και αρκούν για την εκτίμηση των τιμών πιθανότητας p και την πραγματοποίηση του ελέγχου.

Το μεθοδολογικό πλαίσιο που αναπτύχθηκε από τους Leung, Mei και Zhang (2000b) αποτελεί έναν στατιστικά τεκμηριωμένο τρόπο για την πραγματοποίηση του ελέγχου της εμφάνισης χωρικής αυτοσυσχέτισης στα εκτιμηθέντα σφάλματα της γεωγραφικά σταθμισμένης παλινδρόμησης. Παρόλα αυτά, η χρησιμοποίησή του δεν έχει καθιερωθεί στις εμπειρικές εφαρμογές κάτι που οφείλεται κυρίως στην κριτική που έχει δεχθεί σε σχέση με τις δύο βασικές προϋποθέσεις, της κανονικότητας των σφαλμάτων και της αμεροληψίας της εκτιμηθείσας τιμής για κάθε γεωγραφικό σημείο, που σπάνια ικανοποιούνται στην πράξη. Επιπρόσθετα, λαμβάνοντας υπ' όψιν ότι η στατιστική I του Moran αποτελεί ολικό μέτρο χωρικής αυτοσυσχέτισης αλλά τα σφάλματα στα οποία εφαρμόζεται προκύπτουν από την εκτίμηση τοπικών υποδειγμάτων, δεν είναι σίγουρο ότι η προτεινόμενη μεθοδολογία θα εντοπίσει την ύπαρξη τοπικής χωρικής αυτοσυσχέτισης.

3.5 Η συμπεριφορά της γεωγραφικά σταθμισμένης παλινδρόμησης στην ανίχνευση της χωρικής ετερογένειας και η αποτελεσματικότητα των ελέγχων βελτίωσης

Η μέθοδος της γεωγραφικά σταθμισμένης παλινδρόμησης έχει επιλεγθεί σε πολλές εμπειρικές εργασίες ως η κατάλληλη τεχνική για την αντιμετώπιση του προβλήματος της χωρικής ετερογένειας μέσω της εκτίμησης τοπικών υποδειγμάτων. Ωστόσο, παρά την αποδοχή της από μεγάλη μερίδα ερευνητών, αρκετοί είναι αυτοί που ασκούν κριτική σε σχέση με την ικανότητα της μεθόδου να διακρίνει μια πραγματικά μη χωρικά σταθερή σχέση, ανάμεσα στην εξαρτημένη και στις ανεξάρτητες μεταβλητές, από μια ολική σταθερή σχέση. Επιπρόσθετα, η τεχνική έχει γενικότερα κατηγορηθεί για την εγκυρότητά της ως στατιστικής μεθόδου και ως βάση για διεξαγωγή επίσημης στατιστικής συμπερασματολογίας. Ειδικότερα, αξίζει να αναφερθεί η εργασία των Farber και Páez (2007) που με τη βοήθεια προσομοίωσης μελέτησαν τη συμπεριφορά της διαδικασίας της διασταυρωτικής επικύρωσης (*cross validation*), που χρησιμοποιείται συχνά στην εφαρμογή της μεθόδου, καταλήγοντας στο συμπέρασμα ότι οι προκύπτουσες τιμές για το εύρος επηρεάζονται σε δραματικό βαθμό από συγκεκριμένες γεωγραφικές περιοχές που ασκούν μεγαλύτερες επιδράσεις από άλλες περιοχές. Η συμπεριφορά αυτή οδηγεί στον προσδιορισμό χαμηλών τιμών για το εύρος αυξάνοντας τη μεταβλητότητα των εκτιμηθέντων τοπικών συντελεστών και βελτιώνοντας φαινομενικά την ερμηνευτική ικανότητα του υποδείγματος.

Οι Wheeler και Tiefelsdorf (2005) έχουν δείξει ότι η μέθοδος πολλές φορές καταλήγει στην εκτίμηση τοπικών συντελεστών που παρουσιάζουν υψηλό βαθμό συσχέτισης, ακόμη και στις περιπτώσεις οι μεταβλητές στο ολικό υπόδειγμα είναι ασυσχέτιστες, επισημαίνοντας ότι οι συνέπειες του προβλήματος της πολυσυγραμμικότητας είναι πιο έντονες στα τοπικά υποδείγματα σε σχέση με το ολικό υπόδειγμα. Κατά συνέπεια, η πολυσυγραμμικότητα στα τοπικά υποδείγματα αυξάνει τη μεταβλητότητα των εκτιμηθέντων συντελεστών και προκαλεί την εμφάνιση ψευδών χωρικών προτύπων. Το αποτέλεσμα αυτό συνεπάγεται ότι η εκτίμηση ενός υποδείγματος με τη γεωγραφικά σταθμισμένη παλινδρόμηση μπορεί να οδηγήσει σε παραπλανητικά αποτελέσματα ακόμη και όσο αφορά τη κατεύθυνση της σχέσης μεταξύ των μεταβλητών σε τοπικό επίπεδο κάτι που έχει περεταίρω διερευνηθεί από τους Wheeler (2007 και 2009), Wheeler και Calder (2007) και Páez, Farber και Wheeler (2011) προτείνοντας διαγνωστικά εργαλεία και μεθόδους αντιμετώπισης του προβλήματος.

Αντίθετα με τη μέθοδο των ελαχίστων τετραγώνων που έχει χρησιμοποιηθεί για πολλά χρόνια από τους ερευνητές και που περιλαμβάνει ένα ευρύτερο αναγνωρισμένο πλαίσιο στατιστικής συμπερασματολογίας και μεγάλο αριθμό έγκυρων ελέγχων, η γεωγραφικά σταθμισμένη παλινδρόμηση είναι μια σχετικά καινούρια τεχνική που εφαρμόζεται χωρίς κάποιο καθιερωμένο πλαίσιο. Οι έλεγχοι που εξετάζουν τη συνεισφορά της τεχνικής αυτής στη βελτίωση της ερμηνευτικής ικανότητας ενός υποδείγματος αρκετές φορές δεν αναφέρονται στις εμπειρικές εφαρμογές όπως, επίσης, τα εκτιμηθέντα σφάλματα δεν ελέγχονται για την ύπαρξη χωρικής αυτοσυσχέτισης. Το γεγονός αυτό οφείλεται σε μεγάλο βαθμό στην αμφισβήτηση της αξιοπιστίας των προτεινόμενων ελέγχων λόγω των υποθέσεων που έχουν στην εφαρμογή τους και στον προσεγγιστικό τρόπο με τον οποίο ακολουθούν τις στατιστικές κατανομές. Παραδείγματος χάριν, η εγκυρότητα των στατιστικών F_1 και F_2 καθώς και της διαδικασίας ελέγχου χωρικής αυτοσυσχέτισης στα σφάλματα, που παρουσιάστηκε στην προηγούμενη ενότητα, βασίζονται αποκλειστικά στις διεξαχθείσες προσομοιώσεις των εργασιών των Leung, Mei και Zhang (2000a) και Leung, Mei και Zhang (2000b) που έχουν πραγματοποιηθεί για μικρά δείγματα και μικρό αριθμό επαναλήψεων. Για τις υπόλοιπες στατιστικές που έχουν παρουσιαστεί, η μελέτη της βιβλιογραφίας δεν αποκάλυψε την ύπαρξη δημοσιευμένων προσομοιώσεων που να ασχολούνται με τη συμπεριφορά τους.

Στην ενότητα αυτή εξετάζεται, με τη βοήθεια προσομοίωσης, η συμπεριφορά της γεωγραφικά σταθμισμένης παλινδρόμησης στην αντιμετώπιση την ύπαρξης χωρικής ετερογένειας σε ένα απλό οικονομετρικό υπόδειγμα και στην εκτίμηση των πραγματικών τιμών των συντελεστών του. Επιπρόσθετα, μελετώνται οι συμπεριφορές του εύρους και των συντελεστών προσδιορισμού ως κριτήρια διάκρισης της τοπικής από την ολική σχέση για δύο μεταβλητές. Ταυτόχρονα, διερευνάται η αποτελεσματικότητα των στατιστικών που εφαρμόζονται στον έλεγχο της αύξησης της ερμηνευτικής ικανότητας ενός υποδείγματος μέσω της εκτίμησης τοπικών υποδειγμάτων και ελέγχονται τα εκτιμηθέντα σφάλματα για την εμφάνιση χωρικής αυτοσυσχέτισης.

Η εκτίμηση ενός υποδείγματος με τη μέθοδο της γεωγραφικά σταθμισμένης παλινδρόμησης προϋποθέτει την ύπαρξη γεωγραφικών συντεταγμένων, καθώς η τεχνική δεν μπορεί να εφαρμοστεί θεωρώντας ως κριτήριο γειτονίας την ύπαρξη φυσικών συνόρων μεταξύ δύο περιοχών, κάτι που ήταν επαρκές στις προσομοιώσεις του προηγούμενου κεφαλαίου. Για να αντιμετωπιστεί αυτό το πρόβλημα, οι γεωγραφικές συντεταγμένες

παρήχθησαν επάνω σε ένα ομαλό τετραγωνικό πλέγμα (*regular grid*) διαστάσεων 10x10, 20x20 και 30x30, με κλίμακες για τους δύο άξονες 0 έως 20, 0 έως 40 και 0 έως 60 αντίστοιχα, παρέχοντας δείγματα 100, 400 και 900 παρατηρήσεων. Κάθε τετράγωνο του πλέγματος θεωρείται ότι είναι μια γεωγραφική περιοχή, όπως παρουσιάζεται στο Διάγραμμα 3.1 για την περίπτωση των 100 παρατηρήσεων, και οι συντεταγμένες των κέντρων τους αποτελούν τις γεωγραφικές συντεταγμένες στη διαδικασία προσομοίωσης. Κατά συνέπεια, εάν στο σημείο i οι γεωγραφικές συντεταγμένες ορίζονται ως (z_{i1}, z_{i2}) τότε οι δυνατές τιμές που λαμβάνουν θα είναι:

$$\begin{aligned} z_{i1}, z_{i2} &= 1, 3, 5, \dots, 19 && \text{όταν } N = 100 \\ z_{i1}, z_{i2} &= 1, 3, 5, \dots, 39 && \text{όταν } N = 400 \\ z_{i1}, z_{i2} &= 1, 3, 5, \dots, 59 && \text{όταν } N = 900 \end{aligned}$$

Για τις ανάγκες της προσομοίωσης ορίστηκε ένα απλό γραμμικό υπόδειγμα παλινδρόμησης με μεταβλητούς συντελεστές της μορφής:

$$y_i = \beta_{i0} + \beta_{i1}x_{i1} + \varepsilon_i \quad (3.30)$$

το οποίο χρησιμοποιείται για την παραγωγή της εξαρτημένης μεταβλητής θεωρώντας διαφορετικές περιπτώσεις για τους συντελεστές του που οδηγούν σε τέσσερα διαφορετικά υποδείγματα που ανταποκρίνονται τόσο στην απουσία χωρικής ετερογένειας όσο και στην εμφάνισή της με έντονα χαρακτηριστικά. Συγκεκριμένα, θεωρήθηκαν οι έξι περιπτώσεις:

- I) Οι συντελεστές στο υπόδειγμα (3.30) είναι ίσοι με $\beta_{i0} = 5$ και $\beta_{i1} = 15$, δηλαδή είναι σταθεροί σε κάθε γεωγραφικό σημείο. Στην περίπτωση αυτή το υπόδειγμα γίνεται:

$$y_i = 5 + 15x_{i1} + \varepsilon_i$$

και εκφράζει την ύπαρξη ολικής σχέσης μεταξύ εξαρτημένης και ανεξάρτητης μεταβλητής. Στην πραγματικότητα πρόκειται για ένα απλό οικονομετρικό υπόδειγμα που η κλασική μέθοδος των ελαχίστων τετραγώνων είναι η πιο κατάλληλη τεχνική για να εκτιμηθεί.

- II) Οι συντελεστές στο υπόδειγμα (3.30) λαμβάνουν δύο διαφορετικές τιμές ανάλογα με τη γεωγραφική περιοχή που ανήκει κάθε παρατήρηση εκφράζοντας διακριτή χωρική ετερογένεια, όπως προκύπτει από την πρώτη γεωγραφική συντεταγμένη z_{i1} , δηλαδή:

$$\beta_{i_0} = 10 \quad \text{και} \quad \beta_{i_1} = 20 \quad \text{εάν} \quad z_{i_1} < \frac{m}{2}$$

$$\beta_{i_0} = 15 \quad \text{και} \quad \beta_{i_1} = 30 \quad \text{διαφορετικά}$$

όπου με m συμβολίζεται η μέγιστη τιμή της κλίμακας του οριζοντίου άξονα σε κάθε πλέγμα. Με τον τρόπο αυτό, σε κάθε πλέγμα σχηματίζονται δύο χωρικά καθεστώτα (*spatial regimes*) που αντιστοιχούν σε δύο ολικά υποδείγματα που θα μπορούσαν να εκτιμηθούν με την κλασική μέθοδο των ελαχίστων τετραγώνων περιλαμβάνοντας ψευδομεταβλητές (*dummy variables*).

- III) Οι συντελεστές του υποδείματος (3.30) είναι γραμμικές συναρτήσεις της πρώτης γεωγραφικής συντεταγμένης z_{i_1} , εκφράζοντας την ύπαρξη συνεχούς χωρικής ετερογένειας χαμηλής μορφής και τοπικής σχέσης των δύο μεταβλητών ως εξής:

$$\beta_{i_0} = 5 + 2z_{i_1} \quad \text{και} \quad \beta_{i_1} = 15 + 5z_{i_1}$$

Η γεωγραφικά σταθμισμένη παλινδρόμηση θεωρείται η πιο κατάλληλη μέθοδος για την εκτίμηση αυτού του είδους της σχέσης.

- IV) Οι συντελεστές στο υπόδειγμα (3.30) σχηματίζονται ως γραμμικές συναρτήσεις και των δύο γεωγραφικών συντεταγμένων z_{i_1} και z_{i_2} δηλώνοντας τοπική σχέση μεταξύ των δύο μεταβλητών με παρουσία υψηλής χωρικής ετερογένειας ως εξής:

$$\beta_{i_0} = 5 + 2z_{i_1} + 3z_{i_2} \quad \text{και} \quad \beta_{i_1} = 15 + 5z_{i_1} + 7z_{i_2}$$

Και στην περίπτωση αυτή, η γεωγραφικά σταθμισμένη παλινδρόμηση πρέπει να ανταποκρίνεται στον εντοπισμό των τοπικών σχέσεων της εξαρτημένης με την ανεξάρτητη μεταβλητή.

Οι τιμές της ανεξάρτητης μεταβλητής προέρχονται από ομοιόμορφη $U(0, 10)$ κατανομή και αφού πολλαπλασιαστούν με το συντελεστή β_{i_1} αθροίζονται με το σταθερό όρο β_{i_0} σχηματίζοντας το συστηματικό μέρος του υποδείματος. Στη συνέχεια, στο συστηματικό μέρος προστίθεται στις τιμές των τυχαίων σφαλμάτων που προέρχονται από $N(0,1)$ κατανομή και παράγονται οι τιμές της εξαρτημένης μεταβλητής. Η διαδικασία της προσομοίωσης πραγματοποιήθηκε στο στατιστικό πρόγραμμα R με τη βοήθεια των πακέτων SPGWR και SPDEP που έχουν αναπτυχθεί από τους Bivand και Yu (2014) και Bivand (2015). Εκτελέστηκαν 10.000 επαναλήψεις και για κάθε μία από τις τέσσερις περιπτώσεις εκτιμήθηκε το υπόδειγμα (3.30) με τη μέθοδο των ελαχίστων τετραγώνων και με τη μέθοδο της γεωγραφικά σταθμισμένης παλινδρόμησης με διτετράγωνη και Γκαουσιανή συνάρτηση

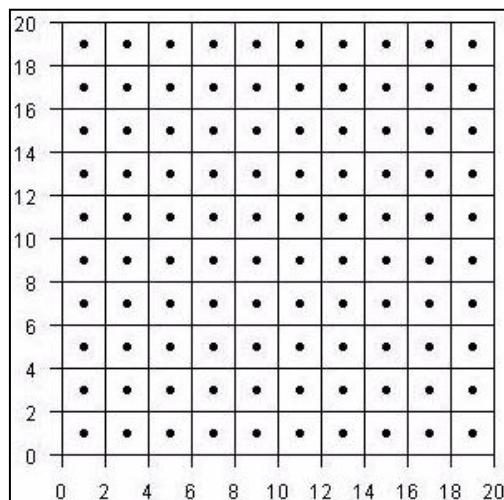
στάθμισης σταθερού πυρήνα. Το άριστο εύρος καθορίστηκε μέσω διασταυρωτικής επικύρωσης (*cross validation*). Κατεγράφησαν, οι μέσες τιμές των συντελεστών προσδιορισμού από την εκτίμηση των υποδειγμάτων με τις δύο μεθόδους εκτίμησης και τα ποσοστά απόρριψης των ελέγχων που εξετάζουν τη βελτίωση στην ερμηνευτική ικανότητα. Σε όλους τους ελέγχους χρησιμοποιήθηκε επίπεδο σημαντικότητας 5%. Επιπρόσθετα, διερευνήθηκαν η συμπεριφορά του εύρους σε κάθε υπόδειγμα και οι εκτιμηθείσες τιμές των σφαλμάτων για ύπαρξη χωρικής αυτοσυσχέτισης. Για να διαπιστωθεί η ικανότητα της γεωγραφικά σταθμισμένης παλινδρόμησης να εντοπίζει τα πραγματικά χωρικά πρότυπα, αποδίδοντας τις πραγματικές τιμές των συντελεστών του υποδείγματος σε κάθε περιοχή, κατεγράφησαν για κάθε τοπικό υπόδειγμα οι εκτιμηθείσες τιμές των τοπικών συντελεστών β_{i0} και β_{i1} και οι μέσες τιμές τους σε κάθε σημείο παρουσιάζονται διαγραμματικά.

Στον Πίνακα 3.1 περιλαμβάνονται οι μέσες τιμές των συντελεστών προσδιορισμού που προέκυψαν από την προσομοίωση όταν εκτιμώνται τα τέσσερα υποδείγματα εφαρμόζοντας τη μέθοδο των ελαχίστων τετραγώνων και τη γεωγραφικά σταθμισμένη παλινδρόμηση. Όπως φαίνεται σε αυτόν τον πίνακα, για το υπόδειγμα I, δηλαδή όταν υπάρχει ολική σχέση εξάρτησης ανάμεσα στην εξαρτημένη και την ανεξάρτητη μεταβλητή, οι δύο μέθοδοι αποδίδουν τους ίδιους μέσους συντελεστές προσδιορισμού κάτι που φανερώνει ότι η γεωγραφικά σταθμισμένη παλινδρόμηση δεν επηρεάζει την ερμηνευτική ικανότητα του υποδείγματος. Στο υπόδειγμα II, το οποίο μπορεί να εκτιμηθεί με την κλασσική μέθοδο των ελαχίστων τετραγώνων χρησιμοποιώντας ψευδομεταβλητές, η γεωγραφικά σταθμισμένη παλινδρόμηση οδηγεί σε ελαφρά μεγαλύτερους συντελεστές προσδιορισμού. Για τα υποδείγματα III και IV που εκφράζουν τοπική σχέση, οι συντελεστές προσδιορισμού που προκύπτουν από την εκτίμηση τοπικών υποδειγμάτων μέσω της γεωγραφικά σταθμισμένης παλινδρόμησης είναι μεγαλύτεροι σε σημαντικό βαθμό δείχνοντας ότι η τεχνική αυτή βελτιώνει την ερμηνευτική ικανότητα των δύο υποδειγμάτων. Κατά συνέπεια, όταν οι τιμές των συντελεστών προσδιορισμού των δύο μεθόδων είναι πολύ κοντά, ο ερευνητής έχει ενδείξεις ότι δεν υπάρχει χωρική ετερογένεια στη σχέση μεταξύ των μεταβλητών του υποδείγματος και τα αποτελέσματα της μεθόδου των ελαχίστων τετραγώνων είναι επαρκή. Αντίθετα, στην περίπτωση που ο συντελεστής προσδιορισμού της γεωγραφικά σταθμισμένης παλινδρόμησης είναι αισθητά υψηλότερος από αυτόν της απλής μεθόδου των ελαχίστων τετραγώνων υπάρχουν σαφείς ενδείξεις παρουσίας χωρικής ετερογένειας που εμφανίζεται σε τοπικό επίπεδο και επομένως υποδεικνύει την ανωτερότητα της πρώτης μεθόδου εκτίμησης.

Πίνακας 3.1

Συντελεστές προσδιορισμού μεθόδων ελαχίστων τετραγώνων και γεωγραφικά σταθμισμένης παλινδρόμησης και ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% των ελέγχων βελτίωσης για τα τέσσερα υποδείγματα και όλα τα μεγέθη δείγματος

| Πυρήνας | | Διτετράγωνος | | | | Γκαουσιανός | | | |
|-------------------|----------------------|--------------|-----------------|--------------------|-------------------|-------------|-----------------|--------------------|-------------------|
| N | Υπόδειγμα | Ολική Σχέση | Χορικό Καθεστώς | Χαμηλή Ετερογένεια | Υψηλή Ετερογένεια | Ολική Σχέση | Χορικό Καθεστώς | Χαμηλή Ετερογένεια | Υψηλή Ετερογένεια |
| | | I | II | III | IV | I | II | III | IV |
| 100 | R ² – OLS | 0,999 | 0,842 | 0,530 | 0,621 | 0,999 | 0,842 | 0,530 | 0,621 |
| | R ² – GWR | 0,999 | 0,991 | 0,996 | 0,997 | 0,999 | 0,993 | 0,997 | 0,998 |
| Ποσοστά Απόρριψης | ANOVA – BCM | 25,07 | 100 | 100 | 100 | 19,02 | 100 | 100 | 100 |
| | ANOVA – FBC | 4,38 | 98,81 | 100 | 100 | 2,43 | 92,99 | 100 | 100 |
| | F – RATIO | 1,67 | 99,25 | 100 | 100 | 1,39 | 96,41 | 100 | 100 |
| | F1 | 0,03 | 100 | 100 | 100 | 0,02 | 100 | 100 | 100 |
| | F2 | 4,18 | 79,65 | 99,96 | 100 | 2,7 | 62,19 | 98,58 | 98,69 |
| 400 | R ² – OLS | 0,999 | 0,843 | 0,469 | 0,593 | 0,999 | 0,843 | 0,469 | 0,593 |
| | R ² – GWR | 0,999 | 0,996 | 0,999 | 0,999 | 0,999 | 0,997 | 0,999 | 0,999 |
| Ποσοστά Απόρριψης | ANOVA – BCM | 37,80 | 100 | 100 | 100 | 32,85 | 100 | 100 | 100 |
| | ANOVA – FBC | 4,95 | 99,80 | 100 | 100 | 3,08 | 99,05 | 100 | 100 |
| | F – RATIO | 0,09 | 99,85 | 100 | 100 | 0,15 | 99,58 | 100 | 100 |
| | F1 | 0,00 | 100 | 100 | 100 | 0,00 | 100 | 100 | 100 |
| | F2 | 7,71 | 98,91 | 100 | 100 | 6,23 | 79,94 | 100 | 100 |
| 900 | R ² – OLS | 0,999 | 0,843 | 0,447 | 0,583 | 0,999 | 0,843 | 0,447 | 0,583 |
| | R ² – GWR | 0,999 | 0,997 | 1,000 | 1,000 | 0,999 | 0,998 | 1,000 | 1,000 |
| Ποσοστά Απόρριψης | ANOVA – BCM | 41,04 | 100 | 100 | 100 | 35,82 | 100 | 100 | 100 |
| | ANOVA – FBC | 4,95 | 99,99 | 100 | 100 | 3,18 | 99,78 | 100 | 100 |
| | F – RATIO | 0,02 | 100 | 100 | 100 | 0,02 | 99,88 | 100 | 100 |
| | F1 | 0,00 | 100 | 100 | 100 | 0,00 | 100 | 100 | 100 |
| | F2 | 8,35 | 99,66 | 100 | 100 | 6,78 | 88,8 | 100 | 100 |



Διάγραμμα 3.1

Οι γεωγραφικές συντεταγμένες σε ένα 10x10 ομαλό πλέγμα (n=100)

Η επιλογή της καλύτερης μεθόδου εκτίμησης, συγκρίνοντας τους δύο συντελεστές προσδιορισμού, ενισχύεται περαιτέρω εφόσον εξεταστούν οι τιμές τους εύρους που έχουν υπολογιστεί από τη διαδικασία της διασταυρωτικής επικύρωσης και που οι μέσες, ελάχιστες και μέγιστες τιμές του παρουσιάζονται στον Πίνακα 3.2. Επιπρόσθετα, ο πίνακας αυτός περιλαμβάνει και το ολικό εύρος (*global bandwidth*) που αντιστοιχεί σε κάθε μέγεθος δείγματος. Το ολικό εύρος υπολογίζεται ως η Ευκλείδεια απόσταση μεταξύ των γεωγραφικών συντεταγμένων των κέντρων του κάτω αριστερά και του επάνω δεξιά τετραγώνων στο ομαλό πλέγμα.¹⁰ Εφόσον, η τιμή του εύρους που έχει καταλήξει η διασταυρωτική επικύρωση είναι πολύ κοντά στο ολικό εύρος, ο ερευνητής έχει ενδείξεις απουσίας χωρικής ετερογένειας στους συντελεστές του υποδείγματος και ύπαρξης ολικής σχέσης, ενώ μικρές τιμές του εύρους φανερώνουν χωρική ετερογένεια και τη καταλληλότητα της γεωγραφικά σταθμισμένης παλινδρόμησης.

Από τον Πίνακα 3.2 διαπιστώνεται ότι το μέσο εύρος για το υπόδειγμα I, που αντιστοιχεί σε ολική σχέση, είναι πολύ κοντά στο ολικό εύρος και κατά απόρροια η κλασική μέθοδος των ελαχίστων τετραγώνων είναι η πιο κατάλληλη για την εκτίμησή του. Από τη άλλη πλευρά, η εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης είναι αναγκαία στα τοπικά υποδείγματα III και IV καθώς οι μέσες τιμές του εύρους είναι αισθητά μικρότερες από το ολικό εύρος. Είναι αξιοσημείωτο ότι στο υπόδειγμα II που στην πραγματικότητα αναφέρεται σε ολική σχέση το μέσο εύρος είναι ιδιαίτερα χαμηλό και μάλιστα μικρότερο από αυτά των υποδειγμάτων III και IV κάτι που δεν θα έπρεπε να συμβαίνει.¹¹ Το γεγονός αυτό μπορεί να οφείλεται σε αδυναμία της διασταυρωτικής επικύρωσης να προσδιορίσει το σωστό εύρος, κάτι που έχει επισημανθεί και στους Farber και Páez (2007), εξ' αιτίας του μηχανισμού παραγωγής των συντελεστών του υποδείγματος που σε αυτή την περίπτωση δεν αποτελούν συνεχή συνάρτηση των γεωγραφικών συντεταγμένων. Τα αποτελέσματα αυτά ισχύουν ανεξάρτητα από τη συνάρτηση στάθμισης και το μέγεθος του δείγματος με τη διαφορά ότι με τον Γκαουσιανό πυρήνα το μέσο εύρος είναι χαμηλότερο.

¹⁰ Για το ομαλό πλέγμα των 100 παρατηρήσεων στο Διάγραμμα 3.1 το ολικό εύρος είναι η απόσταση μεταξύ των συντεταγμένων (1, 1) και (19, 19) και υπολογίζεται ως $\sqrt{(19-1)^2 + (19-1)^2} = 25,455844$.

¹¹ Είναι αναγκαίο να αναφερθεί ότι κατά τη διαδικασία της προσομοίωσης, σε κάποιες επαναλήψεις, η εκτίμηση του υποδείγματος II παρουσίασε προβλήματα που οφείλονται στον υπολογισμό του εύρους από τη διασταυρωτική επικύρωση το οποίο είτε δεν ήταν δυνατόν να προσδιοριστεί είτε ήταν τόσο μικρό ώστε να μην μπορεί να ακολουθήσει η εκτίμηση με τη γεωγραφικά σταθμισμένη παλινδρόμηση. Τα προβλήματα αυτά, που εμφανίστηκαν με τον Γκαουσιανό πυρήνα, αφορούν πολλές περιπτώσεις στο πλέγμα των 100 παρατηρήσεων και τρεις στο πλέγμα των 400 παρατηρήσεων οι οποίες αφαιρέθηκαν από την προσομοίωση και αντικαταστάθηκαν από περισσότερες επαναλήψεις μέχρι να συμπληρωθεί ο αριθμός των 10.000 επαναλήψεων.

Πίνακας 3.2
Μέσες, ελάχιστες και μέγιστες τιμές εύρους για τα τέσσερα
υποδείγματα και όλα τα μεγέθη δείγματος

| Πυρήνας | | Διτετράγωνος | | | | Γκαουσιανός | | | |
|---------|----------------|--------------|-----------------|--------------------|-------------------|-------------|-----------------|--------------------|-------------------|
| N | Υπόδειγμα | Ολική Σχέση | Χωρικό Καθεστώς | Χαμηλή Ετερογένεια | Υψηλή Ετερογένεια | Ολική Σχέση | Χωρικό Καθεστώς | Χαμηλή Ετερογένεια | Υψηλή Ετερογένεια |
| | | I | II | III | IV | I | II | III | IV |
| 100 | Ολικό Εύρος | 25,456 | | | | 25,456 | | | |
| | Μέσο Εύρος | 22,061 | 4,816 | 5,105 | 5,104 | 19,285 | 1,606 | 1,757 | 1,755 |
| | Ελάχιστο Εύρος | 4,244 | 2,012 | 3,219 | 3,283 | 1,574 | 0,450 | 1,038 | 0,980 |
| | Μέγιστο Εύρος | 25,456 | 10,380 | 7,701 | 7,953 | 25,456 | 3,695 | 2,641 | 2,542 |
| 400 | Ολικό Εύρος | 53,740 | | | | 53,740 | | | |
| | Μέσο Εύρος | 46,677 | 4,655 | 5,419 | 5,417 | 40,403 | 1,537 | 1,963 | 1,964 |
| | Ελάχιστο Εύρος | 9,634 | 2,007 | 4,755 | 4,536 | 2,907 | 0,488 | 1,576 | 1,544 |
| | Μέγιστο Εύρος | 53,740 | 8,346 | 7,284 | 7,290 | 53,740 | 2,928 | 2,394 | 2,398 |
| 900 | Ολικό Εύρος | 82,024 | | | | 82,024 | | | |
| | Μέσο Εύρος | 70,713 | 4,752 | 5,613 | 5,605 | 61,807 | 1,531 | 2,100 | 2,099 |
| | Ελάχιστο Εύρος | 12,785 | 2,122 | 5,093 | 5,130 | 5,353 | 0,517 | 1,813 | 1,820 |
| | Μέγιστο Εύρος | 82,024 | 7,754 | 7,275 | 7,352 | 82,024 | 2,666 | 2,420 | 2,397 |

Πίνακας 3.3
Μέσες τιμές στατιστικής I και ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5%
με την τεχνική των Leung, Mei και Zhang (2000b) του ελέγχου ύπαρξης χωρικής
αυτοσυσχέτισης στα εκτιμηθέντα σφάλματα της μεθόδου της γεωγραφικά σταθμισμένης
παλινδρόμησης για τα τέσσερα υποδείγματα και όλα τα μεγέθη δείγματος

| Πυρήνας | | Διτετράγωνος | | | | Γκαουσιανός | | | |
|---------|----------------|--------------|-----------------|--------------------|-------------------|-------------|-----------------|--------------------|-------------------|
| N | Υπόδειγμα | Ολική Σχέση | Χωρικό Καθεστώς | Χαμηλή Ετερογένεια | Υψηλή Ετερογένεια | Ολική Σχέση | Χωρικό Καθεστώς | Χαμηλή Ετερογένεια | Υψηλή Ετερογένεια |
| | | I | II | III | IV | I | II | III | IV |
| 100 | Μέση Τιμή I | -0,034 | 0,095 | 0,196 | 0,197 | -0,026 | 0,087 | 0,186 | 0,187 |
| | Ποσ. Απόρριψης | 3,70 | 99,03 | 100 | 99,99 | 3,81 | 99,09 | 100 | 100 |
| 400 | Μέση Τιμή I | -0,008 | 0,101 | 0,168 | 0,168 | -0,006 | 0,094 | 0,180 | 0,181 |
| | Ποσ. Απόρριψης | 4,55 | 99,95 | 100 | 100 | 4,55 | 99,96 | 100 | 100 |
| 900 | Μέση Τιμή I | -0,004 | 0,114 | 0,148 | 0,148 | -0,003 | 0,097 | 0,177 | 0,177 |
| | Ποσ. Απόρριψης | 4,67 | 100 | 100 | 100 | 4,68 | 100 | 100 | 100 |

Η απόφαση της επιλογής μεταξύ ολικού και τοπικού υποδείγματος διευκολύνεται και από την εφαρμογή των ελέγχων συνεισφοράς της γεωγραφικά σταθμισμένης παλινδρόμησης που παρουσιάστηκαν αναλυτικά στην ενότητα 3.3. Τα αποτελέσματα της προσομοίωσης για αυτούς τους ελέγχους, σε επίπεδο σημαντικότητας 5%, περιλαμβάνονται στον Πίνακα 3.1 για τα τέσσερα διαφορετικά υποδείγματα και για όλα τα μεγέθη δείγματος και συναρτήσεις στάθμισης. Επειδή το υπόδειγμα I αντιπροσωπεύει μια απόλυτα ολική σχέση, τα ποσοστά απόρριψης στην περίπτωση αυτή αποτελούν το μέγεθος των ελέγχων και θα αναμενόταν για όλες τις στατιστικές να κυμαίνονται κοντά στο 5%. Εντούτοις, η στατιστική ANOVA-BCM έχει ποσοστά απόρριψης πολύ υψηλότερα από το επίπεδο σημαντικότητας του 5% που τείνουν να αυξάνονται στα μεγαλύτερα δείγματα προσεγγίζοντας το 41,04% στο δείγμα των 900 παρατηρήσεων με διτετράγωνη συνάρτηση στάθμισης ενώ είναι λίγο πιο μικρά με την Γκαουσιανή συνάρτηση στάθμισης. Αντίθετα, τα ποσοστά απόρριψης των ελέγχων με τις στατιστικές F -RATIO και F_1 είναι ιδιαίτερα μικρά, ουσιαστικά μηδέν. Την πιο αποδεκτή συμπεριφορά, όσον αφορά το μέγεθος των ελέγχων, έχουν οι στατιστικές ANOVA-FBC και F_2 με την πρώτη να έχει ποσοστά απόρριψης κοντά στο 5% και τη δεύτερη λίγο υψηλότερα. Γενικά, το μέγεθος των ελέγχων δεν είναι το αναμενόμενο σύμφωνα με τη στατιστική θεωρία γεγονός που οφείλεται στον τρόπο εφαρμογής τους καθώς ακολουθούν προσεγγιστικά τις κατανομές χ^2 και F . Τα ποσοστά απόρριψης και των πέντε στατιστικών για τα υποδείγματα III και IV είναι σχεδόν 100 φανερώνοντας ότι έχουν εξαιρετική ισχύ όταν η γεωγραφικά σταθμισμένη παλινδρόμηση είναι η αποδεκτή μέθοδος εκτίμησης. Παρόμοια συμπεριφορά εντοπίζεται και για το υπόδειγμα II κάτι αναμενόμενο αφού και οι συντελεστές προσδιορισμού εσφαλμένα δείχνουν τη γεωγραφικά σταθμισμένη παλινδρόμηση ως τη καλύτερη τεχνική εκτίμησης.

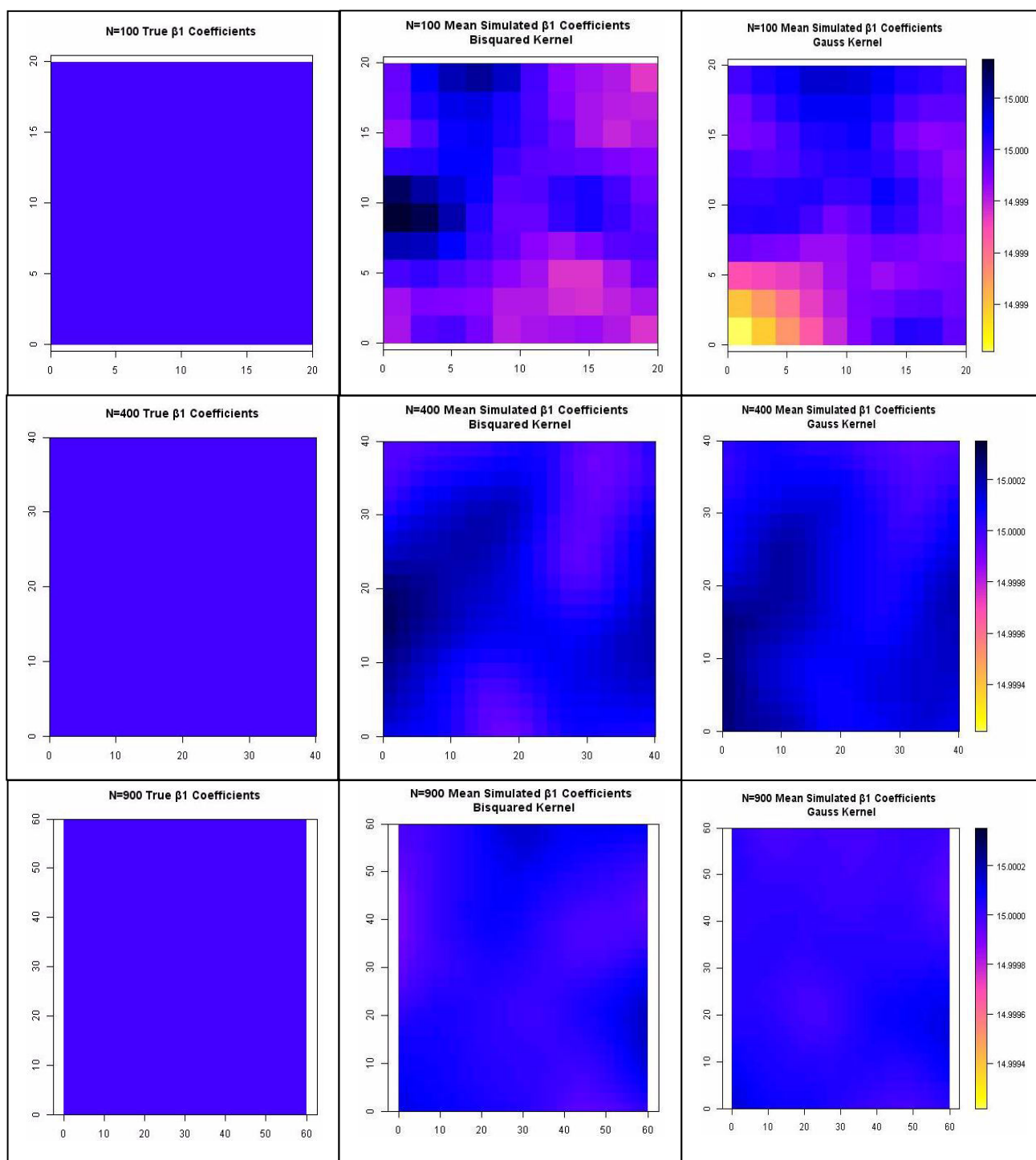
Τα εκτιμηθέντα σφάλματα από την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης θα πρέπει να είναι ανεξάρτητα χωρίς προβλήματα χωρικής αυτοσυσχέτισης. Ο Πίνακας 3.3 παρουσιάζει τις μέσες τιμές της στατιστικής I του Moran χρησιμοποιώντας το κριτήριο του πύργου και τα ποσοστά απόρριψης του αντίστοιχου ελέγχου με τη μεθοδολογία των Leung, Mei και Zhang (2000b) που προέκυψαν από την προσομοίωση. Όπως διαπιστώνεται, όταν η εκτίμηση για το ολικό υπόδειγμα I γίνεται με τη γεωγραφικά σταθμισμένη παλινδρόμηση τα εκτιμηθέντα σφάλματα δεν αντιμετωπίζουν προβλήματα χωρικής αυτοσυσχέτισης όπως δηλώνεται από τις σχεδόν μηδενικές μέσες τιμές της στατιστικής I που έχουν καταγραφεί και τα ποσοστά απόρριψης που είναι κοντά στο επίπεδο

σημαντικότητας του 5%. Ωστόσο, για το υπόδειγμα II και ιδιαίτερα για τα τοπικά υποδείγματα III και IV οι μέσες τιμές της I είναι υψηλές και τα ποσοστά απόρριψης του ελέγχου 100%.¹² Το αποτέλεσμα αυτό είναι δύσκολο να εξηγηθεί και δείχνει ότι η εκτίμηση με τη γεωγραφικά σταθμισμένη παλινδρόμηση μπορεί να οδηγήσει σε συσχετισμένα σφάλματα.

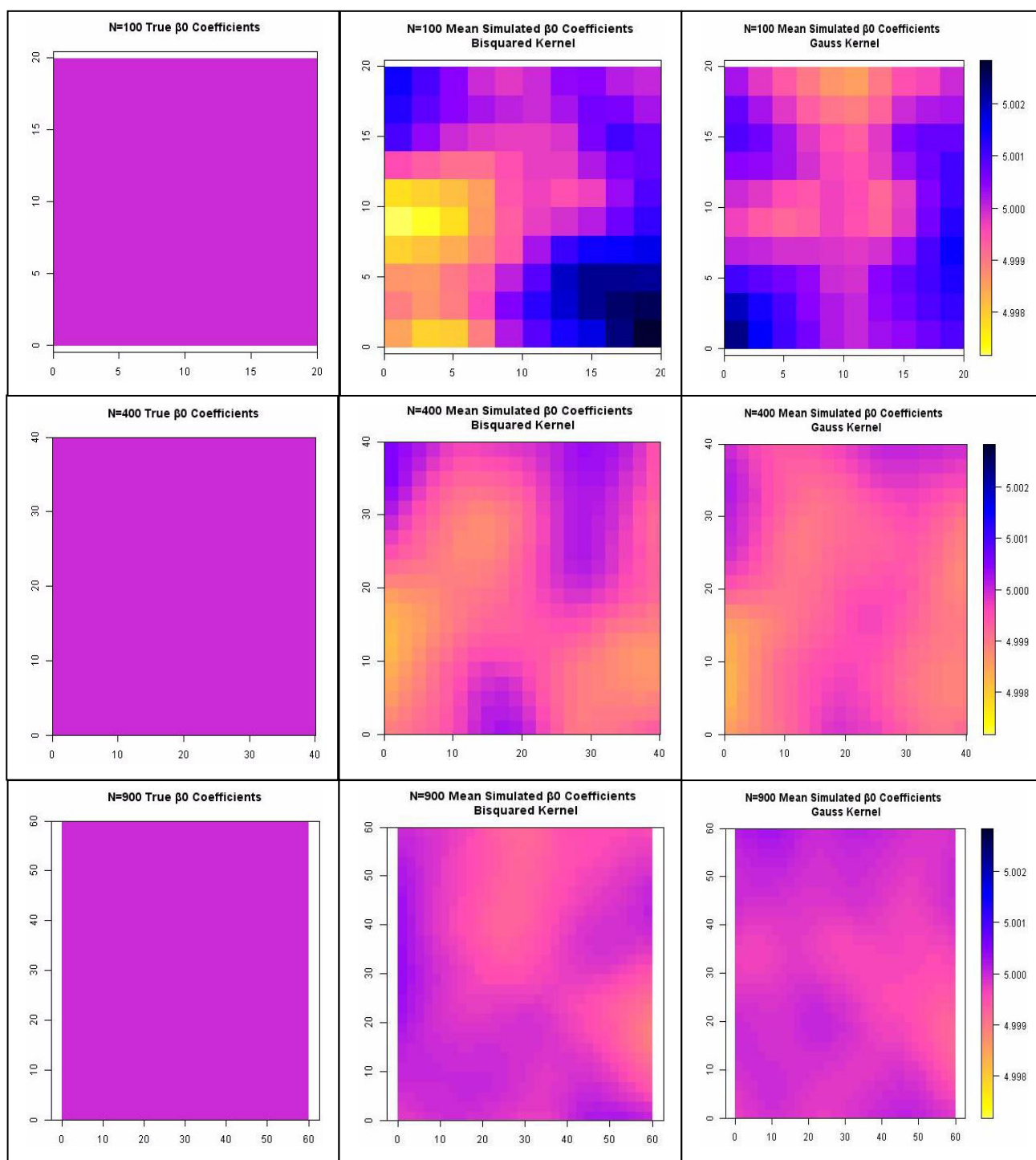
Μια πιθανή εξήγηση που θα μπορούσε να δοθεί είναι ότι η συμπεριφορά αυτή μπορεί να οφείλεται σε αποτυχία της στατιστικής I και των αντίστοιχων ελέγχων να εξετάσουν αξιόπιστα τα σφάλματα για χωρική αυτοσυσχέτιση διότι τα σφάλματα αυτά προέρχονται από την εκτίμηση τοπικών υποδειγμάτων ενώ η στατιστική I ελέγχει για ολική χωρική αυτοσυσχέτιση. Ο υπολογισμός τοπικών στατιστικών I του Moran και η διενέργεια ελέγχων για τοπική χωρική αυτοσυσχέτιση, όπως έχει προταθεί από τον Anselin (1995), ίσως οδηγεί σε διαφορετικά αποτελέσματα. Όμως, αυτή η διαδικασία είναι δύσκολο να μελετηθεί και να παρουσιαστεί σε μία προσομοίωση επειδή σε κάθε περιοχή υπολογίζεται η στατιστική και πραγματοποιείται ο έλεγχος αυτόνομα. Ένας άλλος λόγος που θα μπορούσε να προκαλεί αυτή τη συμπεριφορά είναι ότι τα υποδείγματα αυτά έχουν εκ της παραγωγής τους χωρική ετερογένεια και κατά συνέπεια τα εκτιμηθέντα σφάλματα ενδέχεται να μην έχουν σταθερή διακύμανση με αποτέλεσμα την εσφαλμένη συμπεριφορά των ελέγχων. Επιπρόσθετα, σε αντίθεση με τις προσομοιώσεις του προηγούμενου κεφαλαίου όπου η μήτρες χωρικών σταθμίσεων ήταν δεδομένες, για τον υπολογισμό της στατιστικής I και την πραγματοποίηση των ελέγχων χωρικής αυτοσυσχέτισης στα σφάλματα των εξεταζόμενων υποδειγμάτων ορίζεται αυθαίρετα το κριτήριο του πύργου χωρίς να είναι σίγουρο ότι ανταποκρίνεται στην πραγματική δομή των σφαλμάτων και κατά πόσο επηρεάζει τα αποτελέσματα.

