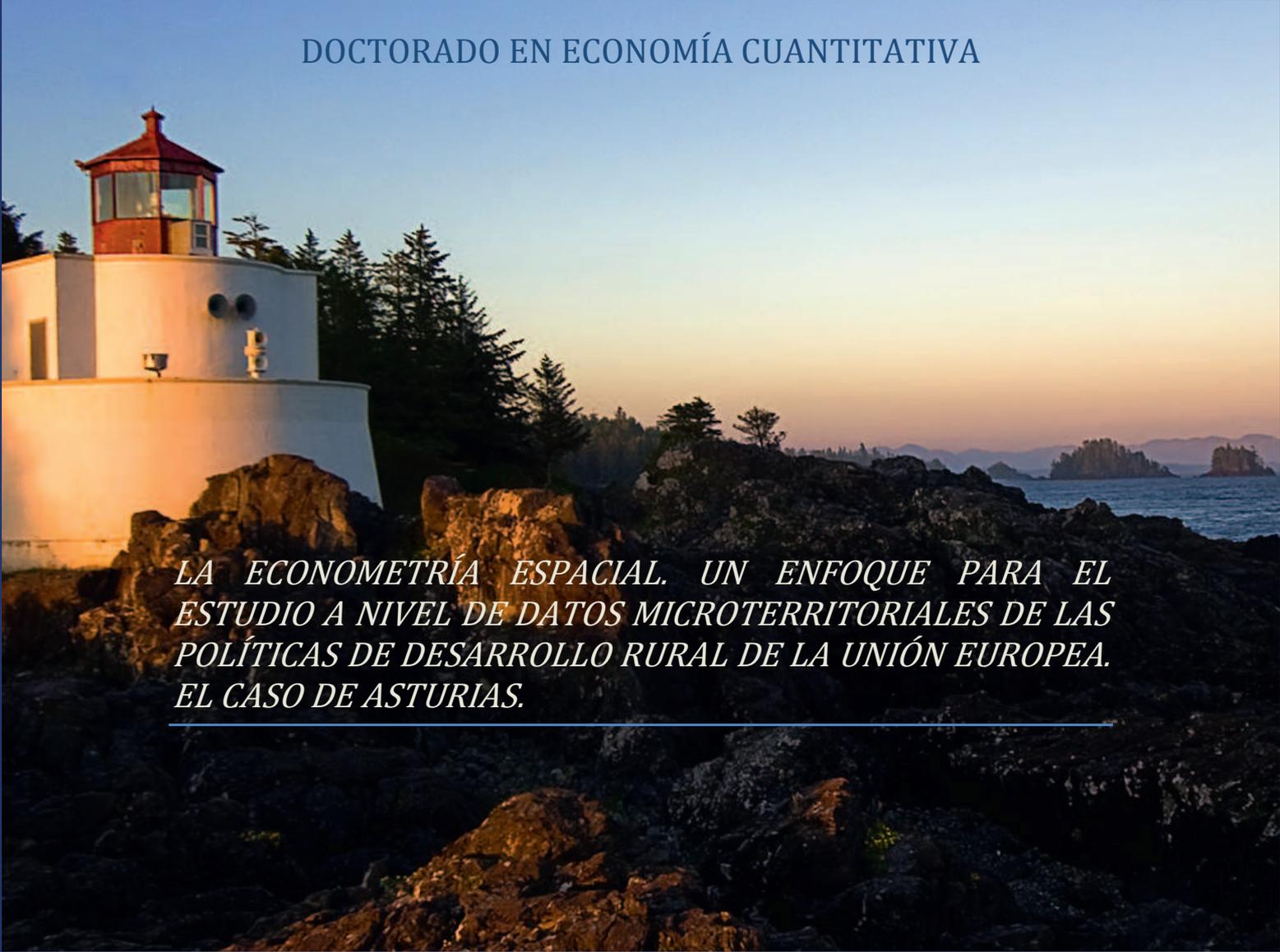




UNIVERSIDAD DE MÁLAGA

DOCTORADO EN ECONOMÍA CUANTITATIVA



*LA ECONOMETRÍA ESPACIAL. UN ENFOQUE PARA EL ESTUDIO A NIVEL DE DATOS MICROTERRITORIALES DE LAS POLÍTICAS DE DESARROLLO RURAL DE LA UNIÓN EUROPEA. EL CASO DE ASTURIAS.*

---

*TESIS DOCTORAL*

ÓSCAR LUIS ALONSO CIENFUEGOS



Publicaciones y  
Divulgación Científica

AUTOR: Óscar Luis Alonso Cienfuegos

 <http://orcid.org/0000-0002-9335-7759>

EDITA: Publicaciones y Divulgación Científica. Universidad de Málaga



Esta obra está bajo una licencia de Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional:

<http://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>

Cualquier parte de esta obra se puede reproducir sin autorización pero con el reconocimiento y atribución de los autores.

No se puede hacer uso comercial de la obra y no se puede alterar, transformar o hacer obras derivadas.

Esta Tesis Doctoral está depositada en el Repositorio Institucional de la Universidad de Málaga (RIUMA): [riuma.uma.es](http://riuma.uma.es)

---

**UNIVERSIDAD DE MÁLAGA**  
**FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES**  
**DEPARTAMENTO DE TEORÍA E HISTORIA ECONÓMICA**



**TESIS DOCTORAL**

**LA ECONOMETRÍA ESPACIAL. UN ENFOQUE PARA EL ESTUDIO A NIVEL DE DATOS  
MICROTERRITORIALES DE LA EFICACIA DE LAS POLÍTICAS DE DESARROLLO  
RURAL COFINANCIADAS POR LA UNIÓN EUROPEA. EL CASO DE ASTURIAS.**

**Presentada por**

**Óscar Luis Alonso Cienfuegos**

**Dirigida por el doctor**

**José Luis Torres Chacón**

**Octubre de 2015**



## DEDICATORIA

A mi familia, especialmente a Anabel, mi otra mitad, a mis hijos, a mis padres, que además de ser los mejores padres han sido los mejores profesores, y a mis abuelos, que aunque ya no están aquí siempre me acompañan.

## AGRADECIMIENTOS

Todo mi agradecimiento al director de esta tesis, José Luis Torres Chacón, por su desinteresada ayuda, sus valiosísimas aportaciones, su enorme paciencia, y sobre todo por su comprensión y capacidad de adaptación.

Mis más sinceras gracias a Marcelino Rodríguez Pérez, Jefe del Servicio Presupuestario de la Secretaría General Técnica de la Consejería de Agroganadería y Recursos Autóctonos, por su eficacia e interés para reunir los datos municipales de ejecución de fondos públicos de la PAC, fundamentales para este estudio, que tan amablemente me ha proporcionado. Se ha notado el buen hacer, la eficiencia y la desinteresada capacidad de servicio público propia de los buenos agentes del antiguo Servicio de Extensión Agraria.



*“Veréis llanuras bélicas y páramos de asceta  
-no fue por esos campos el bíblico jardín-;  
son tierras para el águila, un trozo de planeta  
por donde cruza errante la sombra de Caín”*

del poema “Por Tierras de España”, Antonio Machado.



---

**ÍNDICE**


---

<b>CAPÍTULO 1. INTRODUCCIÓN.</b>	<b>11</b>
1.1. Motivación.	11
1.2. Contenidos de la tesis.	12
1.3. La Política Agraria Común (PAC).	15
<b>CAPÍTULO 2. ENFOQUE Y TÉCNICAS DE ECONOMETRÍA ESPACIAL.</b>	<b>23</b>
2.1. Enfoque espacial. Repaso histórico.	23
2.2. “Problema” espacial.	31
2.3. Análisis exploratorio de datos espaciales.	37
2.4. Análisis confirmatorio de datos espaciales.	45
➤ 2.4.1. Modelos.	45
➤ 2.4.2. Estimación y contrastes.	63
2.5. Programas y paquetes informáticos.	79
<b>CAPÍTULO 3. ENFOQUE METODOLÓGICO, VARIABLES Y FUENTES DE INFORMACIÓN ESTADÍSTICA.</b>	<b>93</b>
3.1. Posibles enfoques metodológicos y revisión de la literatura reciente.	93
3.2. El problema del nivel de agregación de los datos espaciales.	98
3.3. Demarcación territorial de la Unión Europea con fines estadísticos.	100
3.4. Variables consideradas y fuentes de información estadística.	102
➤ 3.4.1. Fuentes de información estadística.	102
➤ 3.4.2. Variables consideradas.	105

---

<b>CAPÍTULO 4. ESTIMACIÓN Y MODELO ECONÓMÉRICO.</b>	<b>115</b>
<b>4.1. Análisis exploratorio de datos espaciales (AEDE).</b>	<b>115</b>
➤ 4.1.1. Análisis univariante.	115
➤ 4.1.2. Resultados del AEDE.	137
<b>4.2. Análisis confirmatorio.</b>	<b>140</b>
➤ 4.2.1. Autocorrelación espacial.	142
➤ 4.2.2. Heterogeneidad espacial.	159
➤ 4.2.3. Resultados del análisis confirmatorio.	180
<b>CAPÍTULO 5. CONCLUSIONES Y DISCUSIÓN.</b>	<b>185</b>
<b>5.1. Discusión.</b>	<b>185</b>
<b>5.2. Conclusiones.</b>	<b>193</b>
<b>BIBLIOGRAFÍA.</b>	<b>203</b>
<b>ANEXOS.</b>	<b>215</b>