Οι μέσες τιμές των τοπικών συντελεστών που εκτιμήθηκαν με τη μέθοδο της γεωγραφικά σταθμισμένης παλινδρόμησης σε κάθε γεωγραφική περιοχή απεικονίζονται στα ψηφιδωτά (*raster*) που περιλαμβάνονται στα Διαγράμματα 3.2 – 3.5 για κάθε ένα από τα τέσσερα υποδείγματα. Συγκεκριμένα, σε αυτά τα διαγράμματα τα ψηφιδωτά στην αριστερή στήλη δείχνουν τις πραγματικές τιμές των συντελεστών και το χωρικό πρότυπο που σχηματίζεται με κάθε μέγεθος δείγματος. Τα ψηφιδωτά στη μεσαία και δεξιά στήλη απεικονίζουν τις μέσες εκτιμηθείσες τιμές των συντελεστών με την εφαρμογή αντίστοιχα, διτετράγωνης και Γκαουσιανής συνάρτησης στάθμισης. Σε όλα τα ψηφιδωτά του ίδιου διαγράμματος έχει χρησιμοποιηθεί κοινή χρωματική κλίμακα με σκοπό τη συγκρισιμότητά τους.

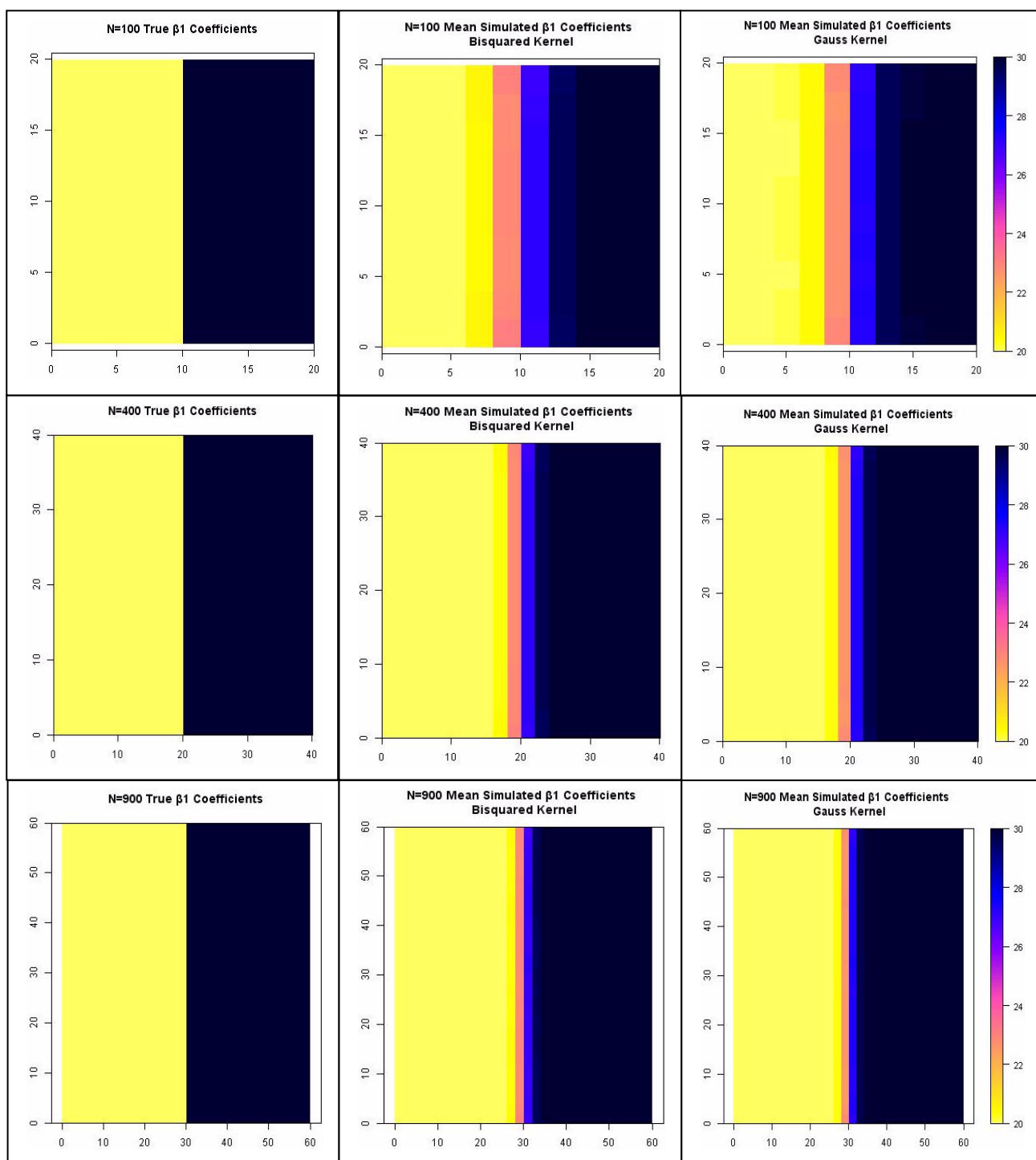
¹² Η διενέργεια των ελέγχων με την προσέγγιση της τυχαιοποίησης ή της κανονικότητας οδηγεί σε λίγο μικρότερα αλλά εξίσου υψηλά ποσοστά απόρριψης.



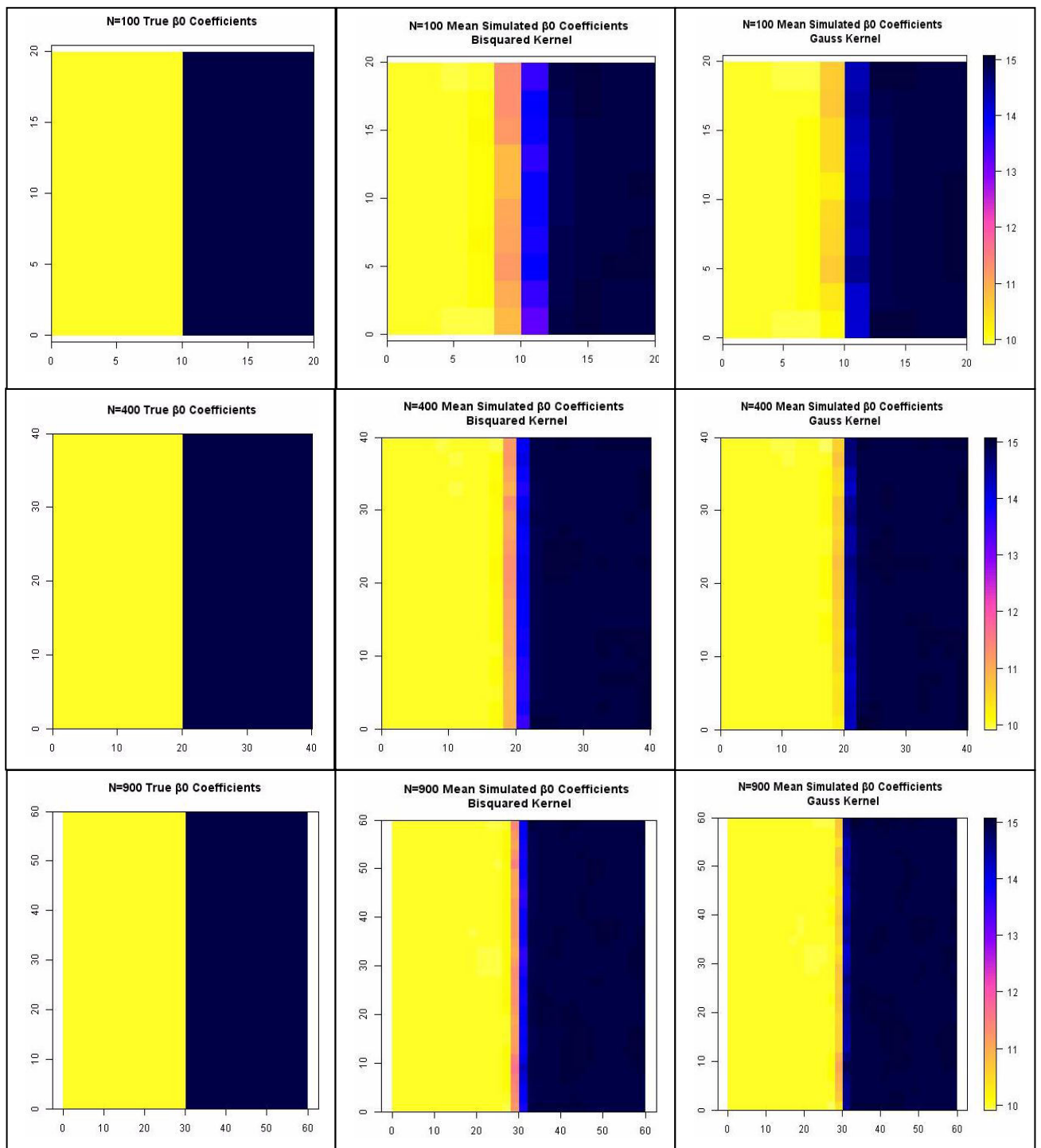
Διάγραμμα 3.2 (Α)
Ψηφιδωτά πραγματικών και προσομοιωμένων μέσων τοπικών εκτιμηθέντων τιμών από την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης για όλα τα μεγέθη δείγματος και τις δύο συναρτήσεις σταθμίσεων
Υπόδειγμα Ι: Ολική Σχέση
Συντελεστής β_{1i}



Διάγραμμα 3.2 (B)
Ψηφιδωτά πραγματικών και προσομοιωμένων μέσων τοπικών εκτιμηθέντων τιμών από την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης για όλα τα μεγέθη δείγματος και τις δύο συναρτήσεις σταθμίσεων
Υπόδειγμα I: Ολική Σχέση Συντελεστής β_{i0}

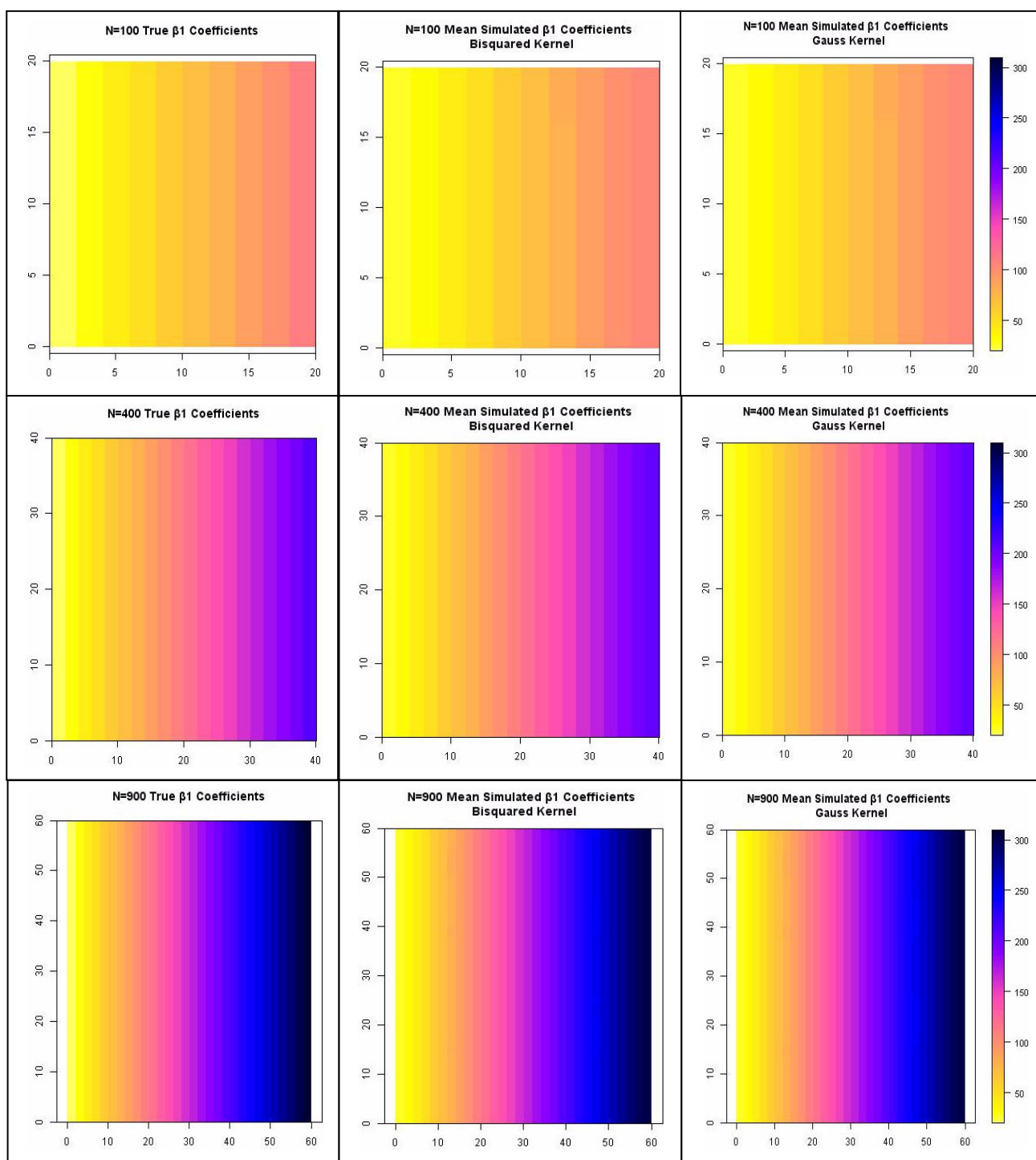


Διάγραμμα 3.3 (Α)
Ψηφιδωτά πραγματικών και προσομοιωμένων μέσων τοπικών εκτιμηθέντων τιμών από την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης για όλα τα μεγέθη δείγματος και τις δύο συναρτήσεις σταθμίσεων
Υπόδειγμα ΙΙ: Χωρικό Καθεστώς
Συντελεστής β_{1i}



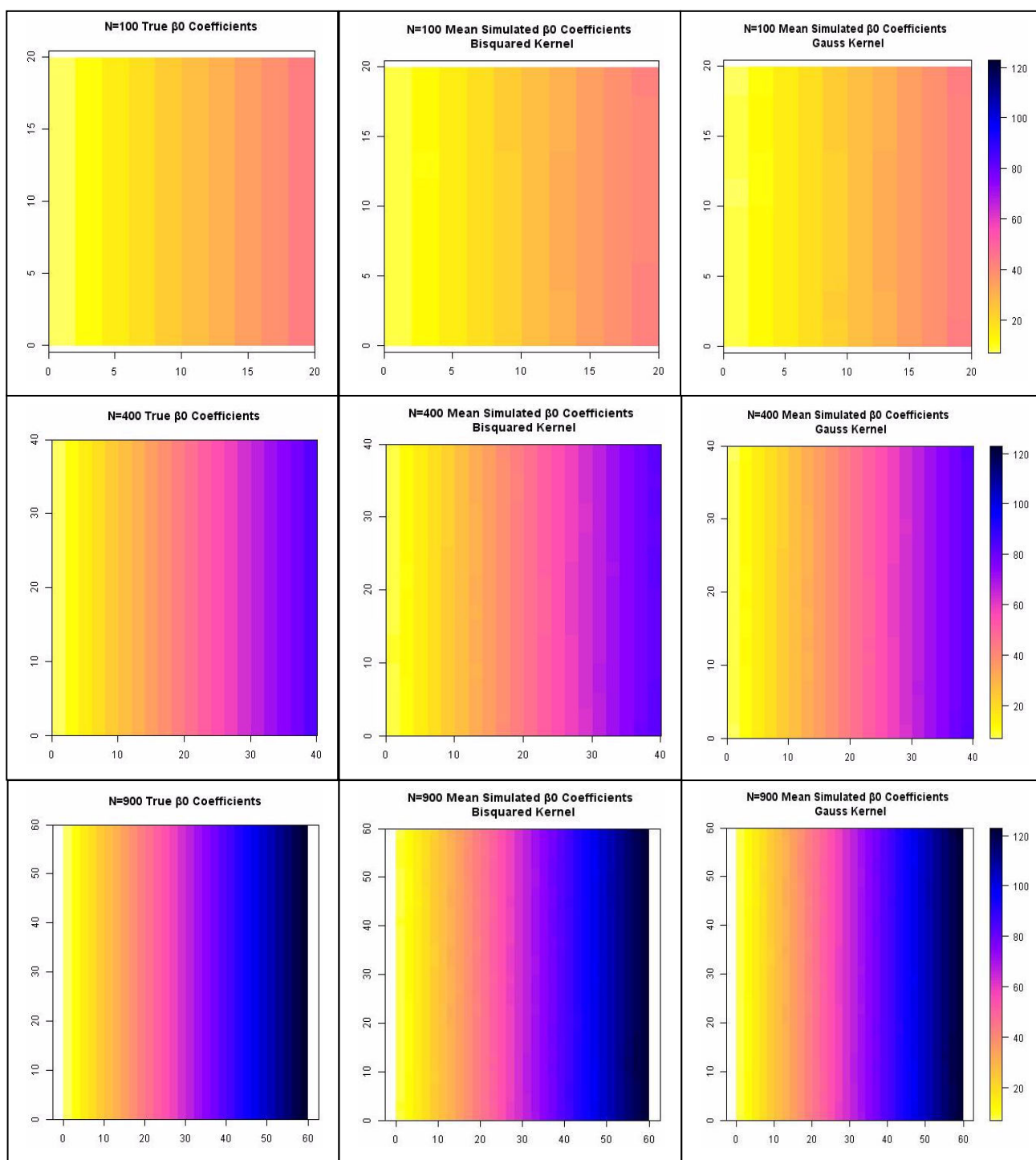
Διάγραμμα 3.3 (B)

**Ψηφιδωτά πραγματικών και προσομοιωμένων μέσων τοπικών εκτιμηθέντων τιμών από την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης για όλα τα μεγέθη δείγματος και τις δύο συναρτήσεις σταθμίσεων
Υπόδειγμα II: Χωρικό Καθεστώς
Συντελεστής β_{i0}**



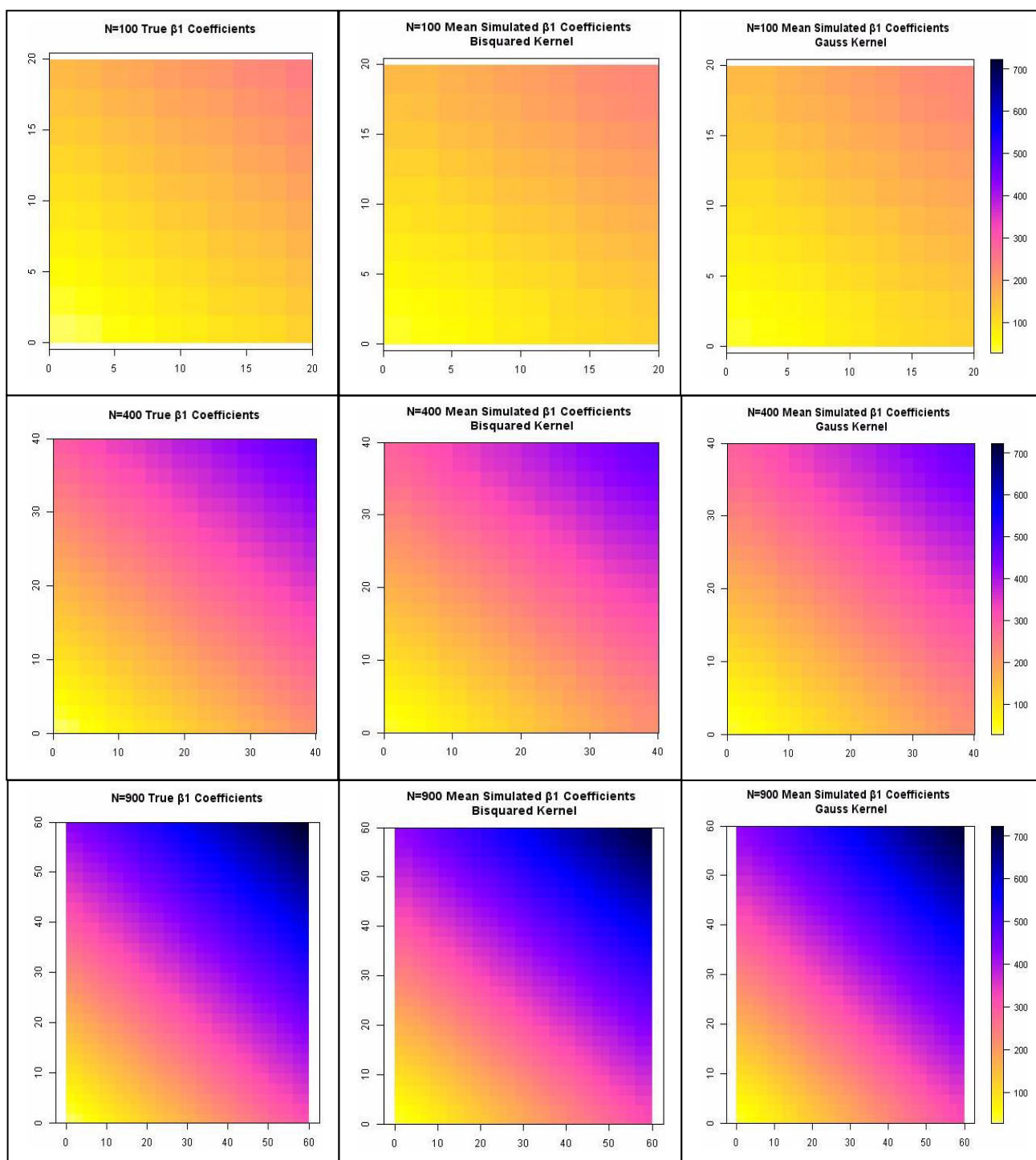
Διάγραμμα 3.4 (Α)

**Ψηφιδωτά πραγματικών και προσομοιωμένων μέσων τοπικών εκτιμηθέντων τιμών από την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης για όλα τα μεγέθη δείγματος και τις δύο συναρτήσεις σταθμίσεων
Υπόδειγμα ΙΙΙ: Χαμηλή Χωρική Ετερογένεια
Συντελεστής β_{11}**



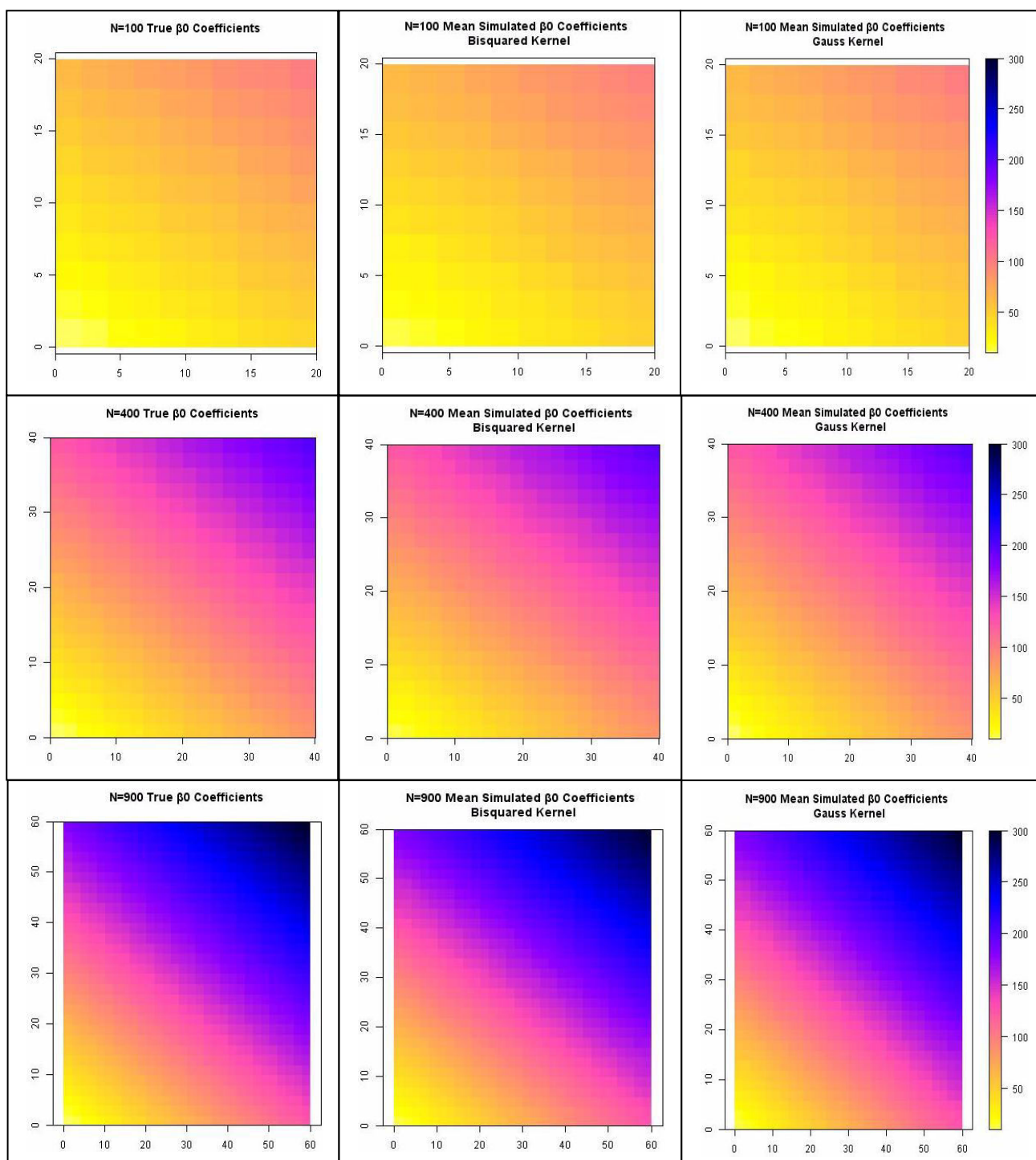
Διάγραμμα 3.4 (B)

Ψηφιδωτά πραγματικών και προσομοιωμένων μέσων τοπικών εκτιμηθέντων τιμών από την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης για όλα τα μεγέθη δείγματος και τις δύο συναρτήσεις σταθμίσεων
Υπόδειγμα III: Χαμηλή Χωρική Ετερογένεια
Συντελεστής β_{i0}



Διάγραμμα 3.5 (Α)

Ψηφιδωτά πραγματικών και προσομοιωμένων μέσων τοπικών εκτιμηθέντων τιμών από την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης για όλα τα μεγέθη δείγματος και τις δύο συναρτήσεις σταθμίσεων
Υπόδειγμα IV: Υψηλή Χωρική Ετερογένεια
Συντελεστής β_{11}



Διάγραμμα 3.5 (B)

Ψηφιδωτά πραγματικών και προσομοιωμένων μέσων τοπικών εκτιμηθέντων τιμών από την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης για όλα τα μεγέθη δείγματος και τις δύο συναρτήσεις σταθμίσεων
Υπόδειγμα IV: Υψηλή Χωρική Ετερογένεια
Συντελεστής β_{i0}

Η βασική διαπίστωση από τη μελέτη των προηγούμενων διαγραμμάτων είναι ότι η γεωγραφικά σταθμισμένη παλινδρόμηση ανταποκρίνεται ικανοποιητικά στον εντοπισμό των χωρικών προτύπων και στην απόδοση των σωστών τιμών στους συντελεστές κάθε υποδείγματος. Ακόμη και για την περίπτωση που στην πραγματικότητα υπάρχει ολική σχέση μεταξύ των δύο εξεταζόμενων τιμών, οι συντελεστές που εκτιμώνται από την τεχνική για τα τοπικά υποδείγματα είναι πολύ κοντά στις πραγματικές ολικές τιμές τους, όπως φαίνεται από τα ψηφιδωτά στο Διάγραμμα 3.2. Ανάλογο συμπέρασμα προκύπτει, σύμφωνα με το Διάγραμμα 3.3, και για το υπόδειγμα χωρικού καθεστώτος όπου οι δύο διαφορετικές ολικές τιμές για τον σταθερό όρο και τον συντελεστή της ανεξάρτητης μεταβλητής εντοπίζονται από τη γεωγραφικά σταθμισμένη παλινδρόμηση. Για τα δύο τοπικά υποδείγματα φαίνεται ότι τα χωρικά πρότυπα ανιχνεύονται από τη μέθοδο καθώς τα ψηφιδωτά στις μεσαίες και δεξιές στήλες στα Διαγράμματα 3.4 και 3.5 αντικατοπτρίζουν σε μεγάλο βαθμό τα ψηφιδωτά με τις πραγματικές τιμές στις πρώτες στήλες των ίδιων διαγραμμάτων. Τα συμπεράσματα αυτά ισχύουν τόσο στα μικρά όσο και στα μεγάλα δείγματα και δεν επηρεάζονται από τον τύπο της συνάρτησης στάθμισης. Αξίζει να αναφερθεί ότι από την προσομοίωση υπολογίστηκαν ψευδό-*t* στατιστικές διαιρώντας κάθε τοπικό συντελεστή με το τυπικό του σφάλμα και διενεργήθηκαν έλεγχοι στατιστικής σημαντικότητας. Επειδή, σε όλες τις περιπτώσεις, τα ποσοστά απόρριψης αυτών των ελέγχων προσεγγίζουν το 100% δεν παρουσιάζονται αναλυτικά.

Συνοψίζοντας τα αποτελέσματα της προσομοίωσης, διαπιστώνεται ότι η σύγκριση των συντελεστών προσδιορισμού της απλής μεθόδου των ελαχίστων τετραγώνων και της γεωγραφικά σταθμισμένης παλινδρόμησης καθώς και του ολικού εύρους με το εύρος που έχει υπολογιστεί από τη διασταυρωτική επικύρωση για τα τοπικά υποδείγματα, συμβάλλουν αρκετές φορές στην επιλογή της σωστής μεθόδου εκτίμησης. Εξίσου, ο ερευνητής μπορεί να λαμβάνει υπ' όψιν τα αποτελέσματα από την εφαρμογή των ελέγχων συνεισφοράς της γεωγραφικά σταθμισμένης παλινδρόμησης μολονότι δεν έχουν όλοι την αναμενόμενη συμπεριφορά ως προς το μέγεθος σύμφωνα με τη στατιστική θεωρία. Ωστόσο, σε κάποιες περιπτώσεις τα κριτήρια μπορεί εσφαλμένα να προτείνουν τη γεωγραφικά σταθμισμένη παλινδρόμηση, όπως συμβαίνει με το υπόδειγμα χωρικού καθεστώτος. Ερωτήματα εγείρονται για την εμφάνιση χωρικής αυτοσυσχέτισης στα σφάλματα των τοπικών υποδειγμάτων και για τον τρόπο που θα πρέπει αυτή να διαπιστωθεί και στατιστικά να ελεγχθεί εάν πραγματικά υπάρχει. Όσον αφορά τους συντελεστές των υποδειγμάτων, διαπιστώνεται ότι η μέθοδος της

γεωγραφικά σταθμισμένης παλινδρόμησης επιτυγχάνει ακριβείς εκτιμήσεις για τις τιμές τους ανεξάρτητα από το εάν η πραγματική σχέση εξάρτησης μεταξύ των δύο μεταβλητών είναι σε ολικό ή σε τοπικό επίπεδο.

Είναι αυτονόητο ότι τα συμπεράσματα της προσομοίωσης αφορούν αποκλειστικά τα υποδείγματα που χρησιμοποιήθηκαν και δεν σημαίνει ότι θα ισχύουν απαραίτητα και σε άλλες μορφές υποδειγμάτων. Αυτό τονίζεται διότι σε αντίθεση με τις εργασίες που εξετάζουν τη χωρική και τη χρονική αυτοσυσχέτιση, όπου υπάρχουν καθιερωμένα θεωρητικά υποδείγματα που χρησιμοποιούνται στις προσομοιώσεις, στις εργασίες για τη γεωγραφικά σταθμισμένη παλινδρόμηση υπάρχει μεγάλη ποικιλία των τρόπων με τους οποίους συμπεριλαμβάνεται η χωρική ετερογένεια στα υποδείγματα. Η ανομοιομορφία αυτή οφείλεται στο γεγονός πως η χωρική ετερογένεια είναι πιο σύνθετη έννοια από τη χωρική αυτοσυσχέτιση και θα μπορούσε να επιδρά στους συντελεστές ενός υποδείγματος με ποικίλους τρόπους μέσω συναρτήσεων των γεωγραφικών συντεταγμένων. Τελειώνοντας, επισημαίνεται ότι τα υποδείγματα που χρησιμοποιήθηκαν στις προσομοιώσεις περιέχουν μόνο μία ανεξάρτητη μεταβλητή. Ενδεχομένως σε υποδείγματα με περισσότερες ανεξάρτητες μεταβλητές τα αποτελέσματα να επηρεάζονται λόγω εμφάνισης τοπικής πολυσυγγραμμικότητας, όπως έχουν δείξει οι Wheeler και Tiefelsdorf (2005).

3.6 Η συμπεριφορά της γεωγραφικά σταθμισμένης παλινδρόμησης όταν υπάρχουν επιδράσεις χωρικής εξάρτησης

Η γεωγραφικά σταθμισμένη παλινδρόμηση αποτελεί την τεχνική που χρησιμοποιείται τις περισσότερες φορές για την αντιμετώπιση της ύπαρξης χωρικής ετερογένειας στη σχέση εξάρτησης των μεταβλητών ενός υποδείγματος παλινδρόμησης. Ωστόσο, αρκετές φορές η διάκριση της χωρικής ετερογένειας από τη χωρική εξάρτηση είναι δύσκολη διότι οι δύο μορφές επιδράσεων παρουσιάζουν ομοιότητες στον τρόπο με τον οποίο εμφανίζονται σε μια ανάλυση κάτι που οφείλεται στη διαστρωματική φύση των δεδομένων που δεν κάνει καθαρό ποια μορφή χωρικών επιδράσεων δημιουργεί το πρόβλημα. Για τον εντοπισμό της χωρικής ετερογένειας δεν υπάρχει κάποια καθιερωμένη διαδικασία ελέγχων, όπως συμβαίνει με τη χωρική εξάρτηση, αλλά συνήθως χρησιμοποιούνται οι κλασσικοί έλεγχοι ετεροσκεδαστικότητας που ελέγχουν τη σταθερότητα της διακύμανση των σφαλμάτων σε ένα υπόδειγμα. Πολλές φορές, οι έλεγχοι αυτοί επηρεάζονται από την παρουσία χωρικής

αυτοσυσχέτισης στα σφάλματα και εσφαλμένα εντοπίζουν το πρόβλημα της ετεροσκεδαστικότητας. Κατά συνέπεια, ο ερευνητής μπορεί να εκλάβει σε ένα υπόδειγμα λανθασμένα τις επιδράσεις χωρικής εξάρτησης ως επιδράσεις χωρικής ετερογένειας και να προχωρήσει στη αντιμετώπισή τους εφαρμόζοντας γεωγραφικά σταθμισμένη παλινδρόμηση. Εκτός της εφαρμογής της τεχνικής λόγω εσφαλμένης αντίληψης του τύπου των επιδράσεων, οι Brunsdon, Fotheringham, και Charlton (1999) αναφέρουν πως η μέθοδος μπορεί να αντιμετωπίσει αποτελεσματικά και το πρόβλημα της χωρικής αυτοσυσχέτισης. Στην ενότητα αυτή εξετάζονται με τη βοήθεια προσομοίωσης οι συνέπειες της εκτίμησης, με τη γεωγραφικά σταθμισμένη παλινδρόμηση, τοπικών υποδειγμάτων στην περίπτωση που υπάρχει ολική σχέση εξάρτησης μεταξύ δύο μεταβλητών που περιλαμβάνει επιδράσεις χωρικής εξάρτησης όπως εμφανίζονται με το υπόδειγμα χωρικού σφάλματος (*Spatial Error Model – SEM*) και με το υπόδειγμα χωρικής υστέρησης (*Spatial Lag Model – SLM*).

Όπως στις προσομοιώσεις στο Κεφάλαιο 2, το υπόδειγμα χωρικού σφάλματος παράγεται από τη σχέση:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \lambda\mathbf{W})^{-1} \mathbf{u}$$

και το υπόδειγμα χωρικής υστέρησης από τη σχέση:

$$\mathbf{y} = (\mathbf{I} - \rho\mathbf{W})^{-1} \mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \rho\mathbf{W})^{-1} \boldsymbol{\varepsilon}$$

όπου \mathbf{y} είναι το $(n \times 1)$ διάνυσμα της εξαρτημένης μεταβλητής y , \mathbf{X} είναι μία $(n \times 2)$ μήτρα που αποτελείται από μία ανεξάρτητη μεταβλητή x_1 και μία στήλη με μονάδες για την εκτίμηση του σταθερού όρου, $\boldsymbol{\beta}$ το (2×1) διάνυσμα των δύο συντελεστών που θεωρούνται σταθεροί για όλες τις περιοχές λαμβάνοντας τις τιμές $\beta_0 = 5$ και $\beta_1 = 15$ και \mathbf{W} η $(n \times n)$ τυποποιημένη κατά γραμμή μήτρα χωρικών σταθμίσεων που κατασκευάστηκε με το κριτήριο του πύργου επάνω στο ομαλό τετραγωνικό πλέγμα. Η ανεξάρτητη μεταβλητή προέρχεται από ομοιόμορφη $U(0, 10)$ κατανομή και τα $(n \times 1)$ διανύσματα των σφαλμάτων \mathbf{u} και $\boldsymbol{\varepsilon}$ από $N(\mathbf{0}, \mathbf{I})$ κατανομή. Χρησιμοποιήθηκαν δείγματα 100, 400 και 900 παρατηρήσεων και για τους συντελεστές λ και ρ θεωρήθηκαν οι τιμές 0,2, 0,5, 0,8 και 0,9.¹³ Εκτελέστηκαν 10.000 επαναλήψεις.

¹³ Η τιμή 0,0 για τα λ και ρ δεν περιλαμβάνεται στην προσομοίωση διότι αντιστοιχεί στο ολικό υπόδειγμα \mathbf{I} που παρουσιάστηκε στην προηγούμενη ενότητα.

Πίνακας 3.4

Συντελεστές προσδιορισμού μεθόδων ελαχίστων τετραγώνων και γεωγραφικά σταθμισμένης παλινδρόμησης και ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% των ελέγχων βελτίωσης για όταν ισχύει το υπόδειγμα χωρικού σφάλματος για όλα τα μεγέθη δείγματος

| Πυρήνας | | Διτετράγωνος | | | | Γκαουσιανός | | | |
|-------------------|----------------------|--------------|-------|-------|-------|-------------|-------|-------|-------|
| <i>N</i> | λ | 0,2 | 0,5 | 0,8 | 0,9 | 0,2 | 0,5 | 0,8 | 0,9 |
| 100 | R ² – OLS | 0,999 | 0,999 | 0,999 | 0,998 | 0,999 | 0,999 | 0,999 | 0,998 |
| | R ² – GWR | 1,000 | 1,000 | 1,000 | 1,000 | 0,999 | 1,000 | 1,000 | 1,000 |
| Ποσοστά Απόρριψης | ANOVA – BCM | 44,48 | 83,14 | 99,55 | 99,97 | 33,98 | 69,8 | 96,21 | 99,17 |
| | ANOVA – FBC | 16,30 | 65,86 | 98,97 | 99,94 | 10,29 | 54,3 | 95,64 | 99,04 |
| | F – RATIO | 9,02 | 54,39 | 98,13 | 99,93 | 8,47 | 53,03 | 95,7 | 99,13 |
| | F1 | 0,50 | 20,87 | 94,81 | 99,72 | 0,42 | 20,22 | 93,05 | 98,96 |
| | F2 | 12,85 | 50,41 | 92,24 | 95,44 | 8,04 | 29,6 | 63,48 | 76,26 |
| 400 | R ² – OLS | 0,999 | 0,999 | 0,999 | 0,998 | 0,999 | 0,999 | 0,999 | 0,998 |
| | R ² – GWR | 0,999 | 1,000 | 1,000 | 1,000 | 0,999 | 1,000 | 1,000 | 1,000 |
| Ποσοστά Απόρριψης | ANOVA – BCM | 58,46 | 90,30 | 99,87 | 99,98 | 48,09 | 77,21 | 97,95 | 99,74 |
| | ANOVA – FBC | 20,37 | 77,77 | 99,56 | 99,98 | 15,52 | 69,58 | 97,77 | 99,72 |
| | F – RATIO | 4,07 | 65,17 | 99,41 | 99,97 | 5,37 | 66,08 | 97,77 | 99,72 |
| | F1 | 0,09 | 50,39 | 99,41 | 99,97 | 0,14 | 54,13 | 97,77 | 99,72 |
| | F2 | 25,07 | 79,98 | 99,61 | 99,98 | 21,38 | 70,87 | 97,73 | 99,71 |
| 900 | R ² – OLS | 0,999 | 0,999 | 0,999 | 0,998 | 0,999 | 0,999 | 0,999 | 0,998 |
| | R ² – GWR | 0,999 | 1,000 | 1,000 | 1,000 | 0,999 | 1,000 | 1,000 | 1,000 |
| Ποσοστά Απόρριψης | ANOVA – BCM | 61,61 | 91,8 | 99,95 | 100 | 51,09 | 79,15 | 98,19 | 99,8 |
| | ANOVA – FBC | 22,10 | 79,26 | 99,75 | 99,98 | 17,23 | 71,77 | 98,02 | 99,79 |
| | F – RATIO | 2,07 | 68,85 | 99,50 | 99,97 | 3,67 | 70,22 | 98,02 | 99,79 |
| | F1 | 0,04 | 66,08 | 99,50 | 99,97 | 0,07 | 69,13 | 98,02 | 99,79 |
| | F2 | 27,53 | 81,41 | 99,80 | 99,99 | 23,19 | 72,68 | 98,02 | 99,79 |

Πίνακας 3.5

Μέσες, ελάχιστες και μέγιστες τιμές εύρους για όταν ισχύει το υπόδειγμα χωρικού σφάλματος

| Πυρήνας | | Διτετράγωνος | | | | Γκαουσιανός | | | |
|----------|----------------|--------------|--------|--------|--------|-------------|--------|--------|--------|
| <i>N</i> | λ | 0,2 | 0,5 | 0,8 | 0,9 | 0,2 | 0,5 | 0,8 | 0,9 |
| 100 | Ολικό Εύρος | 25,456 | | | | 25,456 | | | |
| | Μέσο Εύρος | 18,796 | 11,062 | 5,410 | 4,946 | 15,943 | 8,675 | 2,693 | 1,902 |
| | Ελάχιστο Εύρος | 3,253 | 3,152 | 2,111 | 2,260 | 1,095 | 0,793 | 0,830 | 0,852 |
| | Μέγιστο Εύρος | 25,456 | 25,456 | 25,456 | 25,456 | 25,456 | 25,456 | 25,456 | 25,456 |
| 400 | Ολικό Εύρος | 53,740 | | | | 53,740 | | | |
| | Μέσο Εύρος | 38,784 | 16,637 | 5,260 | 4,914 | 33,034 | 15,209 | 2,830 | 1,780 |
| | Ελάχιστο Εύρος | 4,972 | 3,858 | 3,390 | 3,386 | 1,774 | 1,485 | 1,191 | 1,165 |
| | Μέγιστο Εύρος | 53,740 | 53,740 | 53,740 | 34,780 | 53,740 | 53,740 | 53,740 | 53,740 |
| 900 | Ολικό Εύρος | 82,024 | | | | 82,024 | | | |
| | Μέσο Εύρος | 58,276 | 21,458 | 5,261 | 4,906 | 50,078 | 21,072 | 3,249 | 1,790 |
| | Ελάχιστο Εύρος | 7,084 | 4,887 | 3,529 | 3,630 | 2,661 | 1,472 | 1,284 | 1,284 |
| | Μέγιστο Εύρος | 82,024 | 82,024 | 82,024 | 57,157 | 82,024 | 82,024 | 82,024 | 82,024 |

Ο Πίνακας 3.4 περιέχει τις μέσες τιμές των συντελεστών προσδιορισμού από τις εκτιμήσεις με τις μεθόδους των ελαχίστων τετραγώνων και τη γεωγραφικά σταθμισμένη παλινδρόμηση του γραμμικού υποδείγματος που έχει εξαρτημένη μεταβλητή την y και ανεξάρτητη τη μεταβλητή x_1 όταν στην πραγματικότητα ισχύει το υπόδειγμα χωρικού σφάλματος καθώς και τα ποσοστά απόρριψης για επίπεδο σημαντικότητας 5% των ελέγχων που εξετάζουν τη συνεισφορά της δεύτερης τεχνικής. Από αυτόν τον πίνακα προκύπτουν δύο βασικά συμπεράσματα. Το πρώτο συμπέρασμα είναι ότι οι μέσοι συντελεστές προσδιορισμού των δύο μεθόδων εκτίμησης είναι ίδιοι ανεξάρτητα από την τιμή του λ , το μέγεθος του δείγματος και τη συνάρτηση στάθμισης. Το αποτέλεσμα αυτό δείχνει ότι η γεωγραφικά σταθμισμένη παλινδρόμηση δεν συνεισφέρει ουσιαστικά στην αύξηση της ερμηνευτικής ικανότητας του υποδείγματος. Το δεύτερο συμπέρασμα αφορά τα ποσοστά απόρριψης των πέντε ελέγχων τα οποία αυξάνουν καθώς αυξάνονται οι τιμές του λ . Ειδικότερα, για τιμές του $\lambda = 0,2$ και $0,5$ τα ποσοστά απόρριψης κυμαίνονται σε μέτρια επίπεδα που είναι υψηλότερα στα μεγαλύτερα δείγματα. Για τιμές του $\lambda = 0,8$ και $0,9$ τα ποσοστά απόρριψης είναι ιδιαίτερα υψηλά προσεγγίζοντας ουσιαστικά το 100 %. Αυτό σημαίνει ότι όταν ο ερευνητής εφαρμόζει γεωγραφικά σταθμισμένη παλινδρόμηση, ενώ στην πραγματικότητα υπάρχει ολική σχέση εξάρτησης των δύο μεταβλητών με επιδράσεις χωρικής εξάρτησης που αντιστοιχούν σε ένα υπόδειγμα χωρικού σφάλματος, τότε οι έλεγχοι, για μεγάλες τιμές του λ , θα δείχνουν πάντα ότι υπάρχει βελτίωση στα αποτελέσματα και θα θεωρούν καλύτερα τα τοπικά υποδείγματα.