---

## CAPÍTULO 1. INTRODUCCIÓN

---

### 1.1. MOTIVACIÓN.

Mi experiencia profesional como economista, de aproximadamente doce años en lo referente programas de desarrollo rural basados en el enfoque LEADER, me ha dado una visión del desarrollo local en territorios rurales un tanto crítica (incluyendo por supuesto también la autocrítica). Los distintos periodos de programación en los que he participado, intentando cooperar desde mi puesto de trabajo en el desarrollo económico de mi comarca, me han servido para acumular experiencias y contrastar la evolución de lo que en principio era una iniciativa comunitaria con una importante vocación de innovación, con carácter piloto y fundamentalmente basada en proyectos demostrativos, hasta el actual enfoque LEADER, que es una metodología que se integra perfectamente en las estrategias de desarrollo local participativo multifondo, así como en los planes de desarrollo rural de los estados miembros (en nuestro caso de los gobiernos de las distintas comunidades autónomas). El enfoque ascendente, participativo, multisectorial y territorial, que aprovecha fondos sectoriales para el desarrollo integral de los territorios, basándose en la diversificación, da lugar en ocasiones a ciertas contradicciones, tanto en el diseño de las estrategias como en su ejecución. Este enfoque ascendente y participativo no siempre es real en la práctica y lo que sí es evidente es, que al integrarse dentro de los planes de desarrollo rural, se ve totalmente condicionado por las decisiones estratégicas y de política económica que toman los gobiernos en lo referente a los fondos europeos de desarrollo rural que se les concede. Me parece que, por tanto, un punto de partida para entender su evolución es medir el posible impacto o repercusión que puedan tener, no solo los fondos destinados a estos programas, sino todos aquellos fondos de la misma naturaleza que se destinan al mismo contexto rural en el que hasta ahora se han enmarcado. En el periodo de programación 2007-2013 se puso en marcha el eje LEADER, cuarto eje del FEADER. Es un eje transversal que puede servir para ejecutar los tres restantes o parte de los mismos, dependiendo de las decisiones políticas y estratégicas del gobierno correspondiente. Hasta entonces LEADER había sido una iniciativa comunitaria y aunque el enfoque era el mismo en lo básico, su dependencia de la administración de la comunidad autónoma era menor. Me parece por tanto procedente analizar como un primer paso, el contexto y los resultados

de los fondos cofinanciados por el FEADER en mi comunidad autónoma para poder conocer el alcance de las acciones y planes ejecutados. Esto me servirá para complementar el conocimiento obtenido desde mi puesto de trabajo con el que resulte de este estudio. Pretendo poder así contrastar de forma empírica aquello que aunque a veces pueda parecer cierto, en realidad, quizá solo lo sea a medias.

## 1.2. CONTENIDOS DE LA TESIS.

Los contenidos del presente trabajo se centran en contrastar la posible contribución de los fondos cofinanciados por FEADER a las variables que miden dos de los principales objetivos de las políticas de desarrollo rural, que son, la fijación de población en el medio rural, para frenar el éxodo que se produce a zonas urbanas desde los años 60, y la creación de empleo. Estas dos variables están muy relacionadas, ya que la actividad económica y los puestos de trabajo son determinantes para que los ciudadanos puedan desarrollar un proyecto vital que fije población en estas zonas y evite su actual declive. En un primer paso se explica la **evolución de la PAC**, contexto necesario en el que se enmarca el periodo objeto de estudio, que será el periodo de programación 2007-2013, el último ejecutado hasta la fecha. Su desarrollo ha marcado las políticas de desarrollo rural y conviene describir la diferenciación entre los actuales dos pilares, que concretan los dos enfoques fundamentales de desarrollo rural, uno el enfoque sectorial de ayuda directa a la actividad, y otro el enfoque territorial de ayuda directa a las zonas con una base rural y un enfoque multisectorial, no solo del sector primario.

El estudio económico de datos distribuidos geográficamente necesariamente debe contar con técnicas de econometría espacial que consideren los posibles efectos de autocorrelación y heterogeneidad espacial, que de ser obviados pueden hacer que los resultados del análisis econométrico sean sesgados, al perder las propiedades estadísticas requeridas para ser aceptados como válidos. Se repasa por tanto, en un segundo paso **la historia y metodología de la econometría espacial y se contextualiza dentro de la ciencia regional** mencionando también otros posibles enfoques teóricos dentro de la misma. La econometría espacial y cualquier estudio científico dependen de **las herramientas y software informático** existente para poder desarrollar y cuantificar las propuestas teóricas. Se repasan los programas, bibliotecas y paquetes informáticos

específicos de econometría espacial disponibles y los utilizados para este estudio. En nuestro caso tienen especial relevancia los desarrollados en código abierto por GeoDa Center, liderado por Luc Anselin, que utilizan el aporte de la comunidad científica para poner a disposición de los usuarios bibliotecas como PySal, con paquetes como GeoDa Space o GeoDa que implementan los avances recientes en econometría espacial, además de las funcionalidades que ya tenían, fruto de la larga trayectoria desarrollada en este ámbito desde el pionero SpaceStat.

La posible **metodología** a aplicar a este tipo de estudios requiere un repaso de la literatura y la elección del método más apropiado al caso que nos ocupa, teniendo en cuenta la información disponible, las peculiaridades de la región y las políticas impulsadas en el último periodo de programación ejecutado. Por esto se desarrolla, en un tercer paso, un repaso a las distintas metodologías aplicadas hasta la fecha por otros autores así como de la desarrollada para este estudio, que se basa en la realización de una colección de modelos que permitan aplicar herramientas de econometría espacial para estudiar la significatividad y sentido económico de las inversiones o gasto público realizado por la administración regional con fondos de la PAC.

En el cuarto paso se analizan y describen las **variables y fuentes de información estadística**, que condicionan totalmente este tipo de estudios. Previamente se explica el problema del nivel de agregación de los datos espaciales y de cómo puede condicionar la validez de este estudio, que necesariamente se ha de hacer con datos microterritoriales, tanto por el objeto de estudio como por las peculiaridades administrativas y geográficas del Principado de Asturias, con concejos en algunos casos excesivamente pequeños y con una orografía de montaña que condiciona su actividad, especialmente la del sector primario. Se explica la demarcación estadística a nivel territorial de la UE, ya que también es condicionante estadístico y económico del estudio planteado. La justificación de las variables seleccionadas se basa en los indicadores comunes de contexto que se plantean en el reglamento FEADER para analizar variables de tipo socioeconómico y contextualizar los territorios rurales, así como para medir su evolución y la repercusión de las políticas de desarrollo rural. Por eso se describen y explican junto con las variables seleccionadas. Se trata de un estudio de datos de corte transversal, donde la unidad espacial es la de concejo y donde se tiene

en cuenta las distintas unidades económicas supramunicipales que puedan existir para plantear soluciones a posibles problemas de heterogeneidad espacial

En un quinto paso se plantea el **estudio econométrico** que se divide en dos partes, por un lado el análisis exploratorio de los datos espaciales, en el que se valora la posible existencia de autocorrelación y heterogeneidad espaciales, y por otro el análisis confirmatorio donde ya se estiman los modelos en función de los resultados obtenidos en el análisis espacial. Se plantearán cuatro modelos básicos que darán lugar a variaciones de los mismos en función de las variables explicativas utilizadas y de la muestra, con o sin valores extremos (que además se corresponden con zonas más “urbanas”). También surgen otros modelos al aplicar métodos para tratar la heterogeneidad espacial, en nuestro caso corregida aplicando modelos de regresiones cambiantes al detectarse una posible inestabilidad paramétrica discreta (*spatial regimes*). Se proponen cuatro variables endógenas, que son las que definen cada modelo, serán la variación de la población en el periodo 2007-2013, la variación del empleo en el mismo periodo y la variación del empleo del sector agrario para ese mismo periodo también. El cuarto modelo es el que tiene como variable dependiente el número de licencias de IAE agraria en el año 2013.

Las variaciones se considera que recogen la dinámica del periodo de programación más que los datos de final de programa y como muchas de las ayudas son anuales y cíclicas nos pareció lo más correcto. Para el empleo nos pareció necesario contrastar tanto el dato para todos los sectores, es decir el total, como el agrario, ya que son fondos agrarios los que sustentan estas políticas y nos parece interesante contrastar si se incide en el empleo agrario o si se diversifica a otros sectores. Para las licencias se usa el dato anual de final de periodo para contrastar el ritmo de creación de nuevas explotaciones ya que es un dato no acumulativo.

La inversión pública se divide en IPDR13.07, que consiste en la inversión del segundo pilar de la PAC, fondos cofinanciados por FEADER, y en la que denominamos IIPIL13.07. Como se dispone de datos de la inversión en el primer pilar de la PAC, es decir los financiados por FEAGA, también se va a modelizar incluyendo esta variable como explicativa, en lugar de IPDR13.07, para ver su comportamiento y poder comparar ambos pilares. Por último se suman ambas variables para estudiar el comportamiento de toda la inversión PAC en Asturias.

Las variaciones propuestas de tres variables distintas de inversión para cada variable dependiente, la prueba con y sin valores extremos y la modelización de la heterogeneidad espacial por el método de regímenes espaciales hace que al final sean 48 modelos o escenarios los que se plantearán en esta tesis doctoral.

La idea fundamental es que si la inversión pública financiada o cofinanciada por fondos de la PAC (incluyendo la parte nacional y regional) ha contribuido a mejorar los datos de población, de empleo total, de empleo agrario y de licencias de actividad agraria, en Asturias, se reflejará en los modelos planteados con coeficientes significativos y positivos. Si estos no son significativos no habrán tenido relevancia, y si son significativos y negativos habrán servido para lo contrario a lo que perseguían. El resto de variables se seleccionan por criterios tanto de teoría económica como por pertenecer a algunos de los grupos de indicadores comunes de contexto propuestos por la UE para medir el desarrollo rural, así como por criterios estadísticos y econométricos.

El origen necesario para entender y explicar el desarrollo rural en Asturias en los últimos años pasa por la Política Agraria Común (PAC).

### 1.3. LA POLÍTICA AGRARIA COMÚN (PAC).

El origen de la PAC se sitúa en la firma del Tratado de Roma en el año 1957, donde Alemania Occidental, Francia, Países Bajos, Bélgica, Italia y Luxemburgo buscan crear un marco común para la actividad agrícola y la revitalización de sus economías. *“En 1958, y siguiendo lo establecido en el artículo 43 del Tratado de Roma, se iniciaron, en la Conferencia de Stressa, los trabajos preparatorios para el establecimiento de la PAC. Allí se concluyó que la agricultura era una parte esencial de la economía y de la sociedad europea y que la PAC debía fomentar la mejora de las estructuras agrarias, el crecimiento de la productividad, contribuir al desarrollo del comercio interno, al equilibrio entre la oferta y la demanda y proteger el modelo de agricultura familiar europeo garantizando unas remuneraciones equitativas a los factores de producción.”* (García y García, 2005).

Se debe tener en cuenta el marco temporal en el surge, con una Europa occidental marcada por la postguerra y con una mínima producción agrícola. Los objetivos iniciales eran el abastecimiento de los mercados y el desarrollo del tejido productivo agrario europeo. Se perseguía un aumento de la producción y productividad agrarias y una protección al agricultor frente a las oscilaciones del mercado. Para lograr sus objetivos se fija un sistema de subvenciones que garantizan precios elevados a los agricultores e incentivan la producción. Los pilares en los que se asienta son:

- Unidad de mercado por la que se garantiza la libre circulación.
- Preferencia comunitaria frente a productos de terceros países.
- Solidaridad financiera por la que existe igualdad de acceso de todos los agricultores europeos a los recursos comunitarios.