Στον Πίνακα 3.5 περιλαμβάνονται τα αποτελέσματα που αφορούν το εύρος που προέκυψε από τη διαδικασία της διασταυρωτικής επικύρωσης. Είναι φανερό ότι καθώς αυξάνεται η τιμή του λ , δηλαδή όταν η χωρική αυτοσυσχέτιση στα σφάλματα του υποδείγματος χωρικού σφάλματος γίνεται εντονότερη, το μέσο εύρος μειώνεται σημαντικά και είναι χαμηλότερο από το ολικό εύρος. Η εξήγηση που μπορεί να δοθεί για αυτή τη συμπεριφορά είναι ότι όταν υπάρχει μεγάλη χωρική αυτοσυσχέτιση στα σφάλματα του υποδείγματος εμφανίζεται ως χωρική ετερογένεια επηρεάζοντας τη διαδικασία της διασταυρωτικής επικύρωσης που επιλέγει μικρό εύρος. Το γεγονός αυτό συνεπάγεται την εκτίμηση σε κάθε σημείο των συντελεστών του υποδείγματος με μικρό αριθμό παρατηρήσεων.

Οι μέσες τιμές της στατιστικής I του Moran και τα ποσοστά απόρριψης του ελέγχου χωρικής αυτοσυσχέτισης για τα εκτιμηθέντα σφάλματα περιέχονται στον Πίνακα 3.6. Ο έλεγχος χωρικής αυτοσυσχέτισης έχει πραγματοποιηθεί με την τεχνική των Leung, Mei και Zhang (2000b) αλλά και με τον κλασικό τρόπο υποθέτοντας τυχαιοποίηση. Όπως

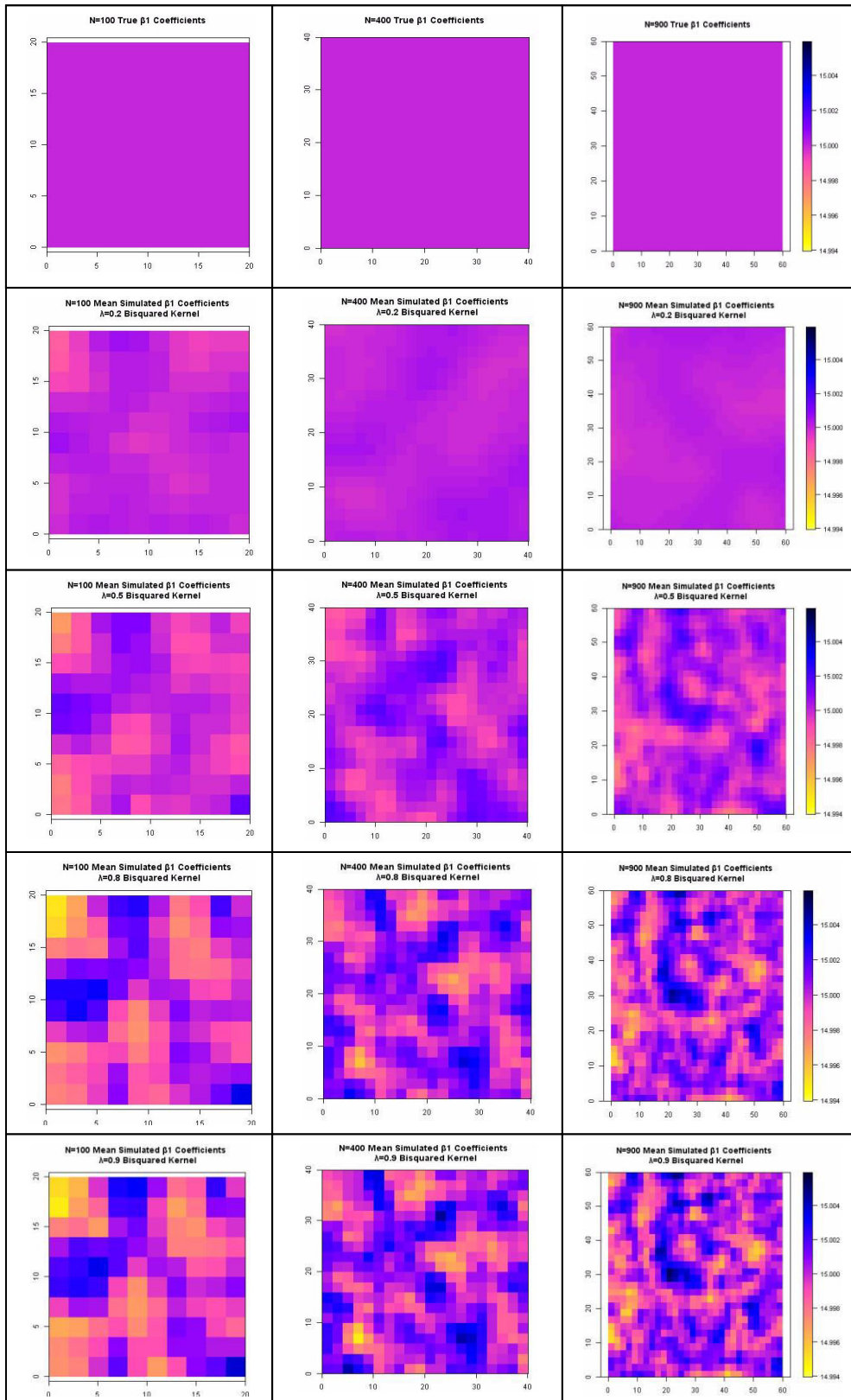
διαπιστώνεται, η στατιστική I παρουσιάζει ιδιόμορφη συμπεριφορά. Συγκεκριμένα, οι μέσες τιμές της I είναι χαμηλές αλλά όχι μηδενικές και δεν ακολουθούν κάποια σταθερή πορεία σε σχέση με τις τιμές του λ . Παρά τις μικρές τιμές που έχουν αυτές οι στατιστικές, οι έλεγχοι για την ύπαρξη χωρικής αυτοσυσχέτισης τόσο με την προσέγγιση των Leung, Mei και Zhang (2000b) όσο και με την υπόθεση της τυχαιοποίησης απορρίπτουν τη μηδενική τους υπόθεση σε αξιοσημείωτα ποσοστά υποδεικνύοντας την εμφάνιση του προβλήματος. Ιδιαίτερα στα μεγάλα δείγματα των 900 παρατηρήσεων και οι δύο τρόποι εφαρμογής του ελέγχου δείχνουν έντονη παρουσία χωρικής αυτοσυσχέτισης. Ωστόσο, πρέπει να σημειωθεί ότι η ερμηνεία του πίνακα αυτού δεν είναι εύκολη και δημιουργεί ερωτήματα για την εγκυρότητα των ελέγχων και τον τρόπο με τον οποίον πρέπει να πραγματοποιείται ο έλεγχος χωρικής αυτοσυσχέτισης στα σφάλματα της γεωγραφικά σταθμισμένης παλινδρόμησης.

Πίνακας 3.6

Μέσες τιμές στατιστικής I και ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% με την τεχνική των Leung, Mei και Zhang (2000b) και την υπόθεση της τυχαιοποίησης του ελέγχου ύπαρξης χωρικής αυτοσυσχέτισης στα εκτιμηθέντα σφάλματα της μεθόδου της γεωγραφικά σταθμισμένης παλινδρόμησης για όταν ισχύει το υπόδειγμα χωρικού σφάλματος

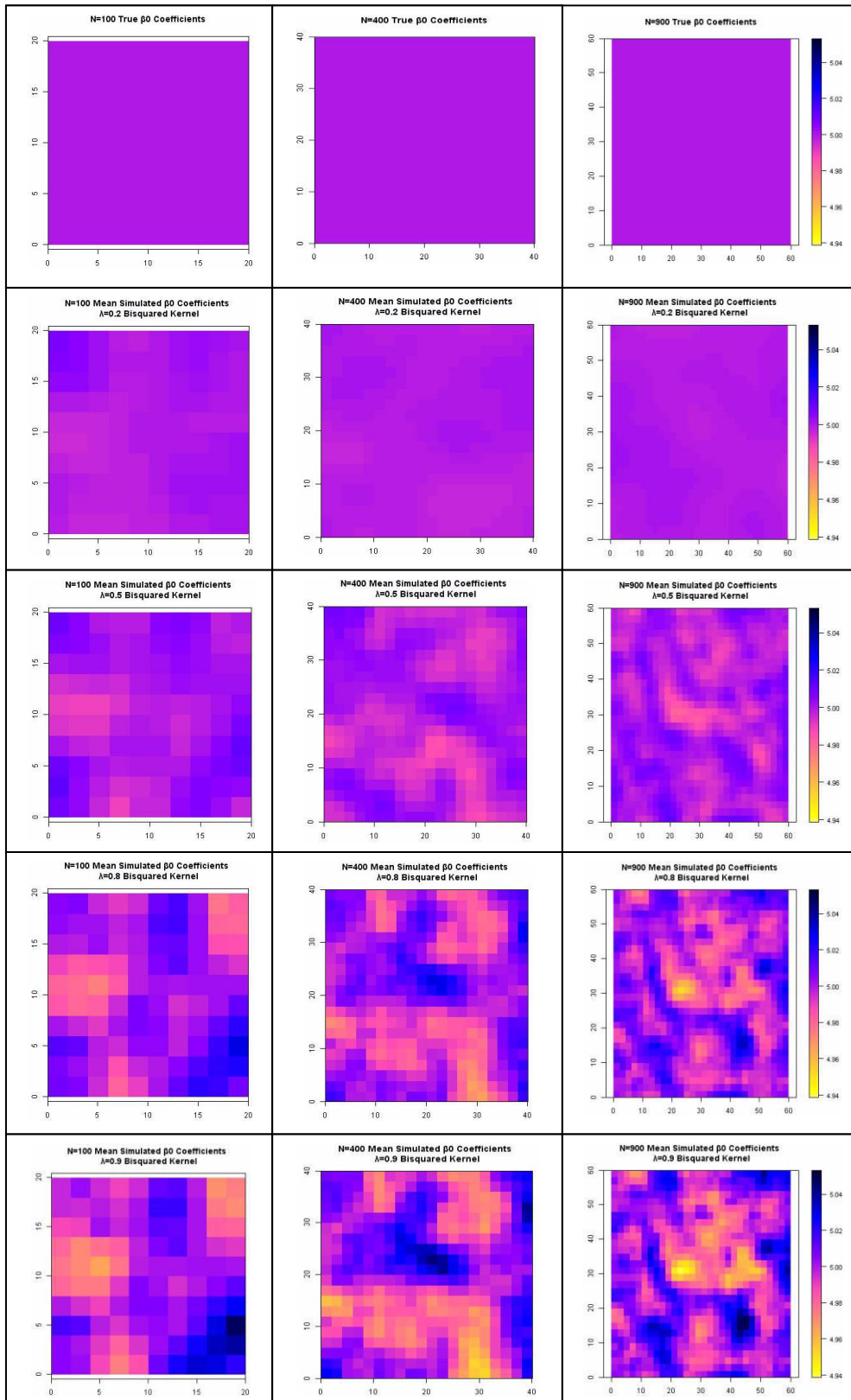
| Πυρήνας | | Διτετράγωνος | | | | Γκαουσιανός | | | |
|---------|-------------------------|--------------|-------|-------|-------|-------------|-------|-------|-------|
| N | λ | 0,2 | 0,5 | 0,8 | 0,9 | 0,2 | 0,5 | 0,8 | 0,9 |
| 100 | Μέση Τιμή I | 0,041 | 0,077 | 0,052 | 0,071 | 0,053 | 0,092 | 0,053 | 0,065 |
| | Ποσ. Απόρριψης LMZ | 33,08 | 91,41 | 99,64 | 99,94 | 34,36 | 92,46 | 99,76 | 99,96 |
| | Ποσ. Απόρ. Τυχαιοποίηση | 10,48 | 30,95 | 11,55 | 17,74 | 14,51 | 36,21 | 10,29 | 12,8 |
| 400 | Μέση Τιμή I | 0,086 | 0,110 | 0,055 | 0,082 | 0,087 | 0,106 | 0,047 | 0,064 |
| | Ποσ. Απόρριψης LMZ | 87,30 | 100 | 100 | 100 | 87,84 | 100 | 100 | 100 |
| | Ποσ. Απόρ. Τυχαιοποίηση | 67,89 | 54,83 | 29,55 | 68,04 | 69,51 | 45,44 | 16,70 | 43,00 |
| 900 | Μέση Τιμή I | 0,095 | 0,106 | 0,056 | 0,083 | 0,095 | 0,099 | 0,045 | 0,063 |
| | Ποσ. Απόρριψης LMZ | 99,56 | 100 | 100 | 100 | 99,53 | 100 | 100 | 100 |
| | Ποσ. Απόρ. Τυχαιοποίηση | 96,63 | 57,11 | 66,15 | 96,31 | 95,8 | 45,36 | 30,92 | 78,26 |

Τα ψηφιδωτά (*raster*) στα Διαγράμματα 3.6 και 3.7 απεικονίζουν τις μέσες εκτιμηθείσες τιμές ανά περιοχή για τους τοπικές συντελεστές β_1 και β_0 που προέκυψαν από τη γεωγραφικά σταθμισμένη παλινδρόμηση με διτετράγωνη και Γκαουσιανή συνάρτηση στάθμισης, αντίστοιχα, για όλες τις τιμές του λ και μεγέθη δείγματος. Καθώς οι πραγματικές τιμές των δύο συντελεστών του υποδείγματος στην προσομοίωση ήταν 15 και 5, δηλαδή σταθερές σε κάθε περιοχή, τα διαγράμματα αυτά δεν θα πρέπει επιδεικνύουν έντονη μεταβλητότητα. Γενικά, το συμπέρασμα αυτό ισχύει ανεξάρτητα από το λ , το μέγεθος δείγματος και τη συνάρτησης στάθμισης. Η διαφοροποίηση που παρατηρείται για τιμές του λ ίσες με 0,8 και 0,9 δεν είναι πραγματική και οφείλεται στη χρωματική κλίμακα, καθώς το εύρος (*range*) των εκτιμηθέντων τιμών των συντελεστών είναι πολύ μικρό.



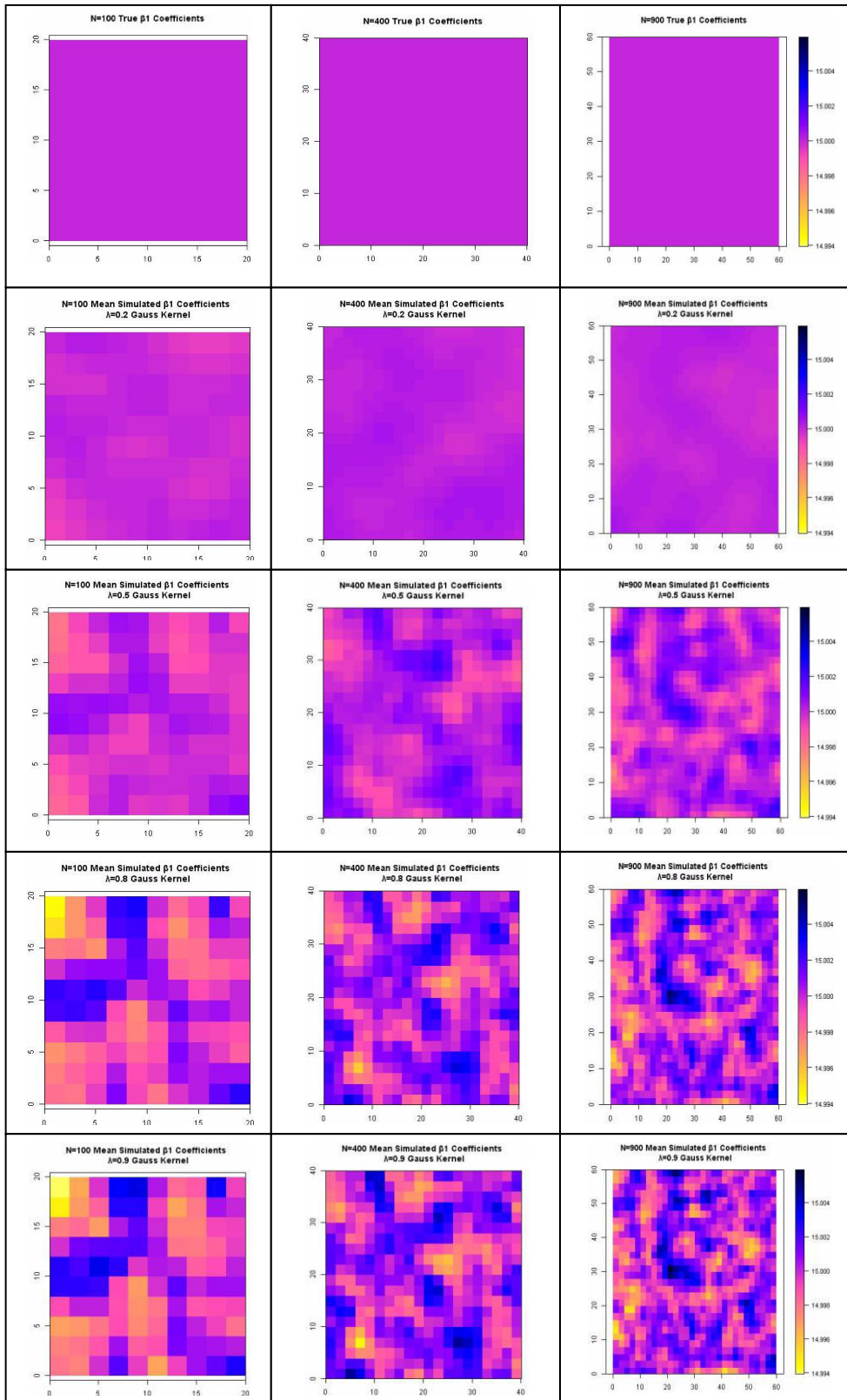
Διάγραμμα 3.6 (Α)

Ψηφιδωτά πραγματικών και προσομοιωμένων μέσων τοπικών εκτιμηθέντων τιμών από την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης στο Υπόδειγμα Χωρικού Σφάλματος με Διτετράγωνη συνάρτηση στάθμισης για όλα τα μεγέθη δείγματος και τιμές του λ Συντελεστής β_{1i}



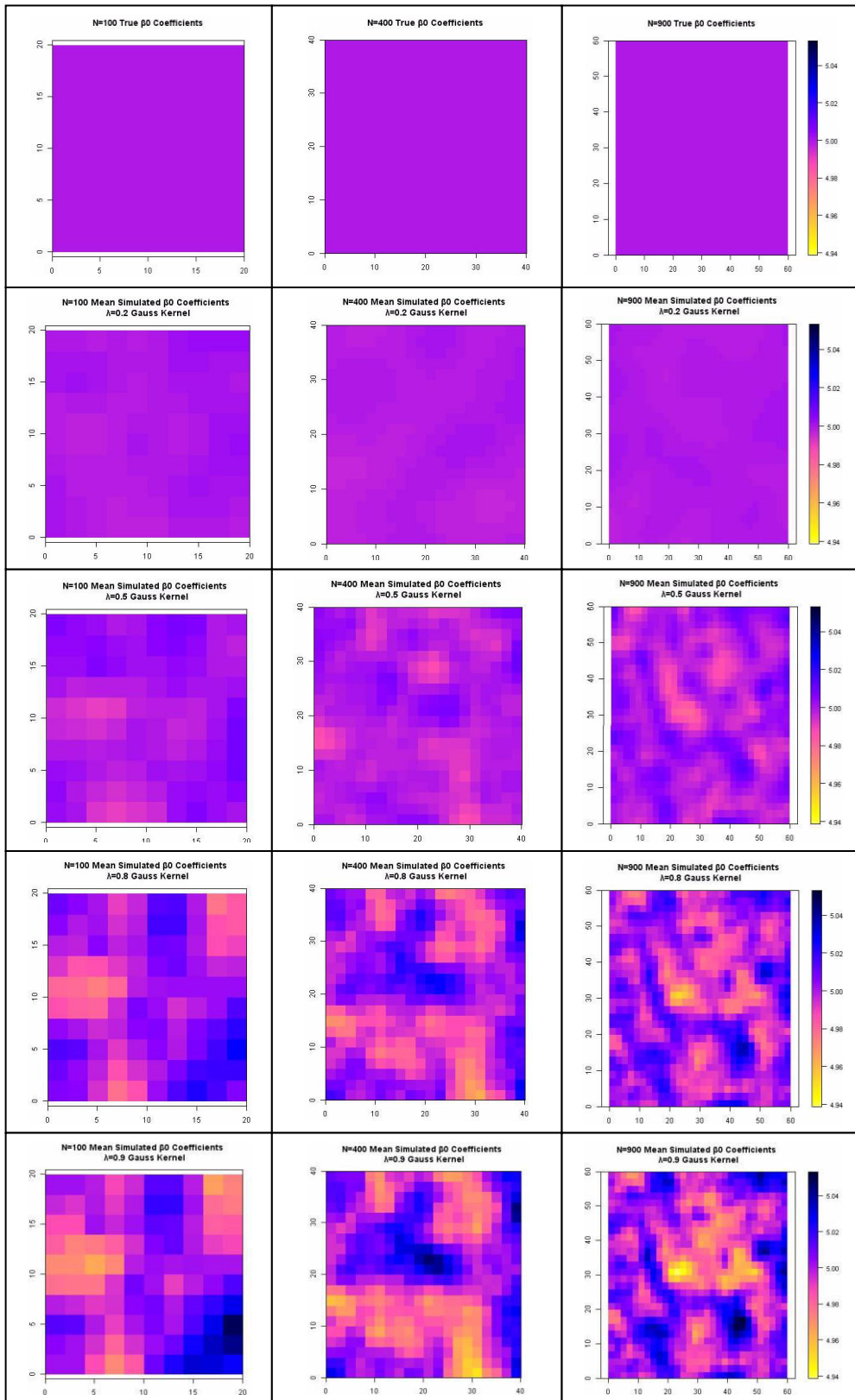
Διάγραμμα 3.6 (B)

Ψηφιδωτά πραγματικών και προσομοιωμένων μέσων τοπικών εκτιμηθέντων τιμών από την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης στο Υπόδειγμα Χωρικού Σφάλματος με Διτετράγωνη συνάρτηση στάθμισης για όλα τα μεγέθη δείγματος και τιμές του λ . Συντελεστής β_{10}



Διάγραμμα 3.7 (Α)

Ψηφιδωτά πραγματικών και προσομοιωμένων μέσω τοπικών εκτιμηθέντων τιμών από την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης στο Υπόδειγμα Χωρικού Σφάλματος με Γκαουσιανή συνάρτηση στάθμισης για όλα τα μεγέθη δείγματος και τιμές του λ Συντελεστής β_{1i}



Διάγραμμα 3.7 (B)

Ψηφιδωτά πραγματικών και προσομοιωμένων μέσων τοπικών εκτιμηθέντων τιμών από την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης στο Υπόδειγμα Χωρικού Σφάλματος με Γκαουσιανή συνάρτηση στάθμισης για όλα τα μεγέθη δείγματος και τιμές του λ Συντελεστής β_{i0}

Η προσομοίωση επαναλήφθηκε με την πραγματική σχέση εξάρτησης των δύο μεταβλητών να είναι το υπόδειγμα χωρικής υστέρησης. Το γραμμικό υπόδειγμα παλινδρόμησης των δύο μεταβλητών, χωρίς τη χωρική υστέρηση της εξαρτημένης μεταβλητής, εκτιμήθηκε με τη μέθοδο των ελαχίστων τετραγώνων και τη γεωγραφικά σταθμισμένη παλινδρόμηση και στον Πίνακα 3.7 συνοψίζονται οι μέσες τιμές των συντελεστών προσδιορισμού από τις δύο μεθόδους και τα ποσοστά απόρριψης των ελέγχων βελτίωσης της εκτίμησης που προέκυψαν από την προσομοίωση. Όπως διαπιστώνεται, οι μέσοι συντελεστές προσδιορισμού της γεωγραφικά σταθμισμένης παλινδρόμησης είναι μεγαλύτεροι από τους αντίστοιχους της μεθόδου των ελαχίστων τετραγώνων σε όλες τις περιπτώσεις. Επιπρόσθετα, για μεγαλύτερες τιμές του αυτοπαλίνδρομου συντελεστή ρ η διαφορά μεταξύ των δύο συντελεστών προσδιορισμού γίνεται εντονότερη υποδεικνύοντας τη γεωγραφικά σταθμισμένη παλινδρόμηση ως καλύτερη μέθοδο εκτίμησης. Το αποτέλεσμα αυτό οφείλεται κυρίως στην πτώση της μέσης τιμής του συντελεστή προσδιορισμού του ολικού υποδείγματος, κάτι αναμενόμενο να συμβαίνει διότι εμφανίζεται μεροληψία λόγω παράληψης της χωρικής υστέρησης από τις ανεξάρτητες μεταβλητές του υποδείγματος, που είναι πιο σοβαρή όταν αυξάνεται η τιμή του ρ . Τα ποσοστά απόρριψης των πέντε ελέγχων βελτίωσης είναι πάρα πολύ υψηλά και προσεγγίζουν το 100% ακόμη και για μικρές τιμές του ρ ανεξάρτητα του μεγέθους του δείγματος και τη συνάρτηση στάθμισης. Τα αντίστοιχα ποσοστά απόρριψης για το υπόδειγμα χωρικού σφάλματος, όπως παρουσιάστηκε στον Πίνακα 3.4, ήταν μικρότερα ιδιαίτερα στις χαμηλές τιμές του λ . Κατά συνέπεια, όταν στην πραγματικότητα η εξάρτηση των δύο μεταβλητών προέρχεται μέσω του υποδείγματος χωρικής υστέρησης, οι έλεγχοι συνεισφοράς θα δείχνουν πάντα καταλληλότερη της γεωγραφικά σταθμισμένη παλινδρόμηση ανεξάρτητα της τιμής του ρ .

Οι τιμές του εύρους που υπολογίστηκαν μέσω της διασταυρωτικής επικύρωσης περιέχονται στον Πίνακα 3.8. Από αυτόν τον πίνακα φαίνεται ότι οι μέσες τιμές του εύρους για όλες τις τιμές του ρ είναι πολύ χαμηλές και πάντα σημαντικά μικρότερες από το ολικό εύρος με τις χαμηλότερες τιμές του να παρατηρούνται όταν εφαρμόζεται η Γκαουσιανή συνάρτηση στάθμισης. Επομένως, ο ερευνητής θα επιλέγει πάντα ως καταλληλότερη τη γεωγραφικά σταθμισμένη παλινδρόμηση. Αξίζει να αναφερθεί ότι, όπως περιγράφεται στον Πίνακα 3.5, στην περίπτωση του υποδείγματος χωρικού σφάλματος οι τιμές του εύρους είναι πιο υψηλές και μόνο για τιμές του λ ίσες με 0,8 και 0,9 παρατηρούνται ιδιαίτερα χαμηλές τιμές κάτι που δεν ισχύει στο υπόδειγμα χωρικής υστέρησης.

Πίνακας 3.7

Συντελεστές προσδιορισμού μεθόδων ελαχίστων τετραγώνων και γεωγραφικά σταθμισμένης παλινδρόμησης και ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% των ελέγχων βελτίωσης για όταν ισχύει το υπόδειγμα χωρικής υστέρησης για όλα τα μεγέθη δείγματος

| Πυρήνας | | Διτετράγωνος | | | | Γκαουσιανός | | | |
|-------------------|----------------------|--------------|-------|-------|-------|-------------|-------|-------|-------|
| <i>N</i> | ρ | 0,2 | 0,5 | 0,8 | 0,9 | 0,2 | 0,5 | 0,8 | 0,9 |
| 100 | R ² – OLS | 0,988 | 0,919 | 0,708 | 0,545 | 0,988 | 0,919 | 0,708 | 0,545 |
| | R ² – GWR | 0,996 | 0,980 | 0,965 | 0,967 | 0,995 | 0,981 | 0,972 | 0,975 |
| Ποσοστά Απόρριψης | ANOVA – BCM | 99,44 | 99,99 | 100 | 100 | 94,23 | 98,36 | 99,73 | 99,91 |
| | ANOVA – FBC | 98,81 | 99,97 | 100 | 100 | 93,38 | 98,36 | 99,73 | 99,91 |
| | F – RATIO | 98,36 | 99,93 | 100 | 100 | 93,73 | 98,35 | 99,73 | 99,91 |
| | F1 | 88,94 | 99,74 | 100 | 100 | 79,85 | 97,93 | 99,73 | 99,91 |
| | F2 | 96,70 | 99,93 | 99,93 | 99,92 | 83,74 | 95,22 | 93,99 | 92,56 |
| 400 | R ² – OLS | 0,989 | 0,920 | 0,703 | 0,525 | 0,989 | 0,920 | 0,703 | 0,525 |
| | R ² – GWR | 0,996 | 0,980 | 0,966 | 0,969 | 0,995 | 0,981 | 0,972 | 0,976 |
| Ποσοστά Απόρριψης | ANOVA – BCM | 99,40 | 99,95 | 100 | 100 | 94,55 | 98,50 | 99,89 | 99,97 |
| | ANOVA – FBC | 98,40 | 99,84 | 100 | 100 | 94,09 | 98,42 | 99,88 | 99,97 |
| | F – RATIO | 97,83 | 99,76 | 99,99 | 100 | 94,08 | 98,42 | 99,88 | 99,97 |
| | F1 | 97,83 | 99,76 | 99,99 | 100 | 94,08 | 98,42 | 99,88 | 99,97 |
| | F2 | 98,64 | 99,85 | 100 | 100 | 94,11 | 98,42 | 99,88 | 99,97 |
| 900 | R ² – OLS | 0,989 | 0,921 | 0,705 | 0,524 | 0,989 | 0,921 | 0,705 | 0,524 |
| | R ² – GWR | 0,996 | 0,980 | 0,966 | 0,969 | 0,996 | 0,981 | 0,972 | 0,976 |
| Ποσοστά Απόρριψης | ANOVA – BCM | 99,56 | 99,92 | 99,99 | 100 | 95,40 | 98,61 | 99,88 | 99,99 |
| | ANOVA – FBC | 98,46 | 99,72 | 99,99 | 100 | 94,86 | 98,51 | 99,88 | 99,99 |
| | F – RATIO | 97,71 | 99,6 | 99,97 | 100 | 94,86 | 98,51 | 99,88 | 99,99 |
| | F1 | 97,71 | 99,60 | 99,97 | 100 | 94,86 | 98,51 | 99,88 | 99,99 |
| | F2 | 98,70 | 99,81 | 99,99 | 100 | 94,86 | 98,51 | 99,88 | 99,99 |

Πίνακας 3.8

Μέσες, ελάχιστες και μέγιστες τιμές εύρους για όταν ισχύει το υπόδειγμα χωρικής υστέρησης

| Πυρήνας | | Διτετράγωνος | | | | Γκαουσιανός | | | |
|----------|----------------|--------------|--------|--------|-------|-------------|--------|--------|--------|
| <i>N</i> | ρ | 0,2 | 0,5 | 0,8 | 0,9 | 0,2 | 0,5 | 0,8 | 0,9 |
| 100 | Ολικό Εύρος | 25,456 | | | | 25,456 | | | |
| | Μέσο Εύρος | 5,834 | 5,270 | 4,944 | 4,848 | 3,473 | 2,317 | 1,795 | 1,689 |
| | Ελάχιστο Εύρος | 4,025 | 3,551 | 3,244 | 3,248 | 1,377 | 1,182 | 1,058 | 1,006 |
| | Μέγιστο Εύρος | 25,456 | 25,456 | 7,484 | 7,597 | 25,456 | 25,456 | 25,456 | 25,456 |
| 400 | Ολικό Εύρος | 53,740 | | | | 53,740 | | | |
| | Μέσο Εύρος | 6,204 | 5,248 | 4,892 | 4,814 | 4,979 | 2,697 | 1,770 | 1,663 |
| | Ελάχιστο Εύρος | 4,673 | 4,413 | 4,081 | 4,056 | 1,731 | 1,591 | 1,392 | 1,325 |
| | Μέγιστο Εύρος | 53,740 | 53,740 | 29,358 | 6,361 | 53,740 | 53,740 | 53,740 | 53,740 |
| 900 | Ολικό Εύρος | 82,024 | | | | 82,024 | | | |
| | Μέσο Εύρος | 6,592 | 5,353 | 4,882 | 4,798 | 5,972 | 3,040 | 1,792 | 1,645 |
| | Ελάχιστο Εύρος | 4,856 | 4,642 | 4,050 | 4,132 | 1,868 | 1,681 | 1,490 | 1,436 |
| | Μέγιστο Εύρος | 82,024 | 82,024 | 66,460 | 5,398 | 82,024 | 82,024 | 82,024 | 82,024 |

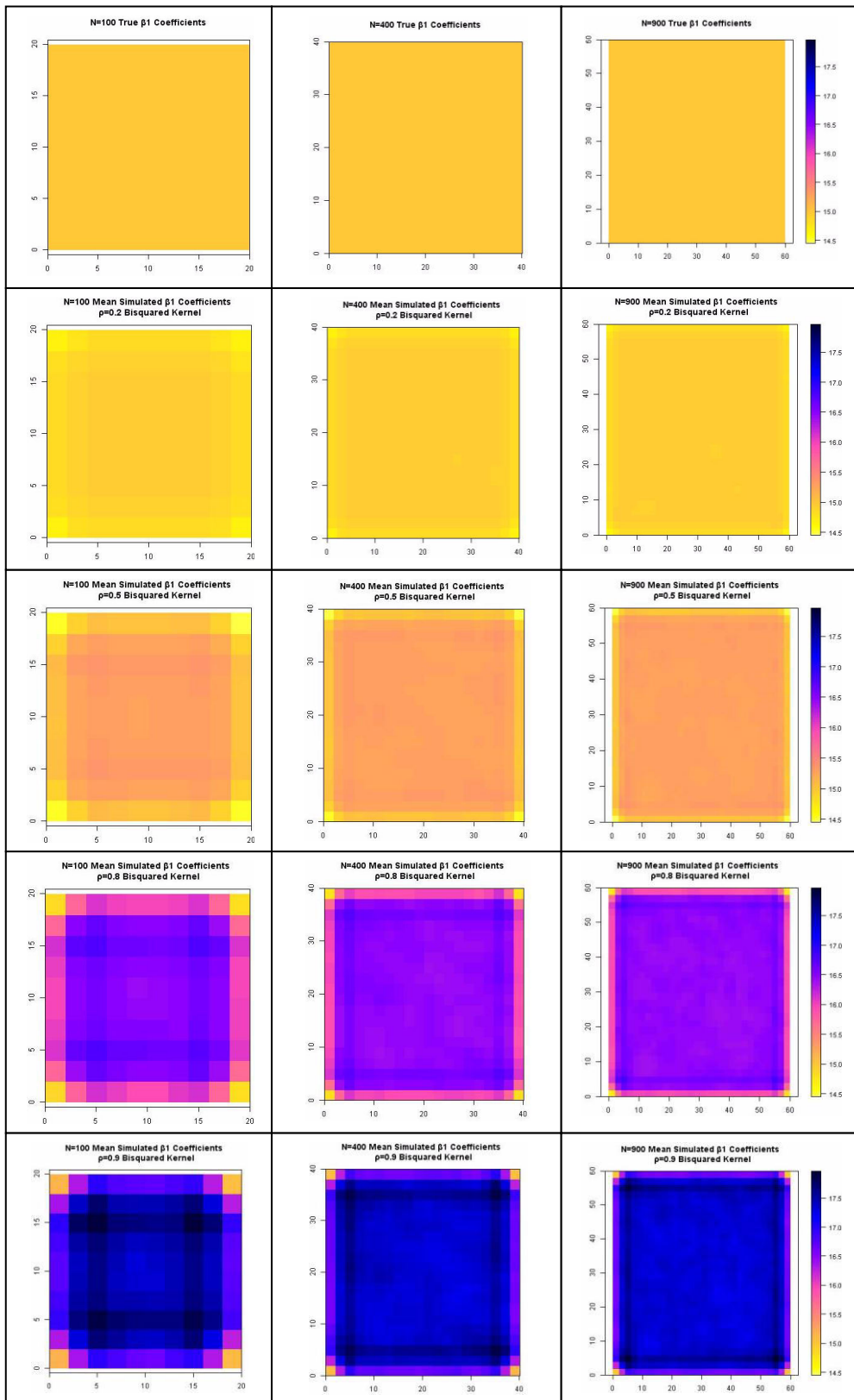
Πίνακας 3.9

Μέσες τιμές στατιστικής I και ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% με την τεχνική των Leung, Mei και Zhang (2000b) και την υπόθεση της τυχαιοποίησης του ελέγχου ύπαρξης χωρικής αυτοσυσχέτισης στα εκτιμηθέντα σφάλματα της μεθόδου της γεωγραφικά σταθμισμένης παλινδρόμησης για όταν ισχύει το υπόδειγμα χωρικής υστέρησης

| Πυρήνας | | Διτετράγωνος | | | | Γκαουσιανός | | | |
|---------|-------------------------|--------------|--------|--------|--------|-------------|--------|--------|--------|
| N | ρ | 0,2 | 0,5 | 0,8 | 0,9 | 0,2 | 0,5 | 0,8 | 0,9 |
| 100 | Μέση Τιμή I | -0,275 | -0,215 | -0,129 | -0,087 | -0,257 | -0,215 | -0,140 | -0,100 |
| | Ποσ. Απόρριψης LMZ | 0,05 | 2,01 | 40,23 | 67,27 | 1,05 | 3,21 | 40,06 | 68,35 |
| | Ποσ. Απόρ. Τυχαιοποίηση | 95,48 | 80,86 | 37,33 | 19,32 | 92,46 | 86,69 | 42,09 | 21,67 |
| 400 | Μέση Τιμή I | -0,265 | -0,206 | -0,121 | -0,079 | -0,249 | -0,205 | -0,130 | -0,090 |
| | Ποσ. Απόρριψης LMZ | 1,66 | 0,96 | 80,21 | 98,66 | 5,15 | 2,23 | 84,51 | 99,3 |
| | Ποσ. Απόρ. Τυχαιοποίηση | 99,14 | 100 | 89,04 | 55,87 | 98,78 | 100 | 96,25 | 68,74 |
| 900 | Μέση Τιμή I | -0,261 | -0,203 | -0,119 | -0,077 | -0,248 | -0,202 | -0,127 | -0,088 |
| | Ποσ. Απόρριψης LMZ | 2,27 | 0,55 | 97,82 | 99,98 | 5,13 | 1,63 | 99,13 | 100 |
| | Ποσ. Απόρ. Τυχαιοποίηση | 99,96 | 100 | 99,72 | 85,78 | 99,98 | 100 | 100 | 96,04 |

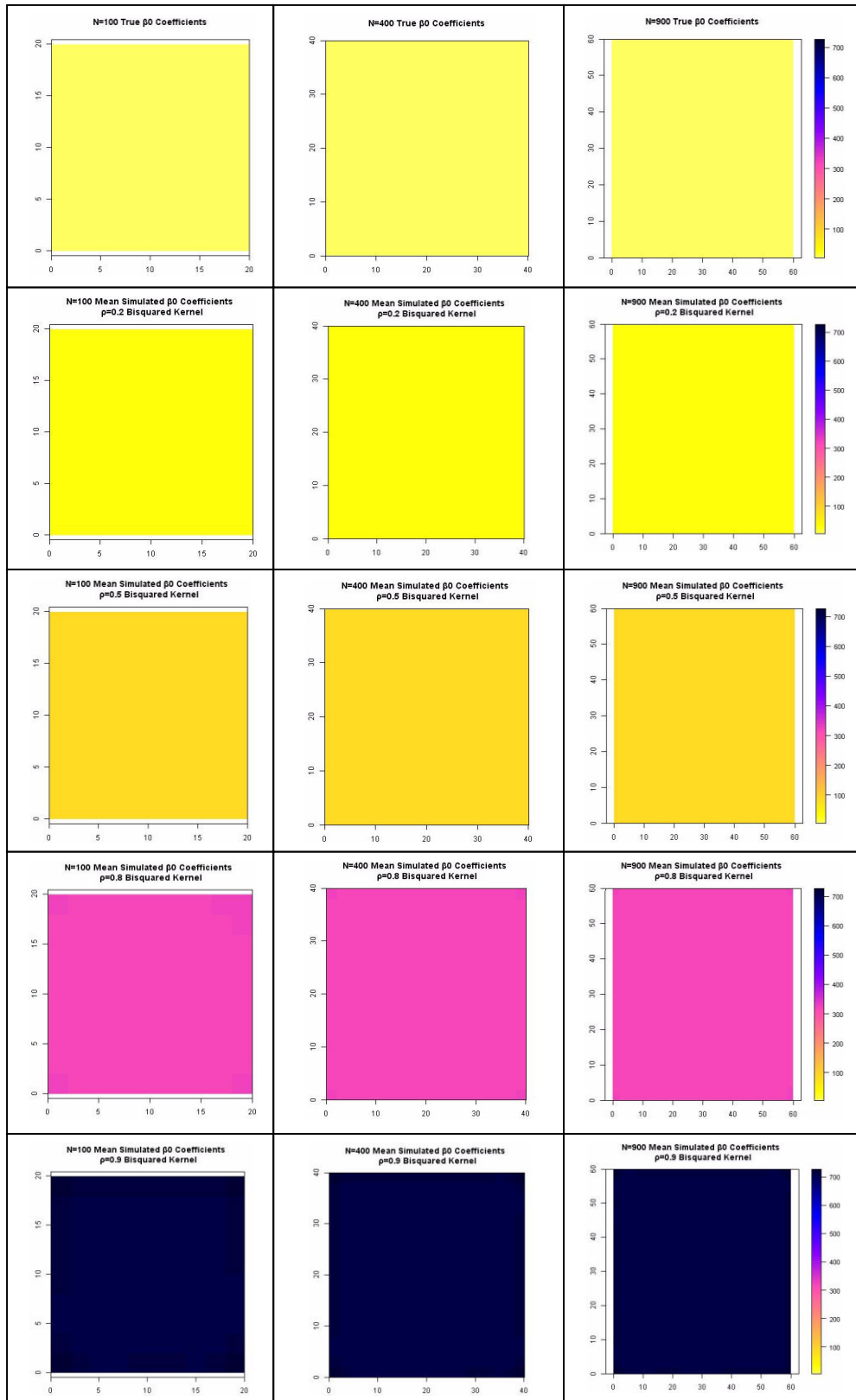
Στον Πίνακα 3.9 περιλαμβάνονται οι μέσες τιμές της στατιστικής I του Moran για τα εκτιμηθέντα σφάλματα των τοπικών υποδειγμάτων χρησιμοποιώντας το κριτήριο του πύργου και τα ποσοστά απόρριψης των ελέγχων χωρικής αυτοσυσχέτισης. Το πιο παράξενο αποτέλεσμα, που παρουσιάζεται σε αυτόν τον πίνακα, είναι η εμφάνιση αρνητικής χωρικής αυτοσυσχέτισης στα σφάλματα που είναι ισχυρή για μικρές τιμές του ρ και γίνεται ασθενέστερη για μεγαλύτερες τιμές του. Εάν ληφθούν υπ' όψιν οι σοβαρές συνέπειες που επέρχονται στην ανάλυση παλινδρόμησης όταν εσφαλμένα εκτιμάται ένα απλό οικονομετρικό υπόδειγμα με τη μέθοδο των ελαχίστων τετραγώνων, ενώ στην πραγματικότητα ισχύει το υπόδειγμα χωρικής υστέρησης, και ότι η βάση της γεωγραφικά σταθμισμένης παλινδρόμησης είναι η κλασσική μέθοδος των ελαχίστων τετραγώνων, τότε το πρόβλημα της χωρικής αυτοσυσχέτισης στα σφάλματα είναι σε κάποιο βαθμό αναμενόμενο. Εν τούτοις, δεν μπορεί εύκολα να εξηγηθεί γιατί η αυτοσυσχέτιση είναι αρνητική και γιατί γίνεται ηπιότερη καθώς αυξάνει το ρ . Ιδιόμορφη είναι και η συμπεριφορά των ελέγχων χωρικής αυτοσυσχέτισης. Ειδικότερα, τα ποσοστά απόρριψης του έλεγχου με την υπόθεση της τυχαιοποίησης είναι υψηλά και προσεγγίζουν το 100% στα μεγάλα δείγματα υποδεικνύοντας το πρόβλημα της αυτοσυσχέτισης. Αντίθετα, τα ποσοστά απόρριψης με την τεχνική των Leung, Mei και Zhang (2000b) είναι πολύ χαμηλά στην ισχυρή αυτοσυσχέτιση και γίνονται μεγαλύτερα στην ασθενή αυτοσυσχέτιση.¹⁴ Τα αποτελέσματα αυτά θέτουν σε αμφισβήτηση για ακόμη μία φορά την εγκυρότητα των ελέγχων χωρικής αυτοσυσχέτισης στην περίπτωση της γεωγραφικά σταθμισμένης παλινδρόμησης.

¹⁴ Οι Leung, Mei και Zhang (2000b) στην σελίδα 887 αναφέρουν ότι ο έλεγχος δεν έχει ικανοποιητική ισχύ στην περίπτωση της μέτριας αρνητικής χωρικής αυτοσυσχέτισης κάτι που ίσως εξηγεί αυτά τα αποτελέσματα.



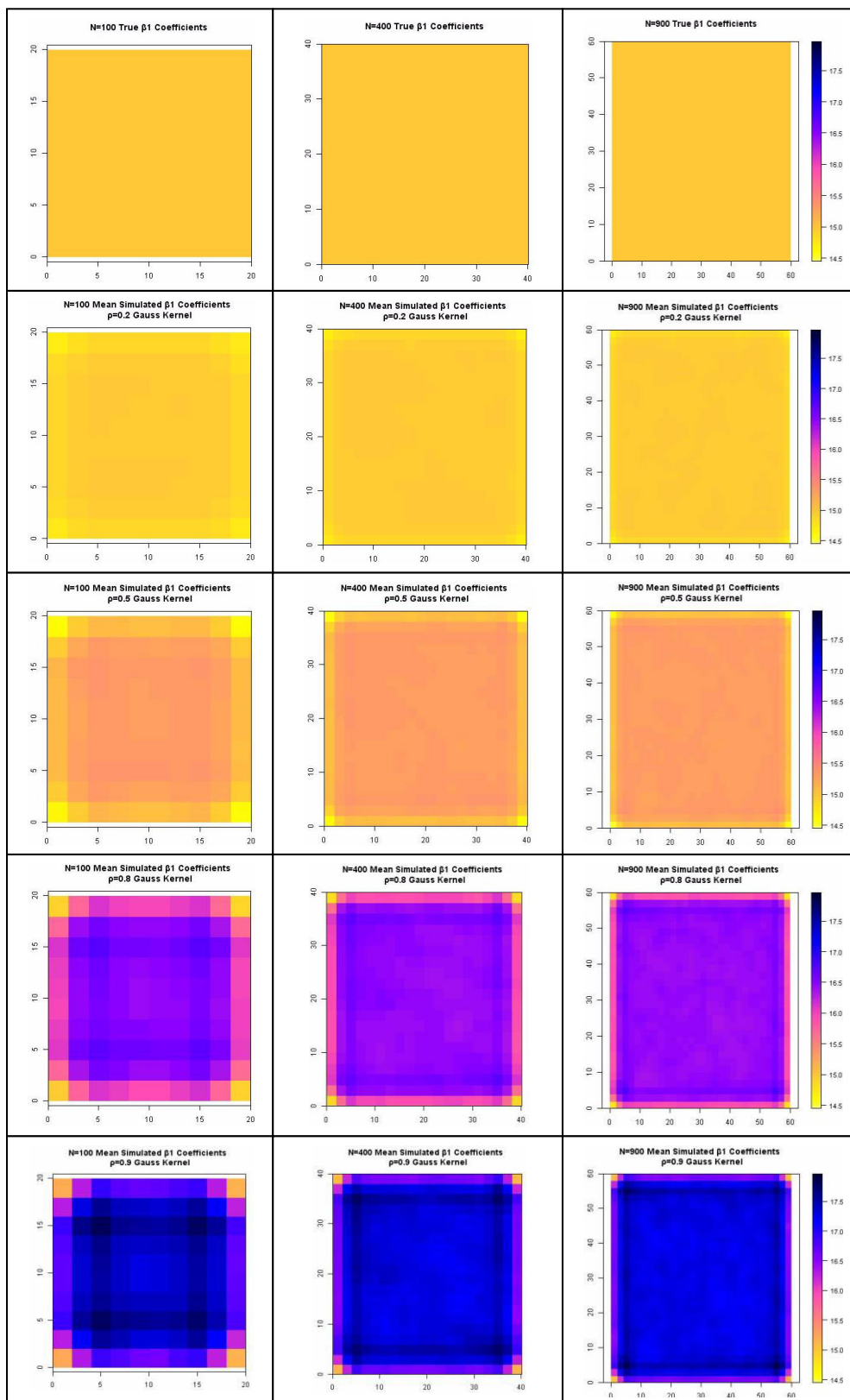
Διάγραμμα 3.8 (Α)

Ψηφιδωτά πραγματικών και προσομοιωμένων μέσων τοπικών εκτιμηθέντων τιμών από την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης στο Υπόδειγμα Χωρικής Υστέρησης με Διτετράγωνη συνάρτηση στάθμισης για όλα τα μεγέθη δείγματος και τιμές του ρ Συντελεστής β_{1i}



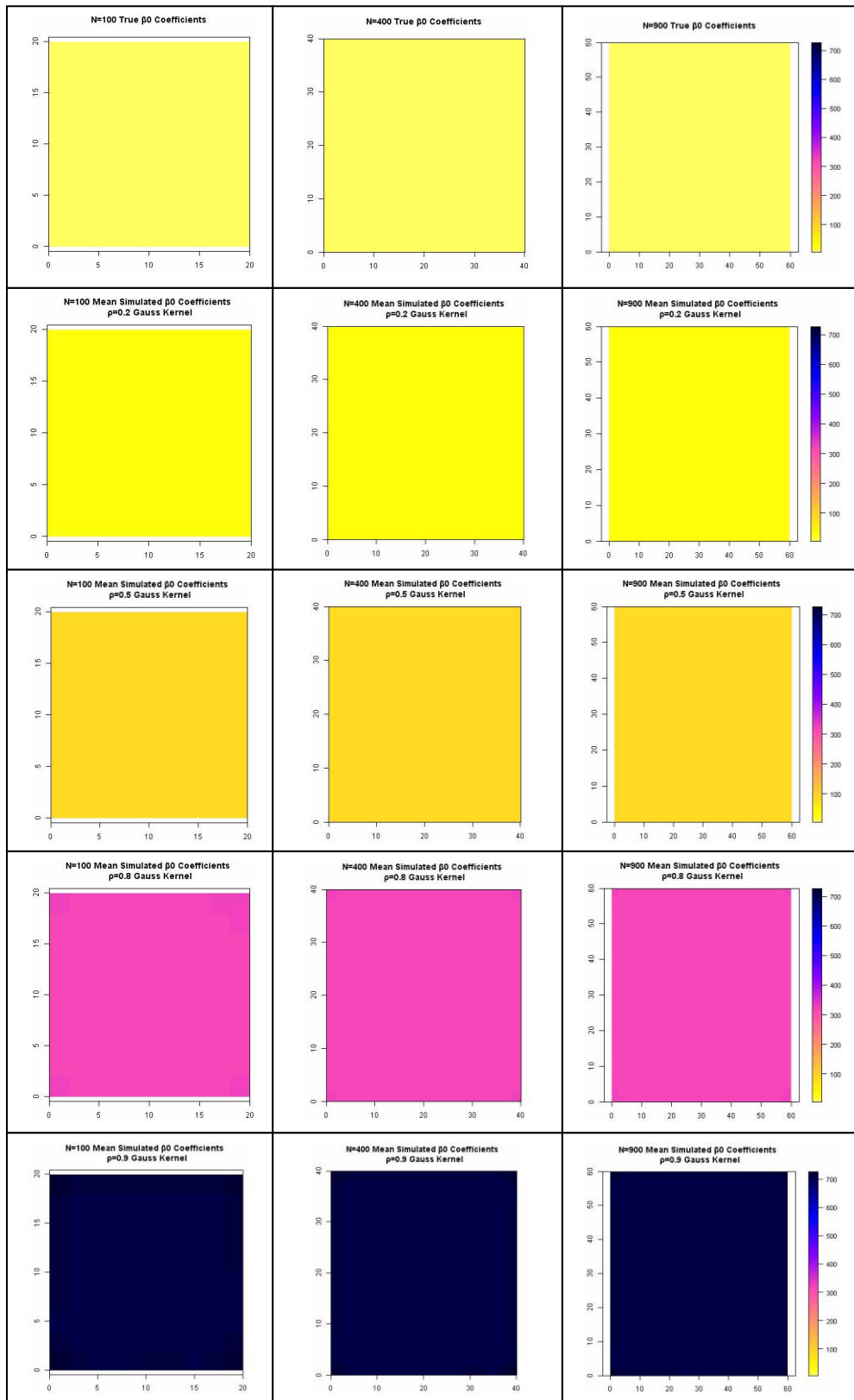
Διάγραμμα 3.8 (B)

Ψηφιδωτά πραγματικών και προσομοιωμένων μέσων τοπικών εκτιμηθέντων τιμών από την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης στο Υπόδειγμα Χωρικής Υστέρησης με Διτετράγωνη συνάρτηση στάθμισης για όλα τα μεγέθη δείγματος και τιμές του ρ Συντελεστής β_{i0}



Διάγραμμα 3.9 (A)

Ψηφιδωτά πραγματικών και προσομοιωμένων μέσων τοπικών εκτιμηθέντων τιμών από την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης στο Υπόδειγμα Χωρικής Υστέρησης με Γκαουσιανή συνάρτηση στάθμισης για όλα τα μεγέθη δείγματος και τιμές του ρ Συντελεστής β_{i1}



Διάγραμμα 3.9 (B)

Ψηφιδωτά πραγματικών και προσομοιωμένων μέσων τοπικών εκτιμηθέντων τιμών από την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης στο Υπόδειγμα Χωρικής Υστέρησης με Γκαουσιανή συνάρτηση στάθμισης για όλα τα μεγέθη δείγματος και τιμές του ρ Συντελεστής β_{i0}

Οι μέσοι εκτιμηθέντες τοπικοί συντελεστές που προέκυψαν από τη γεωγραφικά σταθμισμένη παλινδρόμηση, όταν ισχύει το υπόδειγμα χωρικής υστέρησης, απεικονίζονται στα ψηφιδωτά (*raster*) των Διαγραμμάτων 3.8 και 3.9 για τη διτετράγωνη και τη Γκαουσιανή συνάρτηση στάθμισης, αντίστοιχα. Η παρατήρηση των διαγραμμάτων αποκαλύπτει μια τελείως διαφορετική εικόνα για τη συμπεριφορά των δύο συντελεστών σε σχέση με την περίπτωση που η πραγματική σχέση εξάρτησης είναι το υπόδειγμα χωρικού σφάλματος. Συγκεκριμένα, από το Διάγραμμα 3.8 (Α) φαίνεται ότι για δεδομένες τιμές του ρ και μεγέθη δείγματος, οι μέσες εκτιμηθείσες τιμές για τον συντελεστή β_1 της ανεξάρτητης μεταβλητής δεν παρουσιάζουν σημαντική χωρική μεταβλητότητα φανερώνοντας την ύπαρξη ολικής σχέσης. Ωστόσο, ενώ στις χαμηλές τιμές του συντελεστή χωρικής υστέρησης ρ οι μέσοι τοπικοί συντελεστές λαμβάνουν τιμές πολύ κοντά στο 15, που είναι η πραγματική τιμή του συντελεστή β_1 που χρησιμοποιήθηκε στην προσομοίωση, για μεγάλες τιμές του ρ διαπιστώνεται αύξηση στις μέσες τιμές των τοπικών συντελεστών που ξεπερνάνε το 17,5. Το αποτέλεσμα αυτό δείχνει πως όταν ισχύει το υπόδειγμα χωρικής υστέρησης, οι εκτιμήσεις της γεωγραφικά σταθμισμένης παλινδρόμησης είναι μεροληπτικές σε αντίθεση με το υπόδειγμα χωρικού σφάλματος που δεν παρατηρήθηκε ιδιαίτερη μεροληψία. Κατά συνέπεια, η εκτίμηση τοπικών υποδειγμάτων με την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης δεν αντιμετωπίζει το πρόβλημα της μεροληψίας που οφείλεται στην παράλειψη της χωρικής υστέρησης της εξαρτημένης μεταβλητής από το υπόδειγμα και οδηγεί σε μεροληπτικές εκτιμήσεις και χωρική αυτοσυσχέτιση στα σφάλματα όπως είναι γνωστό ότι θα συνέβαινε και στην περίπτωση που η εκτίμηση γινόταν με την απλή μέθοδο των ελαχίστων τετραγώνων.

Η μεροληψία είναι πιο μεγάλη για τον σταθερό όρο β_0 του υποδείγματος όπως απεικονίζουν τα ψηφιδωτά στο Διάγραμμα 3.8 (Β). Διαπιστώνεται ξανά η απουσία χωρικής ετερογένειας στους εκτιμηθέντες τοπικούς σταθερούς όρους οι οποίοι είναι σταθεροί σε όλα τα γεωγραφικά σημεία για δεδομένα ρ δηλώνοντας την ύπαρξη ολικής σχέσης αλλά οι τιμές τους απέχουν από την τιμή του 5 που θεωρήθηκε στην προσομοίωση. Μάλιστα, οι εκτιμηθέντες τιμές τους αυξάνονται σημαντικά στις μεγάλες τιμές του ρ προσεγγίζοντας το 700 κάτι που φανερώνει ότι η αυτοσυσχέτιση στην εξαρτημένη μεταβλητή, μέσω της χωρικής υστέρησης που έχει αγνοηθεί, εκφράζεται με τους σταθερούς όρους. Από τα Διαγράμματα 3.9 (Α) και 3.9 (Β), φαίνεται ότι η ίδια συμπεριφορά για τους δύο συντελεστές του υποδείγματος υπάρχει και όταν χρησιμοποιείται η Γκαουσιανή συνάρτηση στάθμισης. Επιπρόσθετα, και για τα δύο υποδείγματα που μελετήθηκαν στην προσομοίωση διενεργήθηκαν έλεγχοι στατιστικής

σημαντικότητας για τους τοπικούς συντελεστές μετά από τον υπολογισμό ψευδό-*t* στατιστικών με ποσοστά απόρριψης που προσεγγίζουν το 100%.

Το τελικό συμπέρασμα από τις διεξαχθείσες προσομοιώσεις είναι ότι ο ερευνητής θα πρέπει να είναι ιδιαίτερα προσεκτικός όταν εφαρμόζει την τεχνική της γεωγραφικά σταθμισμένης παλινδρόμησης για τη διερεύνηση της σχέσης δύο μεταβλητών όταν υπάρχουν επιδράσεις χωρικής εξάρτησης. Ο εντοπισμός του σωστού τύπου των χωρικών επιδράσεων που εμφανίζονται στην ανάλυση, δηλαδή η διάκριση μεταξύ χωρικής αυτοσυσχέτισης και χωρικής ετερογένειας είναι αναγκαία διαδικασία. Αυτό τονίζεται διότι αρκετές φορές σε εμπειρικές εφαρμογές, κυρίως στον επιστημονικό τομέα της ποσοτικής γεωγραφίας, η γεωγραφικά σταθμισμένη παλινδρόμηση χρησιμοποιείται αδιακρίτως ως μέσο εκτίμησης κάθε χωρικού υποδείγματος και αντιμετώπισης των κάθε μορφής χωρικών επιδράσεων.

Τα αποτελέσματα της προσομοίωσης έδειξαν ότι οι συνέπειες από την αντιμετώπιση των επιδράσεων χωρικής εξάρτησης ως χωρική ετερογένεια, μέσω της εκτίμησης τοπικών υποδειγμάτων με την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης, είναι διαφορετικές ανάλογα με την αιτία που προκαλεί την εμφάνισή τους. Ποιο ήπιες είναι οι επιπτώσεις όταν οι χωρικές επιδράσεις οφείλονται στο υπόδειγμα χωρικού σφάλματος το οποίο διακρίνεται πιο εύκολα συγκρίνοντας τους συντελεστές προσδιορισμού της απλής μεθόδου των ελαχίστων τετραγώνων και της γεωγραφικά σταθμισμένης παλινδρόμησης ενώ για μέτριες τιμές των λ βοήθεια προσφέρουν και οι έλεγχοι βελτίωσης καθώς και οι τιμές του εύρους που προκύπτουν από τη διασταυρωτική επικύρωση. Αντίθετα, όταν οι χωρικές επιδράσεις έχουν αιτία την ύπαρξη χωρικής υστέρησης στην εξαρτημένη μεταβλητή μέσω του υποδείγματος χωρικής υστέρησης, ο ερευνητής θα έχει πάντα ενδείξεις υπέρ της καταλληλότητας της γεωγραφικά σταθμισμένης παλινδρόμησης.

Η χαρτογράφηση των εκτιμηθέντων τιμών των τοπικών συντελεστών και στις δύο περιπτώσεις διευκολύνουν τον ερευνητή να εντοπίσει την απουσία της χωρικής ετερογένειας και την ύπαρξη ολικής σχέσης. Εφόσον οι τοπικές εκτιμήσεις κρατηθούν για να χρησιμοποιηθούν στην ανάλυση τότε για το υπόδειγμα χωρικού σφάλματος δεν θα έχουν σοβαρή μεροληψία αλλά για το υπόδειγμα χωρικής υστέρησης με μεγάλες τιμές του ρ θα υπάρχει σοβαρό πρόβλημα μεροληψίας όπως συμβαίνει και με την απλή μέθοδο των ελαχίστων τετραγώνων. Ένα άλλο συμπέρασμα που πρέπει να ληφθεί υπ' όψιν στην αξιολόγηση των αποτελεσμάτων είναι η εμφάνιση χωρικής αυτοσυσχέτισης στα εκτιμηθέντα σφάλματα αν και η εγκυρότητα των ελέγχων χωρικής αυτοσυσχέτισης αμφισβητείται.

3.7 Υποφαινομενική συμπεριφορά σε τοπικό επίπεδο από την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης

Στις προηγούμενες ενότητες εξετάστηκαν με τη βοήθεια προσομοίωσης η συμπεριφορά της γεωγραφικά σταθμισμένης παλινδρόμησης, ως προς τα αποτελέσματα των εκτιμήσεων και τη διεξαγωγή των στατιστικών ελέγχων που συγκρίνουν τη βελτίωση που επέρχεται σε σχέση με την εφαρμογή της απλής παλινδρόμησης, θεωρώντας ότι οι εμπλεκόμενες στην παλινδρόμηση μεταβλητές δεν είναι ανεξάρτητες αλλά πραγματικά υπάρχει σχέση είτε στο ολικό επίπεδο είτε σε τοπικό επίπεδο. Η υπόθεση αυτή ήταν η βασική φιλοσοφία και των προσομοιώσεων στις εργασίες των Farber και Páez (2007), Wheeler και Tiefelsdorf (2005), Wheeler (2007 και 2009), Wheeler και Calder (2007) και Páez, Farber και Wheeler (2011).

Στην ενότητα αυτή πραγματοποιούνται μια σειρά προσομοιώσεων σε τελείως διαφορετική βάση. Συγκεκριμένα, θεωρείται ότι οι μεταβλητές που εμπλέκονται στην παλινδρόμηση είναι ανεξάρτητες μεταξύ τους τόσο σε ολικό όσο και σε τοπικό επίπεδο κάτι που δεν έχει ερευνηθεί για την περίπτωση της γεωγραφικά σταθμισμένης παλινδρόμησης, μολονότι αρκετές προσπάθειες έχουν γίνει για τη μελέτη ανεξάρτητων χωρικών μεταβλητών σε ολικό επίπεδο στα πλαίσια της εκτίμησής τους με την κλασική μέθοδο των ελαχίστων τετραγώνων. Ειδικότερα, οι Fingleton (1999), Mur και Trivez (2003) και Agiakloglou, Tsimbos και Tsimpanos (2015) έχουν εξετάσει τη σχέση μεταξύ ανεξάρτητων στάσιμων ή μη στάσιμων χωρικών διαδικασιών σε ολικό επίπεδο και κατέληξαν στο συμπέρασμα ότι η εφαρμογή της μεθόδου των ελαχίστων τετραγώνων οδηγεί σε παραπλανητικά αποτελέσματα, ένα φαινόμενο που είναι γνωστό από την ανάλυση χρονοσειρών ως υποφαινομενική παλινδρόμηση (*spurious regression*) έπειτα από την πρωτοποριακή εργασία των Granger και Newbold (1974). Τα αποτελέσματα των προσομοιώσεων δείχνουν ότι η εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης για την εκτίμηση τοπικών γραμμικών σχέσεων μεταξύ ανεξάρτητων χωρικών διαδικασιών μπορεί να εμφανίσει σε τοπικό επίπεδο υποφαινομενική συμπεριφορά (*local spurious behavior*) παρόμοια με το πρόβλημα της υποφαινομενικής παλινδρόμησης που είναι γνωστό ότι εμφανίζεται σε ολικό επίπεδο.

3.7.1 Τοπική υποφαινομενική συμπεριφορά με στάσιμες χωρικές αυτοπαλίνδρομες διαδικασίες πρώτης τάξης

Η διερεύνηση της ύπαρξης γραμμικής σχέσης σε ολικό επίπεδο μεταξύ δύο μεταβλητών των οποίων οι τιμές προέρχονται από διαφορετικά γεωγραφικά σημεία γίνεται με την εκτίμηση ενός υποδείγματος της μορφής της σχέσης (3.1) με μία ανεξάρτητη μεταβλητή, δηλαδή:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (3.31)$$

όπου $\varepsilon_i \sim iid N(0, \sigma^2)$ και το $i = 1, 2, \dots, n$ συμβολίζει τα γεωγραφικά σημεία στα οποία έχει συλλεχθεί η κάθε παρατήρηση. Το υπόδειγμα (3.31) αποκαλείται συχνά ως ολικό υπόδειγμα (*global model*) διότι εξετάζει τη σχέση των δύο μεταβλητών σε ολικό επίπεδο υποθέτοντας ότι παραμένει σταθερή σε όλη την περιοχή μελέτης και η εκτίμησή του επιτυγχάνεται στα πλαίσια των καθιερωμένων οικονομετρικών τεχνικών με τη βοήθεια της μεθόδου των ελαχίστων τετραγώνων. Όταν η σχέση εξάρτησης δεν είναι σταθερή σε όλες τις γεωγραφικές περιοχές το ολικό υπόδειγμα μπορεί να γραφεί στη μορφή τοπικού υποδείγματος (*local model*) επιτρέποντας στους συντελεστές τους να έχουν διαφορετικές τιμές, ανάλογα με την περιοχή από όπου προέρχεται η κάθε παρατήρηση, δηλώνοντας διαφορετική κατεύθυνση αλλά και ένταση στη σχέση των ανεξάρτητων μεταβλητών με την εξαρτημένη μεταβλητή. Το τοπικό υπόδειγμα που μπορεί να χρησιμοποιηθεί για διερεύνηση της ύπαρξης γραμμικής σχέσης σε τοπικό επίπεδο μεταξύ δύο μεταβλητών και που αντιστοιχεί στο ολικό υπόδειγμα (3.31) ορίζεται ως εξής:

$$y_i = \beta_{i0} + \beta_{i1} x_i + \varepsilon_i \quad (3.32)$$

και πρέπει να εκτιμηθεί χωριστά σε κάθε γεωγραφικό σημείο χρησιμοποιώντας εξειδικευμένες τεχνικές χωρικής ανάλυσης, όπως η μέθοδος της γεωγραφικά σταθμισμένης παλινδρόμησης.

Η σύγκριση ανάμεσα στο τοπικό ή στο ολικό υπόδειγμα και η επιλογή του καλύτερου, όσον αφορά την προσαρμογή στα δεδομένα της ανάλυσης, γίνεται μέσω των συντελεστών προσδιορισμού των δύο υποδειγμάτων αλλά και την πραγματοποίηση στατιστικών ελέγχων που παρουσιάστηκαν αναλυτικά σε προηγούμενες ενότητες. Εφόσον ο ερευνητής επιλέξει το τοπικό υπόδειγμα, το επόμενο βήμα είναι να καθορίσει σε τι βαθμό υπάρχει, σε τοπικό επίπεδο, γραμμική σχέση μεταξύ των δύο μεταβλητών. Η απόφαση αυτή στηρίζεται στους ελέγχους με τις τιμές των τοπικών ψευδό- t στατιστικών που υπολογίζονται διαιρώντας τους εκτιμηθέντες τοπικούς συντελεστές με τα τοπικά τους σφάλματα. Εάν, για επίπεδο

σημαντικότητας 5%, περισσότερες από 5% των απόλυτων τιμών των n τοπικών ψευδό- t στατιστικών είναι μεγαλύτερες της κριτικής τιμής του ελέγχου, τότε θεωρείται ότι υπάρχει τοπική σχέση ανάμεσα στις δύο μεταβλητές.

Με βάση αυτή τη πρακτική, αρκετές φορές στις εμπειρικές εφαρμογές που γίνεται χρήση της τεχνικής της γεωγραφικά σταθμισμένης παλινδρόμησης εκτιμάται αρχικά ένα ολικό υπόδειγμα με τη μέθοδο των ελαχίστων τετραγώνων. Όταν κάποια από τις ανεξάρτητες μεταβλητές εμφανίζεται να είναι στατιστικά μη σημαντική, τότε συνηθίζεται να παραμένει στην ανάλυση προσδοκώντας ότι σε τοπικό επίπεδο, σε κάποιες από τις γεωγραφικές περιοχές, θα γίνει στατιστικά σημαντική. Εφόσον οι περιοχές που η μεταβλητή είναι στατιστικά σημαντική σε τοπικό επίπεδο είναι περισσότερες από το επίπεδο σημαντικότητας του 5%, σύμφωνα με τις τοπικές ψευδό- t στατιστικές, δηλαδή όσο το αναμενόμενο σφάλμα τύπου I, τότε θεωρείται ότι επηρεάζει σε τοπικό επίπεδο την εξαρτημένη μεταβλητή και παραμένει στην ανάλυση. Σε αντίθετη περίπτωση το ολικό υπόδειγμα αναθεωρείται αφαιρώντας τη συγκεκριμένη μεταβλητή και η διαδικασία επαναλαμβάνεται.

Για να εξεταστεί με την εφαρμογή προσομοίωσης η εμφάνιση υποφαινομενικών παλινδρομήσεων σε τοπικό επίπεδο χρησιμοποιήθηκε, όπως και στις προηγούμενες προσομοιώσεις αυτού του κεφαλαίου, ένα ομαλό τετραγωνικό πλέγμα (*regular grid*) διαστάσεων 10x10, 20x20 και 30x30, με κλίμακες για τους δύο άξονες 0 έως 20, 0 έως 40 και 0 έως 60 αντίστοιχα, που οδηγεί σε δείγματα 100, 400 και 900 παρατηρήσεων. Οι συντεταγμένες του κέντρου κάθε τετραγώνου του πλέγματος αποτέλεσαν και πάλι τις γεωγραφικές συντεταγμένες για τη διαδικασία της προσομοίωσης. Επάνω στο ομαλό πλέγμα παρήχθησαν δύο ανεξάρτητες στάσιμες χωρικές αυτοπαλίνδρομες διαδικασίες πρώτης τάξης, SAR(1), y και x σύμφωνα με τις σχέσεις $y = (I - \phi W)^{-1} \varepsilon_y$ και $x = (I - \phi W)^{-1} \varepsilon_x$, όπου τα σφάλματα ε_y και ε_x είναι χωρικές διαδικασίες λευκού θορύβου $N(0, I)$ ανεξάρτητες μεταξύ τους και ο αυτοπαλίνδρομος συντελεστής ϕ λαμβάνει τις τιμές 0,0, 0,2, 0,5, 0,8 και 0,9. Η $(n \times n)$ μήτρα χωρικών σταθμίσεων W , που εκφράζει την χωρική αλληλεπίδραση ανάμεσα στις γειτονικές περιοχές, ορίστηκε σύμφωνα με τα κριτήρια του πύργου (*rook contiguity*) στην τυποποιημένη κατά γραμμές μορφή της χωρίς διόρθωση τόρος.¹⁵ Πραγματοποιήθηκαν 10.000 επαναλήψεις.

¹⁵ Επισημαίνεται ότι τα στοιχεία της μήτρας W που χρησιμοποιούνται στον ορισμό των SAR(1) διαδικασιών είναι πάντα δεδομένα και δεν έχουν καμία σχέση με τις τιμές w_{ij} που προκύπτουν από τις συναρτήσεις στάθμισης κατά τη διαδικασία κάθε τοπικής εκτίμησης με τη γεωγραφικά σταθμισμένη παλινδρόμηση.

Πίνακας 3.10

Συντελεστές προσδιορισμού μεθόδων ελαχίστων τετραγώνων και γεωγραφικά σταθμισμένης παλινδρόμησης και ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% των ελέγχων βελτίωσης για δύο ανεξάρτητες SAR(1) διαδικασίες για όλα τα μεγέθη δείγματος

| Πυρήνας | | Διτετράγωνος | | | | | Γκαουσιανός | | | | |
|-------------------|----------------------|--------------|-------|-------|-------|-------|-------------|-------|-------|-------|-------|
| <i>N</i> | ϕ | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 |
| 100 | R ² – OLS | 0,010 | 0,011 | 0,014 | 0,033 | 0,062 | 0,010 | 0,011 | 0,014 | 0,033 | 0,062 |
| | R ² – GWR | 0,058 | 0,111 | 0,350 | 0,733 | 0,848 | 0,044 | 0,094 | 0,360 | 0,787 | 0,889 |
| Ποσοστά Απόρριψης | ANOVA – BCM | 24,34 | 44,01 | 86,32 | 99,86 | 99,99 | 18,26 | 34,28 | 75,01 | 98,63 | 99,89 |
| | ANOVA – FBC | 4,31 | 16,42 | 70,30 | 99,41 | 99,77 | 2,54 | 11,22 | 60,55 | 97,06 | 98,74 |
| | F – RATIO | 1,50 | 8,46 | 57,27 | 98,77 | 99,90 | 1,38 | 8,60 | 57,97 | 98,29 | 99,73 |
| | F1 | 0,04 | 0,41 | 23,05 | 96,61 | 99,85 | 0,04 | 0,36 | 23,14 | 96,54 | 99,88 |
| | F2 | 4,24 | 13,40 | 57,59 | 86,70 | 89,45 | 2,97 | 9,15 | 35,82 | 63,42 | 67,45 |
| 400 | R ² – OLS | 0,002 | 0,003 | 0,003 | 0,009 | 0,019 | 0,002 | 0,003 | 0,003 | 0,009 | 0,019 |
| | R ² – GWR | 0,014 | 0,037 | 0,326 | 0,727 | 0,845 | 0,011 | 0,034 | 0,367 | 0,804 | 0,901 |
| Ποσοστά Απόρριψης | ANOVA – BCM | 36,93 | 59,37 | 93,86 | 99,97 | 100 | 31,41 | 48,96 | 82,06 | 99,46 | 99,99 |
| | ANOVA – FBC | 4,93 | 22,02 | 84,73 | 99,90 | 100 | 2,7 | 17,05 | 77,08 | 99,46 | 99,99 |
| | F – RATIO | 0,11 | 3,87 | 73,94 | 99,84 | 100 | 0,11 | 5,42 | 75,12 | 99,46 | 99,99 |
| | F1 | 0,00 | 0,12 | 61,24 | 99,84 | 100 | 0,00 | 0,13 | 65,35 | 99,46 | 99,99 |
| | F2 | 7,81 | 27,12 | 86,42 | 99,94 | 100 | 6,20 | 22,76 | 77,83 | 99,2 | 99,29 |
| 900 | R ² – OLS | 0,001 | 0,001 | 0,002 | 0,004 | 0,009 | 0,001 | 0,001 | 0,002 | 0,004 | 0,009 |
| | R ² – GWR | 0,006 | 0,018 | 0,334 | 0,721 | 0,842 | 0,005 | 0,019 | 0,380 | 0,805 | 0,902 |
| Ποσοστά Απόρριψης | ANOVA – BCM | 40,30 | 62,92 | 95,02 | 100 | 100 | 35,88 | 52,75 | 84,49 | 99,59 | 100 |
| | ANOVA – FBC | 5,46 | 23,24 | 85,95 | 99,97 | 100 | 3,29 | 19,14 | 79,42 | 99,59 | 100 |
| | F – RATIO | 0,01 | 2,11 | 78,18 | 99,94 | 100 | 0,02 | 4,07 | 78,72 | 99,59 | 100 |
| | F1 | 0,00 | 0,03 | 76,21 | 99,94 | 100 | 0,00 | 0,05 | 78,25 | 99,59 | 100 |
| | F2 | 8,96 | 29,23 | 87,84 | 99,97 | 100 | 7,23 | 25,18 | 79,96 | 99,59 | 100 |

Πίνακας 3.11

Μέσες, ελάχιστες και μέγιστες τιμές εύρους για δύο ανεξάρτητες SAR(1) διαδικασίες

| Πυρήνας | | Διτετράγωνος | | | | | Γκαουσιανός | | | | |
|----------|----------------|--------------|--------|--------|--------|--------|-------------|--------|--------|--------|--------|
| <i>N</i> | ϕ | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 |
| 100 | Ολικό Εύρος | 25,456 | | | | | 25,456 | | | | |
| | Μέσο Εύρος | 22,024 | 18,701 | 10,380 | 4,943 | 4,397 | 19,132 | 15,721 | 7,535 | 1,987 | 1,531 |
| | Ελάχιστο Εύρος | 4,304 | 3,308 | 2,982 | 2,057 | 2,058 | 1,660 | 1,025 | 0,777 | 0,588 | 0,514 |
| | Μέγιστο Εύρος | 25,456 | 25,456 | 25,456 | 25,456 | 25,456 | 25,456 | 25,456 | 25,456 | 25,456 | 25,456 |
| 400 | Ολικό Εύρος | 53,740 | | | | | 53,740 | | | | |
| | Μέσο Εύρος | 46,528 | 38,001 | 13,540 | 4,887 | 4,546 | 40,640 | 32,253 | 12,191 | 1,870 | 1,458 |
| | Ελάχιστο Εύρος | 8,404 | 5,330 | 3,490 | 3,101 | 3,029 | 3,109 | 2,129 | 1,400 | 0,991 | 0,922 |
| | Μέγιστο Εύρος | 53,740 | 53,740 | 53,740 | 53,740 | 6,343 | 53,740 | 53,740 | 53,740 | 53,740 | 53,740 |
| 900 | Ολικό Εύρος | 82,024 | | | | | 82,024 | | | | |
| | Μέσο Εύρος | 71,054 | 57,374 | 16,647 | 4,874 | 4,559 | 61,611 | 48,630 | 16,042 | 1,894 | 1,436 |
| | Ελάχιστο Εύρος | 12,950 | 6,926 | 4,799 | 3,287 | 3,226 | 4,592 | 2,497 | 1,487 | 1,134 | 1,050 |
| | Μέγιστο Εύρος | 82,024 | 82,024 | 82,024 | 52,471 | 5,324 | 82,024 | 82,024 | 82,024 | 82,024 | 1,952 |

Τα υποδείγματα (3.31) και (3.32) εκτιμήθηκαν χρησιμοποιώντας αντίστοιχα τις μεθόδους των ελαχίστων τετραγώνων (OLS) και τη γεωγραφικά σταθμισμένη παλινδρόμηση (GWR) με διτετράγωνη και Γκαουσιανή συνάρτηση στάθμισης σταθερού πυρήνα και οι μέσες τιμές των συντελεστών προσδιορισμού των δύο υποδειγμάτων παρουσιάζονται στον Πίνακα 3.10. Όπως φαίνεται σε αυτόν τον πίνακα, η εξειδίκευση του τοπικού υποδείγματος είναι πιο αποδεκτή από ότι η εξειδίκευση του ολικού υποδείγματος, καθώς οι μέσες τιμές των συντελεστών προσδιορισμού της γεωγραφικά σταθμισμένης παλινδρόμησης είναι πάντοτε μεγαλύτερες από τις μέσες τιμές των συντελεστών προσδιορισμού της μεθόδου των ελαχίστων τετραγώνων για όλα τα μεγέθη δείγματος και για όλες τις τιμές της αυτοπαλίνδρομης παραμέτρου ϕ . Μάλιστα, η διαφορά μεταξύ των μέσων τιμών των δύο συντελεστών προσδιορισμού γίνεται μεγαλύτερη σε σημαντικό βαθμό όσο αυξάνονται οι τιμές της αυτοπαλίνδρομης παραμέτρου για δεδομένο μέγεθος δείγματος και παραμένει κατά προσέγγιση αμετάβλητη ανεξάρτητα του δείγματος και της συνάρτησης στάθμισης παρέχοντας ισχυρότερες ενδείξεις για την καταλληλότητα του τοπικού υποδείγματος.

Επιπρόσθετα, ο Πίνακας 3.10 περιλαμβάνει τα ποσοστά απόρριψης, σε επίπεδο σημαντικότητας 5%, των πέντε ελέγχων που εξετάζουν την συνεισφορά της γεωγραφικά σταθμισμένης παλινδρόμησης. Όπως διαπιστώνεται από τον πίνακα αυτό και οι πέντε έλεγχοι συμπεριφέρονται με παρόμοιο τρόπο υποδεικνύοντας την εξειδίκευση του υποδείγματος (3.32) με την έννοια ότι η εκτίμησή του βελτιώνει τα αποτελέσματα. Οι ενδείξεις των ελέγχων υπέρ των τοπικών υποδειγμάτων ενισχύονται σημαντικά καθώς αυξάνονται οι τιμές της αυτοπαλίνδρομης παραμέτρου ανεξάρτητα από το μέγεθος του δείγματος και της συνάρτησης στάθμισης. Στην πραγματικότητα, για $\phi = 0,9$ τα ποσοστά απόρριψης είναι 100% και η εξειδίκευση τοπικών υποδειγμάτων δηλώνεται πάντοτε από τους ελέγχους.

Εκτός από τη σύγκριση των δύο συντελεστών προσδιορισμού και τα αποτελέσματα των στατιστικών ελέγχων, η τιμή του εύρους (*bandwidth*) που καθορίζεται από τη διαδικασία της διασταυρωτικής επικύρωσης (*cross validation*) μπορεί να αποτελέσει κριτήριο επιλογής του σωστού υποδείγματος. Μια χαμηλή τιμή του εύρους φανερώνει ότι η εξειδίκευση του τοπικού υποδείγματος είναι πιο κατάλληλη ενώ μια υψηλή τιμή του εύρους προτείνει το ολικό υπόδειγμα ως πιο σωστό. Ο Πίνακας 3.11 παρουσιάζει τις μέσες τιμές του εύρους που προέκυψαν από την προσομοίωση των 10.000 επαναλήψεων μαζί με τις ελάχιστες και μέγιστες τιμές του καθώς και το ολικό εύρος (*global bandwidth*). Η μελέτη αυτού του πίνακα δείχνει ότι οι τιμές του εύρους, ως κριτήριο επιλογής υποδείγματος, προτείνουν την

εξειδίκευση του τοπικού υποδείγματος όταν αυξάνει η τιμή της αυτοπαλίνδρομης παραμέτρου, ανεξάρτητα από το μέγεθος του δείγματος και τη συνάρτηση στάθμισης. Η συμπεριφορά αυτή είναι σε συμφωνία με τη συμπεριφορά των στατιστικών ελέγχων και των συντελεστών προσδιορισμού.

Ωστόσο, είναι αναγκαίο να τονιστεί ότι η επιλογή της τιμής του εύρους, μολονότι είναι μια παράμετρος που δεν συνδέεται άμεσα με το υπόδειγμα, είναι ουσιαστική στη διαδικασία της εξειδίκευσης διότι καθορίζει τον αριθμό των παρατηρήσεων που είναι απαραίτητες και που θα χρησιμοποιηθούν στις εκτιμήσεις σε τοπικό επίπεδο. Κατά συνέπεια, μια υψηλή τιμή του εύρους, κοντά στην ολική τιμή του, θα υποδεικνύει ότι υπάρχει μικρή ή και καθόλου χωρική μεταβλητότητα στους συντελεστές του υποδείγματος και επομένως ότι το ολικό υπόδειγμα είναι το πιο κατάλληλο για να εκτιμηθεί. Αντίθετα, μια χαμηλή τιμή του εύρους θα υποδεικνύει ότι υπάρχει μεγάλη χωρική μεταβλητότητα στους συντελεστές και άρα ότι πρέπει να εκτιμηθεί το τοπικό υπόδειγμα. Επιπρόσθετα, μια χαμηλή τιμή του εύρους, που επιλέγεται μέσα από τη διασταυρωτικής επικύρωσης, θα εξαναγκάσει τη διαδικασία εκτίμησης των τοπικών υποδειγμάτων να χρησιμοποιήσει μικρό αριθμό γειτονικών παρατηρήσεων. Το γεγονός αυτό μπορεί να οδηγήσει στην εμφάνιση ψεύτικης χωρικής μεταβλητότητας και συνεπώς σε μη αξιόπιστα αποτελέσματα. Επιπλέον, είναι ενδιαφέρον να επισημανθεί ότι η τιμή του εύρους έχει στενή σχέση με το συντελεστή προσδιορισμού της γεωγραφικά σταθμισμένης παλινδρόμησης καθώς, όπως περιγράφεται στους Πίνακες 3.10 και 3.11, η ερμηνευτική ικανότητα αυξάνεται όταν η τιμή του εύρους μειώνεται, κάτι που έχει αναφερθεί και στην εργασία των Farber και Páez (2007).¹⁶

Ο Πίνακας 3.12 παρουσιάζει τη συχνότητα των τοπικών ποσοστών απορρίψεων σε κάθε επανάληψη, που σχηματίστηκε από την προσομοίωση των 10.000 επαναλήψεων, για τον έλεγχο της μηδενικής υπόθεσης ότι ο συντελεστής της ανεξάρτητης μεταβλητής στο υπόδειγμα (3.32) ισούται με το μηδέν σύμφωνα με τις απόλυτες τιμές των τοπικών ψευδό-*t* στατιστικών. Ειδικότερα, η συχνότητα των τοπικών ποσοστών απόρριψης υπολογίζεται ως εξής. Σε κάθε επανάληψη, *n* ψευδό-*t* στατιστικές υπολογίζονται από την εκτίμηση του υποδείγματος (3.32) για όλες τις *n* περιοχές και οι απόλυτες τιμές τους συγκρίνονται με την κριτική τιμή του 1,96 για επίπεδο σημαντικότητας 5%, αποδίδοντας ένα ποσοστό απορρίψεων για τη συγκεκριμένη επανάληψη που θεωρείται ως μία συχνότητα. Παραδείγματος χάριν, εάν η μηδενική υπόθεση ότι ο συντελεστής β_{1i} ισούται με το μηδέν

¹⁶ Βλέπε Farber και Páez (2007) σελίδα 380.