En 1962 se crea el FEOGA (Fondo Europeo de Orientación y Garantía Agraria) y en 1963 se divide en dos secciones la de “Orientación” (FEOGA-O) y la de “Garantía” (FEOGA-G). La Sección de Garantía se destinaba a la política de precios y mercados, es decir es la que se ocupa de las restituciones a la exportación y de sufragar las intervenciones de las regulaciones de los mercados. La Sección de Orientación es la encargada de financiar la política estructural y el desarrollo de territorios rurales. Al contrario de la sección de garantía que se financia exclusivamente con fondos comunes, la sección de orientación se basa en el principio de cofinanciación y se ponen por una parte fondos comunes y por otra parte fondos de los estados miembros.

*“La PAC se asienta sobre dos pilares muy desiguales: la política de regulación de precios y mercados, que ha sido la más importante en cuanto a recursos absorbidos y actuaciones realizadas, y la política de estructuras (que en el futuro se quiere que sea una política de desarrollo rural), que ha tenido escasa entidad, primero, porque los recursos a ella asignados han sido insuficientes y, segundo, porque ha exigido la cofinanciación de los Estados miembros, lo que ha provocado que se mantenga en un segundo plano.”(García y García, 2005)*

Al principio se lograron sus objetivos pero ya en los años 70 y 80 se empiezan a plantear problemas. En 1968 se publicó el “Plan Mansholt”, que señalaba las limitaciones que tienen en el tiempo estas políticas de precios y mercados. Hasta 1970

era la única política europea y absorbía la mayor parte de su presupuesto. En los años siguientes el presupuesto aumentó, aumentaron los excedentes y se pusieron de manifiesto algunas de las limitaciones y fallos previstos, lo que supuso la adopción constante de nuevas medidas y reformas para su desarrollo. Algunos ejemplos serían los programas de retirada de tierras cultivables o el establecimiento de cuotas de producción.

En 1992 se produjo reforma MacSharry, para pasar a lo que se puede considerar una nueva etapa en la forma de gestionar la política agraria. Se perseguía incidir en el equilibrio entre la producción y el medio ambiente y atribuir al agricultor la doble función de productor de alimentos y protector del medio ambiente. Se produce un giro importante hacia una sensibilidad más ecologista y concienciada en el mantenimiento del medio.

*“Hasta esa fecha, la política agraria había sido, como ya se ha visto, básicamente una política de precios y mercados, cuyos efectos indeseados trataron de corregirse, desde finales de la década de 1970, con diversas medidas que no cuestionaban, al menos abiertamente, sus fundamentos. Desde 1992 esta política va a ser sustituida progresivamente por otra de ayudas directas a las rentas en la que se otorga un creciente protagonismo a la política de desarrollo rural, más amplia y completa en sus pretensiones que su predecesora la política de estructuras, cuyos resultados, hasta la fecha, habían sido muy limitados.” (García Grande, 2005).*

Algunos de los aspectos relevantes de esta reforma serían un mayor control del gasto comunitario, una distribución equitativa del presupuesto entre todos los estados miembros, una producción de calidad respetuosa con el medio ambiente y, por último, el desarrollo rural, que, entre otras cosas, persigue diversificar las actividades del medio rural más allá de la agricultura. Se sustituye el sistema de protección al agricultor a través del control de los precios por el sistema de ayuda directa a las rentas.

La consecución de algunos de los objetivos propuestos con la reforma de 1992 llevo a profundizar más en algunos de sus aspectos con la reforma de 1999, que se enmarcó en la Agenda 2000. Se fijan los dos pilares, el primero de política de rentas y mercados, apoyando a la renta de los agricultores mediante la consolidación de los pagos directos que se introdujeron en la reforma MacSharry y el segundo de políticas ambientales y de

desarrollo rural. El medio rural se entiende como un territorio que genera actividad, modo de vida y cultura más allá de la agricultura y la ganadería. Se aplica el concepto territorial frente al sectorial del primer pilar y se establecen programas específicos para cada zona, distinguiendo entre zona rural y no rural. Aparece la metodología o enfoque LEADER. *“Aunque la reforma de 1999 aporta menos novedades que su predecesora, es interesante llamar la atención sobre dos aspectos: la apuesta cada vez más decidida de la Unión Europea por una política de rentas que sustituya la tradicional política de precios y el fortalecimiento de lo que se ha dado en llamar segundo pilar de la PAC, es decir, el desarrollo rural.”* (García Grande, 2005).

En 2003 se produce la mayor reforma de la PAC desde su creación, es la llamada Reforma Fischler. *“De todas las reformas que la PAC ha experimentado hasta la fecha, la de 2003 ha sido la más profunda sin lugar a dudas, la que marca el principio del fin de la PAC tal como se conoce en la actualidad. A partir de ahora, el protagonismo absoluto lo adquiere la política de ayudas directas a las rentas, desvinculadas de la producción, un paso previo, parece que, a la instauración de una única política de desarrollo rural.”* (García Grande, 2005).

Se endurecen las medidas de la agenda 2000. Se da prioridad al consumidor y por tanto se tiene en cuenta lo que demanda el mercado. Salvo algunas excepciones se produce el desacoplamiento entre las ayudas y la producción pasando a una ayuda única y directa por explotación para el mantenimiento de la renta del agricultor. Se plantean nuevos parámetros para la disminución paulatina de las ayudas (modulación), y se someten a eco-condicionalidad, es decir a cumplir una serie de normas ambientales por parte del agricultor para cobrarlas. Se fija una disciplina financiera que garantice no sobrepasar un límite fijado para 2013. Se le da más relevancia y peso al desarrollo rural.

*“La actual política de desarrollo rural de la UE no surgió de la nada, sino que se ha ido forjando paulatinamente, por la cansina introducción de instrumentos de política agraria que, frente al predominio del componente sectorial, han intentado poner mayor énfasis en la vertiente territorial.”* (Colino y Martínez, 2005)

En 2009 se produce el llamado Chequeo Médico que fija parámetros para ajustar la reforma de 2003. Se mantiene y consolida el pago único y se apuesta por su simplificación administrativa. La modulación continúa aumentando sus porcentajes y se

avanza en el desacoplamiento total de las ayudas en sectores a los que no se había aplicado.

La tendencia hacia una política agraria basada en el desarrollo rural, se perfila como el camino a recorrer por la PAC. Ya en 2005 decía García Grande (2005): *“La PAC del futuro será básicamente una política de desarrollo rural que otorgará una protección selectiva al sector”*, y parece que esa tendencia se consolida.

Para el próximo periodo de programación 2014-2020 se mantiene la estructura actual de la PAC con 2 pilares, y se plantea una política de importancia estratégica para la seguridad alimentaria, el equilibrio territorial y el medio ambiente. En el primer pilar se sustituye el pago único por los pagos directos, sistema de pago basado exclusivamente en la hectárea al que se le suman otros por buenas prácticas medioambientales, jóvenes agricultores, pequeños agricultores y limitaciones naturales. Supone un pago por superficie sin tener en cuenta las referencias históricas del pago único. Las medidas de competitividad y mercado persiguen estabilizar los mercados y garantizar un nivel de vida equitativo a la población agrícola, así como garantizar la igualdad de acceso a las mercancías y de trato de los compradores. Desaparecen las cuotas y los derechos de producción.

En cuanto al segundo pilar la política de desarrollo rural complementa los pagos directos y las medidas de mercado y contribuye de este modo a conseguir los objetivos de la PAC. La herramienta financiera para conseguirlo es el FEADER. Sus objetivos son:

1. Fomentar la competitividad de la agricultura.
2. Garantizar la gestión sostenible de los recursos naturales y la acción por el clima.
3. Lograr un desarrollo territorial equilibrado de las economías y comunidades rurales incluyendo la creación y la conservación del empleo.

Estos objetivos se incluyen en las 6 prioridades de desarrollo rural fijadas por la Unión Europea:

1. Promover la transferencia de conocimientos y la innovación.
2. Fomentar la competitividad.

3. Mejorar la organización de la cadena de distribución de alimentos y la gestión de riesgos.
4. Restaurar, preservar y mejorar los ecosistemas.
5. Promover la eficiencia de los recursos y alentar el paso a una economía hipocarbónica.
6. Fomentar la inclusión social, la reducción de la pobreza y el desarrollo económico en las zonas rurales.

La gestión del FEADER deberá coordinarse con la de los fondos estructurales y el fondo de cohesión.

El enfoque del segundo pilar parece que no solo se consolida, sino que adquiere cada vez mayor relevancia y peso dentro de la PAC. A pesar de haber sido hasta ahora el pilar menos favorecido en cuanto a inversiones, estas han ido aumentando paulatinamente. Por esto es por lo que parece conveniente estudiar los posibles logros que hasta la fecha han tenido estas políticas de desarrollo rural en aquellos territorios rurales en los que se han implementado. Nos parece conveniente estudiar los resultados obtenidos con la aplicación de la política de desarrollo rural, haciendo un estudio sobre la posible relación de las variables sobre las que intenta incidir, siempre que además dicha relación tenga sentido desde el punto de vista de la teoría económica. El enfoque de estas políticas es territorial y dado que estas se aplican, en el caso del estado español, a nivel de comunidades autónomas, entendemos que se requiere una metodología que analice datos espaciales y territoriales contemplando así los posibles problemas que puedan surgir del manejo de este tipo de datos. Este es el enfoque de la econometría espacial. Los informes realizados para el seguimiento y cuantificación de los distintos planes de desarrollo rural (PDR) se basan en sistemas de indicadores prefijados, algunos comunes y otros específicos, que se enmarcan en informes que los presentan de forma descriptiva. Se suelen comparar con valores también prefijados de los mismos, para evaluar su grado de ejecución. Se interpretan como exitosos aquellos programas que consiguen alto grado de ejecución. También tienen esa consideración si tienen indicadores de volumen de inversión, empleos asociados, número de empresas, etc elevados. El enfoque de este estudio pretende ser distinto al descrito, ya que persigue cuantificar la posible relación o no entre las acciones realizadas con enfoque territorial y estructural cofinanciadas con fondos de desarrollo rural de la UE y las variables

económicas de los territorios en los que se aplican, para contrastar si existe o no relación entre las mismas, y en qué medida, si es el caso, para poder así cuantificar sus posibles logros.

### **La econometría espacial.**

En un territorio como la Comunidad Autónoma del Principado de Asturias, de carácter uniprovincial, con un territorio relativamente pequeño, al menos en comparación con otras comunidades autónomas de España, con una orografía difícil, predominantemente montañosa, que presenta territorios de distintas características, históricamente poco relacionados entre sí, entre otras cosas debido a esas dificultades orográficas, que además presenta una distribución administrativa atomizada en municipios o concejos muy pequeños, al menos en algunos casos, parece razonable pensar que las relaciones económicas que se produzcan en ellos tengan un alto grado de interconexión, a la vez que la suficiente entidad para ser consideradas individualmente para ser objeto de estudio. No se consigue el mismo grado de desarrollo económico en el área central de Asturias, más urbanizada, que en las alas, más rurales. No es el mismo desarrollo el que experimentan las zonas rurales de costa que las zonas rurales de interior, e incluso se dan importantes diferencias entre el ala oriental y el ala occidental.

Para estudiar la evolución de dichas variables nos parece relevante el enfoque de la econometría espacial, que tiene en cuenta la posible existencia de autocorrelación y heterogeneidad espaciales, y en caso de que se den estos fenómenos aplicar la metodología necesaria para la estimación de modelos econométricos válidos, posibilitando así el manejo de datos a nivel de desagregación microterritorial.



## CAPÍTULO 2. ENFOQUE Y TÉCNICAS DE ECONOMETRÍA ESPACIAL.

---

### 2.1. ENFOQUE ESPACIAL. REPASO HISTÓRICO.

Podemos considerar la econometría espacial como una especialización de la econometría general que se ocupa del tratamiento de los datos distribuidos en el espacio. En palabras de Chasco (2003) *“incluye el conjunto de técnicas de especificación, estimación, contraste y predicción necesarias para el tratamiento de los datos espaciales”*.

El nombre concreto de “econometría espacial” se utilizó por primera vez para denominar a estas técnicas por Paelinck y Klaasen (1979) en un texto que titulan “Spatial Econometrics”, aunque los problemas causados por la dependencia y heterogeneidad espacial, en las técnicas de análisis econométrico tradicional, se detectan por otros autores en un origen temporal anterior. Ya Student (1914) los menciona en su obra, según Chasco (2003) y Moreno y Vayá (2000).

Es en la década de los cincuenta, cuando aparecen los primeros índices que detectan la presencia de autocorrelación espacial, concretamente en los trabajos de Moran (1948) y Geary (1954), y quizás estos se puedan considerar los primeros avances que sientan las bases de una subdisciplina independiente dentro de la econometría general. Es también destacable, en esa década, la aportación de Isard, W. con obras como Location and Space Economy (1956) o Methods of Regional Analysis (1960), que se ocupan del papel del factor espacio en la concreción de procesos económicos, y del efecto espacial en el tratamiento de datos estadísticos, entre otras cosas.

Podemos considerar entonces, que hasta los años setenta los avances producidos se basan fundamentalmente en la detección del problema espacial y a partir de esas fechas, en los años ochenta y noventa, se produce el desarrollo importante de las técnicas de econometría espacial. Se inicia con los trabajos de Cliff y Ord (1981). Posteriormente, merecen especial mención, por su contribución importante al desarrollo de la metodología de análisis espacial, los trabajos de Anselin (1980, 1988) y Arbia (1989).

Es considerado manual de referencia el libro “Spatial Econometrics: Methods and Models” de Luc Anselin (1988).

Durante los años noventa empiezan a aparecer numerosas publicaciones en revistas entre las que destacan Anselin y Florax (1995, 2000) y Ficher y Getis (1997). A partir del año 2000 es destacable el trabajo de Anselin, Florax y Rey (2004) “Advances in Spatial Econometrics”, Arbia (2006) “Spatial econometrics” y Le Sage y Pace (2009) “Introduction to Spatial Econometrics” como manuales de referencia donde se recogen las distintas técnicas y modelos de econometría espacial.

Con carácter general, se puede decir que las técnicas de estadística espacial se centran en el estudio de los datos espaciales, mientras que las de econometría espacial se ocupan de las peculiaridades que causa el espacio en el análisis estadístico de los modelos de la ciencia regional (Anselin, 1988).

Los datos espaciales presentan unas características particulares que deben ser consideradas para su tratamiento y utilización en la modelización de fenómenos económicos, ya que en múltiples ocasiones la utilización de técnicas econométricas “tradicionales”, o que solo tienen en cuenta el carácter temporal de las variables, invalidan los resultados al no cumplir los supuestos fundamentales exigidos por estas técnicas. Estas “peculiaridades” que causa el espacio en el análisis estadístico de la Ciencia Regional es lo que a juicio de algunos autores no ha sido tenido en cuenta lo suficientemente en el desarrollo de la econometría en los últimos años, a pesar de su importante evolución, y sobre todo si lo comparamos con el desarrollo que ha tenido la dimensión “temporal” frente a la “espacial”.

Podemos considerar que el desarrollo de la ciencia regional es el contexto en el que se desarrolla la econometría espacial, siendo ésta, a la ciencia económica regional, lo mismo que la “econometría general” a la teoría económica. Según Pérez Pineda (2005) *“el hecho de que la economía convencional no pudiera explicar problemas derivados de la variable espacio, promovería la evolución de la Geografía Económica y la Economía Regional, que incluían el estudio del espacio en el análisis económico, generando respuestas y predicciones para la economía. En concreto, la evolución de la Economía Regional crearía un intento de Teoría Economía Espacial en los sesenta con la obra de W. Isard, con obras como Location and Space Economy (1956) o Methods of*

*Regional Analysis (1960), que explicaría el papel del factor espacio en la determinación de procesos económicos”.*

*“El siguiente cuadro, resume la tradición intelectual preocupada por la inclusión del espacio como variable del análisis económico”*

**Cuadro 2.1. Evolución de la teoría económica espacial.**

PERIODO	AUTORES
1800-1929	Von Thünen (1826), Weber (1909), Hotelling (1929)
1930-1950	Christaller (1933), Lösch (1940), Florence (1948), Hoover (1948), Perroux (1950)
1951-1970	Harris (1954), Isard (1956), Meyer (1966), Alonso (1967), Lowry (1964), Pred (1966)
1971-1990	Teitz (1970, 2001), Richardson (1973), Henderson (1974), Balassa (1980), Anselin (1988,1992)
1991	Glaeser et al (1992), Krugman (1991, 1995, 2000), Porter (1990, 1996), Storper (1997), Gallup, Sachs, Mellinger (1998), Venables (1999, 2000), Fujita (2000), Ludema (2002)

Fuente: Pérez Pineda (2005)

A su juicio la Geografía Económica y la Economía Regional se han consolidado y mejorado hasta dar lugar a la denominada Nueva Geografía Económica que se ocupa del estudio de la aglomeración en el espacio de la actividad económica y de la convergencia regional.

Según Krugman (1995) la geografía económica se ocupa del “*estudio de la localización de la actividad económica*”, lo que hace considerar, con un papel preponderante, la variable espacio, aunque también tiene en cuenta otras variables de índole geográfica, como los factores naturales y sociales, y también la relación del hombre con los mismos.

Según Pérez Pineda (2005) “ *... la consideración del espacio en el estudio económico, avanzaría de la geografía económica a la conformación paralela de la economía regional, que intenta acercarse más a la economía que a la geografía a través de su*

*relación con temas de economía internacional, el papel de los costos de transacción, nuevas técnicas que permitan modelizar el papel del espacio etc”.*

Es interesante analizar el sentido del término “regional” cuando nos referimos a la economía regional, ya que el estudio de aspectos económicos en distintos territorios, y su posible interrelación, va a estar condicionado por lo que se entienda por región y por los factores políticos, sociales y geográficos que los definan y que sean relevantes para la ciencia económica. Aquí podríamos distinguir, a grandes rasgos, entre un enfoque de economía internacional, que estudia la cooperación y la formación de alianzas entre países, y otro enfoque, que sería el que tiende a estudiar regiones dentro de los países, cualquiera que sea el tamaño o denominación de las mismas, tales como autonomías, estados, ciudades, municipios, etc. Este segundo enfoque también se encarga del estudio de las relaciones entre subregiones de distintos países, por lo que también sería necesario para determinar o explicar dinámicas de carácter internacional.

Según Pérez Pineda (2005) la economía regional *“está marcada por altibajos e imprecisiones sobre su situación conceptual. El nombre de economía regional surgiría de la relación que guarda esta materia con el estudio del espacio y su vínculo directo con la ciencia regional, que originalmente se pensó en llamar ciencia espacial. Sin embargo por la efervescencia de los programas espaciales (la denominada “carrera espacial”) realizados entre los estadounidenses y soviéticos a finales de los cincuenta y principios de los sesenta, se consideró que se generaría cierta confusión con el “espacio exterior”, por lo que se escogería el termino de ciencia regional como el mejor termino (Isard 2001)”.*

A pesar de esta trayectoria no homogénea y cierto desinterés por el enfoque regional en esos años posteriores a la segunda guerra mundial se puede decir que algunos autores sentaron ciertas bases para crear un cuerpo teórico de la llamada ciencia regional. Según Pérez Pineda (2005) *“grandes obras como las de Isard (1956), Perroux (1954) o Meyer (1966), destacarían por lograr sentar las bases de un cuerpo teórico regional que descendía de las teorías de la localización. El auge se perdería hacia finales de los sesenta por los factores antes señalados y por el cuestionamiento de si la ciencia regional era una disciplina o un campo de estudio, y en consecuencia si podía considerarse como ciencia. A principios de los setenta resurgiría el interés por lo regional, es en esta etapa que la economía regional, definida en sus orígenes como “la*

*teoría económica de la separación espacial” (Meyer 1966) se despegaría como tal de su hermana la geografía económica y obtendría su lugar dentro del cuerpo de la ciencia regional. En particular, obras como las de Richardson (1973) esquematizarían la estructura de la ciencia regional “moderna” y los componentes de la economía regional, dividiendo a la ciencia regional en tres áreas: 1- teorías referentes a la localización (relacionado con geografía), 2- la economía urbana (relacionado con planeación) y 3- la economía regional (relacionado con economía)”.*

A partir de los 90, y como consecuencia de una mejora evidente en los instrumentos de análisis económico que tienen en cuenta el factor espacio, se produce una revitalización de las ideas y temas de estudio que se enmarcaron dentro de la geografía económica y la economía regional, que se “fusionan” y evolucionan a la denominada como **nueva geografía económica**. Esta trata de analizar y explicar por qué se producen determinados tipos de concentraciones (o aglomeraciones) económicas en espacios geográficos concretos.