για την περίπτωση $n = 100$ απορρίπτεται 11 φορές στις 100 τοπικές εκτιμήσεις σε μία επανάληψη, τότε το ποσοστό απορρίψεων 11% αναφέρεται ως μία τιμή συχνότητας στο διάστημα (10, 20]. Κατά συνέπεια, ο Πίνακας 3.12 κατασκευάστηκε συλλέγοντας όλες αυτές τις συχνότητες. Αξίζει να αναφερθεί ότι η εφαρμογή διόρθωσης Bonferroni ή κάποιας διόρθωσης παρόμοιας φιλοσοφίας για την ταυτόχρονη πραγματοποίηση πολλαπλών συγκρίσεων δεν χρησιμοποιείται σε αυτή την περίπτωση, καθώς η μέθοδος Bonferroni εκτός από ότι είναι πολύ συντηρητική υποθέτει ανεξάρτητους ελέγχους κάτι που δεν ισχύει στη γεωγραφικά σταθμισμένη παλινδρόμηση όπως αναφέρουν οι Fotheringham, Brunson και Charlton (2002).¹⁷

Υποφαινομενικές παλινδρομήσεις μαζί με αυτοσυσχετισμένα σφάλματα εμφανίζονται στην ανάλυση χρονοσειρών για δύο ανεξάρτητες στάσιμες αυτοπαλίνδρομες διαδικασίες πρώτης τάξης, AR(1), όπως, μεταξύ άλλων, έχουν δείξει οι Granger, Hyung, και Jeon (2001) και ο Agiakloglou (2009). Σε αυτή την περίπτωση, η μηδενική υπόθεση της απουσίας γραμμικής σχέσης μεταξύ δύο ανεξάρτητων μεταβλητών θα απορρίπτεται σύμφωνα με την τιμή της αυτοπαλίνδρομης παραμέτρου προσεγγίζοντας το ποσοστό του 50% για τιμές ίσες με 0,9 ανεξάρτητα από το μέγεθος του δείγματος. Αυτή η ιδιαίτερη συμπεριφορά μπορεί επίσης να παρατηρηθεί και στη χωρική ανάλυση για δύο ανεξάρτητες στάσιμες χωρικές αυτοπαλίνδρομες διαδικασίες πρώτης τάξης, SAR(1), όπως έχουν δείξει ο Fingleton (1999) και Agiakloglou, Tsimbos και Tsimpanos (2015). Η μηδενική υπόθεση θα απορρίπτεται πιο συχνά καθώς αυξάνεται η τιμή της χωρικά αυτοπαλίνδρομης παραμέτρου, κάτι που συμβαίνει και στην περίπτωση των χρονοσειρών, και οι τιμές της στατιστικής I του Moran θα υποδεικνύουν την ύπαρξη χωρικά αυτοσυσχετισμένων σφαλμάτων όπως η στατιστική των Durbin και Watson στην ανάλυση χρονοσειρών.

Το πρόβλημα των υποφαινομενικών σχέσεων στην ανάλυση με τη γεωγραφικά σταθμισμένη παλινδρόμηση τείνει να εμφανίζεται με παρόμοιο τρόπο. Σε κάθε εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης, ο ερευνητής θα έχει ενδείξεις ότι υπάρχει γραμμική σχέση σε τοπικό επίπεδο επειδή η μηδενική υπόθεση θα απορρίπτεται σε ένα σημαντικό ποσοστό περιοχών μεγαλύτερο από το αναμενόμενο του 5%, ένα αποτέλεσμα που έχει σχέση με την τιμή της χωρικά αυτοπαλίνδρομης παραμέτρου. Πιο συγκεκριμένα,

¹⁷ Βλέπε Fotheringham, Brunson και Charlton (2002) σελίδες 249-250. Γενικά, πρέπει να αναφερθεί ότι δεν υπάρχει κάποιος κοινά αποδεκτός τρόπος για τη διόρθωση του επιπέδου σημαντικότητας, λόγω των πολλαπλών ελέγχων, στη γεωγραφικά σταθμισμένη παλινδρόμηση. Κάποιες προτάσεις, σε μεμονωμένες εργασίες, που έχουν εμφανιστεί δεν έχουν ακόμη υιοθετηθεί ευρέως στις εμπειρικές εφαρμογές και δεν είναι αποδεδειγμένη η αξιοπιστία τους.

Πίνακας 3.12
Συχνότητα τοπικών ποσοστών απορρίψεων σε επίπεδο σημαντικότητας 5% για δύο
ανεξάρτητες SAR(1) διαδικασίες

| Πυρήνας | | Διτετράγωνος | | | | | Γκαουσιανός | | | | |
|----------|-----------|--------------|------|------|------|------|-------------|------|------|------|------|
| <i>N</i> | φ | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 |
| 100 | [0, 5] | 7871 | 7048 | 4021 | 534 | 56 | 8351 | 7576 | 4582 | 705 | 85 |
| | (5, 10] | 463 | 800 | 1984 | 1680 | 576 | 268 | 591 | 1810 | 1872 | 788 |
| | (10, 20] | 680 | 1061 | 2613 | 4892 | 3889 | 495 | 829 | 2360 | 4906 | 4181 |
| | (20, 40] | 496 | 659 | 1093 | 2778 | 5208 | 363 | 514 | 891 | 2411 | 4753 |
| | (40, 80] | 328 | 302 | 228 | 111 | 270 | 185 | 200 | 176 | 67 | 189 |
| | (80, 100] | 162 | 130 | 61 | 5 | 1 | 338 | 290 | 181 | 39 | 4 |
| 400 | [0, 5] | 8054 | 7088 | 2490 | 16 | 0 | 8497 | 7526 | 3037 | 43 | 0 |
| | (5, 10] | 433 | 907 | 3917 | 462 | 8 | 280 | 726 | 3970 | 816 | 25 |
| | (10, 20] | 558 | 1086 | 3090 | 8135 | 3543 | 397 | 911 | 2607 | 8313 | 5389 |
| | (20, 40] | 512 | 563 | 354 | 1385 | 6449 | 356 | 438 | 173 | 811 | 4585 |
| | (40, 80] | 272 | 220 | 101 | 1 | 0 | 158 | 132 | 54 | 1 | 0 |
| | (80, 100] | 171 | 136 | 48 | 1 | 0 | 312 | 267 | 159 | 16 | 1 |
| 900 | [0, 5] | 8121 | 7138 | 1773 | 2 | 0 | 8495 | 7512 | 2221 | 36 | 0 |
| | (5, 10] | 408 | 939 | 5442 | 57 | 0 | 272 | 804 | 5477 | 202 | 0 |
| | (10, 20] | 568 | 1045 | 2451 | 9426 | 2999 | 414 | 886 | 2067 | 9610 | 5925 |
| | (20, 40] | 444 | 529 | 209 | 514 | 7001 | 319 | 388 | 45 | 148 | 4075 |
| | (40, 80] | 293 | 226 | 87 | 1 | 0 | 184 | 154 | 51 | 0 | 0 |
| | (80, 100] | 166 | 123 | 38 | 0 | 0 | 316 | 256 | 139 | 4 | 0 |

περισσότερα υποφαινομενικά αποτελέσματα θα εμφανίζονται στην ανάλυση όσο αυξάνεται η τιμή της αυτοπαλίνδρομης παραμέτρου μια συμπεριφορά που είναι όμοια στην ανάλυση χρονοσειρών και στη χωρική ανάλυση σε ολικό επίπεδο. Ωστόσο, για υψηλές τιμές της χωρικά αυτοπαλίνδρομης παραμέτρου το πρόβλημα της υποφαινομενικής συμπεριφοράς θα υπάρχει πάντα καθώς, όπως φαίνεται στον Πίνακα 3.12, σχεδόν το 100% των τοπικών ποσοστών απορρίψεων ανήκει σε διαστήματα ανώτερα του διαστήματος [0, 5]. Ισοδύναμα, για διαδικασίες λευκών θορύβου και για χαμηλές και μέτριες τιμές της αυτοπαλίνδρομης παραμέτρου, αντίστοιχα, 20%, 30% και περίπου 60% των τοπικών ποσοστών απορρίψεων ανήκουν σε διαστήματα ανώτερα του διαστήματος [0, 5], φανερώνοντας λιγότερη υποφαινομενική συμπεριφορά. Το μέγεθος του δείγματος επηρεάζει τη συμπεριφορά των ελέγχων μόνο για μέτριες και υψηλές τιμές της χωρικά αυτοπαλίνδρομης παραμέτρου αλλάζοντας την πυκνότητα της κατανομής της συχνότητας. Με άλλα λόγια, καθώς αυξάνει το μέγεθος του δείγματος η συχνότητα των τοπικών ποσοστών απορρίψεων συγκεντρώνεται σε συγκεκριμένα διαστήματα.

Ως επακόλουθο, αξίζει να διερευνηθεί η συμπεριφορά των ελέγχων με τις ψευδό-*t* στατιστικές όσον αφορά το επικρατέστερο διάστημα της συχνότητας των τοπικών ποσοστών απορρίψεων. Για διαδικασίες λευκού θορύβου και για μικρές τιμές της αυτοπαλίνδρομης

παραμέτρου επικρατέστερο είναι το διάστημα $[0, 5]$, ενώ για τιμές του $\varphi = 0,5$ και $0,8$ επικρατέστερο είναι το διάστημα $(5, 10]$ και $(10, 20]$ αντίστοιχα, ανεξάρτητα από τη συνάρτηση στάθμισης. Εν' τούτοις, για $\varphi = 0,9$ το επικρατέστερο διάστημα εξαρτάται από τη συνάρτηση στάθμισης και είναι το $(20, 40]$ χρησιμοποιώντας τη διτετράγωνη συνάρτηση στάθμισης και το $(10, 20]$ χρησιμοποιώντας τη Γκαουσιανή συνάρτηση στάθμισης, φανερώνοντας ότι περισσότερες υποφαινομενικές ενδείξεις θα εμφανίζονται με την εφαρμογή διτετράγωνου πυρήνα στη συνάρτηση στάθμισης. Επομένως, το επικρατέστερο διάστημα της συχνότητας των τοπικών ποσοστών απορρίψεων θα αυξάνεται όταν αυξάνεται η τιμή της χωρικά αυτοπαλίνδρομης παραμέτρου ανεξάρτητα του πυρήνα στη συνάρτηση στάθμισης.

Είναι πιθανόν μια λογική εξήγηση της επίδοσης του ελέγχου τοπική γραμμικής σχέσης να οφείλεται στις τιμές του εύρους. Όπως αναφέρθηκε προηγουμένως, το εύρος μειώνεται όταν αυξάνεται η τιμή της χωρικά αυτοπαλίνδρομης παραμέτρου. Κατά συνέπεια, ο μεγάλος αριθμός τοπικών απορρίψεων προκύπτει ταυτόχρονα με πολύ χαμηλές τιμές του εύρους, δείχνοντας ότι τα αποτελέσματα είναι ακόμη πιο ύποπτα, επειδή το χαμηλό εύρος προκαλεί πλασματική μεταβλητότητα που επηρεάζει την εφαρμογή του ελέγχου. Πραγματικά, όταν το εύρος είναι πολύ χαμηλό, το υπόδειγμα εκτιμάται σε τοπικό επίπεδο με πολύ μικρό αριθμό γειτονικών σημείων κάτι που προκαλεί αστάθεια στις εκτιμήσεις των παραμέτρων και συνεπώς αναξιόπιστους ελέγχους, λαμβάνοντας μάλιστα υπ' όψιν, ότι αυτοί οι έλεγχοι βασίζονται σε ψευδό- t στατιστικές που δεν τεκμηριώνεται μαθηματικά η εφαρμογή τους.

Τελειώνοντας, τα εκτιμηθέντα σφάλματα από τη γεωγραφικά σταθμισμένη παλινδρόμηση εντοπίζονται να παρουσιάζουν χωρική αυτοσυσχέτιση σύμφωνα με τη στατιστική I του Moran, που υπολογίστηκε με το κριτήριο του πύργου, και την εφαρμογή του ελέγχου με την προσέγγιση των Leung, Mei και Zhang (2000b), κάτι που ισχύει και στην ανάλυση χρονοσειρών και στη χωρική ανάλυση σε ολικό επίπεδο. Όπως διαπιστώνεται από τον Πίνακα 3.13, ο οποίος περιγράφει τις μέσες τιμές της στατιστικής I που προέκυψαν από τις 10.000 επαναλήψεις και τα ποσοστά απόρριψης του ελέγχου της μηδενικής υπόθεσης απουσίας χωρικής αυτοσυσχέτισης στα εκτιμηθέντα σφάλματα, η υποφαινομενική συμπεριφορά συνδέεται με χωρική αυτοσυσχέτιση στα σφάλματα. Συγκεκριμένα, περισσότερες ενδείξεις χωρικής αυτοσυσχέτισης στα σφάλματα θα εμφανίζονται όταν η τιμή της χωρικά αυτοπαλίνδρομης παραμέτρου αυξάνεται για δεδομένο δείγμα καθώς και όταν

Πίνακας 3.13

Μέσες τιμές στατιστικής I και ποσοστά απόρριψης σε επίπεδο σημαντικότητας 5% με την τεχνική των Leung, Mei και Zhang (2000b) του ελέγχου ύπαρξης χωρικής αυτοσυσχέτισης στα εκτιμηθέντα σφάλματα της μεθόδου της γεωγραφικά σταθμισμένης παλινδρόμησης για δύο ανεξάρτητες SAR(1) διαδικασίες

| Πυρήνας | | Διτετράγωνος | | | | | Γκαουσιανός | | | | |
|---------|----------------|--------------|-------|-------|-------|-------|-------------|-------|-------|-------|-------|
| N | φ | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 | 0,0 | 0,2 | 0,5 | 0,8 | 0,9 |
| 100 | Μέση Τιμή I | -0,034 | 0,038 | 0,056 | 0,018 | 0,032 | -0,027 | 0,049 | 0,064 | 0,014 | 0,032 |
| | Ποσ. Απόρριψης | 3,75 | 32,31 | 91,53 | 99,64 | 99,92 | 3,85 | 33,61 | 92,4 | 99,63 | 99,98 |
| 400 | Μέση Τιμή I | -0,008 | 0,084 | 0,075 | 0,030 | 0,053 | -0,006 | 0,086 | 0,069 | 0,011 | 0,033 |
| | Ποσ. Απόρριψης | 4,3 | 87,59 | 100 | 100 | 100 | 4,42 | 88,15 | 100 | 100 | 100 |
| 900 | Μέση Τιμή I | -0,003 | 0,094 | 0,069 | 0,034 | 0,057 | -0,003 | 0,094 | 0,062 | 0,011 | 0,033 |
| | Ποσ. Απόρριψης | 4,56 | 99,59 | 100 | 100 | 100 | 4,53 | 99,57 | 100 | 100 | 100 |

αυξάνεται το μέγεθος του δείγματος για δεδομένη τιμή της χωρικά αυτοπαλινδρομης παραμέτρου ιδιαίτερα για μικρές τιμές της.¹⁸

Στην ενότητα αυτή εξετάστηκε με τη βοήθεια προσομοίωσης η εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης όταν οι μεταβλητές που εμπλέκονται στο υπόδειγμα είναι ανεξάρτητες μεταξύ τους. Το κίνητρο της προσομοίωσης ήταν η πρακτική πολλών ερευνητών, σε εμπειρικές εργασίες, να περιλαμβάνουν σε τοπικό επίπεδο μεταβλητές που έχουν μη σημαντικές επιδράσεις στο ολικό επίπεδο ελπίζοντας ότι θα γίνουν σημαντικές σε κάποιες περιοχές. Η στρατηγική αυτή εμπεριέχει ακόμη μεγαλύτερο κίνδυνο καθώς η απόφαση σχετικά με το εάν οι επιπλέον μεταβλητές επηρεάζουν σημαντικά την εξαρτημένη μεταβλητή εκτελείται με ψευδό- t στατιστικές. Επομένως, παρουσιάζει ιδιαίτερο ενδιαφέρον να διερευνηθεί το ποσοστό που μια ανεξάρτητη μεταβλητή εντοπίζεται να είναι στατιστικά σημαντική σε τοπικό επίπεδο, λαμβάνοντας υπ' όψιν ότι μια μεταβλητή παραμένει στο υπόδειγμα όταν το ποσοστό των περιοχών στις οποίες είναι στατιστικά σημαντική υπερβαίνει το 5% των συνολικών περιοχών στο δείγμα.

Για το λόγο αυτό, θεωρήθηκαν δύο ανεξάρτητες μεταβλητές τόσο σε ολικό όσο και σε τοπικό επίπεδο που παρήχθησαν ως SAR(1) διαδικασίες με σκοπό να εξεταστεί εάν εμφανίζεται γραμμική σχέση μεταξύ τους σε τοπικό επίπεδο. Δυστυχώς, η συμπεριφορά των ψευδό- t στατιστικών για τον έλεγχο της μηδενικής υπόθεσης μη ύπαρξης σχέσης δεν ήταν καθόλου ικανοποιητική διότι οδήγησε σε ισχυρότερες ενδείξεις υποφαινομενικής συμπεριφοράς σε τοπικό επίπεδο από ότι εμφανίζεται στο ολικό επίπεδο, όπως έχουν

¹⁸ Αξίζει να αναφερθεί ότι η εφαρμογή του ελέγχου χωρικής αυτοσυσχέτισης με τον κλασικό τρόπο, υποθέτοντας τυχαιοποίηση, αποδίδει μικρότερα ποσοστά απόρριψης από αυτά που παρουσιάζονται στον Πίνακα 3.13. Ωστόσο, επειδή η συμπεριφορά του είναι παράξενη και δύσκολο να ερμηνευθεί, τα ποσοστά απόρριψης δεν περιλαμβάνονται στον πίνακα.

επισημάνει ο Fingleton (1999) και οι Agiakloglou, Tsimbos και Tsimpanos (2015), άσχετα από τον πυρήνα στη συνάρτηση στάθμισης. Το πρόβλημα γίνεται ακόμη πιο σοβαρό επειδή όλοι οι έλεγχοι βελτίωσης, οι συντελεστές προσδιορισμού και οι τιμές του εύρους υποδεικνύουν εκτίμηση σε τοπικό επίπεδο.

Από τα αποτελέσματα της προσομοίωσης προέκυψε επιπρόσθετα, ότι το φαινόμενο της τοπική υποφαινομενικής παλινδρόμησης (*local spurious regression*), το οποίο παρουσιάζει παρόμοια συμπεριφορά με την ανάλυση χρονοσειρών και τη χωρική ανάλυση για δύο ανεξάρτητες AR(1) και SAR(1) διαδικασίες αντίστοιχα, σχετίζεται επίσης και με την εμφάνιση χωρικής αυτοσυσχέτισης στα σφάλματα. Παρά το γεγονός ότι οι μέσες τιμές της στατιστικής I του Moran είναι χαμηλές, τα ποσοστά απόρριψης του ελέγχου της μηδενικής υπόθεσης της απουσίας χωρικής αυτοσυσχέτισης με την προσέγγιση των Leung, Mei και Zhang (2000b) είναι πολύ υψηλά και έχουν σχέση με την τιμή της χωρικά αυτοπαλινδρομης παραμέτρου. Ωστόσο, η βασική διαφορά σε σχέση με την ανάλυση χρονοσειρών και τη χωρική ανάλυση σε ολικό επίπεδο είναι ότι ο ερευνητής δεν έχει διέξοδο να αποφύγει αυτή την παγίδα. Συγκεκριμένα, στην ανάλυση χρονοσειρών ο ερευνητής μπορεί να εφαρμόσει κάποια από τις εναλλακτικές μορφές εκτίμησης του υποδείγματος που έχουν προταθεί από τον Agiakloglou (2013) για να εξαλείψει την αυτοσυσχέτιση και την υποφαινομενικότητα. Αντίστοιχα, στη χωρική ανάλυση, οι Agiakloglou, Tsimbos και Tsimpanos (2015) έχουν δείξει ότι η εφαρμογή των κλασικών LM ελέγχων χωρικής εξάρτησης και εξειδίκευσης προτείνουν άλλες μορφές εκτίμησης του υποδείγματος που δεν οδηγούν σε υποφαινομενικά αποτελέσματα και σε χωρική αυτοσυσχέτιση στα σφάλματα.

Κατά συνέπεια, η γεωγραφικά σταθμισμένη παλινδρόμηση πρέπει να εφαρμόζεται με προσοχή σε μια χωρική ανάλυση και τα αποτελέσματα που προκύπτουν να ερμηνεύονται πολύ προσεχτικά ακόμη και στην περίπτωση που οι έλεγχοι βελτίωσης δείχνουν ότι η εκτίμηση σε τοπικό επίπεδο είναι καλύτερη. Μεγαλύτερη προσοχή πρέπει να δίνεται, για να αποφευχθούν τοπικές υποφαινομενικές παλινδρομήσεις, όταν υψηλές τιμές του συντελεστή προσδιορισμού της γεωγραφικά σταθμισμένης παλινδρόμησης εμφανίζονται ταυτόχρονα με χαμηλές τιμές του εύρους. Στην πραγματικότητα, οι χαμηλές τιμές του εύρους καθιστούν τη τεχνική πολύ αναξιόπιστη.

3.8 Ανακεφαλαίωση

Στο κεφάλαιο αυτό παρουσιάστηκε το μεθοδολογικό πλαίσιο της γεωγραφικά σταθμισμένης παλινδρόμησης που χρησιμοποιείται ευρέως στις εμπειρικές εφαρμογές για την αντιμετώπιση του προβλήματος της χωρικής ετερογένειας μέσω της εκτίμησης τοπικών υποδειγμάτων. Η φιλοσοφία της τεχνικής είναι η εκτίμηση υποδειγμάτων με διαφορετικούς συντελεστές, για κάθε σημείο του χώρου, χρησιμοποιώντας υποσύνολα του δείγματος που αποτελούνται από κοντινές παρατηρήσεις, με διαφορετικό βαθμό στάθμισης σύμφωνα με τη γεωγραφική τους απόσταση που προσδιορίζεται από την εφαρμογή μιας συνάρτησης στάθμισης. Ταυτόχρονα, η μέθοδος αυτή υπολογίζει τοπικά σφάλματα που χρησιμοποιούνται στον ορισμό ψευδό-*t* στατιστικών και τον έλεγχο της τοπικής στατιστικής σημαντικότητας των συντελεστών. Η αξιολόγηση των αποτελεσμάτων και η επιλογή μεταξύ της κλασικής μεθόδου των ελαχίστων τετραγώνων και της γεωγραφικά σταθμισμένης παλινδρόμησης γίνεται μέσω συντελεστών προσδιορισμού αλλά και με τη βοήθεια στατιστικών ελέγχων που συγκρίνουν τα αθροίσματα των τετραγώνων των εκτιμηθέντων σφαλμάτων των δυο μεθόδων εκτίμησης στη λογική των ελέγχων ανάλυσης διακύμανσης και του πηλίκου δύο διακυμάνσεων.

Στη συνέχεια πραγματοποιήθηκαν μια σειρά από προσομοιώσεις με υποδείγματα μίας ανεξάρτητης μεταβλητής με σκοπό να διερευνηθούν η συμπεριφορά της μεθόδου της γεωγραφικά σταθμισμένης παλινδρόμησης στον εντοπισμό της χωρικής ετερογένειας και των ελέγχων που εξετάζουν τη συνεισφορά της. Τα αποτελέσματα έδειξαν ότι η σύγκριση των συντελεστών προσδιορισμού του ολικού υποδείγματος με αυτόν των τοπικών υποδειγμάτων και της τιμής του εύρους που επιλέγει η διασταυρωτική επικύρωση με το ολικό εύρος μπορούν να βοηθήσουν στην επιλογή της σωστής μεθόδου εκτίμησης και στη διάκριση της χωρικής ετερογένειας από μια σταθερή σχέση. Επιπρόσθετα, οι έλεγχοι συνεισφοράς, παρά τα προβλήματα που δείχνουν ότι έχουν μερικές φορές, μπορούν επικουρικά να χρησιμοποιηθούν για την επιλογή της σωστής μεθόδου εκτίμησης. Εν' τούτοις, σε κάποιες περιπτώσεις τα κριτήρια μπορεί εσφαλμένα να προτείνουν τη γεωγραφικά σταθμισμένη παλινδρόμηση, όπως συμβαίνει με το υπόδειγμα χωρικού καθεστώτος, κάτι που σημαίνει ότι ο ερευνητής θα πρέπει να είναι ιδιαίτερα προσεκτικός όταν εφαρμόζει την τεχνική καθώς δεν μπορεί να αποδειχθεί ότι πάντα οδηγεί στα σωστά συμπεράσματα. Οι εκτιμηθείσες τιμές των συντελεστών των υποδειγμάτων είναι πολύ κοντά στις πραγματικές τους τιμές και κατά συνέπεια η χαρτογράφησή τους μπορεί να συμβάλει στον εντοπισμό των χωρικών προτύπων

και της χωρικής ετερογένειας. Ένα αξιοσημείωτο αποτέλεσμα που προέκυψε από την προσομοίωση είναι η εμφάνιση χωρικής αυτοσυσχέτισης στα εκτιμηθέντα σφάλματα που όμως δεν μπορεί να τεκμηριωθεί αν όντως υπάρχει ή οφείλεται σε ακαταλληλότητα του τρόπου διεξαγωγής των ελέγχων.

Επιπλέον, μελετήθηκαν οι επιπτώσεις από την εκτίμηση με τη γεωγραφικά σταθμισμένη παλινδρόμηση τοπικών υποδειγμάτων όταν στην πραγματικότητα η σχέση εξάρτησης μεταξύ των δύο μεταβλητών είναι σε ολικό επίπεδο αλλά περιλαμβάνει χωρικές επιδράσεις εξάρτησης που αγνοούνται. Τα αποτελέσματα των προσομοιώσεων έδειξαν ότι όταν ισχύει το υπόδειγμα χωρικού σφάλματος, οι συνέπειες από τη λανθασμένη επιλογή μεθόδου εκτίμησης είναι λιγότερο σοβαρές και η διάκριση της εσφαλμένης ταυτοποίησης μπορεί να γίνει πιο εύκολα συγκρίνοντας τις τιμές των συντελεστών προσδιορισμού και, για μέτριες τιμές των λ , χρησιμοποιώντας το εύρος και τα αποτελέσματα των ελέγχων συνεισφοράς. Ωστόσο, οι επιπτώσεις από την εκτίμηση με τη γεωγραφικά σταθμισμένη παλινδρόμηση είναι πολύ πιο σοβαρές όταν ισχύει το υπόδειγμα χωρικής υστέρησης, καθώς η διάκριση της εσφαλμένης ταυτοποίησης είναι δύσκολο να γίνει διότι οι συντελεστές προσδιορισμού, το εύρος και τα αποτελέσματα των ελέγχων υποστηρίζουν την εκτίμηση τοπικών υποδειγμάτων ενώ παρουσιάζεται και μεροληψία στις εκτιμήσεις των τοπικών συντελεστών.

Στην τελευταία ενότητα του κεφαλαίου θεωρήθηκε ότι οι μεταβλητές του υποδείγματος είναι ανεξάρτητες μεταξύ τους τόσο σε ολικό όσο και σε τοπικό επίπεδο με σκοπό να μελετηθεί η συμπεριφορά της γεωγραφικά σταθμισμένης παλινδρόμησης σε αυτή την περίπτωση. Τα αποτελέσματα της προσομοίωσης, όταν εμπλέκονται στην παλινδρόμηση δύο ανεξάρτητες στάσιμες χωρικές αυτοπαλίνδρομες διαδικασίες πρώτης τάξης, SAR(1), φανερώνουν ότι μπορούν να εμφανιστούν τοπικές υποφαινομενικές παλινδρομήσεις (*local spurious regression*), όπως είναι γνωστό ότι συμβαίνει στο ολικό επίπεδο της ανάλυσης, καθώς οι τοπικοί συντελεστές της ανεξάρτητης μεταβλητής είναι στατιστικά σημαντικοί, σύμφωνα με τις ψευδό-*t* στατιστικές, σε αξιοσημείωτο ποσοστό, οι στατιστικοί έλεγχοι υποδεικνύουν ότι η εκτίμηση σε τοπικό επίπεδο είναι καλύτερη ενώ ταυτόχρονα τα σφάλματα παρουσιάζουν χωρική αυτοσυσχέτιση. Η παγίδα αυτή δεν μπορεί να αποφευχθεί, σε αντίθεση με το ίδιο πρόβλημα στο ολικό επίπεδο όπου η εφαρμογή των κλασικών LM ελέγχων χωρικής εξάρτησης και εξειδίκευσης οδηγούν σε εκτιμήσεις χωρικών οικονομετρικών υποδειγμάτων με τα οποία η υποφαινομενική συμπεριφορά αντιμετωπίζεται. Κατά συνέπεια, ο ερευνητής θα πρέπει να είναι ενήμερος ότι το πρόβλημα της υποφαινομενικής

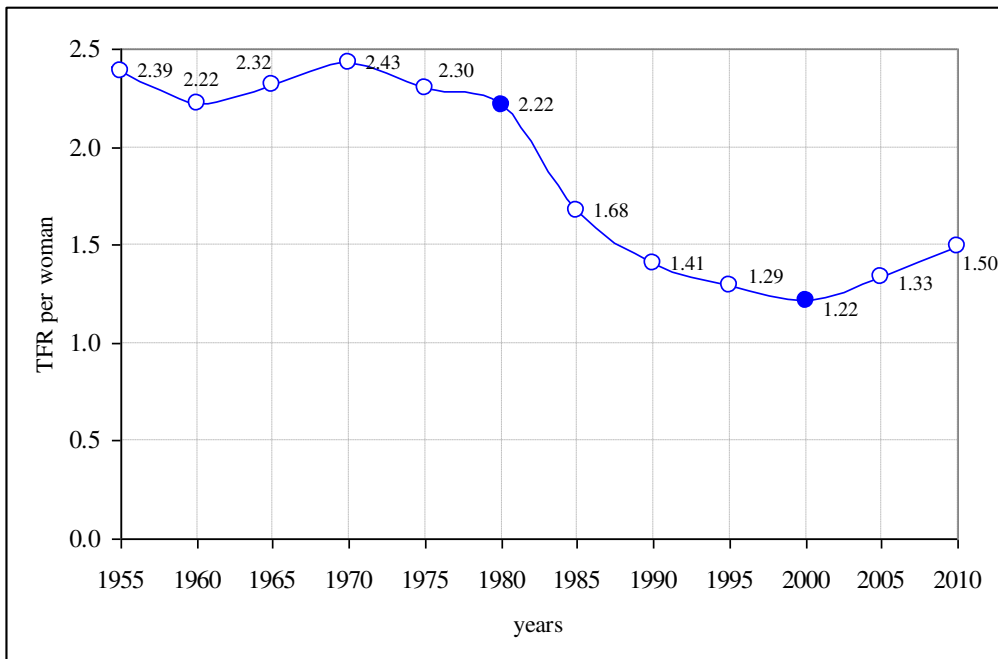
παλινδρόμησης μπορεί να εμφανιστεί και στα τοπικά υποδείγματα που εκτιμώνται με τη μέθοδο της γεωγραφικά σταθμισμένης παλινδρόμησης και να είναι ιδιαίτερα προσεχτικός στην ερμηνεία και αξιολόγηση των αποτελεσμάτων, ιδιαίτερα όταν προκύπτουν υψηλές τιμές για τον συντελεστή προσδιορισμού με ταυτόχρονα πολύ χαμηλές τιμές του εύρους.

ΚΕΦΑΛΑΙΟ 4

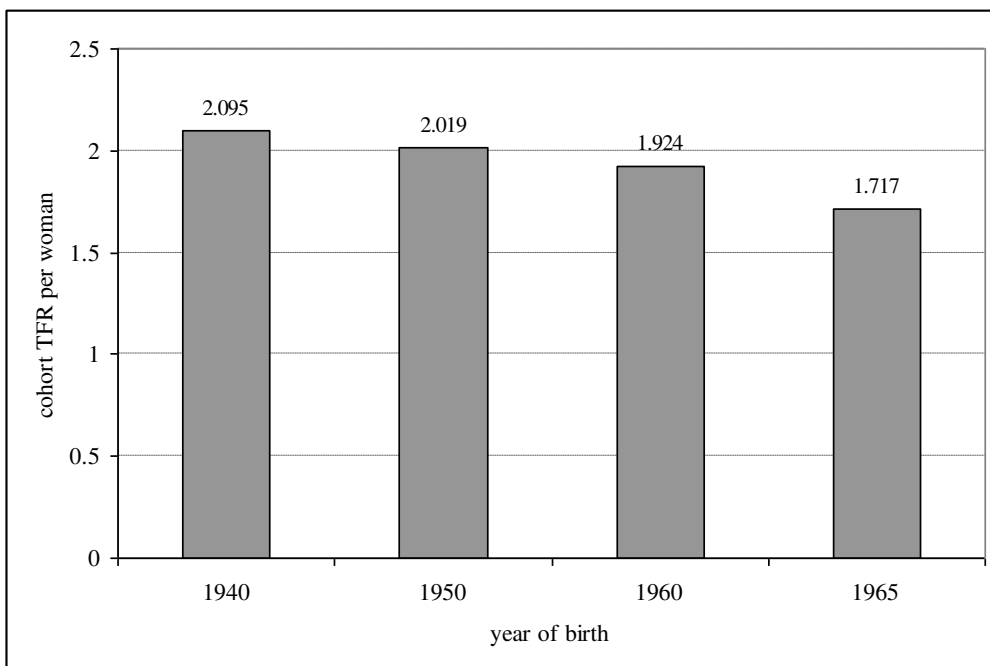
ΔΙΕΡΕΥΝΗΣΗ ΤΩΝ ΧΩΡΙΚΩΝ ΔΙΑΦΟΡΟΠΟΙΗΣΕΩΝ ΤΩΝ ΕΠΙΔΡΑΣΕΩΝ ΣΤΗ ΓΟΝΙΜΟΤΗΤΑ ΤΟΥ ΠΛΗΘΥΣΜΟΥ ΤΗΣ ΕΛΛΑΔΑΣ ΣΕ ΤΟΠΙΚΟ ΔΙΟΙΚΗΤΙΚΟ ΕΠΙΠΕΔΟ

4.1 Εισαγωγή

Κατά τη διάρκεια της μεταπολεμικής περιόδου στην Ελλάδα έχουν παρατηρηθεί σημαντικές μειώσεις στο επίπεδο της γονιμότητας του πληθυσμού ως συνέπεια ουσιαστικών κοινωνικό – οικονομικών αλλαγών και μετασχηματισμών στη στάση, στα πρότυπα και στη συμπεριφορά που αφορούν τη δημιουργία της οικογένειας και την ανατροφή των παιδιών. Η πορεία της γονιμότητας του πληθυσμού της Ελλάδας της περιόδου 1955 – 2010, όπως εκφράζεται από το δείκτη ολικής γονιμότητας (*Total Fertility Rate – TFR*), παρουσιάζεται αναλυτικά στο Διάγραμμα 4.1, από το οποίο διαπιστώνεται ότι μέχρι το 1970 ο δείκτης ολικής γονιμότητας (TFR) κυμαίνονταν γύρω από το επίπεδο των 2,3 παιδιών ανά γυναίκα αλλά από το 1970 ακολουθεί μία αξιοσημείωτη πτωτική τάση που επιταχύνεται τις δεκαετίες του 1980 και του 1990. Το 2000 σημειώνεται η χαμηλότερη τιμή του δείκτη ολικής γονιμότητας (Kohler, Billari και Ortega, 2002) που καταλήγει στο επίπεδο των 1,22 παιδιών ανά γυναίκα. Τα επόμενα χρόνια παρατηρείται μία μέτρια αύξηση της γονιμότητας που κυρίως αποδίδεται στην παρουσία των μεταναστών (Tsimbos, 2008) αλλά οι τιμές του δείκτη παραμένουν κάτω από το επίπεδο πληθυσμιακής αναπλήρωσης των 2,1 παιδιών ανά γυναίκα. Θεωρώντας τη συνολική αξιολόγηση των προτύπων γονιμότητας και τις τάσεις που παρατηρούνται στην Ευρωπαϊκή Ένωση (Bijak, Kupiszewska και Kupiszewski, 2008 και Coleman, 2005) καθώς και των δυναμικών επιδράσεων της οικονομικής κρίσης στην Ελλάδα οι προοπτικές για ανάκαμψη στο επίπεδο πληθυσμιακής αναπλήρωσης κρίνονται δυσοίωνες.

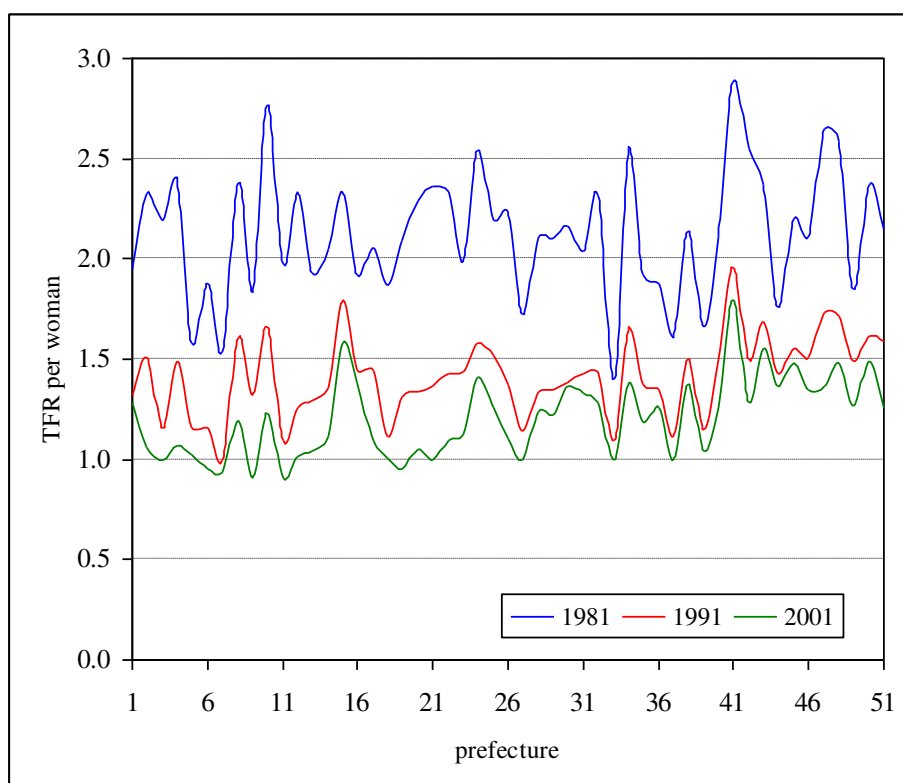


Διάγραμμα 4.1
Δείκτης Ολικής Γονιμότητας (TFR) ανά γυναίκα της Ελλάδας
της περιόδου 1955 – 2010



Διάγραμμα 4.2
Δείκτες Ολικής Γονιμότητας (TFR) ανά γυναίκα της Ελλάδας
για τις κοορτες γεννήσεων των γυναικών της περιόδου 1940 – 1965,
πηγή: Frejka and Sardon (2004)

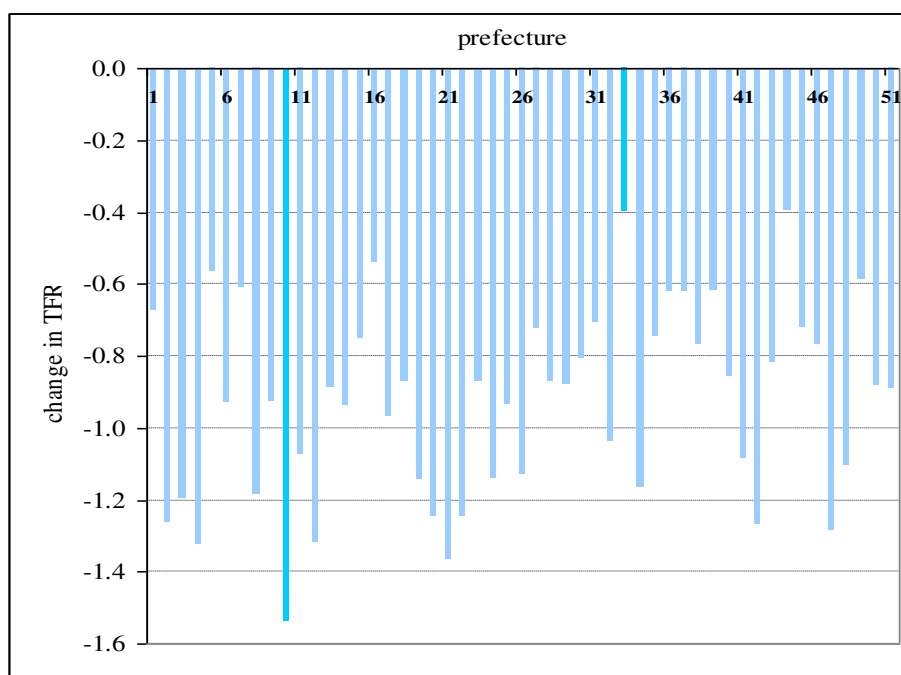
Επιπρόσθετα, όπως φαίνεται στο Διάγραμμα 4.2, η διατήρηση των επιπέδων της γονιμότητας σε χαμηλά επίπεδα για μεγάλο χρονικό διάστημα έχει οδηγήσει σε πτώση του δείκτη ολικής γονιμότητας για τις κοορτές των γεννήσεων των γυναικών της περιόδου 1940 – 1965. Παρόλο που σημάδια δεύτερης δημογραφικής μετάβασης (Lesthaeghe και Kaa, 1986 και Sobotka, 2008) δεν είναι ξεκάθαρα ορατά στην Ελλάδα η αναβολή της μητρότητας και της τεκνοποίησης είναι φανερή. Ο γάμος ως θεσμός δεν είναι τόσο σταθερός όπως ήταν τα παλιότερα χρόνια αλλά το ποσοστό των εξωσυζυγικών γεννήσεων στην Ελλάδα (4% το 2000 και 7% το 2010) παραμένει ακόμη σε χαμηλό επίπεδο σε σχέση με τις άλλες Ευρωπαϊκές χώρες.



Διάγραμμα 4.3
Δείκτες Ολικής Γονιμότητας (TFR) ανά γυναίκα σε επίπεδο νομών
της Ελλάδας για τα έτη 1981, 1991 και 2001

Εξετάζοντας τα δεδομένα της Ελλάδας στο επίπεδο των 51 νομών (NUTS 3) διαπιστώνεται ότι οι μεταβολές της γονιμότητας σε τοπικό επίπεδο, που παρουσιάζονται στο Διάγραμμα 4.3, έχουν συντελεστεί προς τις ίδιες κατευθύνσεις με τις εθνικές τάσεις. Συγκεκριμένα, το έτος 1981 ο δείκτης ολικής γονιμότητας (TFR) σε 30 από τους 51 νομούς της χώρας ήταν επάνω από το επίπεδο πληθυσμιακής αναπλήρωσης αλλά τα έτη 1991 και

2001 σε όλους τους νομούς παρατηρείται γονιμότητα μικρότερη των 2 παιδιών ανά γυναίκα. Ειδικότερα, το 2001 36 από τις 51 διοικητικές περιφέρειες έχουν τα πιο χαμηλά επίπεδα γονιμότητας που έχουν παρατηρηθεί με δείκτη ολικής γονιμότητας μικρότερο του 1,3. Όπως φαίνεται στο Διάγραμμα 4.4, μεταξύ των ετών 1981 και 2001 η πτώση του δείκτη ολικής γονιμότητας κυμαίνεται από 0,4 στο νομό του Κιλκίς στη βόρεια Ελλάδα, έως 1,54 στο νομό Αχαΐας στην Πελοπόννησο. Για όλους τους νομούς η πτώση στους δείκτες γονιμότητας ήταν περισσότερο αισθητή κατά τη διάρκεια της περιόδου 1981 – 1991. Σύγκλιση στις περιφερειακές διαφορές της γονιμότητας έχουν παρατηρηθεί διαχρονικά ως αποτέλεσμα περισσότερων ή λιγότερων κοινών δημογραφικών μεταβολών και ιδεατών αλλαγών στα γεωγραφικά και διοικητικά τμήματα της χώρας.



Διάγραμμα 4.4
Μεταβολές στους Δείκτες Ολικής Γονιμότητας (TFR) ανά γυναίκα
σε επίπεδο νομών της Ελλάδας την περίοδο 1981 – 2001

Τα τελευταία χρόνια έχει γίνει κοινά αποδεκτό από τους ερευνητές ότι χώρος και πληθυσμός είναι στενά συνδεδεμένες έννοιες μέσω μίας έμφυτης πολυπρόσωπης διαδικασίας διασποράς δημογραφικών, κοινωνικό – οικονομικών και πολιτιστικών παραγόντων και αξιών. Οι γεωγραφικές μελέτες που αφορούν τη γονιμότητα έχουν ιδιαίτερο ενδιαφέρον καθώς συμβάλουν στην περιγραφή των τοπικών επιπέδων αναπαραγωγής και πρακτικών, στην ανάχνευση των επιπτώσεων των κοινωνικών αποτελεσμάτων στην τεκνοποίηση και στη

χάραξη πολιτικών σε περιφερειακό επίπεδο. Όμως, μόλις πρόσφατα οι δημογράφοι έχουν αρχίσει να εξετάζουν συστηματικά τα χωρικά πρότυπα της γονιμότητας και να μελετάνε τους βαθύτερους μηχανισμούς που οδηγούν στις περιφερειακές διαφοροποιήσεις στο επίπεδό της. Στην κατεύθυνση αυτή έχουν συμβάλλει, μεταξύ άλλων παραγόντων, η αυξανόμενη διάθεση χωρικών δεδομένων, η τεχνολογική εξέλιξη των ηλεκτρονικών υπολογιστών, η ανάπτυξη λογισμικών χωρικής ανάλυσης και η εισαγωγή εξειδικευμένων στατιστικών τεχνικών (de Castro, 2007).

Παρά το σύντομο χρονικό διάστημα που η χωρική διάσταση της γονιμότητας απασχολεί τους ερευνητές για τις περισσότερες Ευρωπαϊκές χώρες έχουν ήδη δημοσιευτεί εργασίες που αναλύουν τις χωρικές της ιδιαιτερότητες κάτι που δεν έχει γίνει για την Ελλάδα. Στο κεφάλαιο αυτό εξειδικεύεται ένα οικονομετρικό υπόδειγμα που προσπαθεί να ερμηνεύσει τις χωρικές διαφοροποιήσεις και τα πρότυπα στις σχέσεις μεταξύ του επιπέδου της γονιμότητας του πληθυσμού της Ελλάδας και ενός αριθμού επιλεγμένων κοινωνικό – οικονομικών δεικτών σε τοπικό διοικητικό επίπεδο λαμβάνοντας υπ' όψιν την παρουσία της χωρικής ετερογένειας.