Según Krugman (1998) “...se considera una estrategia de modelación hacia un enfoque que se concentra en los efectos del papel del tamaño de mercado en la generación de conexiones que fomenten la concentración geográfica de un lado y de otro en las fuerzas opuestas de los factores inmóviles que actúan en contra de tal concentración.”

Ofrece un marco teórico que explica la concentración de la actividad económica en el espacio y como las diferencias existentes en los distintos territorios explican sus diferencias económicas.

El equilibrio se alcanza por la acción de dos fuerzas antagónicas. Por un lado estaría la fuerza centrípeta, denominada de demanda o de tamaño de mercado, que es la responsable de la aglomeración geográfica. Por otro lado estaría la centrífuga, responsable de la dispersión, en la que influye la competencia entre empresas, que tienden a aprovechar ventajas de la región con mercados de menor tamaño.

Según Pérez Pineda (2005) “su estructura formal sugiere: modelos de equilibrio general, distribuciones geográficas endógenas de la población, oferta y demanda, agentes que maximizan su riqueza en términos de un lugar dadas las distribuciones existentes (población, oferta, etc.) y dadas las decisiones de otros agentes, competencia

*imperfecta, rendimientos crecientes, fuerzas centrífugas y centrípetas que explican la convergencia o divergencia de los lugares analizados (ciudades, regiones, industrias, etc.) su concentración y aglomeración; modelos con dos regiones y dos tipos de industrias, una inmóvil con competencia perfecta y otra móvil con competencia imperfecta (al interior de un país serían; periferia agrícola-inmóvil y centro manufacturero-móvil, respectivamente), diferencias salariales entre regiones que dependerán de la distribución de manufacturas entre esas regiones”.*

Efectivamente, son puntos clave el asumir la existencia de **rendimientos crecientes**, que hacen a su vez que nos encontremos con mercados de **competencia imperfecta**, así como la existencia de unos **costes de transporte** que determinan ventajas o desventajas de localización. La **movilidad de los factores productivos y los consumidores** sería otro factor clave para explicar los procesos económicos de concentración en el espacio. (Fujita y Krugman, 2004).

Los modelos propuestos por Fujita, Krugman y Venables (1999) son tres:

- 1- modelos regionales
- 2- modelos de sistemas urbanos
- 3- modelos internacionales.

El primero se correspondería con el **modelo centro-periferia** introducido por Krugman (1991) en el que se suponen dos regiones, dos sectores (agricultura y manufacturas) y dos tipos de trabajo (agricultores y trabajadores).

El sector manufacturero produce distintos tipos de un producto, cada uno de los cuales es producido por una empresa con economías de escala crecientes. Utiliza el factor trabajo (trabajadores) como único input. El sector agrario produce un solo bien sin diferenciar, utiliza también factor trabajo como único input (agricultores) y presenta rendimientos constantes. La movilidad de trabajadores y agricultores y el coste de transporte son dos factores importantes que presentan también diferencias entre los dos sectores. Los trabajadores del sector manufacturero tienen libertad de movimientos entre regiones, al contrario que los agricultores que son inmóviles, y la producción

agraria se transporta sin coste entre regiones mientras que las manufacturas si que tienen un coste de transporte.

La fuerza centrípeta, responsable de la aglomeración, va a depender, entre otras cosas, del grado de economías de escala y del tamaño y número de empresas establecidas en la región. Una economía mayor hace, además, que se tenga una mayor variedad de bienes de consumo y salarios mas altos atrayendo esto a más trabajadores. Por otro lado los productores van a querer aprovechar las economías de escala y la reducción de costes de transporte, produciendo donde el mercado es mayor, y se van a situar en esa región. Estas dos tendencias serían, a grandes rasgos, las que darían origen a la “causalidad circular” que generaría la fuerza centrípeta. La fuerza centrífuga se originaría de un modo más sencillo debido fundamentalmente a la inmovilidad de los agricultores. (Fujita, Krugman y Venables (1999)). Si la centrípeta supera a la centrífuga tendrá lugar un patrón centro-periferia donde la producción de manufacturas se concentrará en una sola región. Son factores favorecedores de esta situación los costes de transporte bajos, alto gasto en manufacturas y que estas estén muy diferenciadas. De todas formas esta concentración no tiene porqué ocurrir y existen condiciones en las que se obtiene una estabilidad con una equitativa distribución regional de la actividad económica, lo mismo que existen otras condiciones bajo las que también se alcanza una situación de equilibrio, sostenida en el tiempo, con una producción concentrada en una sola región.

El **modelo de sistema urbano** incorpora una visión que sustituye a la del modelo anterior de dos regiones y plantea una distribución uniforme de la actividad económica en el espacio, pudiendo los trabajadores localizarse donde deseen y escoger también el tipo de ocupación. Hay un cambio en el bien agrícola que ahora utiliza ambos factores de producción, capital y trabajo, a la vez que tiene que soportar costes de transporte, al igual que las manufacturas. La fuerza centrífuga se origina por la inmovilidad de la tierra, único factor inmóvil del modelo. Se parte de una ciudad, donde se concentra la producción industrial, y de una zona agrícola que la rodea, que se encuentran en equilibrio. El crecimiento económico hace que llegue un punto en el que sea interesante para algunas industrias, debido a la lejanía de la zona externa de la parte agrícola, trasladar su actividad industrial a una nueva localización. De esto surge una nueva ciudad donde se reproduce la situación, dando lugar a otras nuevas. En este proceso es importante la existencia de mercados potenciales que atraen industria y crean nuevas

áreas de actividad económica e industrial, definiendo un nuevo escenario y creando, a su vez, de forma paralela, otros nuevos mercados potenciales, que generarán también, nuevas concentraciones industriales haciendo así que la situación evolucione. Llegado a un punto suficiente, la distancia y el tamaño de las ciudades tiende a estabilizarse, como resultado de la acción de fuerzas centrípetas y centrífugas, alcanzando un nuevo equilibrio.

Un **tercer enfoque** sería el que también considera la especialización productiva que tiene lugar en algunas zonas, frente a los dos planteamientos anteriores, que consideran que aquellos puntos de concentración de la actividad generan todo tipo de industria. La realidad es que en algunos lugares se produce una especialización en un reducido número de bienes industriales o ramas de actividad. Esto es difícil de justificar, sobre todo teniendo en cuenta la existencia del supuesto de inmovilidad del factor trabajo, así como la no inclusión de factores de difícil cuantificación, pero de gran importancia en la localización de actividades. Es necesario, para corregir esto, restar importancia a la acumulación de factores en la aglomeración de la actividad. Para ello se considera una estructura vertical de la producción, en la que unos sectores “de arriba” producen para otros que están “debajo”, asumiendo, para las empresas de ambos, rendimientos crecientes y costes de transporte. Los incentivos de los productores para instalar su actividad vienen determinados por su mayor mercado o por la cercanía de sus proveedores, lo que hace que los dos sectores tiendan a ubicarse en la misma localización. Un enfoque que simplifique este planteamiento básico asumiendo que los sectores de arriba y abajo producen los mismos bienes, es decir lo producido se usa como input de otros bienes iguales, permite alcanzar un modelo algebraicamente isomórfico (con pequeñas reinterpretaciones en la simbología) al de centro periferia, tal como demuestran Krugman y Veneables (1995), que explica las razones que tiene la población para concentrarse y que explica por qué las industrias se concentran a su vez en determinadas concentraciones de población (Fujita y Krugman, 2004).

Los propios Fujita y Krugman (2004), manifiestan su opinión sobre algunas críticas negativas a la nueva geografía económica por parte de algunos científicos regionales y economistas urbanos:

Fujita: *“Bueno, se trata de un caso de “paraíso invadido”. Para los científicos regionales y economistas urbanos el área de la geografía económica/ economía*

---

*espacial ha constituido un paraíso donde, desde finales de los años cincuenta, exclusivamente ellos participaban de la realización de modelos matemáticos en relación al comportamiento de la economía espacial. Durante un largo periodo de tiempo el paraíso estaba protegido de forma natural con respecto a los economistas pertenecientes a la corriente dominante, quienes consideraron a la geografía económica como terreno poco favorable,..., ya que no se plantea el problema de la estructura de mercado en el espacio en el caso de la existencia de rendimientos crecientes. Sin embargo, con el tiempo, un economista perteneciente a tal corriente, Paul Krugman, llegó al paraíso a principios de los noventa trayendo consigo la nueva geografía económica. Esto se consideró, en cierto sentido, como una invasión. Sin embargo la llegada de la nueva geografía económica constituyó un menor problema para los científicos regionales y economistas urbanos, ya que ellos mismos eran recién llegados (ciencia regional a finales de 1950 y economía urbana a finales de 1960) al paraíso original de la geografía económica. Ellos abrazaron rápidamente a la nueva geografía económica como una nueva aportación al campo de la economía espacial.*

Krugman: “De hecho, nosotros consideramos, en cierta medida, que nuestro trabajo es como una continuación, quizás incluso validación, del sueño de Isard de retomar el espacio como núcleo de la ciencia económica.”

## 2.2. “PROBLEMA” ESPACIAL

Según Mur y Angulo (2008) “*un modelo econométrico espacial se ajusta al concepto general de modelo, como representación simplificada de la realidad, que incorpora una serie de especificaciones sobre aspectos fundamentales del mismo para convertirlo en un instrumento de investigación operativo. Su peculiaridad radica en la existencia de efectos espaciales, los cuales pueden agruparse en elementos de dependencia espacial o en elementos de heterogeneidad espacial. La interpretación de ambos conceptos es la usual. Por dependencia entenderemos que se quiebra el supuesto de independencia entre las observaciones, y por heterogeneidad que la hipótesis de estabilidad estructural no es admisible*”.

Existen cinco “principios fundamentales” que proponen Paelinck y Klaassen (1979) que sirven de referencia en la evolución y definición de los modelos de econometría espacial y sientan ciertas bases en el desarrollo de los mismos. Estos serían:

- Interdependencia espacial: la dependencia recíproca entre las diferentes unidades de análisis debe ser tomada en cuenta e incorporada a los modelos.
- Asimetría en las relaciones espaciales
- *Alotopía*: la causa de un fenómeno espacial tiene su origen en otro lugar del espacio.
- No linealidad. Se diferencia entre interacción *ex ante* y *ex post*, lo que suele dar lugar a especificaciones de los modelos no lineales.
- Topología: Inclusión de variables de distancias entre dos localizaciones, coordenadas, densidades, superficies, etc.

Según Chasco (2003) *“dado que la econometría espacial incluye las técnicas econométricas aplicadas a datos y modelos de naturaleza espacio-geográfica, la cuestión que se plantearía es si este tipo de datos contiene tantas peculiaridades como para demandar unas técnicas de investigación específicas. En este sentido, habría que decir, en primer lugar, que la principal característica de estos datos es su naturaleza georreferenciada, es decir, que su posición relativa o absoluta sobre el espacio contiene información valiosa para interpretar las relaciones consideradas. Otra característica básica del espacio geográfico es la **multidireccionalidad** de las relaciones que sobre él se establecen y la **multidimensionalidad**: efectivamente, en un área geográfica no es posible distinguir entre pasado, presente, futuro, sino que todo es presente, todo es pasado y todo es futuro. Estas peculiaridades de los datos geográficos dan lugar a los llamados “efectos espaciales” de dependencia o autocorrelación espacial y heterogeneidad espacial, que deberían ser considerados explícita y adecuadamente en todo ejercicio de modelización y predicción de datos geográficos”*.

Para Anselin (2001) una definición de econometría espacial podría ser *“la parte de la econometría que se ocupa del tratamiento de la interacción espacial (autocorrelación espacial) y la estructura espacial (heterogeneidad espacial) en los modelos de regresión de corte transversal y de datos de panel”*. Según Moreno y Vayá (2000) los denominados efectos espaciales son la heterogeneidad y la dependencia (o autocorrelación) espacial.

La **heterogeneidad espacial** surge por la existencia de variaciones en las relaciones entre las variables objeto de estudio cuando consideramos distintas ubicaciones en el espacio. Puede tener una doble concreción, por un lado en forma de inestabilidad estructural y por otro en forma de heteroscedasticidad. En el primer caso las variables no son homogéneas en el espacio lo que hace que los parámetros y la forma funcional de una regresión cambien a su vez en el mismo, siendo necesario tener en cuenta esta circunstancia, que obligaría a plantear funciones y parámetros flexibles en el espacio. En el caso de la heteroscedasticidad esta es bastante habitual en el contexto espacial, incluso más que en el temporal, y puede deberse tanto a la naturaleza de los datos a utilizar como a errores de especificación de los modelos. Su tratamiento en muchos casos es el mismo que el empleado en la modelización econométrica “no espacial”, aunque a veces requiere técnicas más complejas, según Moreno y Vayá (2000) “*más eficientes, como la expansión espacial de parámetros (Cassetti, 1972) o las regresiones ponderadas geográficamente (Fotheringham et al, 1998)*”.

La **autocorrelación o dependencia espacial** se produce cuando no existe independencia entre las observaciones, siendo este un problema que se puede originar, tanto por la naturaleza de los datos espaciales, como por la forma de obtenerlos. Siempre que el valor de una variable tenga una relación funcional con el valor de esa misma variable en otro lugar del espacio y no dependa, por tanto, exclusivamente de factores propios de la variable en esa región, viéndose así condicionada por lo que ocurre en regiones vecinas, se produce la autocorrelación espacial. La variable espacio es continua y la división del mismo muchas veces arbitraria o al menos realizada con criterios distintos y no coincidentes con las unidades espaciales relevantes para la generación de las variables económicas objeto de estudio. Sería este el caso en el que la forma de obtención de los datos provocaría dependencia espacial. Otras veces la propia naturaleza de los datos y variables provoca este fenómeno. Lo que ocurre en un lugar del espacio suele tener repercusión, ya sea positiva o negativa, en otras regiones que se consideran vecinas. Según Moreno y Vayá (2000) “*la existencia de fenómenos de interacción espacial, de efectos desbordamiento y de jerarquías espaciales pueden tener como consecuencia la aparición de un esquema de autocorrelación espacial*”.

La dependencia puede ser positiva o negativa. El primer caso tiene lugar cuando la aparición o existencia de un determinado fenómeno económico que se produce en una

unidad espacial o región hace que se produzca ese fenómeno en las colindantes o “vecinas”. Cuando lo que se consigue es que no se produzca o que se reduzca estaremos en el segundo caso de autocorrelación espacial negativa.

## Operadores espaciales y matriz de pesos espaciales

La multidireccionalidad es una de las características propias del análisis espacial que va a condicionar el planteamiento del retardo espacial. En el caso del análisis temporal el operador de retardos facilita el manejo matemático de las realizaciones de las variables en distintos momentos de tiempo, y su función es “trasladar” la variable a lo largo del eje temporal, que tiene una única dirección y presenta de forma ordenada en el tiempo la secuencia de los datos de la variable. Además se expresa en potencias del operador  $L$ , tal que:

$$L_j y_t = y_{t-1}$$

La mencionada multidireccionalidad y la no ordenación prefijada de los datos impiden utilizar un operador como el temporal para “trasladar” una variable en el espacio, siendo necesario utilizar una matriz que recoja y ordene esas múltiples relaciones.

La matriz de pesos espaciales (también denominada matriz de contactos o matriz de proximidad espacial) y simbolizada con  $W$ , es una matriz cuadrada de  $N \times N$  (siendo  $N$  el número de unidades espaciales), no estocástica, cuyos elementos ( $w_{i,j}$ ) reflejan la intensidad de la interdependencia entre cada par de regiones  $i,j$  (Moreno y Vayá, 2000).

No existe un criterio único para definir la interdependencia o contigüidad entre regiones, siendo lo más adecuado ajustarse las características concretas de cada caso objeto de análisis. Los más habituales serían los siguientes:

1.- El más habitual, simple y que suele dar mejores resultados es el que atribuye el valor 1 si las regiones  $i,j$  son contiguas y cero si no lo son. Los elementos de la diagonal principal son cero, pues ninguna región puede ser vecina de sí misma. Se utiliza un criterio de contigüidad física de primer orden de tipo binario. También se pueden calcular las matrices de contigüidad de órdenes superiores.

En algunos casos suele normalizarse la matriz  $W$  por filas. Esto es, se divide cada elemento por la suma de fila a la que pertenece. De esta forma, la suma de los pesos asociados a cada área es igual a 1.

Parece razonable suponer que los puntos más próximos entre sí tendrán relaciones más fuertes que con otros más distantes, aunque, también es cierto que la distancia geográfica no tiene por qué ser un elemento determinante en todos los casos, ya que pueden existir otros factores económicos que establezcan dependencias. Por tanto, como ya se ha dicho, la interpretación de la cercanía debe relativizarse siempre al caso concreto que se esté analizando.

2.- Cliff y Ord (1981) se basan en la distancia que separa a dos regiones  $i$  y  $j$ , ( $d_{i,j}$ ) ponderada por el tamaño de la frontera que presenta en común ( $B_{i,j}$ ). En este caso la formulación da lugar a una matriz asimétrica. Los elementos de la matriz  $W$  serían:

$$w_{i,j} = (d_{i,j})^{-a} (B_{i,j})^b$$

Donde  $a$  y  $b$  son parámetros a estimar.

3.- Dacey (1969) propone

$$w_{i,j} = \gamma_{i,j} B_{i,j} \alpha_i$$

donde

$B_{i,j}$  sigue siendo el tamaño de la frontera que dos regiones presentan en común,

$\gamma_{i,j}$  es un factor de contigüidad binario.

$\alpha_i$  es el area de la región  $i$  en relación al área total.

4.- Bodson y Peeters (1975) plantean que la matriz  $W$  considere un criterio de accesibilidad general. Esto es, que tenga en cuenta y combine los distintos canales de comunicación entre regiones. En consecuencia, esta formulación permite redefinir el concepto de vecindad, ya que para que dos regiones lo sean deban existir buenos canales de comunicación entre ellas. La expresión matemática de los elementos  $w_{i,j}$  de la matriz sería:

$$w_{i,j} = \sum K_n \{a/(1+be^{-c_j d_{ij}})\}$$

Siendo:

$K_n$  la importancia relativa de la vía de comunicación  $n$ ,

$N$  el número de vías de comunicación,

$d_{i,j}$  la distancia entre las regiones  $i$  y  $j$ ,

$a$ ,  $b$  y  $c_j$  y son parámetros a estimar.

5.- Case, A., Rosen, H. y Hines, J. (1993) proponen que la matriz  $W$  se sustente en distancias económicas. En este caso, se asumirá que una región es vecina de otra cuando esas distancias, en términos de alguna variable económica operativa, sea pequeña (lo que deberá ser definido por el investigador).

De una forma análoga al operador de retardos temporal, que desplaza la variable en el tiempo, la matriz de pesos espaciales provoca desplazamientos en el espacio. Si la aplicamos a un vector de observaciones de una variable, obtenemos otro vector que será el retardo espacial del primero.

El **operador de retardo espacial** consiste en un promedio ponderado de variables aleatorias en localizaciones vecinas, con unas ponderaciones fijas y dadas de forma exógena (Moreno y Vayá, 2000).

Es decir, el retardo espacial considera la suma ponderada de todos los valores de una variable que pertenecen a un mismo criterio de contigüidad (torre/alfil/reina), en vez de tomar individualmente cada uno de ellos. Los términos de esta suma se obtienen multiplicando las observaciones en cuestión ( $y_i$ ) por sus correspondientes pesos de la matriz de ponderaciones  $W$  del modo siguiente:

$$B^s y_i = \sum_j w_{ij} y_j; \text{ para todo } j \text{ que pertenezca a } J_i$$

donde :

$B^s$ : operador retardo asociado con el criterio de contigüidad  $s$

$j$ : subíndice correspondiente al conjunto  $J_i$  de unidades relacionadas con  $i$ , según el criterio de contigüidad  $s$

$w_{ij}$ : ponderaciones espaciales.

De esta forma, cada elemento del retardo espacial es igual a un promedio ponderado de los valores de la variable  $Y$  en el subgrupo de observaciones vecinas a ella,  $J_i$ , dado que,  $w_{ij} = 0$ , para todo  $j$  que no pertenezca a  $J_i$ . (Chasco, 2003).