4.2 Τα δεδομένα και οι μεταβλητές του υποδείγματος

Τα δεδομένα που χρησιμοποιούνται στην ανάλυση προέρχονται από τα επίσημα στοιχεία που έχουν ανακοινωθεί από την Ελληνική Στατιστική Αρχή (ELSTAT, 2011) για την απογραφή του Ελληνικού πληθυσμού του έτους 2001 και αφορούν τις 325 μονάδες τοπικής αυτοδιοίκησης (Δήμοι) της χώρας. Πρόκειται δηλαδή, για τις νέες διοικητικές μονάδες που έχουν εισαχθεί από το Υπουργείο Εσωτερικών το 2010 στα πλαίσια του προγράμματος Καλλικράτης. Για τις ανάγκες της έρευνας αθροίστηκαν δεδομένα από διαφορετικές χωρικές μονάδες για να προκύψουν οι τιμές των μεταβλητών για το νέο διοικητικό επίπεδο της χώρας αλλά η προέλευσή τους από μία μόνο πηγή εξασφαλίζει ότι στην ανάλυση αποφεύγονται επιδράσεις που προέρχονται από πιθανές αναντιστοιχίες μεταξύ της χωρικής κλίμακας του φαινομένου που μελετάται και τις κλίμακες μέτρησης των μεταβλητών (Anselin, 2001).

Λόγω του μικρού αριθμού ληξιαρχικών στοιχείων σε τοπικό επίπεδο δεν είναι εφικτό να υπολογιστούν ειδικοί κατά ηλικία δείκτες γονιμότητας και κατά συνέπεια δείκτες ολικής γονιμότητας (*Total Fertility Rates – TFR*). Ως εκ τούτου, η γονιμότητα του πληθυσμού

μετρήθηκε με το δείκτη του λόγου του παιδικού προς τον αναπαραγωγικό πληθυσμό *CWR* (*Children Women Ratio*) που ορίζεται ως:

$$CWR = \frac{P_{0-4}}{W_{15-49}} \times 1000$$

όπου με P_{0-4} συμβολίζεται ο αριθμός των παιδιών ηλικίας κάτω των 5 ετών και με W_{15-49} ο αριθμός των γυναικών αναπαραγωγικής ηλικίας, δηλαδή των γυναικών ηλικίας 15 – 49 ετών. Το πλεονέκτημα του δείκτη *CWR* είναι ότι ο υπολογισμός του βασίζεται αποκλειστικά σε στοιχεία απογραφών και αντιπροσωπεύει ένα χρήσιμο μέτρο της γονιμότητας όταν δεν είναι διαθέσιμα ληξιαρχικά δεδομένα για μικρό χωρικό επίπεδο ή όταν οι κατοχυρωμένοι αριθμοί των γεννήσεων είναι αμελητέοι. Εξαιτίας αυτών των χαρακτηριστικών του, ο δείκτης *CWR* έχει σημαντικές εφαρμογές στη χωρική πληθυσμιακή ανάλυση (Rowland, 2006). Το ένα πέμπτο της τιμής του δείκτη μπορεί να θεωρηθεί έμμεση εκτίμηση του γενικού δείκτη γονιμότητας αλλά με κάποια μεροληψία υποεκτίμησης λόγω της παρουσίας της θνησιμότητας. Σύμφωνα με τους Siegel και Swanson (2004) οι μεταβολές στις τιμές του δείκτη *CWR* ανταποκρίνονται στις μεταβολές των ποσοστών τεκνοποίησης όπως εκτιμώνται με τα ληξιαρχικά δεδομένα υποδεικνύοντας ότι υψηλές τιμές του δείκτη δηλώνουν υψηλά επίπεδα γονιμότητας και αντίστροφα.

Παρά την ευκολία στον υπολογισμό και τη χρησιμότητά του στη γεωγραφική πληθυσμιακή έρευνα ο δείκτης *CWR* έχει και ορισμένα μειονεκτήματα, ιδιαίτερα όταν χρησιμοποιείται για συγκρίσεις στο επίπεδο της γονιμότητας. Αυτό οφείλεται στο γεγονός ότι οι διαφοροποιήσεις στις τιμές του δείκτη μπορεί να επηρεάζονται από τη βρεφική και παιδική θνησιμότητα, την εσωτερική μετανάστευση και την πληθυσμιακή δομή των γυναικών που ανήκουν στην αναπαραγωγική ηλικία. Προηγούμενη εμπειρική έρευνα (Guilmoto και Rajan, 2001) έχει δείξει ότι η επίδραση της θνησιμότητας στις τιμές του δείκτη είναι συνήθως περιορισμένη αλλά το αποτέλεσμα της μετανάστευσης δεν είναι πάντα ξεκάθαρο. Επειδή τα ποσοστά της βρεφικής και παιδικής θνησιμότητας είναι πολύ χαμηλά στην Ελλάδα και το αποτέλεσμα της εσωτερικής μετανάστευσης στη γονιμότητα ασήμαντο, οι περιφερειακές διαφορές σε αυτούς τους δύο παράγοντες δεν προσδοκείται να ασκούν ιδιαίτερη επίδραση στο δείκτη. Όμως, οι τιμές του δείκτη μπορεί να επηρεάζονται από τις διαφορές στην ηλικιακή κατανομή των γυναικών αναπαραγωγικής ηλικίας στους τοπικούς πληθυσμούς. Το χαρακτηριστικό αυτό μπορεί να έχει κάποια επίδραση στις εκτιμήσεις του υποδείγματος αλλά σε πολύ μικρό βαθμό καθώς ο συντελεστής μεταβλητότητας της μέσης ηλικίας των γυναικών

στο ηλικιακό κλιμάκιο των 15 – 49 ετών είναι πολύ χαμηλός (0,022). Γενικά, ο δείκτης του λόγου του παιδικού προς τον αναπαραγωγικό πληθυσμό έχει την τάση να υποεκτιμά το πραγματικό επίπεδο γονιμότητας επειδή ο υπολογισμός του βασίζεται μόνο στα παιδιά που έχουν επιβιώσει μέχρι την ημέρα διεξαγωγής της απογραφής. Για το λόγο αυτό, γίνεται η υπόθεση ότι η σχετική μεροληψία του δείκτη προς τα κάτω, που οφείλεται στη θνησιμότητα και σε στατιστικά σφάλματα, είναι η ίδια σε όλες τις γεωγραφικές περιοχές που εμπλέκονται στην ανάλυση.

Ως ερμηνευτικές μεταβλητές του επιπέδου της γονιμότητας θεωρήθηκαν τέσσερις κοινωνικό – οικονομικοί δείκτες. Συγκεκριμένα οι μεταβλητές που επιλέχθηκαν είναι:

1) Ο αριθμός των παντρεμένων γυναικών ηλικίας 15 – 49 ετών ανά 100 γυναίκες αναπαραγωγικής ηλικίας. Σύμφωνα με τα επίσημα στοιχεία της Ελληνικής Στατιστικής Αρχής (ELSTAT, 2011) οι γεννήσεις εκτός γάμου αντιπροσωπεύουν ένα πολύ μικρό ποσοστό του συνολικού αριθμού γεννήσεων που για τη περίοδο 2000 – 2002 υπολογίζεται σε 3%. Κατά συνέπεια, εφόσον οι εκτός γάμου γεννήσεις είναι αμελητέες το ποσοστό των παντρεμένων γυναικών αναμένεται ότι επηρεάζει θετικά τη γονιμότητα.

2) Ο αριθμός των μεταναστών ανά 100 κατοίκους για κάθε μονάδα τοπικής αυτοδιοίκησης. Όπως έχει παρουσιάσει ο Tsimbos (2006 και 2008) από το 1990 οι μετανάστες αποτελούν την κύρια δύναμη καθοδήγησης των πληθυσμιακών αλλαγών στην Ελλάδα που επηρεάζει την ανάπτυξη και την ηλικιακή δομή του πληθυσμού και συνεισφέρει στον αριθμό των γεννήσεων. Αναμένεται ότι το ποσοστό των μεταναστών θα ασκεί θετική επίδραση στο δείκτη του λόγου του παιδικού προς τον αναπαραγωγικό πληθυσμό.

3) Για να εκτιμηθεί στατιστικά το κοινωνικό – οικονομικό περιβάλλον σε κάθε περιοχή υπολογίστηκε ένας συνθετικός δείκτης χαμηλών κοινωνικό – οικονομικών συνθηκών (LSE) με στοιχεία που είναι διαθέσιμα σε επίπεδο τοπικής αυτοδιοίκησης. Η μεθοδολογία που χρησιμοποιήθηκε στην κατασκευή του δείκτη είναι παρόμοια με αυτή του δείκτη ανθρώπινης φτώχειας (*Human Poverty Index*) που έχει προταθεί στο Πρόγραμμα Ανάπτυξης του Οργανισμού Ηνωμένων Εθνών UNDP (1997 και 2006). Ειδικότερα, ο δείκτης υπολογίστηκε ως η μέση τιμή τριών βασικών μεταβλητών, δηλαδή το ποσοστό του πληθυσμού που δεν γνωρίζει γραφή και ανάγνωση, του ποσοστού της ανεργίας και την πιθανότητα ενός μωρού στη γέννησή του να μην επιβιώσει μέχρι την ηλικία των 60 ετών. Για την κατασκευή του δείκτη χρησιμοποιήθηκαν στοιχεία από την απογραφή του πληθυσμού για τον αριθμό των ανέργων και των ατόμων που δεν γνωρίζουν γραφή και ανάγνωση και πιθανότητες επιβίωσης

για τις Ελληνικές μονάδες τοπικής αυτοδιοίκησης που έχουν παρουσιαστεί από τους Tsimbos, Kalogirou και Verropoulou (2011). Σύμφωνα με προηγούμενη έρευνα για την Ελλάδα (Symeonidou *et al.*, 1997) αναμένεται ο δείκτης χαμηλών κοινωνικό – οικονομικών συνθηκών να συσχετίζεται με υψηλά επίπεδα γονιμότητας.

4) Ο αριθμός των πυρηνικών οικογενειών ανά 100 νοικοκυριά.¹ Επειδή δεν υπάρχουν λεπτομερειακά στοιχεία για τα νοικοκυριά τρίτης γενιάς ή για τις δομές των νοικοκυριών εισάγεται αυτή η μεταβλητή για να εκφράσει τον τύπο των οικογενειακών δομών στους τοπικούς πληθυσμούς. Οι πυρηνικές οικογένειες αποτελούν το νέο τύπο οικογενειών ιδιαίτερα στις πόλεις. Το ποσοστό των πυρηνικών οικογενειών, λόγω των εργασιακών περιορισμών των μητέρων στη φροντίδα των παιδιών και γενικότερα λόγω της οικονομικής κατάστασης στις οικογένειες, όπως έχει παρουσιαστεί και στους Burch και Gendell (1970) και Presser και Baldwin (1980), πιστεύεται ότι επηρεάζει αρνητικά το επίπεδο της γονιμότητας.

Η επιλογή των παραπάνω μεταβλητών βασίστηκε στη διαθεσιμότητα και αξιοπιστία των στοιχείων και στη σημασία τους στο Ελληνικό κοινωνικό και πολιτιστικό περιβάλλον. Εκτός από αυτές τις μεταβλητές, στα πλαίσια της διερευνητικής ανάλυσης που προηγήθηκε, εξετάστηκαν ένας αριθμός πρόσθετων δημογραφικών δεικτών αλλά η ενσωμάτωσή τους στο υπόδειγμα οδήγησε σε ασαφή ή μη στατιστικά σημαντικά αποτελέσματα. Μεταξύ αυτών των δεικτών αξίζει να αναφερθεί ότι παρόλο που στις εργασίες των Perez (1991), Pandit (1992), White, Moreno και Guo (1995), Hank (2001) και Kulu (2005) εντοπίζονται σχέσεις ανάμεσα στη γονιμότητα και τη γεωγραφική κινητικότητα, για την Ελλάδα η εσωτερική μετανάστευση αποδείχτηκε ανεπαρκής στο να εξηγήσει τη μεταβλητότητα της γονιμότητας και για το λόγο αυτό δεν περιλήφθηκε στο τελικό υπόδειγμα. Επιπρόσθετα, η ύπαρξη νηπιαγωγείων και παιδικών σταθμών δεν φαίνεται ότι επηρεάζει τα επίπεδα της γονιμότητας ενώ η διαθεσιμότητά τους θα έπρεπε να αποτελεί παράγοντα ενθάρρυνσης της τεκνοποίησης όπως έχει αναφερθεί στον Hank (2001) για την περίπτωση της Γερμανίας. Οι μεταβλητές που δεν συνεισφέρουν στην ερμηνεία της γονιμότητας και δεν βελτιώνουν τα αποτελέσματα της εκτίμησης του τελικού υποδείματος δεν περιλήφθηκαν σε αυτό παρά τη δημογραφική τους σημασία σύμφωνα με τη βιβλιογραφία.

¹ Με τον όρο πυρηνικές οικογένειες (*nuclear families*) εννοούνται οι οικογένειες που αποτελούνται αποκλειστικά από τους γονείς και τα παιδιά χωρίς την παρουσία παππού ή γιαγιάς.

4.3 Η εφαρμοζόμενη μέθοδος χωρικής ανάλυσης

Οι πέντε μεταβλητές του υποδείγματος βασίζονται σε στατιστικά στοιχεία που έχουν συλλεχθεί από τις διαφορετικές διοικητικές μονάδες της χώρας γεγονός που σημαίνει ότι στην εκτίμησή του δεν θα πρέπει να αγνοηθεί η παρουσία των χωρικών επιδράσεων, δηλαδή της χωρικής εξάρτησης και της χωρικής ετερογένειας που εμφανίζονται με αυτό τον τύπο στατιστικών δεδομένων. Καθώς ο κύριος σκοπός της ανάλυσης είναι η ερμηνεία των χωρικών διαφοροποιήσεων της γονιμότητας σε σχέση με τους επιλεγμένους κοινωνικό – οικονομικούς παράγοντες ανά δημοτική περιφέρεια, η μέθοδος που υιοθετήθηκε για την εκτίμηση του υποδείγματος είναι η γεωγραφικά σταθμισμένη παλινδρόμηση (*Geographically Weighted Regression – GWR*) που έχει προταθεί από τους Fotheringham, Brunson και Charlton (2002). Η τεχνική αυτή εκτιμά τοπικά υποδείγματα (*local models*) σε κάθε γεωγραφικό σημείο επιτρέποντας τη διερεύνηση για κάθε δήμο χωριστά της εξάρτησης της γονιμότητας από τις ερμηνευτικές μεταβλητές αντιμετωπίζοντας τις επιδράσεις της χωρικής ετερογένειας ενώ ταυτόχρονα, έχει ευεργετικά αποτελέσματα και στην ύπαρξη χωρικής αυτοσυσχέτισης στα σφάλματα, όπως αναφέρουν οι Brunson, Fotheringham και Charlton (1999).

Η ανάλυση ξεκινάει με τη χαρτογράφηση και τον υπολογισμό των περιγραφικών στατιστικών μέτρων για τις μεταβλητές του υποδείγματος και με τον έλεγχο για χωρική αυτοσυσχέτιση των τιμών τους, με σκοπό να διαπιστωθεί η παρουσία των χωρικών επιδράσεων. Στη συνέχεια, εκτιμάται ένα απλό οικονομετρικό υπόδειγμα από το οποίο προκύπτουν ως βάσεις σύγκρισης οι ολικές (*global*) τιμές των συντελεστών και με την εφαρμογή των κατάλληλων οικονομετρικών ελέγχων συμπεραίνονται οι αδυναμίες του. Το τελευταίο στάδιο της ανάλυσης αποτελείται από την εκτίμηση τοπικών (*local*) υποδειγμάτων σε κάθε δήμο με τη μέθοδο της γεωγραφικά σταθμισμένης παλινδρόμησης με συνάρτηση στάθμισης προσαρμοσμένου διτετράγωνου πυρήνα (*adaptive bi-squared weight function*) και τη χαρτογράφηση των αποτελεσμάτων για την εξαγωγή των συμπερασμάτων. Παρόμοια προσέγγιση για τη μελέτη κοινωνικό – οικονομικών φαινομένων στο γεωγραφικό χώρο έχει ακολουθηθεί στην εργασία των Kalogirou και Hatzichristos (2007) για την ανάλυση του εισοδήματος στο Δήμο της Αθήνας και στις εργασίες των Isik και Pinarcioglu (2006) και Muniz (2009) για τα επίπεδα της γονιμότητας στην Τουρκία και τη Βραζιλία, αντίστοιχα.

Ένα πρόβλημα που προέκυψε κατά τη διαδικασία εκτίμησης των τοπικών υποδειγμάτων ήταν η εμφάνιση τοπικής πολυσυγγραμμικότητας (*local multicollinearity*) των μεταβλητών

με το σταθερό όρο. Η τοπική πολυσυγγραμμικότητα αυτής της αιτίας οφείλεται στην απουσία τοπικής μεταβλητότητας σε μία ή περισσότερες από τις ανεξάρτητες μεταβλητές για κάποιες γειτονικές γεωγραφικά περιοχές λόγω του χαρακτηριστικού του χώρου ότι, μερικές φορές, οι κοινωνικό – οικονομικές συνθήκες είναι παρόμοιες σε πολύ κοντινούς δήμους. Στην περίπτωση αυτή, ο αλγόριθμος στο λογισμικό εκτίμησης προσπαθεί να εντοπίσει το μέγιστο εύρος (*bandwidth*), μέσα από τη διασταυρωτική επικύρωση (*cross validation*) ή την ελαχιστοποίηση του διορθωμένου πληροφοριακού κριτηρίου του Akaike, επιλέγοντας τοπικά δείγματα στα οποία οι τιμές κάποιας από τις ανεξάρτητες μεταβλητές είναι σχεδόν σταθερές προκαλώντας την εμφάνιση του φαινομένου της πολυσυγγραμμικότητας με το σταθερό όρο στα τοπικά υποδείγματα. Κατά συνέπεια, ο αλγόριθμος εκτίμησης αδυνατεί να συγκλίνει και να προσδιορίσει το μέγιστο εύρος. Παρόλο που το πρόβλημα της πολυσυγγραμμικότητας στα τοπικά υποδείγματα που εκτιμώνται με την τεχνική της γεωγραφικά σταθμισμένης παλινδρόμησης έχει επισημανθεί από τους Wheeler και Tiefelsdorf (2005) και τον Wheeler (2007) δεν έχει, μέχρι στιγμής, καθιερωθεί κάποια μεθοδολογία για να αντιμετωπιστεί και οι προτεινόμενες διαδικασίες επίλυσης δεν εύκολα εφαρμόσιμες και δεν περιλαμβάνονται στα λογισμικά χωρικής ανάλυσης.

Ένας πρακτικός τρόπος αντιμετώπισης της πολυσυγγραμμικότητας αυτής της αιτίας στα τοπικά υποδείγματα είναι ο μετασχηματισμός των ανεξάρτητων μεταβλητών αφαιρώντας τις μέσες τιμές τους όπως συχνά γίνεται για την εκτίμηση ενός απλού οικονομετρικού υποδείγματος με τη μέθοδο των ελαχίστων τετραγώνων (Rawling, Pantula και Dickey, 1998) όταν υπάρχει το ίδιο πρόβλημα. Ο μετασχηματισμός αυτός καθιστά τα διανύσματα των ανεξάρτητων μεταβλητών ορθογώνια σε σχέση με το σταθερό όρο και απαλύνει τις συνέπειες της πολυσυγγραμμικότητας, ενώ ταυτόχρονα οι εκτιμηθείσες τιμές των παραμέτρων και τα τυπικά τους σφάλματα δεν αλλάζουν. Επειδή η μέθοδος της γεωγραφικά σταθμισμένης παλινδρόμησης αποτελεί επέκταση της κλασσικής παλινδρόμησης που εκτιμάται με τη διαδικασία των ελαχίστων τετραγώνων, ο μετασχηματισμός των ανεξάρτητων μεταβλητών σε αποκλίσεις από τις μέσες τιμές τους αναμένεται να διορθώσει το πρόβλημα της πολυσυγγραμμικότητας και στα τοπικά υποδείγματα. Για το λόγο αυτό, οι εκτιμήσεις των τοπικών υποδειγμάτων αλλά και του απλού υποδείγματος, για να υπάρχει απόλυτη συγκρισιμότητα των αποτελεσμάτων, γίνονται με τις τέσσερις ανεξάρτητες μεταβλητές να έχουν μετασχηματιστεί με αφαίρεση των μέσων τιμών τους.

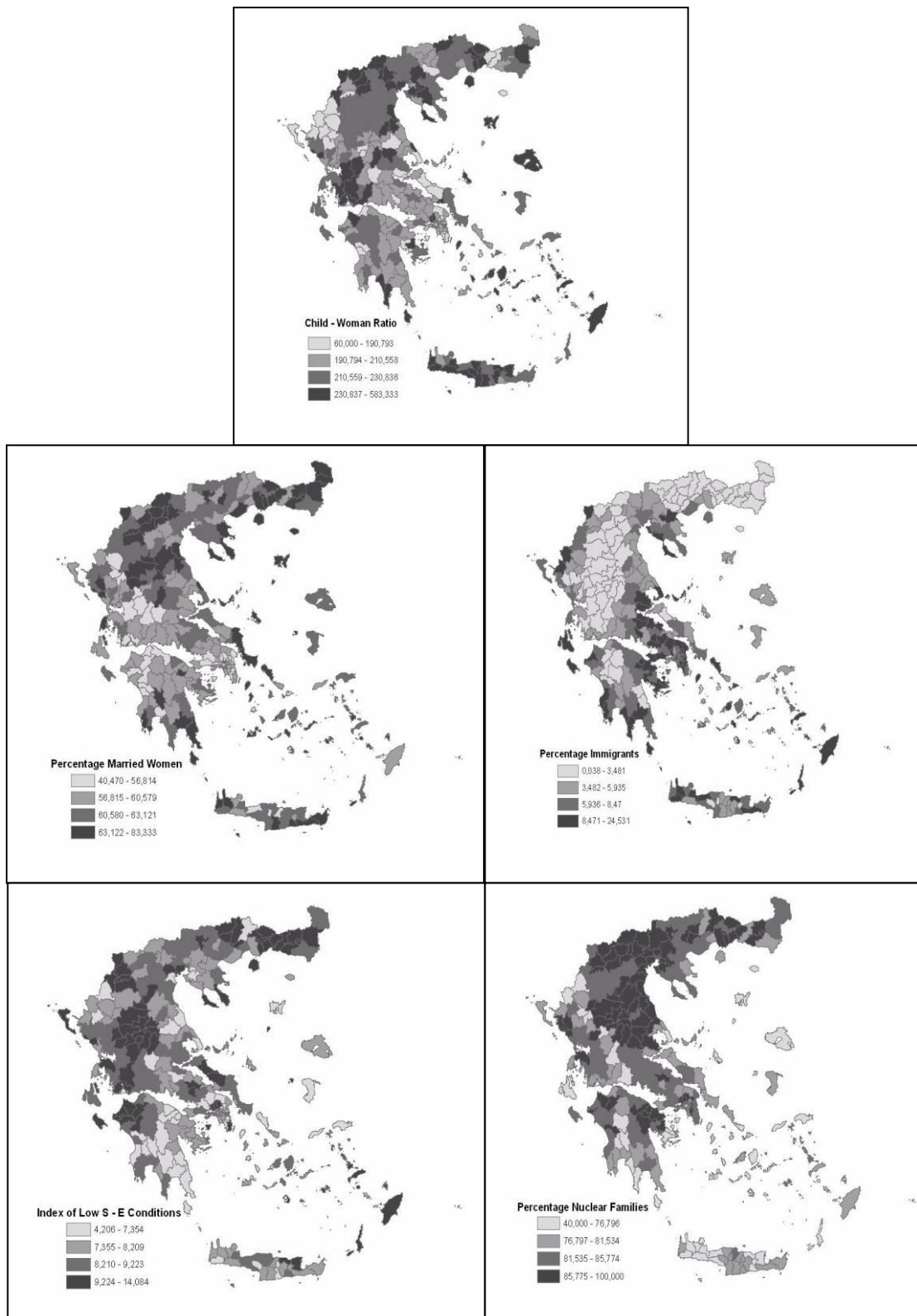
4.4 Περιγραφική στατιστική παρουσίαση των μεταβλητών

Ο Πίνακας 4.1 συνοψίζει τα βασικά περιγραφικά στατιστικά μέτρα των μεταβλητών που περιλαμβάνονται στην ανάλυση και το Διάγραμμα 4.5 απεικονίζει τις χωρικές τους κατανομές σε χάρτες ποσοστιαίων σημείων (*quantile maps*).² Όπως διαπιστώνεται όλες οι μεταβλητές παρουσιάζουν αξιοσημείωτες χωρικές διαφοροποιήσεις. Συγκεκριμένα, οι τιμές του δείκτη *CWR* κυμαίνονται από 60 μέχρι 583 παιδιά ανά 1000 γυναίκες αναπαραγωγικής ηλικίας. Οι μεγαλύτερες τιμές του δείκτη (210 ή υψηλότερες) εντοπίζονται κυρίως στη Βόρεια Ελλάδα, σε περιοχές της Κεντρικής Ελλάδας, στην Κρήτη και στα περισσότερα νησιά της χώρας. Ποσοστά παντρεμένων γυναικών μεγαλύτερα του εθνικού μέσου όρου του 60% παρατηρούνται στη Βόρεια Ελλάδα και στην Κρήτη. Οι περιοχές της χώρας με ποσοστά πυρηνικών οικογενειών μικρότερα του εθνικού μέσου όρου (81%) είναι κυρίως τα νησιά του Αιγαίου και Ιονίου πελάγους, στην Κρήτη και σε δήμους της Πελοποννήσου και της Κεντρικής Ελλάδας. Τα ποσοστά των παντρεμένων γυναικών και των πυρηνικών οικογενειών παρουσιάζουν από όλες τις μεταβλητές του υποδείγματος τους χαμηλότερους συντελεστές μεταβλητότητας που υπολογίζονται κοντά στο 9%. Αντίστοιχα, το ποσοστό των μεταναστών έχει το μεγαλύτερο συντελεστή μεταβλητότητας (62%) που είναι συγκεντρωμένοι κυρίως στις πόλεις. Το φτωχότερο κοινωνικό και οικονομικό περιβάλλον, όπως μετρείται με τον προτεινόμενο συνθετικό δείκτη, εμφανίζεται στη βορειανατολική πλευρά της χώρας, δηλαδή στην περιφέρεια της Θράκης, στην Ήπειρο και στη Βόρεια – Δυτική πλευρά της Πελοποννήσου.

Πίνακας 4.1
Βασικά περιγραφικά στατιστικά μέτρα των μεταβλητών του υποδείγματος

| Μεταβλητή | Min | Median | Max | Mean | SD | CV (%) |
|---|--------|---------|---------|---------|--------|--------|
| Δείκτης <i>CWR</i> | 60,000 | 210,558 | 583,333 | 212,404 | 39,916 | 18,80 |
| Ποσοστό Παντρεμένων Γυναικών | 40,470 | 60,579 | 83,333 | 60,020 | 5,288 | 8,81 |
| Ποσοστό Μεταναστών | 0,038 | 5,935 | 24,531 | 6,471 | 4,017 | 62,08 |
| Δείκτης Χαμηλών Κοινωνικών και Οικονομικών Συνθηκών (LSE) | 4,206 | 8,209 | 14,084 | 8,440 | 1,651 | 19,56 |
| Ποσοστό Πυρηνικών Οικογενειών | 40,000 | 81,534 | 100,000 | 81,078 | 7,396 | 9,12 |

² Ένας χάρτης ποσοστιαίων σημείων (*quantile map*) απεικονίζει χωρικά μια μεταβλητή σε διαστήματα τιμών που περιλαμβάνουν ίσο αριθμό παρατηρήσεων (χωρικών μονάδων). Όταν ο χάρτης κατασκευάζεται με 4 διαστήματα, όπως οι χάρτες του διαγράμματος, τότε τα διαστήματα χωρίζονται σύμφωνα με τα τεταρτημόρια της κατανομής της εξεταζόμενης μεταβλητής.



Διάγραμμα 4.5
Χάρτες ποσοστιαίων σημείων:
Χωρικές κατανομές των μεταβλητών του υποδείγματος

4.5 Ανάλυση χωρικής αυτοσυσχέτισης των μεταβλητών

Η χωρική αυτοσυσχέτιση των πέντε δεικτών του υποδείγματος εξετάστηκε υπολογίζοντας το συντελεστή I του Moran (1948 και 1950). Ένα πρόβλημα που αντιμετωπίζει ο ερευνητής όταν διενεργεί χωρική ανάλυση για την Ελλάδα και οφείλεται στις γεωγραφικές ιδιαιτερότητες της χώρας είναι ο μεγάλος αριθμός χωρικών μονάδων που δεν έχουν φυσικά σύνορα. Οι μονάδες αυτές είναι τα νησιά αλλά και γειτονικές περιοχές που διαχωρίζονται από θαλάσσια σύνορα. Ιδιαίτερα, αξίζει να αναφερθεί η Νότια Στερεά Ελλάδα και η Πελοπόννησος που δεν έχουν φυσικά σύνορα αλλά η μεταξύ τους επικοινωνία είναι εύκολη λαμβάνοντας μάλιστα υπ' όψιν και τη σύνδεση του Ρίο με το Αντίρριο. Η ιδιομορφία του Ελληνικού γεωγραφικού χώρου δεν επιτρέπει τον ορισμό της μήτρας χωρικών σταθμίσεων με βάση τα φυσικά σύνορα των δήμων. Για το λόγο αυτό, η μήτρα χωρικών σταθμίσεων ορίστηκε με το κριτήριο των k κοντινών γειτόνων για διαφορετικές τιμές του k ίσες με 5, 6, 8, 10 και 20 με σκοπό να επιβεβαιωθεί η συνέπεια των αποτελεσμάτων.

Πίνακας 4.2
Ανάλυση χωρικής αυτοσυσχέτισης των μεταβλητών του υποδείγματος:
Τιμές στατιστικής I του Moran για εναλλακτικό αριθμό γειτόνων

| Μεταβλητή | Αριθμός Γειτόνων | | | | | P - Value |
|---|------------------|--------|--------|--------|--------|-----------|
| | $k=5$ | $k=6$ | $k=8$ | $k=10$ | $k=20$ | |
| Δείκτης CWR | 0,2896 | 0,2894 | 0,2648 | 0,2661 | 0,2713 | 0,0001 |
| Ποσοστό Παντρεμένων Γυναικών | 0,5365 | 0,5333 | 0,5218 | 0,5095 | 0,4696 | 0,0001 |
| Ποσοστό Μεταναστών | 0,3730 | 0,3744 | 0,3469 | 0,3173 | 0,2435 | 0,0001 |
| Δείκτης Χαμηλών Κοινωνικών και Οικονομικών Συνθηκών (LSE) | 0,4432 | 0,4246 | 0,3975 | 0,3696 | 0,2918 | 0,0001 |
| Ποσοστό Πυρηνικών Οικογενειών | 0,4768 | 0,4609 | 0,4320 | 0,4151 | 0,3627 | 0,0001 |

Ο Πίνακας 4.2 περιλαμβάνει τις τιμές του συντελεστή I του Moran για τις πέντε μεταβλητές του υποδείγματος που προέκυψαν από την εφαρμογή όλων των κριτηρίων κοντινών γειτόνων.³ Ανεξάρτητα από τον αριθμό των γειτόνων οι τιμές του I παραμένουν αμετάβλητες σε μέγεθος και υποδεικνύουν την ύπαρξη υψηλής θετικής χωρικής αυτοσυσχέτισης που σημαίνει ότι υπάρχουν συμπλέγματα (*clusters*) στη γεωγραφική

³ Ο συντελεστής I του Moran εκτιμήθηκε με τη βοήθεια του λογισμικού χωρικής ανάλυσης GeoDa 0.9.5-i που έχει αναπτυχθεί από τον Anselin (2004). Το λογισμικό αυτό υπολογίζει τιμές p - value για τη στατιστική σημαντικότητα του I που βασίζονται σε διαδικασία τυχαίων μεταθέσεων των τιμών της εξεταζόμενης μεταβλητής στις χωρικές μονάδες και επανα - υπολογισμό της στατιστικής πολλές φορές για να παραχθεί μια κατανομή αναφοράς. Η προσέγγιση αυτή διαφέρει από την υπόθεση της τυχαιοποίησης που περιγράφηκε στο Κεφάλαιο 1 και κρίνεται πιο αξιόπιστη λόγω των ιδιαιτεροτήτων του Ελληνικού γεωγραφικού χώρου.

επικράτεια όμοιων τιμών. Σε όλες τις περιπτώσεις η χωρική αυτοσυσχέτιση είναι στατιστικά σημαντική όπως συμπεραίνεται από τις πολύ χαμηλές τιμές των p-values (0,0001) που έχουν υπολογιστεί με διαδικασία 9999 τυχαίων μεταθέσεων. Σημειώνεται ότι το p-value δεν αλλάζει ιδιαίτερα για τα διαφορετικά κριτήρια κοντινών γειτόνων και ως εκ τούτου στον πίνακα αναγράφεται μόνο μία φορά. Οι μεταβλητές που παρουσιάζουν τη μεγαλύτερη και τη μικρότερη χωρική αυτοσυσχέτιση είναι, αντίστοιχα, το ποσοστό των παντρεμένων γυναικών και το ποσοστό των μεταναστών. Κατά συνέπεια, το πρώτο αποτέλεσμα της ανάλυσης είναι ότι οι τιμές τόσο της εξαρτημένης μεταβλητής όσο και των ερμηνευτικών μεταβλητών του υποδείγματος δεν είναι ανεξάρτητες στο χώρο αλλά συσχετίζονται γεωγραφικά. Η ύπαρξη της χωρικής αυτοσυσχέτισης αμφισβητεί την αξιοπιστία των αποτελεσμάτων από την εκτίμηση του ολικού (*global*) υποδείγματος.

4.6 Τα αποτελέσματα της εκτίμησης του απλού υποδείγματος

Το απλό οικονομετρικό υπόδειγμα εκτιμάται με τη μέθοδο των ελαχίστων τετραγώνων με σκοπό να διερευνηθεί η συνεισφορά των τεσσάρων κοινωνικό – οικονομικών δεικτών στην ερμηνεία της μεταβλητότητας του επιπέδου της γονιμότητας όπως εκφράζεται από το δείκτη του λόγου του παιδικού προς τον αναπαραγωγικό πληθυσμό (*CWR*). Τα αποτελέσματα της εκτίμησης του ολικού υποδείγματος παρουσιάζονται στον Πίνακα 4.3. Το επίπεδο της γονιμότητας εξαρτάται θετικά από το ποσοστό των παντρεμένων γυναικών, το ποσοστό των μεταναστών και το δείκτη χαμηλών κοινωνικό – οικονομικών συνθηκών και αρνητικά με το ποσοστό των πυρηνικών οικογενειών. Όλοι οι συντελεστές του υποδείγματος, περιλαμβάνοντας και το σταθερό όρο, είναι στατιστικά σημαντικοί σε 5% επίπεδο σημαντικότητας και έχουν τα αναμενόμενα πρόσημα. Επειδή όλες οι ανεξάρτητες μεταβλητές έχουν μετασχηματισθεί σε αποκλίσεις από τις μέσες τιμές τους ο σταθερός όρος ισούται με τη μέση τιμή του δείκτη *CWR*. Η τιμή της στατιστικής *F* είναι υψηλή, ενώ ο αριθμός κατάστασης (*MCN – Multicollinearity Condition Number*) για την ύπαρξη προβλήματος πολυσυγγραμμικότητας έχει χαμηλή τιμή φανερώνοντας την απουσία της. Από την άλλη πλευρά, ο συντελεστής προσδιορισμού (R^2) και ο προσαρμοσμένος συντελεστής προσδιορισμού (R^2_{adj}) έχουν χαμηλές τιμές, 0,345 και 0,337, αντίστοιχα. Το αποτέλεσμα αυτό δείχνει ότι το ολικό υπόδειγμα δεν μπορεί να εξηγήσει ένα πολύ σημαντικό ποσοστό της μεταβλητότητας της γονιμότητας.

Πίνακας 4.3
Αποτελέσματα εκτίμησης ολικού υποδείγματος (OLS)

| Μεταβλητή | Συντελεστής | Τυπικό Σφάλμα | t – Στατιστική | P-Value |
|---|-------------|---------------------|----------------|---------|
| Σταθερός όρος | 212,404 | 1,803 | 117,775 | 0,0000 |
| Ποσοστό Παντρεμένων Γυναικών | 4,055 | 0,365 | 11,107 | 0,0000 |
| Ποσοστό Μεταναστών | 1,058 | 0,481 | 2,199 | 0,0286 |
| Δείκτης Χαμηλών Κοινωνικών και Οικονομικών Συνθηκών (LSE) | 3,680 | 1,220 | 3,017 | 0,0028 |
| Ποσοστό Πυρηνικών Οικογενειών | -0,568 | 0,261 | -2,175 | 0,0303 |
| R^2 : 0,345 | | R^2_{adj} : 0,337 | | |
| F – Στατιστική: 42,089 | | MCN: 1,677 | | |
| Άθροισμα Τετραγώνων Καταλοίπων SSE: 338.263 | | AIC: 3.190,330 | | |

Πίνακας 4.4
Στατιστικοί έλεγχοι αυτοσυσχέτισης, ετεροσκεδαστικότητας και κανονικότητας για τα εκτιμηθέντα σφάλματα του ολικού υποδείγματος (OLS)

| Αριθμός Γειτόνων | Στατιστική I του Moran | Τιμή Z | P –value |
|---|------------------------|--------|----------|
| k=5 | 0,2182 | 7,410 | 0,000 |
| k=6 | 0,2188 | 8,152 | 0,000 |
| k=8 | 0,1838 | 7,976 | 0,000 |
| k=10 | 0,1624 | 8,014 | 0,000 |
| k=20 | 0,1524 | 11,238 | 0,000 |
| Breusch-Pagan (1979) Στατιστική Ετεροσκεδαστικότητας: 1.219,877 | | | 0,000 |
| White (1980) Στατιστική Ετεροσκεδαστικότητας: 234,767 | | | 0,000 |
| Koenker-Basset (1982) Στατιστική Ετεροσκεδαστικότητας: 156,943 | | | 0,000 |
| Jarque-Bera (1987) Έλεγχος Κανονικότητας: 2.549,827 | | | 0,000 |

Η ομοσκεδαστικότητα των σφαλμάτων του εκτιμηθέντος υποδείγματος εξετάστηκε με βάση τους ελέγχους των Breusch και Pagan (1979), του White (1980) και των Koenker και Basset (1982), ενώ η υπόθεση της κανονικότητας των σφαλμάτων αξιολογήθηκε με τη στατιστική των Jarque και Bera (1987). Τέλος, η παρουσία χωρικής αυτοσυσχέτισης στα σφάλματα διερευνήθηκε με τον έλεγχο της στατιστική I του Moran (1950), όπως έχει προταθεί η εφαρμογή του από τους Cliff και Ord (1972) για τα κατάλοιπα παλινδρόμησης, χρησιμοποιώντας τις πέντε μήτρες χωρικών σταθμίσεων που έχουν αναφερθεί.

Τα αποτελέσματα των ελέγχων συνοψίζονται στον Πίνακα 4.4 και αποκαλύπτουν ότι τα εκτιμηθέντα σφάλματα δεν έχουν σταθερή διακύμανση, παρουσιάζουν θετική χωρική αυτοσυσχέτιση και δεν ακολουθούν κανονική κατανομή. Κατά συνέπεια, είναι φανερό ότι παραβιάζονται οι βασικές υποθέσεις της μεθόδου των ελαχίστων τετραγώνων και το θεώρημα Gauss Markov δεν ισχύει. Οι διακυμάνσεις των εκτιμηθέντων συντελεστών είναι υποεκτιμημένες και η απουσία κανονικότητας για τις τιμές των σφαλμάτων καθιστά οποιαδήποτε στατιστική συμπερασματολογία αναξιόπιστη. Η εμφάνιση των προβλημάτων

της ετεροσκεδαστικότητας και της αυτοσυσχέτισης στα σφάλματα του υποδείγματος υποδεικνύουν την παρουσία χωρικών επιδράσεων. Η σχέση εξάρτησης της γονιμότητας και των τεσσάρων ερμηνευτικών μεταβλητών δεν είναι σταθερή για όλους τους δήμους της χώρας.

Η μέθοδος των ελαχίστων τετραγώνων δεν είναι κατάλληλη σε αυτή την περίπτωση για την εκτίμηση του υποδείγματος διότι αγνοεί τη χωρική πλευρά των δεδομένων. Μία δυνατή λύση του προβλήματος είναι η εφαρμογή της μη παραμετρικής μεθόδου της γεωγραφικά σταθμισμένης παλινδρόμησης (GWR) των Fotheringham, Brunson και Charlton (2002) για την εκτίμηση ξεχωριστών τοπικών υποδειγμάτων (local models) σε κάθε δήμο και την αντιμετώπιση των επιπτώσεων από τις χωρικές επιδράσεις.

4.7 Τα αποτελέσματα της εφαρμογής της Γεωγραφικά Σταθμισμένης Παλινδρόμησης

Η ποσοτική σχέση μεταξύ του δείκτη *CWR* και των επιλεγμένων ερμηνευτικών μεταβλητών επαναξιολογήθηκε σε τοπικό επίπεδο με τη βοήθεια της γεωγραφικά σταθμισμένης παλινδρόμησης (GWR) με συνάρτηση στάθμισης προσαρμοσμένου διτετράγωνου πυρήνα (*adaptive bi-squared weight function*).⁴ Για να καθοριστούν οι αποστάσεις μεταξύ των περιοχών χρησιμοποιήθηκαν τα γεωγραφικά κεντροειδή κάθε δήμου. Ο εντοπισμός του μέγιστου εύρους έγινε με την ελαχιστοποίηση του διορθωμένου πληροφοριακού κριτηρίου του Akaike, AIC_C , που έχει προταθεί από τους Hurvich, Simonoff, and Tsai (1998). Η εφαρμογή της διαδικασίας ελαχιστοποίησης καθόρισε 109 κοντινούς γείτονες (δήμους) να περιληφθούν στην εκτίμηση κάθε τοπικού υποδείγματος.⁵ Ο Πίνακας 4.5 παρουσιάζει τους πέντε περιληπτικούς αριθμούς (*5-number summary statistic*) της κατανομής των εκτιμηθέντων τοπικών συντελεστών και τα βασικά αποτελέσματα της διαδικασίας εκτίμησης.

⁴ Η εκτίμηση των τοπικών υποδειγμάτων με τη μέθοδο της γεωγραφικά σταθμισμένης παλινδρόμησης έγινε με τη βοήθεια του γεωγραφικού συστήματος πληροφοριών ArcGIS 9.3.

⁵ Όταν χρησιμοποιείται συνάρτηση στάθμισης προσαρμοσμένου πυρήνα το μέγιστο εύρος κάθε τοπικού υποδείγματος είναι διαφορετικό και ορίζεται ως η απόσταση μέχρι τον N κοντινό γείτονα ώστε η εκτίμηση σε κάθε σημείο να γίνεται με τον ίδιο αριθμό γειτονικών δήμων.