El operador de retardo espacial (que también se denota como  $W_y$ ) viene a distribuir o acumular en el espacio valores de variables en puntos concretos. Esto, a veces, puede ocasionar problemas de interpretación por lo que se debe buscar siempre el sentido económico a la hora de definir las relaciones espaciales que recojan el “cambio espacial” entre regiones.

### 2.3. ANÁLISIS EXPLORATORIO DE DATOS ESPACIALES.

El análisis exploratorio de datos (AED) estudia asociaciones, comportamientos recurrentes y tendencias en grandes bases de datos, fundamentalmente, pero no tiene en cuenta las peculiaridades espaciales. Esto puede llegar a invalidar sus resultados, sobre todo cuando se dan relaciones de autocorrelación y heterogeneidad espacial.

Según Chasco (2003) *“el análisis exploratorio de datos (AED), a diferencia de los contrastes de hipótesis tradicionales diseñados para la verificación a priori de conexiones entre variables, se utiliza para identificar relaciones sistemáticas entre variables cuando no existen expectativas claras sobre la naturaleza de estas relaciones”*.

*“De este modo, es posible obtener una estructura explicativa de los datos a través de unas técnicas que combinan herramientas de la estadística básica (descriptivos, correlaciones, tablas de frecuencias o de correlación cruzada) con análisis multivariante avanzado, especialmente diseñado para identificar formas en grandes bases de datos (análisis cluster, escalas multidimensionales, análisis logit, correlación canónica, análisis de correspondencias, árboles de clasificación, etc.). Las técnicas de AED suelen estar acompañadas de métodos de visualización gráfica capaces de identificar conexiones, tendencias o sesgos presentes en las bases de datos inicialmente desestructuradas”*

El AEDE para Anselin (1998) es *“una colección de técnicas para describir y visualizar distribuciones espaciales, identificar localizaciones espaciales atípicas u outliers espaciales, descubrir patrones de asociación espacial, clusters o puntos calientes, y sugerir regímenes espaciales u otras formas heterogeneidad espacial”*.

Según Chasco (2003) *“el AEDE tiene un carácter descriptivo (estadístico) más que confirmatorio (econométrico) pero, tras la detección de estructuras espaciales en las variables geográficas, hará posible la formulación de hipótesis previas para la modelización econométrica y, en su caso, posible predicción espacial de nuevos datos. Es decir, el AEDE debe constituir la fase previa a toda modelización econométrica espacial, sobre todo cuando no exista un marco formal o teoría previa acerca del fenómeno que se pretende explicar”*.

Las distintas técnicas que se pueden emplear en al AEDE, son varias y pueden ser clasificadas según distintos criterios, pero la mayoría de autores consideran, por un lado, los métodos gráficos y por otro, los contrastes estadísticos, y ambos son complementarios. Se debe considerar si el análisis es univariante o multivariante y si queremos considerar la asociación espacial global o local. En las técnicas gráficas existen dos perspectivas que serían la geoestadística y la reticular o “lattice”, que se diferencian, según Anselin (1998), en la forma en que la proximidad espacial es formalizada. La geoestadística supone procesos espaciales continuos, que hacen usar la distancia métrica para organizar las observaciones y la lattice supone observaciones de carácter discreto, siendo clave el concepto de vecino espacial que conduce a la matriz de pesos espaciales y a las variables espaciales retardadas. Para ambos enfoques algunos autores (Anselin (1998), Moreno y Vayá (2000) y Chasco (2003)) distinguen entre técnicas que se usan para representar la distribución espacial, de otras que reflejan la asociación espacial y un tercer bloque para la heterogeneidad espacial. Recogen las siguientes **técnicas gráficas**:

- **Función de distribución acumulativa.** Refleja la distribución espacial, desde un enfoque geoestadístico para el caso univariante, lo mismo que el **diagrama/mapa de puntos** y el **mapa de círculos**.
- **Diagrama/mapa de caja** representa la distribución espacial, desde un enfoque lattice para el caso univariante.

- **Diagrama dispersión-caja** muestra la distribución espacial bajo la perspectiva lattice en el caso multivariante.
- **Cartograma y gráfico de coordenadas paralelas** muestran la distribución espacial bajo la perspectiva lattice y geoestadística en el caso multivariante.
- **Diagrama de dispersión espacialmente retardado, nube del variograma y diagrama de caja del variograma** recogen la asociación espacial global bajo la aproximación geoestadística.
- **Puntos atípicos en el diagrama de caja del variograma y puntos atípicos en la nube del variograma** recogen la asociación espacial local bajo la aproximación geoestadística.
- **Diagrama/mapa de dispersión de Moran, mapa de contigüidades espaciales y gráfico del retardo espacial** muestran la asociación espacial global bajo el enfoque lattice.
- **Mapas LISA, diagrama de caja LISA y puntos atípicos en el diagrama de dispersión de Moran**, muestran la asociación espacial local bajo el enfoque lattice.
- **Nube del variograma multivariante** representa la asociación espacial bajo el enfoque geoestadístico para el caso multivariante.
- **Diagrama de dispersión multivariante de Moran** representa la asociación espacial bajo enfoque lattice en el caso multivariante.
- **Mapa del histograma de frecuencias** estudia la heterogeneidad espacial con la perspectiva lattice lo mismo que el **diagrama de dispersión**.

Con los **contrastes estadísticos** se persigue detectar el grado de asociación espacial global o local. Los test de **autocorrelación espacial global** tratan de contrastar la independencia o no de la distribución de una variable en la totalidad de un área territorial, mientras que los de **autocorrelación espacial local**, se centran en la detección de relaciones espaciales en zonas o subzonas concretas del territorio,

midiendo su aportación a las características de independencia de la variable en todo el espacio considerado.

Los test más conocidos serían, para la autocorrelación espacial global, el test **I de Moran**, del que se deriva el **diagrama de dispersión de Moran**, el test **c de Geary** y la **G (d) de Getis y Ord**. También estaría el **Γ de Mantel**. Para la autocorrelación espacial local estarían los **estadísticos locales de Getis y Ord** y los **LISA de Anselin**.

Además están los **diagramas espacio temporales** como sería el **coeficiente de autocorrelación espacio tiempo de Moran** propuesto por Chasco (2010) y los **indicadores LISA para el caso bivalente**. También está el **análisis de correlación espacial para tasas**, que pretende corregir los problemas ocasionados al presentar las variables en ratios. Las técnicas serían la **estandarización empírica bayesiana**, el **método directo de estandarización de tasas** propuesto por Saez y Saurina (2007) y la **adaptación del estadístico I de Moran** propuesto por Assunção y Reis (1999).

## Contrastes de autocorrelación espacial global

### - I de Moran

$$I = \frac{N}{S_0} \frac{\sum_i \sum_j w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^N (y_i - \bar{y})^2}$$

con  $i \neq j$ , donde

$y_i$  representa el valor de la variable cuantitativa y para la región  $i$ ,

$N$  el tamaño muestral,

$w_{ij}$  los pesos espaciales de una matriz de contactos  $W$ ,

$S_0$  la sumatoria de los pesos espaciales.

Si se utiliza una matriz de pesos espaciales estandarizada por filas  $S_0 = N$ .

El valor esperado de la I de Moran es  $-1 / N-1$  por lo que es negativo, depende de N y tiende a cero cuando N aumenta. Además según Cliff y Ord (1981) cuando el tamaño muestral es suficientemente amplio, la expresión estandarizada del test I, que se obtiene restandole a I su valor esperado y dividiendo por la desviación típica, se distribuye como una normal tipificada,  $N(0,1)$ .

### - Diagrama de dispersión de Moran

Del índice de Moran se deriva la siguiente expresión matricial:

$$I = \frac{N y' W y}{S_0 y' y}$$

N es el nº observaciones,

$S_0$  es la suma de todos los elementos de la matriz de pesos espaciales,

y' es el vector de observaciones en desviaciones a la media,

$W y'$  es el retardo espacial asociado a la variable y en desviaciones a la media.

Donde  $S_0$  es igual a N al estandarizar por filas la matriz de pesos espaciales haciendo que estas sumen 1 y la expresión quedaría

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

En este caso I sería la pendiente de una regresión lineal simple de  $W y'$  sobre y, lo que permite interpretarlo como el grado de ajuste de la asociación lineal entre el retardo espacial de una variable expresada en desviaciones con respecto a la media y dicha variable igualmente expresada.

### - c de Geary

$$C = \frac{(N - 1) \sum_i \sum_j w_{ij} (y_i - y_j)^2}{2 S \sum (y_i - \bar{y})^2}$$

Los elementos tienen el mismo significado que en la expresión de la I de Moran. La c de Geary tiene en cuenta el cuadrado de las diferencias entre observaciones en vez de su diferencia con respecto a la media. También, cuando el tamaño muestral es suficientemente amplio, la expresión estandarizada, que se obtiene restándole a c su valor esperado y dividiendo por la desviación típica, se distribuye como una normal tipificada,  $N(0,1)$ .

#### - G (d) de Getis y Ord

$$G(d) = \frac{\sum \sum w_{ij}(d) y_i y_j}{\sum \sum y_i y_j}$$

En este estadístico se introduce el concepto de distancia como criterio a la hora de establecer relaciones. Para una región serán vecinas aquellas que estén a una distancia igual o inferior a d. Para cada distancia d se construye una matriz de pesos espaciales  $W(d)$ , para  $i \neq j$ , donde  $w_{ij}(d)$  es 1, si i y j están a una distancia igual o inferior a d, y 0 en el caso contrario. Este estadístico solo se puede calcular para variables de valores positivos. En el caso del contraste global se aplica a toda la muestra, al contrario que los estadísticos locales que se calculan para cada punto i. Al igual que los anteriores se trabaja sobre el valor estandarizado.

#### - $\Gamma$ de Mantel

El índice general de asociación matricial  $\Gamma$  se expresa

$$\Gamma = \sum \sum a_{ij} b_{ij}$$

donde  $a_{ij}$  y  $b_{ij}$  son los elementos de dos matrices de similitudes. Es la base metodológica de la I de Moran y la c de Geary. Chasco (2003) "Las medidas de asociación espacial pueden obtenerse, en general, expresando la similitud espacial en una matriz (por ejemplo, una matriz de contigüidades o de pesos espaciales) y la similitud de valores de una variable en otra. Distintas medidas de similitud de valores dan lugar a diferentes índices de asociación espacial. Por ejemplo, para  $a_{ij} = w_{ij}$ :

- $b_{ij} = y_i y_j$ : conduce a una medida parecida al test I de Moran.
- $b_{ij} = (y_i - y_j)^2$ : da lugar a una índice parecido al test c de Geary.