Πίνακας 4.5
Αποτελέσματα εκτίμησης τοπικών υποδειγμάτων (GWR)

| Συντελεστής | Min | Q1 | Median | Q3 | Max |
|---|---|-----------|--|-----------|------------|
| Σταθερός όρος | 195,647 | 206,937 | 209,960 | 214,338 | 238,144 |
| Ποσοστό Παντρεμένων Γυναικών | -0,773 | 2,487 | 3,446 | 3,812 | 5,709 |
| Ποσοστό Μεταναστών | -2,129 | 0,591 | 1,208 | 1,879 | 4,773 |
| Δείκτης Χαμηλών Κοινωνικών και Οικονομικών Συνθηκών (LSE) | -2,699 | 1,781 | 4,194 | 5,756 | 19,059 |
| Ποσοστό Πυρηνικών Οικογενειών | -4,241 | 0,257 | 0,659 | 1,054 | 2,616 |
| Τοπικά R^2 | 0,061 | 0,323 | 0,468 | 0,604 | 0,705 |
| R^2_{GWR} : 0,647 Adj. R^2_{GWR} : 0,586 | AICc: 3.067,818 Κοντινοί Γείτονες: 109 | | Άθροισμα Τετραγώνων Καταλοίπων SSE: 182.037,181 | | |

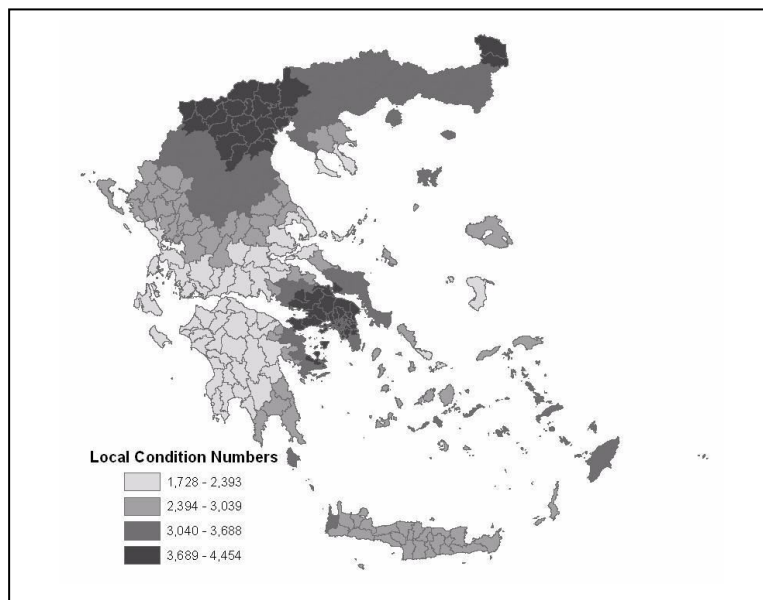
Πίνακας 4.6
Έλεγχος ανάλυση διακύμανσης για τη σύγκριση ολικού υποδείγματος (OLS) και τοπικών υποδειγμάτων (GWR)

| Πηγή | SS | DF | MS | F |
|--------------|-----------|-----------|-----------|----------|
| SSE OLS | 338.262,6 | 5 | | |
| GWR Βελτίωση | 156.225,4 | 32,97 | 4.738,833 | |
| SSE GWR | 182.037,2 | 287,03 | 634,2031 | 7,4721 |

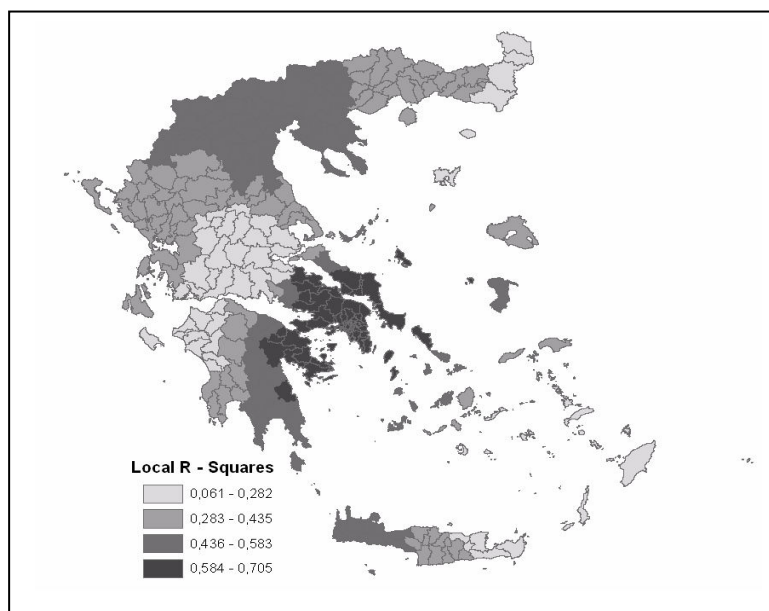
Είναι φανερό ότι η γεωγραφικά σταθμισμένη παλινδρόμηση βελτιώνει σημαντικά την ερμηνευτική ικανότητα του υποδείγματος. Η τιμή του προσαρμοσμένου συντελεστή προσδιορισμού που προέκυψε (0,586) είναι σχεδόν διπλάσια από την αντίστοιχη της εκτίμησης του υποδείγματος με τη μέθοδο των ελαχίστων τετραγώνων (0,337), κάτι που δηλώνει μεγαλύτερη ερμηνευτική ικανότητα, παρόλο που δεν μπορεί να γίνει άμεση σύγκριση. Αξιοσημείωτη είναι και η μείωση που επέρχεται στο άθροισμα των τετραγώνων των καταλοίπων, δηλαδή από 338.263 στο ολικό υπόδειγμα σε 182.037 στα τοπικά υποδείγματα. Η καλύτερη επίδοση των τοπικών υποδειγμάτων υποστηρίζεται και από την υψηλή τιμή της στατιστικής F του ελέγχου ανάλυσης διακύμανσης (*ANOVA Test*) των Fotheringham, Brunsdon και Charlton, (2002), για τη βελτίωση της ερμηνευτικής ικανότητας της γεωγραφικά σταθμισμένης παλινδρόμησης σε σχέση με το απλό υπόδειγμα, που περιλαμβάνεται στον Πίνακα 4.6.

Όπως έχει αναφερθεί, τα εκτιμηθέντα σφάλματα του ολικού υποδείγματος αντιμετωπίζουν το πρόβλημα της θετικής χωρικής αυτοσυσχέτισης. Η στατιστική *I* του Moran για τα τοπικά σφάλματα λαμβάνει πολύ χαμηλές τιμές. Συγκεκριμένα, εάν υπολογιστεί με 5 κοντινούς γείτονες για τον ορισμό της μήτρας χωρικών σταθμίσεων ισούται με 0,0813 ενώ με 20 κοντινούς γείτονες λαμβάνει την τιμή 0,0112. Οι χαμηλές τιμές της *I* φανερώνουν ότι τα τοπικά σφάλματα είναι ασυσχέτιστα. Κατά συνέπεια, η γεωγραφικά σταθμισμένη

παλινδρόμηση έχει αντιμετωπίσει το πρόβλημα της χωρικής αυτοσυσχέτισης παρέχοντας πιο αξιόπιστα αποτελέσματα.



Διάγραμμα 4.6
Χάρτης τοπικών αριθμών κατάστασης για την
ύπαρξη τοπικής πολυσυγγραμμικότητας



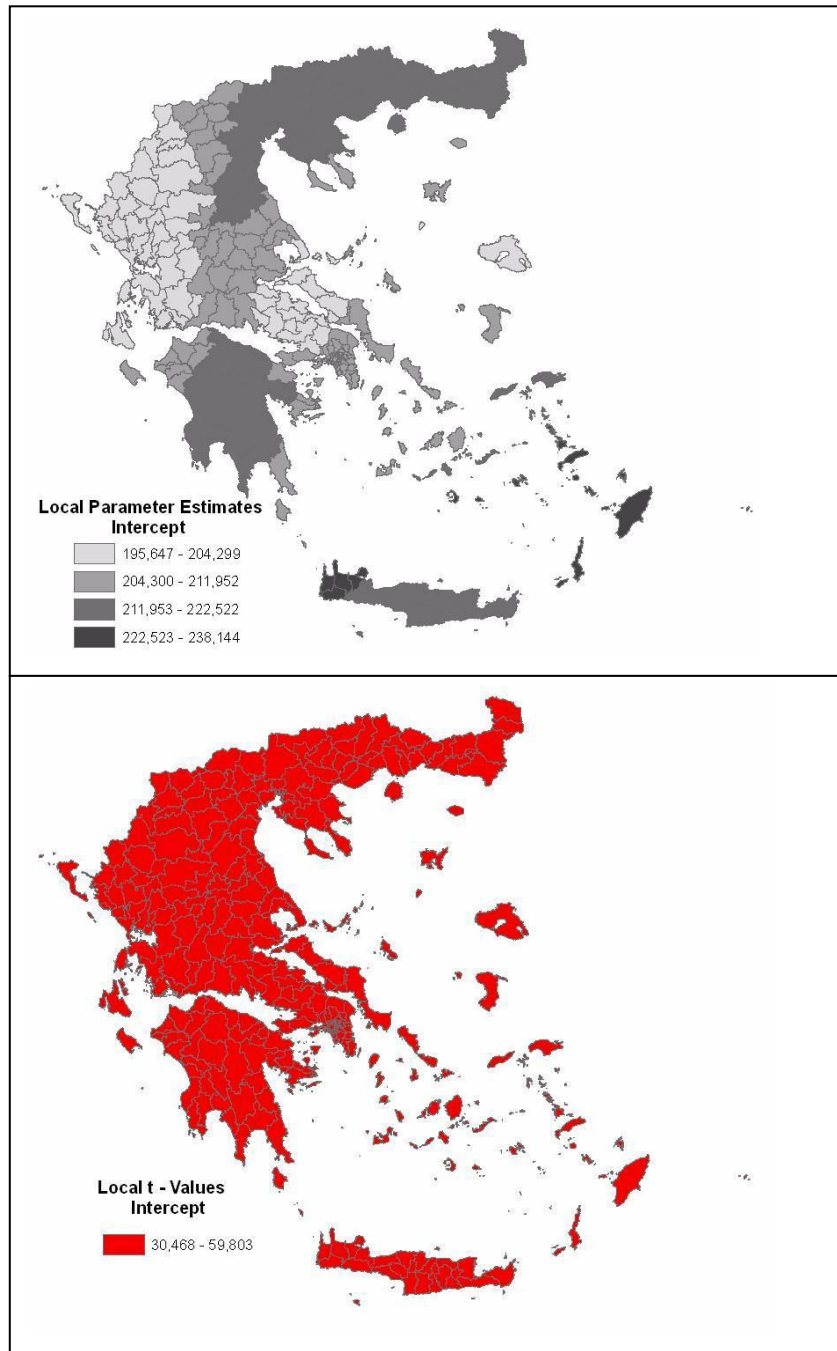
Διάγραμμα 4.7
Χάρτης τοπικών συντελεστών προσδιορισμού

Ο χάρτης στο Διάγραμμα 4.6 απεικονίζει τη χωρική κατανομή των τοπικών αριθμών κατάστασης (*Local Condition Numbers*) για την ανίχνευση του προβλήματος της πολυσυγγραμμικότητας στα τοπικά υποδείγματα. Από τις χαμηλές τιμές αυτού του δείκτη σε όλους τους δήμους συμπεραίνεται ότι δεν εμφανίζεται πρόβλημα τοπικής πολυσυγγραμμικότητας για κανένα τοπικό υπόδειγμα.

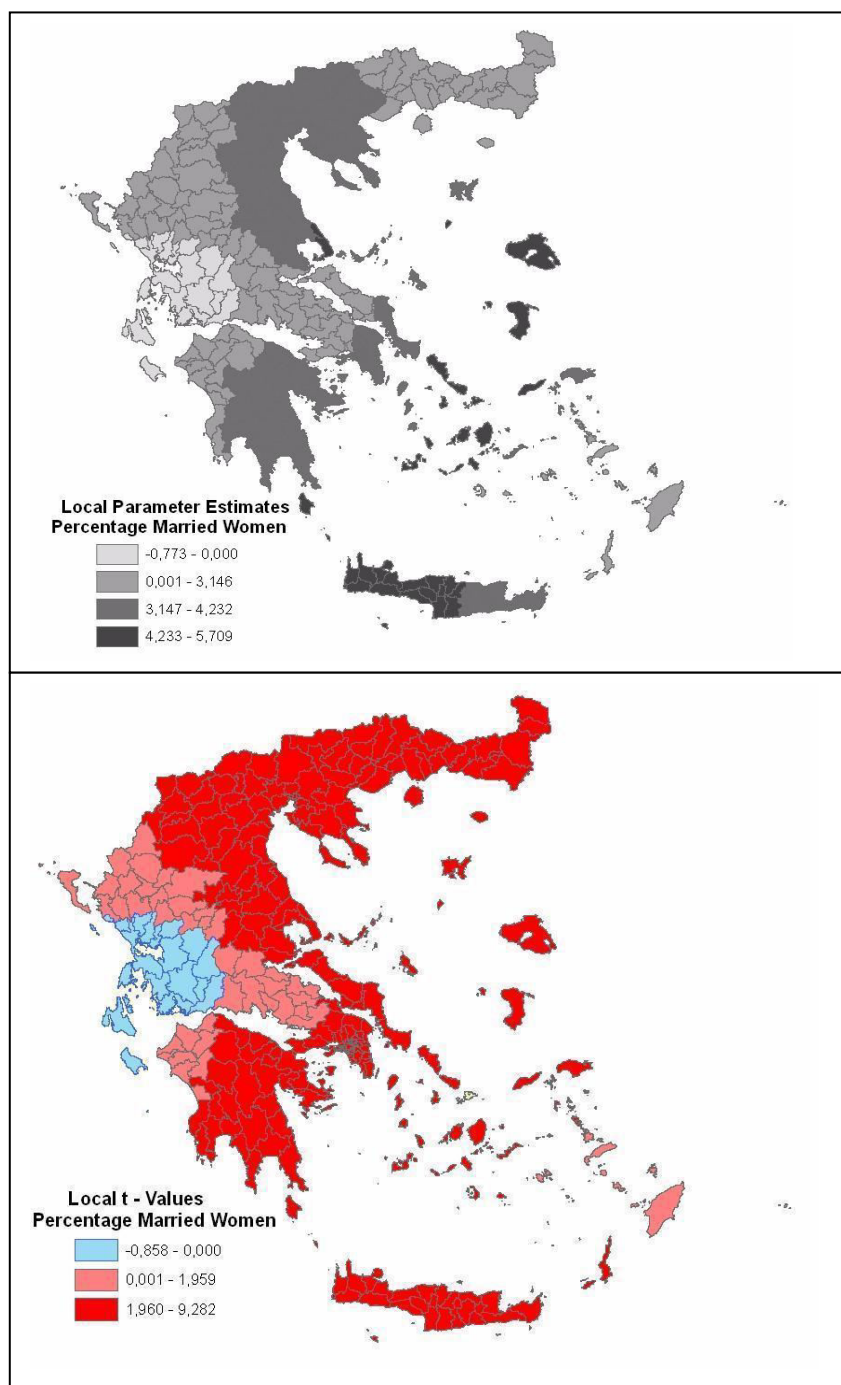
Η χωρική κατανομή των τοπικών συντελεστών προσδιορισμού, που εκφράζουν την τοπική ερμηνευτική ικανότητα του εκτιμηθέντος υποδείγματος, απεικονίζονται στο χάρτη του Διαγράμματος 4.7, ενώ οι πέντε περιληπτικοί αριθμοί της κατανομής τους υπάρχουν στον Πίνακα 4.5. Όπως διαπιστώνεται, η διάμεσος των τοπικών συντελεστών προσδιορισμού (0,468) είναι μεγαλύτερη από το συντελεστή προσδιορισμού του ολικού υποδείγματος (0,345) και από το Διάγραμμα 4.7 είναι προφανές ότι η σχέση εξάρτησης της γονιμότητας από τις τέσσερις μεταβλητές είναι πιο ισχυρή σε κάποιους δήμους. Ειδικότερα, οι υψηλότεροι τοπικοί συντελεστές προσδιορισμού σχηματίζουν ένα σύμπλεγμα τιμών που περιλαμβάνει την περιφέρεια της Αττικής και τις γειτονικές της περιοχές που αποτελούν το κέντρο των οικονομικών δραστηριοτήτων στην Ελλάδα. Το σύμπλεγμα αυτό αποτελείται από τις πιο πυκνοκατοικημένες περιοχές στην Ελλάδα, δηλαδή την πρωτεύουσα της χώρας και τους γειτονικούς μητροπολιτικούς δήμους, καθώς και τις κοντινές περιοχές της Εύβοιας, της Βοιωτίας και της βορειανατολικής Πελοποννήσου. Από την άλλη πλευρά, το υπόδειγμα έχει χαμηλή ερμηνευτική ικανότητα στα νησιά των Δωδεκανήσων, στην ανατολική Κρήτη, στην Κεντρική Ελλάδα (ιδιαίτερα στους νομούς Φθιώτιδας και Ευρυτανίας), σε μερικούς δήμους της περιφέρειας Θεσσαλίας, στη βορειοδυτική Πελοπόννησο και στο νομό Έβρου. Στις υπόλοιπες περιοχές της χώρας οι τοπικοί συντελεστές προσδιορισμού έχουν μέτριες τιμές.

4.8 Σύγκριση των ολικών και τοπικών επιδράσεων των κοινωνικών και οικονομικών παραγόντων στα επίπεδα της γονιμότητας

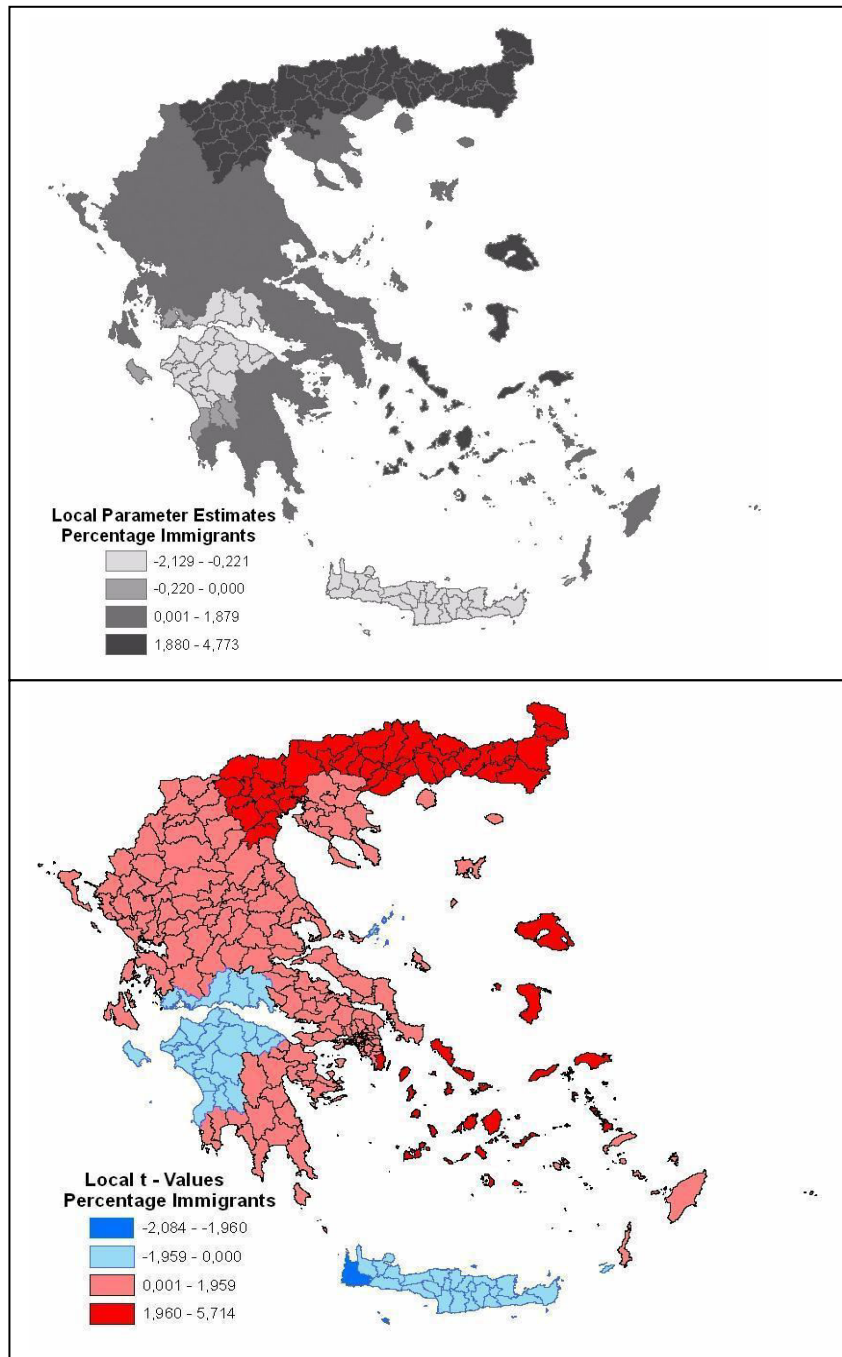
Η στατιστική σημαντικότητα των εκτιμηθέντων τοπικών συντελεστών αξιολογείται υπολογίζοντας τοπικές ψευδό-*t* στατιστικές που προκύπτουν από τη διαίρεσή τους με τα αντίστοιχα τοπικά τυπικά σφάλματα. Η χαρτογράφηση των τοπικών συντελεστών και ψευδό-*t* στατιστικών οδηγούν στον εντοπισμό των χωρικών διαφοροποιήσεων ως προς την κατεύθυνση και το μέγεθος των επιδράσεων των τεσσάρων κοινωνικό – οικονομικών παραγόντων στο επίπεδο της γονιμότητας. Οι χάρτες αυτοί περιέχονται στο Διάγραμμα 4.8.



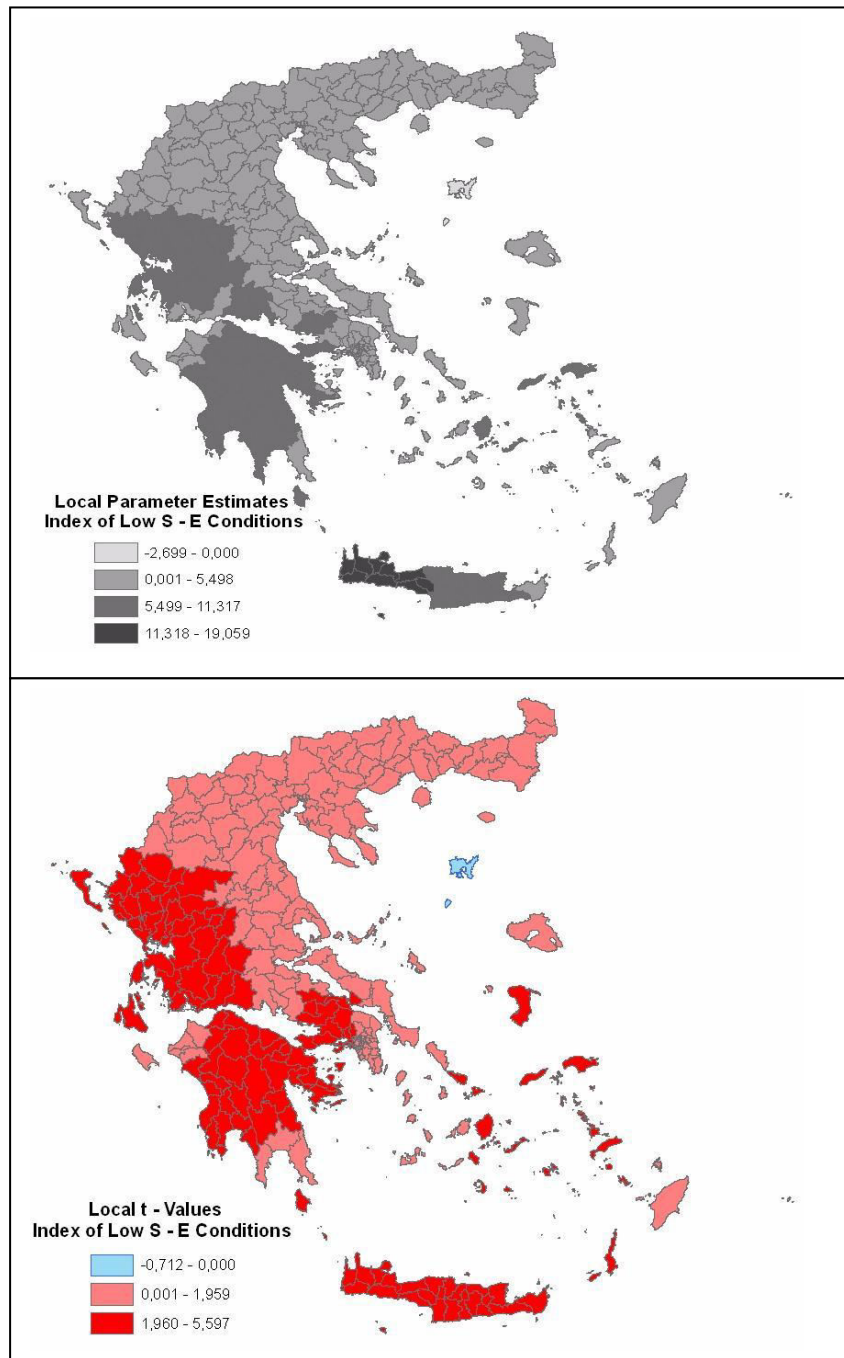
Διάγραμμα 4.8 (Α)
Χάρτες τοπικών συντελεστών και ψευδό-*t* στατιστικών των τοπικών υποδειγμάτων
Σταθερός Όρος



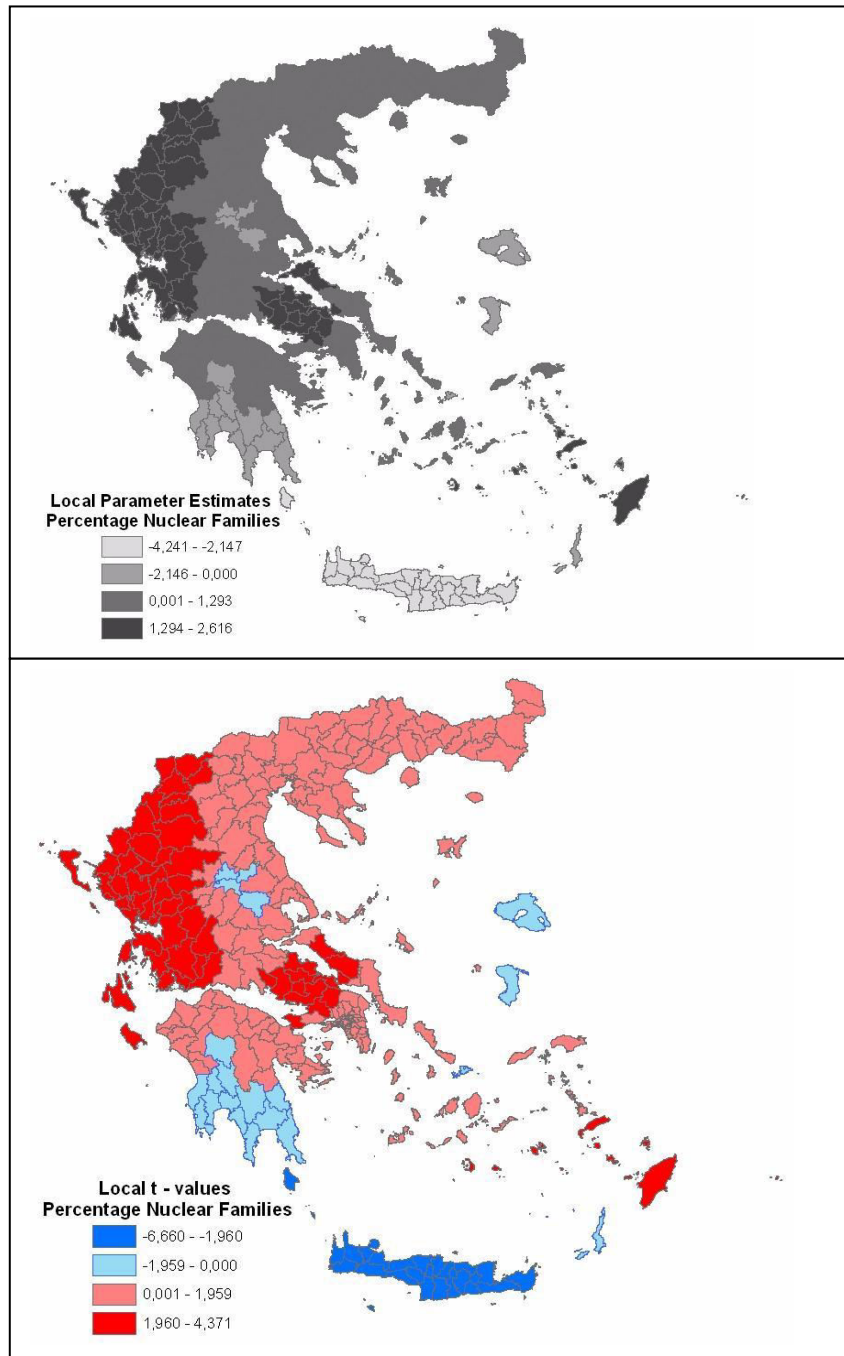
Διάγραμμα 4.8 (B)
Χάρτες τοπικών συντελεστών και ψευδό-*t* στατιστικών των τοπικών υποδειγμάτων
Ποσοστό Παντρεμένων Γυναικών



Διάγραμμα 4.8 (Γ)
Χάρτες τοπικών συντελεστών και ψευδό-*t* στατιστικών των τοπικών υποδειγμάτων
Ποσοστό Μεταναστών



Διάγραμμα 4.8 (Δ)
Χάρτες τοπικών συντελεστών και ψευδό-*t* στατιστικών των τοπικών υποδειγμάτων
Δείκτης Χαμηλών Κοινωνικών και Οικονομικών Συνθηκών (LSE)



Διάγραμμα 4.8 (Ε)
Χάρτες τοπικών συντελεστών και ψευδό-*t* στατιστικών των τοπικών υποδειγμάτων
Ποσοστό Πυρηνικών Οικογενειών

Από το Διάγραμμα 4.8 (Α) προκύπτει ότι οι τοπικοί σταθεροί όροι έχουν θετικές τιμές και είναι στατιστικά σημαντικοί σε ολόκληρη τη χώρα. Οι τιμές τους κυμαίνονται από 195,647 έως 238,144 και είναι κοντά στη μέση τιμή του δείκτη *CWR* λόγω του μετασχηματισμού των ανεξάρτητων μεταβλητών σε αποκλίσεις από τις μέσες τιμές τους. Συγκρίνοντας τα αποτελέσματα της εκτίμησης των τοπικών υποδειγμάτων με τα αποτελέσματα του ολικού υποδείματος όσον αφορά τις επιδράσεις των τεσσάρων κοινωνικό – οικονομικών παραγόντων στο επίπεδο της γονιμότητας προκύπτουν τα εξής συμπεράσματα:

1) Στο ολικό υπόδειγμα το ποσοστό των παντρεμένων γυναικών έχει θετική και στατιστικά σημαντική επίδραση στα επίπεδα της γονιμότητας. Στα τοπικά υποδείγματα, όπως φαίνεται στο Διάγραμμα 4.8 (Β), οι θετικές και στατιστικά σημαντικές επιδράσεις παραμένουν για την πλειοψηφία των δήμων της χώρας με εξαίρεση μία μικρή περιοχή της δυτικής Ελλάδας που εντοπίζεται αρνητική αλλά μη στατιστικά σημαντική σχέση. Όμως, σε μερικούς δήμους των νησιών των Δωδεκανήσων, της Κεντρικής Ελλάδας, της Ηπείρου και της βορειοδυτικής Πελοποννήσου η σχέση γονιμότητας και παντρεμένων γυναικών μολονότι θετική είναι μη στατιστικά σημαντική. Η αρνητική σχέση μεταξύ γονιμότητας και ποσοστού παντρεμένων γυναικών που παρατηρείται σε μικρό αριθμό δήμων κρίνεται μη στατιστικά σημαντική.

2) Στο ολικό υπόδειγμα η σχέση μεταξύ ποσοστού μεταναστών και γονιμότητας είναι θετική και στατιστικά σημαντική. Η εκτίμηση των τοπικών υποδειγμάτων με τη μέθοδο της γεωγραφικά σταθμισμένης παλινδρόμησης αποκαλύπτει, σε κάποιο βαθμό, μια διαφορετική εικόνα. Ειδικότερα, από τους χάρτες στο Διάγραμμα 4.8 (Γ), διαπιστώνεται ότι παρόλο που η θετική σχέση γονιμότητας και μεταναστών παραμένει για το μεγαλύτερο μέρος της χώρας μόνο σε 74 δήμους, που βρίσκονται στη Μακεδονία, τη Θράκη και σε μερικά νησιά του Αιγαίου, οι επιδράσεις είναι στατιστικά σημαντικές. Αρνητική σχέση γονιμότητας και μεταναστών εντοπίζεται στην Κρήτη και στη Βορειοδυτική Πελοπόννησο αλλά είναι μη στατιστικά σημαντική με εξαίρεση δύο δήμους στη δυτική πλευρά της Κρήτης.

3) Τόσο στο ολικό υπόδειγμα όσο και στα τοπικά υποδείγματα οι επιδράσεις των χαμηλών κοινωνικών και οικονομικών συνθηκών, όπως εκφράζονται από τον προτεινόμενο δείκτη *LSE*, στα επίπεδα της γονιμότητας είναι θετικές με εξαίρεση το νησί της Λήμνου. Ωστόσο, από το χάρτη στο Διάγραμμα 4.8 (Δ) φαίνεται ότι οι επιδράσεις αυτές είναι στατιστικά σημαντικές μόνο στους μισούς δήμους της χώρας που βρίσκονται κυρίως στην Πελοπόννησο, τη δυτική Ελλάδα, την Κρήτη και μερικά νησιά του Αιγαίου. Το αποτέλεσμα

αυτό προτείνει ότι, τουλάχιστον σε συνολικό επίπεδο, η σχέση μεταξύ κοινωνικών και οικονομικών παραγόντων και επιπέδου γονιμότητας δεν είναι πάντα ξεκάθαρη και πιθανών εμπλέκονται και άλλοι ποσοτικοί ή ποιοτικοί παράγοντες που τη διαμορφώνουν.

4) Στο σύνολο της χώρας η εκτίμηση του ολικού υποδείγματος δείχνει ότι υπάρχει αρνητική και στατιστικά σημαντική σχέση ανάμεσα στις πυρηνικές οικογένειες και στα επίπεδα της γονιμότητας. Σε τοπικό επίπεδο, οι χάρτες στο Διάγραμμα 4.8 (Ε) φανερώνουν ότι αυτή η διαπίστωση ισχύει μόνο στην Κρήτη. Στη νότια Πελοπόννησο, τέσσερεις δήμους της Θεσσαλίας και μερικά νησιά του Αιγαίου η σχέση είναι αρνητική αλλά μη στατιστικά σημαντική. Σε όλους τους άλλους δήμους της χώρας η σχέση πυρηνικών οικογενειών και γονιμότητας γίνεται θετική και στατιστικά σημαντική μόνο σε 67 από αυτούς που βρίσκονται κυρίως στη δυτική Ελλάδα, την κεντρική Ελλάδα και σε μερικά νησιά των Δωδεκανήσων.

4.9 Συμπεράσματα

Η εκτίμηση της γονιμότητας σε μικρό γεωγραφικό επίπεδο είναι ενδιαφέρουσα διότι περιγράφει τις τοπικές δημογραφικές συνθήκες. Η πληροφορία αυτή γίνεται ακόμη πιο πολύτιμη όταν υπάρχουν επιπλέον διαθέσιμοι κοινωνικό – οικονομικοί δείκτες σε τοπικό επίπεδο όχι μόνο για την ερμηνεία περιφερειακών διαφορών αλλά και για τον εντοπισμό συγκεκριμένων περιοχών και την άσκηση εξειδικευμένων περιφερειακών πολιτικών αντί για τη χάραξη εθνικών πολιτικών οι οποίες μπορεί να μην είναι αποτελεσματικές ή να μη γίνονται ευρέως αποδεκτές στις τοπικές κοινωνίες.

Τα αποτελέσματα της ανάλυσης φανερώνουν ότι υπάρχουν σημαντικές χωρικές διαφοροποιήσεις στα επίπεδα της γονιμότητας μεταξύ αλλά ίσως και εντός των τοπικών διοικητικών μονάδων μολονότι η δεύτερη υπόθεση δεν μπορεί να υποστηριχθεί από τα διαθέσιμα στατιστικά στοιχεία. Αυτές οι χωρικές διαφοροποιήσεις δεν είναι τυχαίες έτσι ώστε οι χωρικές επιδράσεις να παίζουν σημαντικό ρόλο στην εκτίμηση του υποδείγματος. Το χαρακτηριστικό αυτό γίνεται ιδιαίτερα φανερό όταν συγκρίνονται το ολικό υπόδειγμα με τα τοπικά υποδείγματα διότι η κατεύθυνση αλλά και η στατιστική σημαντικότητα των επιδράσεων των επιλεγμένων μεταβλητών στη γονιμότητα σε πολλές περιπτώσεις είναι διαφορετικές. Η εφαρμογή των τοπικών υποδειγμάτων δεν είναι σταθερή σε όλους τους δήμους της χώρας. Η υψηλότερη ερμηνευτική ικανότητα των τοπικών υποδειγμάτων εντοπίζεται στην περιφέρεια της Αττικής και πιο συγκεκριμένα στην Αθήνα και στους

γειτονικούς της δήμους, δηλαδή τις περιοχές που περιλαμβάνουν το κέντρο των οικονομικών δραστηριοτήτων και τον πιο αστικοποιημένο πληθυσμό της χώρας. Οι περιοχές αυτές επιδεικνύουν χαμηλά επίπεδα γονιμότητας αλλά, καθώς ο πληθυσμός τους αντιπροσωπεύει περισσότερο από το μισό πληθυσμό της Ελλάδας, η συμπεριφορά τους τείνει να επηρεάζει ολόκληρο το κοινωνικό – δημογραφικό περιβάλλον. Από την άλλη πλευρά, η χαμηλότερη ερμηνευτική ικανότητα του υποδείγματος υπάρχει στη κεντρική Ελλάδα και κυρίως στα νησιά όπου, εξ' αιτίας των γεωγραφικών τους χαρακτηριστικών, προκύπτουν δυσκολίες στην ανάλυση χωρικών δεδομένων.

Αν ληφθεί υπ' όψιν ότι η εξωσυζυγική γονιμότητα στην Ελλάδα είναι ακόμη πολύ χαμηλή, το ποσοστό των παντρεμένων γυναικών αναπαραγωγικής ηλικίας ασκεί την πιο σημαντική θετική επίδραση στα περιφερειακά επίπεδα γονιμότητας κάτι που αποδεικνύεται τόσο από το ολικό υπόδειγμα όσο και από τα τοπικά υποδείγματα. Σε παρόμοια συμπεράσματα έχουν καταλήξει αλλά για μη Ευρωπαϊκούς πληθυσμούς οι Kohli (1977), Weeks (2010) και Kamata, Iwasawa και Tanaka (2010). Όμως, είναι αμφίβολο εάν ο θεσμός του γάμου θα συνεχίσει να έχει κυρίαρχο ρόλο στο μέλλον εφόσον η Ελλάδα ακολουθήσει σταδιακά τα βήματα της δεύτερης δημογραφικής μετάβασης, όπως παρουσιάζουν οι Lesthaeghe και Kaa (1986) και Sobotka (2008). Μολαταύτα, από την ανάλυση είναι ξεκάθαρο ότι ο γάμος και η υποστήριξη της οικογένειας μέσω οικονομικής βοήθειας, φοροαπαλλαγών και επιδόματα για το τρίτο παιδί πρέπει να έχουν προτεραιότητα στην άσκηση πληθυσμιακής πολιτικής από τις κεντρικές και τοπικές αρχές.

Η δομή των οικογενειών, που εκφράζεται μέσω του ποσοστού των πυρηνικών οικογενειών, θεωρείται παράγοντας επηρεασμού των γεννήσεων και της φροντίδας των παιδιών σε τοπικό επίπεδο. Παρόλο που η πυρηνική οικογένεια γίνεται ο κυρίαρχος τύπος της οικογενειακής δομής στην Ελλάδα, ιδιαίτερα στις αστικές περιοχές, τα τοπικά υποδείγματα δεν έδειξαν μία σαφή εικόνα σχετικά με την επίδρασή της. Το αποτέλεσμα αυτό θα μπορούσε, κατά κύριο λόγο, να αποδοθεί στο θεσμό υποστήριξης της οικογένειας που κυριαρχεί στη χώρα ο οποίος φαίνεται να έχει προστατευτικά αποτελέσματα για τις γεννήσεις λόγω της βοήθειας, οικονομικής και μη, που λαμβάνουν τα νέα ζευγάρια από τους γονείς τους για την ανατροφή των παιδιών.

Το έτος 2001, που αποτελεί το έτος αναφοράς στην ανάλυση, η μετανάστευση ήταν ένα πολύ πρόσφατο φαινόμενο στην Ελλάδα. Κατά συνέπεια, η επίδραση των μεταναστών στα επίπεδα της γονιμότητας εισάγεται έμμεσα μέσω της νεανικής ηλικιακής δομής των

μεταναστών παρά μέσω της άμεσης συνεισφοράς τους σε νέες γεννήσεις. Πρόσφατα στοιχεία ληξιαρχικών καταγραφών δείχνουν ότι ένα σχετικά μεγάλο μέρος των γεννήσεων (17% το 2006) προέρχεται από μητέρες μετανάστριες. Ωστόσο, σύμφωνα με εμπειρική έρευνα που έχει διεξαχθεί από τον Tsimbos (2008) ο αριθμός των μεταναστριών δεν είναι αρκετά μεγάλος για να επηρεάσει τα συνολικά επίπεδα της γονιμότητας αλλά πιο πρόσφατα στατιστικά δεδομένα ίσως στο μέλλον αποκαλύψουν μία διαφορετική εικόνα και πιο σταθερές στατιστικές σχέσεις.

Οι κοινωνικό – οικονομικές διαφοροποιήσεις ερμηνεύουν σε κάποιο βαθμό τις περιφερειακές διακυμάνσεις της γονιμότητας αλλά οι έμφυτες εξαρτήσεις δεν είναι σε όλες τις περιπτώσεις καθарές ή στατιστικά σημαντικές. Οι συνθήκες στέρησης σε τοπικό επίπεδο εκφράζονται με μία περιορισμένη παραλλαγή του γνωστού δείκτη ανθρώπινης φτώχιας που έχει προταθεί από τον Οργανισμό Ηνωμένων Εθνών. Αυτός ο συνθετικός δείκτης, παρά τους περιορισμούς στην εφαρμογή του, έχει αποδειχθεί ένα χρήσιμο μέτρο στην ανάλυση τοπικών θεμάτων που σχετίζονται με την υγεία (Bagheri, Holt και Benwell, 2009) αλλά φαίνεται ότι αυτό δεν ισχύει πάντοτε για την περίπτωση της γονιμότητας. Στην πραγματική ζωή, ένας συνδυασμός οικονομικών, κοινωνικών και πολιτιστικών παραγόντων οδηγούν σε διαφοροποιήσεις στη γονιμότητα μεταξύ τοπικών πληθυσμιακών ομάδων που οι αξίες τους, οι αποφάσεις τους και οι ιδιαίτερες συνθήκες όπως το κόστος ανατροφής των παιδιών ή οι ειδικές δυσκολίες δεν μπορούν να μετρηθούν από έναν απλό συνθετικό δείκτη. Οικισμοί με περισσότερα ή λιγότερα κοινά μετρήσιμα οικονομικά χαρακτηριστικά παρουσιάζουν διαφορετικά επίπεδα γονιμότητας και δυναμικές. Η χρήση των δεικτών ανθρώπινης ανάπτυξης ή ανθρώπινης φτώχιας σε τοπικό διοικητικό επίπεδο μπορεί να αποτελούν αδυναμία στη χωρική ανάλυση της γονιμότητας και πιο λεπτομερειακοί και πολυδιάστατοι δείκτες θα πρέπει να εφαρμοστούν στα δεδομένα της απογραφής του 2011 ή σε άλλα στατιστικά δεδομένα εφόσον γίνουν διαθέσιμα για έρευνα. Για τον ίδιο λόγο, παρουσιάζει εξαιρετικό ενδιαφέρον να ενσωματωθούν στατιστικά στοιχεία από δύο διαδοχικές απογραφές για να εξεταστούν τα επίπεδα, οι διαφορές και οι τάσεις της γονιμότητας σε τοπικό επίπεδο και για τη περίοδο της έντονης οικονομικής κρίσης στην Ελλάδα.

4.10 Ανακεφαλαίωση

Κατά τη διάρκεια του δεύτερου μισού του 20^{ου} αιώνα η Ελλάδα έχει βιώσει σημαντικές μειώσεις στα επίπεδα της γονιμότητας ως αποτέλεσμα αξιοσημείωτων κοινωνικό – οικονομικών μεταβολών που έχουν συντελεστεί τόσο σε εθνικό όσο και σε τοπικό επίπεδο. Το έτος 2000 ο δείκτης ολικής γονιμότητας για ολόκληρη τη χώρα και για το 70% των νομών της βρίσκονταν κάτω από το επίπεδο των 1,3 παιδιών ανά γυναίκα. Ωστόσο, παρά το γεγονός ότι οι περιφερειακές διαφορές στα επίπεδα της γονιμότητας έχουν σταδιακά μειωθεί και παρά την κοινωνικό – οικονομική σύγκλιση των Ελληνικών τοπικών κοινωνιών λόγω της υφιστάμενης ολοκλήρωσης, ακόμη υπάρχουν ουσιαστικές διακυμάνσεις στη γονιμότητα.

Μεγάλο μέρος της βιβλιογραφίας υποδεικνύει ότι τα περισσότερα κοινωνικά φαινόμενα δεν παρουσιάζουν ομοιογένεια στο χώρο αλλά συνήθως επηρεάζονται από χωρικά αποτελέσματα. Όσον αφορά τη γονιμότητα είναι λογικό να αναμένεται ότι όχι μόνο τα επίπεδά της αλλά και οι δομικές της σχέσεις με τους παράγοντες που την καθορίζουν διαφοροποιούνται στο χώρο εξ' αιτίας συγκεκριμένων τοπικών συνθηκών. Η χωρική μη στασιμότητα στα επίπεδα της γονιμότητας θα μπορούσε να αποδοθεί σε έναν αριθμό παραγόντων που περιλαμβάνουν τοπικές στάσεις, αξίες, τρόπους ζωής, αποφάσεις, κοινωνικά χαρακτηριστικά και συναφείς ιδιότητες καθώς επίσης στη γεωγραφική ολοκλήρωση που συνεπάγεται διάχυση όμοιων συμπεριφορών σε γειτονικές περιοχές.

Το κεφάλαιο αυτό συνεισφέρει στην κατανόηση της περιφερειακής διάστασης της γονιμότητας στην Ελλάδα και εξετάζει τα χωρικά πρότυπα και τις σχέσεις μεταξύ γονιμότητας και επιλεγμένων κοινωνικών και δημογραφικών παραγόντων ελέγχοντας την παρουσία χωρικής αυτοσυσχέτισης και έλλειψης σταθερότητας σε επίπεδο δήμων. Η υπόθεση της χωρικής ομοιογένειας στη σχέση της γονιμότητας και των επιλεγμένων ερμηνευτικών μεταβλητών ελέγχτηκε με την εκτίμηση ενός ολικού υποδείγματος παλινδρόμησης για το σύνολο της χώρας με τη μέθοδο των ελαχίστων τετραγώνων και τοπικών υποδειγμάτων σε κάθε δήμο που εκτιμήθηκαν με τη μέθοδο της γεωγραφικά σταθμισμένης παλινδρόμησης.

Όπως παρουσιάστηκε, μολονότι οι συντελεστές του ολικού υποδείγματος έχουν τα αναμενόμενα πρόσημα και είναι στατιστικά σημαντικοί, εντούτοις η ερμηνευτική του ικανότητα είναι περιορισμένη. Επιπρόσθετα, οι διαγνωστικοί έλεγχοι αποκαλύπτουν ότι παραβιάζονται οι βασικές υποθέσεις της μεθόδου των ελαχίστων τετραγώνων για τα σφάλματα του εκτιμηθέντος υποδείγματος τα οποία δεν έχουν σταθερή διακύμανση,

παρουσιάζουν θετική χωρική αυτοσυσχέτιση και δεν κατανέμονται κανονικά. Αυτά τα προβλήματα φανερώνουν την ύπαρξη χωρικής ετερογένειας ως επίπτωση χωρικών επιδράσεων, δηλαδή η σχέση εξάρτησης ανάμεσα στο επίπεδο της γονιμότητας και τις επιλεγμένες ερμηνευτικές μεταβλητές δεν είναι στο χώρο στάσιμη. Κατά συνέπεια, η εκτίμηση ενός ολικού υποδείγματος αποτυγχάνει να απεικονίσει για το σύνολο της χώρας τις βαθύτερες περιφερειακές σχέσεις εξάρτησης της γονιμότητας με τους εξεταζόμενους παράγοντες.

Η αντιμετώπιση των επιπτώσεων της χωρικής ετερογένειας και εξάρτησης των δεδομένων στο υπόδειγμα έγινε με την εφαρμογή της μεθόδου της γεωγραφικά σταθμισμένης παλινδρόμησης με συνάρτηση στάθμισης προσαρμοσμένου διτετράγωνου πυρήνα που οδήγησε στην εκτίμηση τοπικών υποδειγμάτων για κάθε δήμο. Η προσέγγιση αυτή σχεδόν διπλασίασε την ερμηνευτική ικανότητα του υποδείγματος και εξάλειψε το πρόβλημα της χωρικής αυτοσυσχέτισης στα σφάλματα φανερώνοντας την ανωτερότητα των τοπικών υποδειγμάτων σε σύγκριση με το ολικό υπόδειγμα. Η χαρτογράφηση των αποτελεσμάτων αποκάλυψε ότι σε επίπεδο δήμων υπάρχουν σημαντικές διαφοροποιήσεις τόσο στην κατεύθυνση όσο και στην ένταση των επιδράσεων των κοινωνικών και οικονομικών παραγόντων στα επίπεδα της γονιμότητας.

ΒΑΣΙΚΑ ΣΥΜΠΕΡΑΣΜΑΤΑ ΤΗΣ ΔΙΑΤΡΙΒΗΣ

Συνοπτικά, από την παρούσα διατριβή έχουν προκύψει τα εξής κύρια ερευνητικά ευρήματα:

- 1) Οι LM έλεγχοι χωρικής εξάρτησης προσφέρουν σημαντική βοήθεια στον εντοπισμό της ύπαρξης του προβλήματος της χωρικής αυτοσυσχέτισης στα σφάλματα ενός γραμμικού υποδείγματος αλλά η συμπεριφορά τους ως προς τη διάκριση της αιτίας του προβλήματος εξαρτάται από το μέγεθος του δείγματος και σε κάποιες περιπτώσεις από τον ορισμό της χωρικής γειτνίασης μέσω της μήτρας των χωρικών σταθμίσεων καθώς μπορεί να οδηγήσει σε τελείως διαφορετικά αποτελέσματα. Πιο συγκεκριμένα:
 - α) Στα μικρά και μέτρια δείγματα οι έλεγχοι χωρικής εξάρτησης καταφέρνουν ικανοποιητικά να εντοπίζουν την αιτία εμφάνισης του προβλήματος της χωρικής αυτοσυσχέτισης και να επιλέγουν τα κατάλληλα χωρικά οικονομετρικά υποδείγματα κάτι που έχει αναφερθεί και σε παλιότερες γνωστές στη βιβλιογραφία της χωρικής οικονομετρίας εργασίες.
 - β) Η χρησιμοποίηση μεγαλύτερων δειγμάτων δεν βελτιώνει απαραίτητα τα αποτελέσματα αλλά αυξάνει τα ποσοστά απόρριψης των ελέγχων χωρικής εξάρτησης προκαλώντας μερικές φορές ασάφεια ως προς τον εντοπισμό της αιτίας της αυτοσυσχέτισης κάτι που δεν είχε επισημανθεί στις παλιότερες εργασίες διότι οι προσομοιώσεις τους χρησιμοποιούσαν μικρότερα δείγματα. Σε αυτές τις περιπτώσεις η σύγκριση των τιμών των ανθεκτικών στατιστικών LM-LE και LM-EL μπορεί να οδηγήσει στο σωστό υπόδειγμα.
 - γ) Ασυνέπεια στα αποτελέσματα προέκυψε για τις ανθεκτικές στατιστικές LM-LE και LM-EL όταν η αιτία της αυτοσυσχέτισης είναι η ύπαρξη χωρικής υστέρησης στην εξαρτημένη μεταβλητή που έχει παραληφθεί από το υπόδειγμα όπου τα ποσοστά απόρριψης των ελέγχων με τις δύο στατιστικές παρουσιάζουν δραματικές διαφοροποιήσεις μεταξύ των κριτηρίων του πύργου και της βασίλισσας. Επομένως, είναι αναγκαίο ο ερευνητής να ελέγχει τη συνέπεια των αποτελεσμάτων από τους ελέγχους χρησιμοποιώντας διαφορετικές μήτρες χωρικής γειτνίασης πριν αποφασίσει το τελικό χωρικό υπόδειγμα που θα εκτιμήσει.

- 2) Ο έλεγχος για την ύπαρξη χωρικής αυτοσυσχέτισης στα σφάλματα του υποδείγματος χωρικής υστέρησης στην εξαρτημένη μεταβλητή μπορεί να γίνει τόσο με στατιστική I του Moran όσο και με την στατιστική $LM_{\lambda\rho}$ με εξίσου ικανοποιητικά αποτελέσματα. Μάλιστα, στα μικρά δείγματα, η ισχύς της στατιστικής I είναι ανώτερη από την ισχύ της στατιστικής $LM_{\lambda\rho}$ ενώ στα μεγαλύτερα δείγματα οι δύο στατιστικές έχουν παρόμοια ισχύ.
- 3) Μια άλλη αιτία που οδηγεί στην εμφάνιση χωρικής αυτοσυσχέτισης στα σφάλματα ενός γραμμικού υποδείγματος που εκτιμάται με χωρικά δεδομένα είναι η ύπαρξη του φαινομένου της υποφαινομενικής συμπεριφοράς όπως έχει επισημανθεί και από άλλους ερευνητές. Ειδικότερα, θεωρώντας δύο ανεξάρτητες στάσιμες χωρικές αυτοπαλίνδρομες διαδικασίες πρώτης τάξης, SAR(1):
- α) Η συμπεριφορά του φαινομένου αυτού είναι παρόμοια με αυτή που εμφανίζεται στην ανάλυση χρονοσειρών για δύο ανεξάρτητες στάσιμες AR(1) διαδικασίες με τη διαφορά ότι προκύπτουν λίγο πιο μικρά ποσοστά απόρριψης της μηδενικής υπόθεσης του ελέγχου σημαντικότητας της ανεξάρτητης μεταβλητής για κάθε τιμή της αυτοπαλίνδρομης παραμέτρου.
 - β) Υπάρχουν ισχυρές ενδείξεις χωρικής αυτοσυσχέτισης στα σφάλματα του απλού υποδείγματος όπως προκύπτει από τις μέσες τιμές της στατιστικής I του Moran που αυξάνουν καθώς αυξάνεται η τιμή της αυτοπαλίνδρομης παραμέτρου. Επιπρόσθετα, τα ποσοστά απόρριψης της στατιστικής I αυξάνουν καθώς αυξάνεται το μέγεθος του δείγματος ακόμη και για μικρές τιμές της αυτοπαλίνδρομης παραμέτρου.
 - γ) Ο έλεγχος LM-LAG απορρίπτει τη μηδενική του υπόθεση πολύ συχνά υποδεικνύοντας την εκτίμηση του Υποδείγματος Χωρικής Υστέρησης (Spatial Lag Model) με το ποσοστό απόρριψης να αυξάνεται καθώς αυξάνεται η τιμή της αυτοπαλίνδρομης παραμέτρου και το μέγεθος του δείγματος. Παρόμοια αποτελέσματα προκύπτουν και για τον έλεγχο LM-ERR που υποδεικνύει την εκτίμηση του Υποδείγματος Χωρικού Σφάλματος (Spatial Error Model). Ο ανθεκτικός έλεγχος LM-LE δείχνει την εκτίμηση του Υποδείγματος Χωρικής Υστέρησης αλλά σε μικρότερο ποσοστό. Αντίθετα, ο ανθεκτικός έλεγχος LM-EL ποτέ δεν επιλέγει το Υπόδειγμα Χωρικού Σφάλματος. Δηλαδή, οι ανθεκτικοί έλεγχοι

διαλέγουν το απλό οικονομετρικό υπόδειγμα στο οποίο υπάρχει το πρόβλημα της υποφαινομενικής συμπεριφοράς.

- δ) Εάν ο ερευνητής ακολουθήσει τις υποδείξεις από την εφαρμογή των απλών ελέγχων χωρικής εξάρτησης και εκτιμήσει το Υπόδειγμα Χωρικής Υστέρησης ή το Υπόδειγμα Χωρικού Σφάλματος τότε η υποφαινομενική συμπεριφορά εξαφανίζεται. Τα ποσοστά απόρριψης του ελέγχου της στατιστικής σημαντικότητας του συντελεστή της ανεξάρτητης μεταβλητής είναι όλα πολύ κοντά στο επίπεδο σημαντικότητας του 5% για όλες τις τιμές της χωρικά αυτοπαλίνδρομης παραμέτρου και όλα τα μεγέθη δείγματος ενώ τα σφάλματα δεν παρουσιάζουν πρόβλημα χωρικής αυτοσυσχέτισης. Είναι αδιάφορο το εάν θα εκτιμηθεί ένα Υπόδειγμα Χωρικής Υστέρησης ή ένα Υπόδειγμα Χωρικού Σφάλματος.
- ε) Η ανακολουθία στις υποδείξεις μεταξύ των δύο μη ανθεκτικών στατιστικών και των δύο ανθεκτικών στατιστικών θα μπορούσε να χρησιμοποιηθεί από τον ερευνητή ως κριτήριο για τον εντοπισμό του προβλήματος της υποφαινομενικής παλινδρόμησης με χωρικά δεδομένα.
- 4) Η γεωγραφικά σταθμισμένη παλινδρόμηση, δηλαδή η μέθοδος εκτίμησης που συνήθως χρησιμοποιείται στην αντιμετώπιση του προβλήματος της χωρικής ετερογένειας μέσω της εκτίμησης τοπικών υποδειγμάτων, έχει πολλά πλεονεκτήματα αλλά υπάρχουν περιπτώσεις που μπορεί να οδηγήσει σε παραπλανητικά αποτελέσματα. Πιο συγκεκριμένα:
- α) Σε όλα τα υποδείγματα που εξετάστηκαν ανεξάρτητα από το εάν εκφράζουν ύπαρξη ολικής ή τοπικής σχέσης μεταξύ των μεταβλητών, οι εκτιμηθείσες τιμές των συντελεστών τους είναι πολύ κοντά στις πραγματικές τους τιμές. Κατά συνέπεια η χαρτογράφηση αυτών των τιμών μπορεί να βοηθήσει τον ερευνητή στη διάκριση μιας μη σταθερής σχέσης από μια σταθερή σχέση.
- β) Η σύγκριση των συντελεστών προσδιορισμού του ολικού υποδείγματος με αυτόν των τοπικών υποδειγμάτων και της τιμής του εύρους που επιλέγει η διασταυρωτική επικύρωση με το ολικό εύρος μπορούν να βοηθήσουν στην επιλογή της σωστής μεθόδου εκτίμησης. Οι έλεγχοι συνεισφοράς, παρά τα προβλήματα που δείχνουν ότι έχουν μερικές φορές, μπορούν επικουρικά να χρησιμοποιηθούν για την επιλογή της σωστής μεθόδου εκτίμησης.

- γ) Σε κάποιες περιπτώσεις τα κριτήρια μπορεί εσφαλμένα να προτείνουν τη γεωγραφικά σταθμισμένη παλινδρόμηση, όπως συμβαίνει όταν στην πραγματικότητα ισχύει ένα υπόδειγμα χωρικού καθεστώτος, κάτι που σημαίνει ότι ο ερευνητής θα πρέπει να είναι ιδιαίτερα προσεκτικός όταν εφαρμόζει την τεχνική καθώς δεν μπορεί να αποδειχθεί ότι πάντα οδηγεί στα σωστά συμπεράσματα.
- 5) Στην περίπτωση που ο ερευνητής εκλάβει εσφαλμένα τις επιδράσεις χωρικής εξάρτησης ως επιδράσεις χωρικής ετερογένειας και προσπαθήσει να τις αντιμετωπίσει με την εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης και την εκτίμηση τοπικών υποδειγμάτων τότε οι επιπτώσεις είναι διαφορετικές ανάλογα με την αιτία που προκαλεί την εμφάνιση της χωρικής εξάρτησης. Ειδικότερα:
- α) Ποιο ήπιες είναι οι επιπτώσεις όταν οι χωρικές επιδράσεις οφείλονται στο Υπόδειγμα Χωρικού Σφάλματος το οποίο διακρίνεται πιο εύκολα συγκρίνοντας τους συντελεστές προσδιορισμού της απλής μεθόδου των ελαχίστων τετραγώνων και της γεωγραφικά σταθμισμένης παλινδρόμησης ενώ για μέτριες τιμές των χωρικών αυτοπαλινδρομων συντελεστών των σφαλμάτων βοήθεια προσφέρουν και οι έλεγχοι βελτίωσης καθώς και οι τιμές του εύρους που προκύπτουν από τη διασταυρωτική επικύρωση. Οι εκτιμηθείσες τιμές των τοπικών συντελεστών δεν παρουσιάζουν σοβαρή μεροληψία και βρίσκονται όλες κοντά στην ολική τιμή τους χωρίς να παρουσιάζουν έντονη μεταβλητότητα.
- β) Όταν οι χωρικές επιδράσεις οφείλονται στην ύπαρξη χωρικής υστέρησης στην εξαρτημένη μεταβλητή που αγνοείται από τον ερευνητή και δεν προβαίνει στην εκτίμηση του Υποδείγματος Χωρικής Υστέρησης αλλά εφαρμόζει γεωγραφικά σταθμισμένη παλινδρόμηση οι επιπτώσεις θα είναι πολύ πιο σοβαρές. Στην περίπτωση αυτή, η διάκριση της εσφαλμένης ταυτοποίησης είναι δύσκολο να γίνει διότι οι συντελεστές προσδιορισμού, το εύρος και τα αποτελέσματα των ελέγχων θα υποστηρίζουν την εκτίμηση τοπικών υποδειγμάτων ενώ ταυτόχρονα παρουσιάζεται και μεροληψία στις εκτιμήσεις των τοπικών συντελεστών.
- 6) Ο ερευνητής θα πρέπει να είναι ενήμερος ότι εκτός από τα προβλήματα που έχουν αναφερθεί στη βιβλιογραφία για τη γεωγραφικά σταθμισμένη παλινδρόμηση υπάρχει και ο κίνδυνος εμφάνισης του φαινομένου της υποφαινομενικής συμπεριφοράς στα τοπικά

υποδείγματα που εκτιμώνται με την εφαρμογή της. Συγκεκριμένα, για την περίπτωση δύο ανεξάρτητων στάσιμων χωρικών αυτοπαλίνδρομων διαδικασιών πρώτης τάξης, SAR(1):

- α) Το φαινόμενο της υποφαινομενικής παλινδρόμησης σε τοπικό επίπεδο είναι ισχυρότερο από ότι στο ολικό επίπεδο.
- β) Οι μέσες τιμές των συντελεστών προσδιορισμού της γεωγραφικά σταθμισμένης παλινδρόμησης είναι πάντοτε μεγαλύτερες από τις μέσες τιμές των συντελεστών προσδιορισμού της μεθόδου των ελαχίστων τετραγώνων για όλα τα μεγέθη δείγματος και για όλες τις τιμές της αυτοπαλίνδρομης παραμέτρου υποδεικνύοντας τα τοπικά υποδείγματα. Η διαφορά μεταξύ των μέσων τιμών των δύο συντελεστών προσδιορισμού γίνεται μεγαλύτερη όσο αυξάνονται οι τιμές της αυτοπαλίνδρομης παραμέτρου για δεδομένο μέγεθος δείγματος και παραμένει κατά προσέγγιση αμετάβλητη ανεξάρτητα του δείγματος και της συνάρτησης στάθμισης παρέχοντας ισχυρότερες ενδείξεις για την καταλληλότητα του τοπικού υποδείγματος.
- γ) Οι έλεγχοι που εξετάζουν τη συνεισφορά της γεωγραφικά σταθμισμένης παλινδρόμησης δείχνουν την εκτίμηση των τοπικών υποδειγμάτων με την έννοια ότι η εκτίμησή τους βελτιώνει τα αποτελέσματα. Οι ενδείξεις των ελέγχων υπέρ των τοπικών υποδειγμάτων ενισχύονται σημαντικά καθώς αυξάνονται οι τιμές της αυτοπαλίνδρομης παραμέτρου ανεξάρτητα από το μέγεθος του δείγματος και της συνάρτησης στάθμισης. Για τιμή της αυτοπαλίνδρομης παραμέτρου ίσης με 0.9 τα ποσοστά απόρριψης είναι 100% και η εξειδίκευση τοπικών υποδειγμάτων δηλώνεται πάντοτε από τους ελέγχους.
- δ) Οι τιμές του εύρους, ως κριτήριο επιλογής υποδείγματος, προτείνουν την εξειδίκευση του τοπικού υποδείγματος όταν αυξάνει η τιμή της αυτοπαλίνδρομης παραμέτρου, ανεξάρτητα από το μέγεθος του δείγματος και τη συνάρτηση στάθμισης. Η συμπεριφορά αυτή είναι σε συμφωνία με τη συμπεριφορά των στατιστικών ελέγχων και των συντελεστών προσδιορισμού. Η τιμή του εύρους έχει στενή σχέση με το συντελεστή προσδιορισμού της γεωγραφικά σταθμισμένης παλινδρόμησης καθώς η ερμηνευτική ικανότητα αυξάνεται όταν η τιμή του εύρους μειώνεται.
- ε) Σε κάθε εφαρμογή της γεωγραφικά σταθμισμένης παλινδρόμησης, ο ερευνητής θα έχει ενδείξεις ότι υπάρχει γραμμική σχέση σε τοπικό επίπεδο επειδή η μηδενική υπόθεση θα απορρίπτεται σε ένα σημαντικό ποσοστό περιοχών μεγαλύτερο από το

αναμενόμενο του 5%, ένα αποτέλεσμα που έχει σχέση με την τιμή της χωρικά αυτοπαλίνδρομης παραμέτρου.

- στ) Περισσότερα υποφαινομενικά αποτελέσματα θα εμφανίζονται στην ανάλυση όσο αυξάνεται η τιμή της χωρικά αυτοπαλίνδρομης παραμέτρου μια συμπεριφορά που είναι όμοια στην ανάλυση χρονοσειρών και στη χωρική ανάλυση σε ολικό επίπεδο. Για υψηλές τιμές της χωρικά αυτοπαλίνδρομης παραμέτρου το πρόβλημα της υποφαινομενικής συμπεριφοράς θα υπάρχει πάντα διότι σχεδόν το 100% των τοπικών ποσοστών απορρίψεων ανήκει σε διαστήματα ανώτερα του διαστήματος $[0, 5]$.
- ζ) Το μέγεθος του δείγματος επηρεάζει τη συμπεριφορά των ελέγχων μόνο για μέτριες και υψηλές τιμές της χωρικά αυτοπαλίνδρομης παραμέτρου αλλάζοντας την πυκνότητα της κατανομής της συχνότητας των τοπικών ποσοστών απορρίψεων προκαλώντας συγκεντρώση σε συγκεκριμένα διαστήματα. Ο μεγάλος αριθμός τοπικών απορρίψεων προκύπτει ταυτόχρονα με πολύ χαμηλές τιμές του εύρους κάτι που δείχνει ότι τα αποτελέσματα είναι ύποπτα καθώς το χαμηλό εύρος προκαλεί πλασματική μεταβλητότητα που ίσως επηρεάζει την εφαρμογή του ελέγχου.
- η) Τα εκτιμηθέντα σφάλματα από τη γεωγραφικά σταθμισμένη παλινδρόμηση παρουσιάζουν χωρική αυτοσυσχέτιση σύμφωνα με τη στατιστική I του Moran με το κριτήριο του πύργου και την εφαρμογή του ελέγχου με την προσέγγιση των Leung, Mei και Zhang (2000b) κάτι που ισχύει και στην ανάλυση χρονοσειρών και στη χωρική ανάλυση σε ολικό επίπεδο. Περισσότερες ενδείξεις χωρικής αυτοσυσχέτισης στα σφάλματα εμφανίζονται όταν η τιμή της χωρικά αυτοπαλίνδρομης παραμέτρου αυξάνεται για δεδομένο δείγμα καθώς και όταν αυξάνεται το μέγεθος του δείγματος για δεδομένη τιμή της χωρικά αυτοπαλίνδρομης παραμέτρου ιδιαίτερα για μικρές τιμές της.
- θ) Σε σχέση με τις χρονοσειρές και τη χωρική ανάλυση σε ολικό επίπεδο ο ερευνητής δεν έχει διέξοδο να αποφύγει την υποφαινομενική συμπεριφορά. Στην ανάλυση χρονοσειρών η εφαρμογή εναλλακτικών μορφών εκτίμησης του υποδείγματος εξαλείφουν την αυτοσυσχέτιση και την υποφαινομενικότητα. Αντίστοιχα, στη χωρική ανάλυση, η εφαρμογή των κλασικών LM ελέγχων χωρικής εξάρτησης και εξειδίκευσης προτείνουν άλλες μορφές εκτίμησης του υποδείγματος που δεν

οδηγούν σε υποφαινομενικά αποτελέσματα και σε χωρική αυτοσυσχέτιση στα σφάλματα.

- ι) Η γεωγραφικά σταθμισμένη παλινδρόμηση θα πρέπει να εφαρμόζεται με προσοχή σε κάθε χωρική ανάλυση και τα αποτελέσματα που προκύπτουν να ερμηνεύονται πολύ προσεχτικά ακόμη και στην περίπτωση που οι έλεγχοι βελτίωσης δείχνουν ότι η εκτίμηση σε τοπικό επίπεδο είναι καλύτερη. Μεγάλη προσοχή θα πρέπει να δίνεται όταν υψηλές τιμές του συντελεστή προσδιορισμού της γεωγραφικά σταθμισμένης παλινδρόμησης εμφανίζονται ταυτόχρονα με χαμηλές τιμές του εύρους.
- 7) Στο εμπειρικό μέρος της διατριβής έγινε προσπάθεια να εξειδικευτεί ένα υπόδειγμα που ερμηνεύει τις χωρικές διαφοροποιήσεις και τα πρότυπα στις σχέσεις μεταξύ του επιπέδου της γονιμότητας του Ελληνικού πληθυσμού και ενός αριθμού επιλεγμένων κοινωνικό – οικονομικών δεικτών σε τοπικό διοικητικό επίπεδο λαμβάνοντας υπ' όψιν την παρουσία της χωρικής ετερογένειας. Το υπόδειγμα αυτό εκτιμήθηκε σε ολικό επίπεδο με τη μέθοδο των ελαχίστων τετραγώνων και σε τοπικό επίπεδο με τη μέθοδο της γεωγραφικά σταθμισμένης παλινδρόμησης. Ειδικότερα διαπιστώθηκε ότι:
 - α) Η γονιμότητα του πληθυσμού της Ελλάδας όπως εκτιμάται με το δείκτη του λόγου του παιδικού προς τον αναπαραγωγικό πληθυσμό (CWR) παρουσιάζει σημαντικές χωρικές διαφοροποιήσεις μεταξύ των τοπικών διοικητικών μονάδων.
 - β) Στο σύνολο της χώρας (ολικό επίπεδο) η γονιμότητα του πληθυσμού επηρεάζεται θετικά από το ποσοστό των παντρεμένων γυναικών αναπαραγωγικής ηλικίας, το ποσοστό των μεταναστών σε κάθε τοπική διοικητική μονάδα και το χαμηλό κοινωνικό και οικονομικό περιβάλλον και αρνητικά από το ποσοστό των πυρηνικών οικογενειών. Όλες οι μεταβλητές του υποδείγματος παρουσιάζουν υψηλή θετική χωρική αυτοσυσχέτιση. Επιπρόσθετα, τα σφάλματα του ολικού υποδείγματος παρουσιάζουν χωρική αυτοσυσχέτιση και ετεροσκεδαστικότητα.
 - γ) Η υψηλότερη ερμηνευτική ικανότητα των τοπικών υποδειγμάτων εντοπίστηκε στην περιφέρεια της Αττικής και πιο συγκεκριμένα στην Αθήνα και στους γειτονικούς της δήμους. Οι περιοχές αυτές επιδεικνύουν χαμηλά επίπεδα γονιμότητας αλλά, καθώς ο πληθυσμός τους αντιπροσωπεύει περισσότερο από το μισό πληθυσμό της Ελλάδας, η συμπεριφορά τους τείνει να επηρεάζει ολόκληρο το κοινωνικό – δημογραφικό περιβάλλον. Η χαμηλότερη ερμηνευτική ικανότητα του υποδείγματος εντοπίστηκε

στη κεντρική Ελλάδα και κυρίως στα νησιά όπου, εξ' αιτίας των γεωγραφικών τους χαρακτηριστικών, προκύπτουν δυσκολίες στην ανάλυση χωρικών δεδομένων.

- δ) Επειδή η εξωσυζυγική γονιμότητα στην Ελλάδα είναι ακόμη πολύ χαμηλή, το ποσοστό των παντρεμένων γυναικών αναπαραγωγικής ηλικίας ασκεί την πιο σημαντική θετική επίδραση στα περιφερειακά επίπεδα γονιμότητας κάτι που αποδεικνύεται τόσο από το ολικό υπόδειγμα όσο και από τα τοπικά υποδείγματα.
- ε) Η επίδραση των μεταναστών στο μέγεθος της γονιμότητας δεν φαίνεται να είναι σε τοπικό επίπεδο ξεκάθαρη. Το γεγονός αυτό μπορεί να οφείλεται στο ότι τα στοιχεία της έρευνας προέρχονται από την απογραφή πληθυσμού του έτους 2001 περίοδος όπου η μετανάστευση ήταν ακόμη νέο φαινόμενο στην Ελλάδα.
- στ) Παρόλο που οι κοινωνικό – οικονομικές διαφοροποιήσεις ερμηνεύουν σε κάποιο βαθμό τις περιφερειακές διακυμάνσεις της γονιμότητας η σχέση μεταξύ κοινωνικών και οικονομικών παραγόντων και επιπέδου γονιμότητας δεν είναι πάντα ξεκάθαρη και πιθανών εμπλέκονται και άλλοι ποσοτικοί ή ποιοτικοί παράγοντες που τη διαμορφώνουν.
- ζ) Παρόλο που η πυρηνική οικογένεια γίνεται ο κυρίαρχος τύπος της οικογενειακής δομής στην Ελλάδα, τα τοπικά υποδείγματα δεν έδειξαν μία σαφή εικόνα σχετικά με την επίδρασή της κάτι που ίσως οφείλεται στη βοήθεια, οικονομική και μη, που λαμβάνουν τα νέα ζευγάρια από τους γονείς τους για την ανατροφή των παιδιών.

BIBΛΙΟΓΡΑΦΙΑ

- Acosta, P., de Matos, E. and Fava, A. (2006). Monte Carlo properties of spatial dependence LM tests, *Applied Economics Letters*, **13**, 905-910.
- Agiakloglou, C. (2009). Evidence of ARCH(1) errors in the context of spurious regressions, *Communications in Statistics – Simulations and Computation*, **38**, 1803-1810.
- Agiakloglou, C. (2013). Resolving spurious regressions and serially correlated errors, *Empirical Economics*, **45**, 1361-1366.
- Agiakloglou, C., Tsimbos C. and Tsimpanos, A. (2015). Is spurious behaviour an issue for two independent stationary spatial autoregressive SAR(1) processes? *Applied Economics Letters*, **22**, 1372-1377.
- Agiakloglou, C. and Tsimpanos, A. (2012). An alternative approach for testing for linear association for two independent stationary AR(1) processes, *Applied Economics*, **44**, 4799-4803.
- Anselin, L. (1988). *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers, Dordrecht.
- Anselin, L. (1988b). Lagrange Multiplier Test Diagnostics for Spatial Dependence and Spatial Heterogeneity, *Geographical Analysis*, **20**, 1-17.
- Anselin, L. (1995). Local indicators of spatial association – LISA, *Geographical Analysis*, **27**, 93-115.
- Anselin, L. (1996). The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association, in M. Fischer, H. Scholten and D. Unwin, (eds.), *Spatial Analytical Perspectives on GIS*, 111-125, Taylor and Francis, London.
- Anselin L (2001). Spatial effects in econometric practice in environmental and resource economics, *American Journal of Agricultural Economics* **83**, 705-710.
- Anselin, L. (2004). *GeoDa 0.95i Release Notes. Spatial Analysis Laboratory (SAL). Department of Agricultural and Consumer Economics. Urbana-Champaign, IL: University of Illinois.*
- Anselin, L. (2010). Thirty years of spatial econometrics, *papers in Regional Science*, **89**, 3-25.
- Anselin, L. and Bera, A. K. (1998), Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics, in: Ullah, A. and Giles, D., (eds.), *Handbook of Applied Economic Statistics*, 237-289, Marcel Dekker, New York.
- Anselin, L., Bera A., Florax, R. and Yoon, M. (1996). Simple diagnostic tests for spatial dependence, *Regional Science and Urban Economics*, **26**, 77-104.
- Anselin, L. and Florax, R. J. (1995). Small sample properties of tests for spatial dependence in regression models: Some further results, in Anselin, L. and Florax, R. J., (eds.), *New Directions in Spatial Econometrics*, 21-74. Springer-Verlag, Berlin.
- Anselin L. and Griffith D. A. (1988). Do Spatial Effects Really Matter in Regression Analysis? *Papers of Regional Science Association*, **65**, 11-34.

- Anselin L. and Rey, S. (1991). Properties of Tests for Spatial Dependence in Linear Regression Models, *Geographical Analysis*, **23**, 112-131.
- Arbia, G. (2006). *Spatial Econometrics*, Springer, Berlin.
- Bagheri N, Holt A., and Benwell G.L. (2009). Using Geographically Weighted Regression to Validate Approaches for Modelling Accessibility to Primary Health Care, *Applied Spatial Analysis and Policy*, **2**, 177-194.
- Bartlett, M. S. (1935). Some aspects of the time-correlation problems in regard to tests of significance, *Journal of the Royal Statistical Society*, **98**, 536-543.
- Bijak, J, Kupiszewska, D., and Kupiszewski, M. (2008). Replacement Migration Revisited, Simulations of the Effects of Selected Population and Labor Market Strategies for the Ageing Europe, 2002-2052, *Population Research and Policy Review*, **27**, 321-342.
- Bivand, R. (2015). Spdep: Spatial Dependence: Weighting Schemes. Statistics and Models. R package version 0.5-82. <http://CRAN.R-project.org/package=spdep>
- Bivand, R. and Yu, D. (2014). Spgwr: Geographically weighted regression. R package version 0.6-26. <http://CRAN.R-project.org/package=spgwr>
- Breusch, T. and Pagan, A. (1979). A Simple Test of Heteroskedasticity and Random Coefficient Variation, *Econometrica*, **47**, 1287-1294.
- Brunsdon, C., Fotheringham, A. S. and Charlton, M. E. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity, *Geographical Analysis*, **28**, 281-298.
- Brunsdon, C., Fotheringham, A. S. and Charlton, M. E. (1999). Some Notes on Parametric Significance Tests For Geographically Weighted Regression, *Journal of Regional Science*, **39**, 497-524.
- Burch T. K. and Gendell M. (1970). Extended Family Structure and Fertility: Some Conceptual and Methodological Issues, *Journal of Marriage and Family*, **32**, 227-236.
- Burridge, P. (1980). On the Cliff-Ord Test for Spatial Correlation, *Journal of the Royal Statistical Society B*, **42**, 107-108.
- Casetti, E. (1972). Generating Models by the Expansion Method: Applications to Geographic Research, *Geographical Analysis*, **4**, 81-91.
- Casetti, E. (1986). The Dual Expansion Method: An Application for Evaluating the Effects of Population Growth on Development, *IEEE Transactions on Systems, Man, and Cybernetics*, **16**, 29-39.
- Casetti, E. and Jones, J. P. (1987). Spatial Parameter Variation by Orthogonal Trend Surface Expansions: An Application to the Analysis of Welfare Program Participation Rates, *Social Science Research*, **16**, 385-300.
- de Castro, M.C. (2007). Spatial Demography: An Opportunity to Improve Policy Making at Diverse Decision Levels, *Population Research and Policy Review*, **26**, 477-509.
- Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots, *Journal of the American Statistical Association*, **74**, 829-836.

- Cleveland W. S. and Devlin S. J. (1988). Locally weighted regression: an approach to regression analysis by local fitting, *Journal of the American Statistical Association*, **83**, 596-610.
- Cliff, A. and Ord, J. K. (1972). Testing for Spatial Autocorrelation among Regression Residuals, *Geographical Analysis*, **4**, 267-284.
- Cliff, A. D. and Ord, J. K. (1981). *Spatial Processes: Models and Applications*, Pion Ltd, London.
- Coleman, D. (2005). Facing the 21st century, New developments, continuing problems. In: Macura M., McDonald A.L. and Haug W. (eds.), *The new Demographic Regime, Population Challenges and Policy Responses*, Geneva, United Nations, pp. 45-56.
- Elhorst, J. P. (2014). *Spatial Econometrics From Cross-Sectional Data to Spatial Panels*, Springer, Berlin New York Dordrecht London.
- ELSTAT (2011). Statistics [electronic resource]. Greece: Hellenic Statistical Authority. <http://www.statistics.gr>.
- Farber S. and Páez, A.(2007). A systematic investigation of cross-validation in GWR model estimation: empirical analysis and Monte Carlo simulations, *Journal of Geographical Systems*, **9**, 371-396.
- Fingleton, B. (1999). Spurious spatial regression: Some Monte Carlo results with a spatial unit root and spatial cointegration, *Journal of Regional Science*, **39**, 1-19.
- Fotheringham, A. S., and Brunson, C.(1999). Local Forms of Spatial Analysis. *Geographical Analysis*, **31**, 340-358.
- Fotheringham, A. S., Brunson, C., and Charlton, M. E. (2002). *Geographically weighted regression: the analysis of spatially varying relationships*, John Wiley, West Sussex.
- Frejka, T. and Sardón, J.P. (2004). *Childbearing Prospects in Low-Fertility Countries: A Cohort Analysis*, Kluwer Academic Publishers, Dordrecht.
- Granger, C. W. J., Hyung, N. and Jeon, Y. (2001). Spurious regressions with stationary series, *Applied Economics*, **33**, 899-904.
- Granger, C. W. J. and Newbold, P. (1974). Spurious regressions in econometrics, *Journal of Econometrics*, **2**, 111-120.
- Guilmoto C. Z. and Rajan S. I. (2001). Spatial Patterns of Fertility Transition in Indian Districts, *Population and Development Review*, **27**, 713-738.
- Haining R. (1978). The moving average process for spatial interaction, *Transactions of the Institute of British Geographers*, **3**, 202-225.
- Hank K. (2001). Regional Fertility Differences in West Germany: An Overview of the Literature and Recent Descriptive Findings, *International Journal of Population Geography*, **7**, 243-257.
- Hurvich, C. M., Simonoff J. S., and Tsai C-L. (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion, *Journal of the Royal Statistical Society Series B*, **60**, 271-293.
- Imhof J. P (1961). Computing the distribution of quadratic forms in normal variables, *Biometrika*, **48**, 419-426.

- Isik, O. and Pinarcioglu, M. M. (2006). Geographies of transitions: a geographically weighted regression approach to regional fertility differences in Turkey, *European Journal of Population*, **22**, 399-421.
- Jarque, Carlos M. and Bera, Anil K. (1987). A test for normality of observations and regression residuals, *International Statistical Review*, **55**, 163–172.
- de Jong, P., Sprenger, C. and van Veen, F. (1984). On Extreme Values of Moran's I and Geary's c, *Geographical Analysis*, **16**, 17-24.
- Kalogirou, S. and Hatzichristos, T. (2007). A spatial modelling framework for income estimation, *Spatial Economic Analysis*, **2**, 297–316.
- Kamata K., Iwasawa M., and Tanaka K. (2010). *Spatial Variations in Fertility: Geographically Weighted Regression Analysis for Town-and-Village-level TFR in Japan*. Paper presented at the Annual Meeting of Population Association of America, Dallas, Texas, April 15-17 2010. Available on line from: <http://paa2010.princeton.edu/download.aspx?sumbmissionID=100780>.
- Kim, C. W., Phipps T. T. and Anselin L. (2003) Measuring the benefits of air quality improvement: a spatial hedonic approach, *Journal of Environmental Economics and Management*, **45**, 24-39
- Koenker, R. and Bassett, J. (1982). Robust tests for Heteroskedasticity based on Regression Quantiles, *Econometrica*, **50**, 43-61.
- Kohler, H-P., Billari, F.C., and Ortege, J.A. (2002). The emergence of lowest-low fertility in Europe during the 1900s, *Population and Development Review*, **28**, 641-680.
- Kohli, K. L. (1977). Regional variations of fertility in Iraq and factors affecting it, *Journal of Biosocial Science*, **9**, 175-182.
- Kulu H. (2005). Fertility and Spatial Mobility: Evidence from Austria. Rostock: Max Planck Institute for Demographic Research (MPIDR Working Paper, WP 2005-002). Available on line at www.demogr.mpg.de/papers/working/wp-2005-002.pdf
- LeSage, J. and Pace, R. (2009). *Introduction to Spatial Econometrics*, Chapman & Hall/CRC, Boca Raton.
- Lesthaeghe, R. and Van de Kaa, D. (1986). Twee demografische transitie's? In Lesthaeghe, R. and Van de Kaa, D. (eds), *Bevolking: Groei en Krimp*, 9-24, Deventer, Van Loghum Slaterus.
- Leung, Y., Mei, C. and Zhang W. (2000a). Statistical test for spatial nonstationarity based on the geographically weighted regression model, *Environment and Planning A*, **32**, 9-32.
- Leung, Y., Mei, C. and Zhang W. (2000b). Testing for spatial autocorrelation among the residuals of the geographically weighted regression, *Environment and Planning A*, **32**, 871-890.
- Marmol, F. (1995). Spurious regressions between I(d) processes, *Journal of Time series Analysis*, **16**, 313-321.
- Moran, P. A. P. (1948). The interpretation of statistical maps, *Journal of the Royal Statistical Society B*, **10**, 243-51.
- Moran, P. A. P. (1950). Notes on continuous stochastic phenomena, *Biometrika*, **37**, 17-23.

- Muniz J. (2009). Spatial dependence and heterogeneity in ten years fertility decline in Brazil, *Population Review*, **48**, 32–65.
- Mur, J. and Trivez F.J. (2003). Unit roots and deterministic trend in spatial econometrics models, *International Regional Science Review*, **26**, 289-312.
- Newbold, P. and Davies, N. (1978). Error mis-specification and spurious regressions, *International Economic Review*, **19**, 513-519.
- Ord J. K (1975). Estimation Methods for Models of Spatial Interaction, *Journal of the American Statistical Association*, **70**, 120-126.
- Ord J. K (1980). Tests of Significance Using Nonnormal Data, *Geographical Analysis*, **12**, 387-392.
- Paelinck, J. and Klaassen, L. (1979). *Spatial Econometrics*, Saxon House, Farnborough.
- Páez, A., Farber, S. and Wheeler, D. (2011). A simulation-based study of geographically weighted regression as a method for investigating spatially varying relationships, *Environment and Planning A*, **43**, 2992-3010.
- Pandit K. (1992). Regional Fertility Differentials and the Effect of Migration: An Analysis of U.S. State-level Data, *Geographical Analysis*, **24**, 352-364.
- Perez A. E. (1991). An analysis of the effects of fertility on women's spatial mobility in the Philippines, *Asia Pacific Population Journal*, **6**, 35-66.
- Phillips, P. C. B. (1986). Understanding spurious regressions in econometrics, *Journal of Econometrics*, **33**, 311-340.
- Presser H. B. and Baldwin W. (1980). Child- Care as a Constraint on Employment: Prevalence, Correlates, and Bearing on the Work and Fertility Nexus, *American Journal of Sociology*, **85**, 1202-1213.
- Rawling, J. O., Pantula, S. G., and Dickey D. A. (1998). *Applied Regression Analysis: A Research Edition*, Springer, New York.
- Rowland D. T. (2006). *Demographic methods and concepts*, Oxford University Press, Oxford.
- Schabenberger, O. and Gotway, C. A. (2005). *Statistical Methods for Spatial Data Analysis*, Chapman & Hall/CRC, Boca Raton.
- Siegel J. S. and Swanson, D. A. (2004). *The Methods and Materials of Demography*, Elsevier – Academic Press, New York.
- Snow, J. (1855). *On the mode of communication of cholera*, 2nd Edition.
- Sobotka, T. (2008). The diverse faces of the Second Demographic Transition in Europe, *Demographic Research*, **19**, 171-224.
- Symeonidou H., Douligeris V., Kappi H., Magdalinos M., Maratou-Alinbranti L., Balourdos D., Pappas P., Samartzi M. (1997). *Socio-economic factors affecting fertility in Greece and comparative analysis between areas*, National Centre of Social Research, Volume B. In Greek, Summary in English, Athens: EKKE.
- Tiefelsdorf, M. and Boots B. (1995). The Exact Distribution of Moran's I, *Environment and Planning A*, **27**, 985-999.

- Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region, *Economic Geography*, **46**, 234-240.
- Tsimbos, C. (2006). The Impact of Migration on Growth and Ageing of the Population in a New Receiving Country: the Case of Greece, *International Migration*, **44**, 232-254.
- Tsimbos, C. (2008). Immigrant and Native Fertility in Greece: New Estimates and Population Prospects (2005-2025), *Population Review*, **47**, 67-84.
- Tsimbos C., Kalogirou S., and Verropoulou G. (2011). *Life expectancy variation across local authorities in Greece in 2001 and associations with socioeconomic conditions and population density of localities*. Paper presented at the 2011 Annual Conference of the British Society for Population Studies, University of York, York, 7–9 September 2011. Abstract available on line from: www2.lse.ac.uk/socialPolicy/BSPS/annualConference/2011/Abstracts.aspx#generated-subheading7.
- UNDP (1997). *Human Development Report 1997*, Oxford University Press, Oxford.
- UNDP (2006). *Human Development Report 2006, Power, Poverty and the Global Water Crisis*, Palgrave Macmillan, New York.
- Weeks J. R. (2010). Spatial Patterns of Fertility in Rural Egypt. In: L. Anselin and S. J. Rey (eds.). *Advances in Spatial Science: Perspectives on Spatial Analysis*, 235-256, Springer, Berlin.
- Wheeler, D. (2007). Diagnostic tools and a remedial method for collinearity in geographically weighted regression, *Environment and Planning A*, **39**, 2464-2481.
- Wheeler, D. (2009). Simultaneous coefficient penalization and model selection in geographically weighted regression: the geographically weighted lasso, *Environment and Planning A*, **41**, 722-742.
- Wheeler, D. and Calder, C. (2007). An assessment of coefficient accuracy in linear regression models with spatially varying coefficients, *Journal of Geographical Systems*, **9**, 145-166.
- Wheeler, D. and Páez, A. (2010). Geographically Weighted Regression, in Fischer M.M. and Getis A. (eds.), *Handbook of Applied Spatial Analysis: Software Tools, Methods and applications*, pp. 461-486, Springer-Verlag Berlin Heidelberg 2010.
- Wheeler, D. and Tiefelsdorf, M. (2005). Multicollinearity and Correlation Among Local Regression Coefficients in Geographically Weighted Regression, *Journal of Geographical Systems*, **7**, 161-87.
- White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, *Econometrica*, **48**, 817–838.
- White M.J., Moreno L., and Guo S. (1995). The Interrelation of Fertility and Geographic Mobility in Peru: A Hazards Model Analysis, *International Migration Review*, **29**, 492-515.
- Whittle, P. (1954). On stationary processes in the plane, *Biometrika*, **41**, 434-449.
- Yule, G. U. (1926). Why do we sometimes get nonsense-correlations between time-series? A study in sampling and the nature of time-series, *Journal of the Royal Statistical Society*, **89**, 1-64.