## Contrastes de autocorrelación espacial local

Los estadísticos de Getis y Ord serían:

-  $G_i(d)$

$$G_i(d) = \frac{\sum_{j=1}^N w_{ij}(d)y_j}{\sum_{j=1}^N y_j} \quad \forall j \neq i$$

$W_{ij}(d)$  matriz de contactos con  $w_{ij}(d) = 1$  si el punto  $j$  está en un radio igual o inferior a  $d$  del punto  $i$ , y  $w_{ij}(d) = 0$  para el resto. No considera la relación de  $i$  consigo misma, que además es  $w_{ii}(d) = 0$ .

-  $G_i^*(d)$

$$G_i^*(d) = \frac{\sum_{j=1}^N w_{ij}(d)y_j}{\sum_{j=1}^N y_j} \quad \forall j$$

Solo varía, con respecto a la anterior, en que considera también la observación de  $i$  consigo misma, siendo  $w_{ii}(d) = 1$ , lo que hace que solo cambie en la fórmula el número de observaciones del denominador.

Tanto  $G_i(d)$  como  $G_i^*(d)$  presentan la importante limitación de poder ser calculados solo para valores naturales y positivos, y para matrices de contactos binarias y simétricas. Para superar esas limitaciones se proponen la nueva  $G_i(d)$  y la nueva  $G_i^*(d)$ , que surgen de estandarizar a los anteriores.

- Nueva  $G_i(d)$

$$\text{Nueva } G_i(d) = \frac{G_i(d) - E[G_i(d)]}{SD[G_i(d)]} \quad \forall j \neq i$$

- Nueva  $G_i^*(d)$

$$\text{Nueva } G_i^*(d) = \frac{G_i^*(d) - E[G_i^*(d)]}{SD[G_i^*(d)]} \quad \forall j$$

Se considera que los dos estadísticos anteriores se distribuyen normalmente.

Los estadísticos LISA son, según Chasco (2003) “capaces de detectar la contribución de cada región a un indicador de dependencia espacial global (por ejemplo, el estadístico *I* de Moran), así como posibles valores atípicos (“outliers”), con una participación en el estadístico global muy superior o inferior a la media. Este tipo de indicadores permiten la consecución de dos objetivos: la determinación de agrupamientos espaciales locales significativos en torno a un punto concreto del espacio (como los estadísticos  $G_i$  y  $G^*_i$  de Getis y Ord) y la obtención de bolsas de inestabilidad (no estacionariedad) espacial, es decir, la presencia de valores atípicos que también pueden ser visualizados mediante el diagrama de dispersión de Moran”.

**- Estadístico local Gamma.**

$$\Gamma_i = \sum_j a_{ij} b_{ij}$$

Se desarrolla a partir del índice general de asociación matricial,  $\Gamma$ . Se cumple que  $\sum \Gamma_i = \Gamma$ .

**- Estadístico local de Moran**

Se basa en la *I* de Moran y también es un caso particular de  $\Gamma_i$ .

$$I_i = z_i \sum_{j=1}^{J_i} w_{ij} z_j$$

donde  $z_i$  y  $z_j$  son las variables estandarizadas, y el sumatorio es solo para los  $j$  que pertenecen a  $J_i$ , siendo  $J_i$  el conjunto de los valores vecinos a  $i$ . La suma de los estadísticos locales  $I_i$  es el estadístico global *I* de Moran

**- Estadístico local de Geary**

$$C_i = \sum_j w_{ij} (z_i - z_j)^2$$

Se basa en la  $c$  de Geary y, al igual que el anterior, también es un caso particular de  $\Gamma_i$ , donde también  $z_i$  y  $z_j$  son las variables estandarizadas, y el sumatorio es solo para los  $j$  que pertenecen a  $J_i$ , siendo  $J_i$  el conjunto de los valores vecinos a  $i$ . La suma de los estadísticos locales  $c_i$  es el estadístico global  $c$  de Geary.

## 2.4. ANÁLISIS CONFIRMATORIO DE DATOS ESPACIALES.

### 2.4.1 MODELOS

Los fenómenos de dependencia o autocorrelación espacial y heterogeneidad espacial pueden estar presentes en modelos de regresión de distintas formas, siendo entonces necesario tenerlos en cuenta en la especificación de dichos modelos. Las técnicas de econometría espacial permiten diferenciar y utilizar distintos tipos de modelos en función del problema espacial que se deba recoger.

La autocorrelación espacial puede estar presente en un modelo de regresión lineal cuando existen variables (endógenas y/o exógenas) espacialmente correlacionadas, denominándose en este caso “**sustantiva**”, o cuando esta correlación espacial se da en el término de perturbación aleatorio, denominándose entonces “**residual**”.

La heterogeneidad espacial, que se produce por inestabilidad estructural, puede a su vez manifestarse también de dos formas, una a través de residuos con varianza no constante, es decir heteroscedasticidad, y otra cuando se da inestabilidad paramétrica (continua o discreta), es decir cambian los coeficientes del modelo lineal.

Para el fenómeno de autocorrelación espacial Anselin (1988) propuso cuatro especificaciones:

#### 1. Modelo autorregresivo de regresión espacial.

Se utiliza esta expresión en los casos en los que se detecta la existencia de dependencia entre los valores que toma una variable en un lugar concreto ( $i$ ) y los valores que toma la misma variable en otras zonas del espacio ( $j$ ). Es decir la  $Cov(y_i, y_j) \neq 0$ ,

$$y = \rho W y + u$$

$$u \approx N(0, \sigma^2 I)$$

donde :

- $y$  es un vector ( $N \times 1$ ) observaciones de la variable endógena
- $\rho$  coeficiente autorregresivo espacial que recoge la intensidad de la dependencia entre las observaciones muestrales
- $W$  es la matriz de pesos espaciales de la variable  $y$
- $u$  es el término de perturbación aleatoria ruido blanco.

## 2. Modelo básico de regresión espacial.

Cuando existe correlación espacial, pero esta se puede explicar por un conjunto de variables exógenas, se plantea esta expresión. Según Chasco (2003) *“el MBRL será la especificación correcta para una variable que presenta autocorrelación espacial, sólo en el caso de que este efecto espacial esté totalmente explicado por los valores de una o más variables explicativas, es decir, por condicionantes internos referidos a dicho lugar  $i$ . Por tanto, la inclusión en el modelo de un número ( $K$ ) de variables explicativas produce una ausencia de significatividad (a contrastar estadísticamente) en la relación espacial existente entre  $y_i$ ,  $y_j$ , de forma que  $Cov(y_i, y_j) = 0$ . Esta situación implica también la ausencia de autocorrelación espacial residual, siempre y cuando se trate de un modelo en el que se hayan especificado correctamente los determinantes estructurales de la variable endógena.”*

$$y = X \beta + u$$

$$u \approx N(0, \sigma^2 I)$$

donde:

- $X$  es una matriz ( $K, N$ ) de  $K$  variables exógenas y  $N$  observaciones
- $\beta$  vector ( $K, 1$ ) parámetros de las variables exógenas
- $u$  es el término de perturbación aleatoria ruido blanco.

### 3. Modelo del error espacial.

Se utiliza en los casos en que la autocorrelación espacial se traslada al término de perturbación aleatoria. Suele ocurrir cuando se da una incorrecta especificación del modelo básico de regresión espacial, lo que hace que el término de perturbación aleatoria recoja las características propias de dependencia espacial que se dan en la variable endógena y que no son consideradas por las variables explicativas.

Se suelen usar dos especificaciones, la primera considera que el término de perturbación aleatoria se distribuye como un proceso autorregresivo de orden 1, AR (1), y es la más habitual.

$$y = X \beta + u$$

$$u = \lambda W u + \varepsilon$$

$$\varepsilon \approx N(0, \sigma^2 I)$$

donde:

- $X$  es una matriz ( $K, N$ ) de  $K$  variables exógenas y  $N$  observaciones.
- $\beta$  vector ( $K, 1$ ) parámetros de las variables exógenas.
- $u$  es la perturbación aleatoria que sigue un proceso AR (1).
- $\lambda$  es el parámetro autorregresivo.
- $W$  es la matriz de pesos espaciales de  $u$ .
- $\varepsilon$  es el término de perturbación aleatoria ruido blanco.

Una segunda especificación considera que el término de perturbación aleatoria se distribuye como un proceso de medias móviles de orden 1 en vez del autorregresivo del modelo anterior. Su expresión sería:

$$y = X \beta + u$$

$$u = \theta W \varepsilon + \varepsilon$$

$$\varepsilon \approx N(0, \sigma^2 I)$$

donde:

- $X$  es una matriz  $(K,N)$  de  $K$  variables exógenas y  $N$  observaciones.
- $\beta$  vector  $(K, 1)$  parámetros de las variables exógenas.
- $u$  es la perturbación aleatoria que sigue un proceso MA (1).
- $\theta$  es el parámetro de medias móviles.
- $W$  es la matriz de pesos espaciales de  $\varepsilon$ .
- $\varepsilon$  es el término de perturbación aleatoria ruido blanco.

#### 4. Modelo del retardo espacial.

Este modelo se utiliza en los casos en los que el modelo básico de regresión espacial tampoco recoge en su totalidad la autocorrelación espacial a través de las variables explicativas. Esto hace que se incorpore en este modelo una variable dependiente espacialmente retardada que lo recoja.

$$y = \rho W y + X \beta + u$$

$$u \approx N(0, \sigma^2 I)$$

donde:

- $y$  es un vector  $(N \times 1)$  observaciones de la variable endógena.
- $\rho$  coeficiente autorregresivo espacial que recoge la intensidad de la dependencia entre las observaciones muestrales.
- $W$  es la matriz de pesos espaciales de la variable  $y$ .
- $X$  es una matriz  $(K, N)$  de  $K$  variables exógenas y  $N$  observaciones.
- $\beta$  vector  $(K, 1)$  parámetros de las variables exógenas.
- $u$  es el término de perturbación aleatoria ruido blanco.

A partir de estas cuatro especificaciones posteriormente se desarrollaron otras. Las más destacadas son las propuestas por Florax y Folmer (1992). Según Moreno y Vayá (2000) a partir de las sencillas propuestas anteriores se puede llegar a la expresión general más común que sería la del **modelo mixto regresivo de regresión espacial, con perturbaciones aleatorias autorregresivas y heteroscedásticas**

$$y = \rho W_1 y + X \beta_1 + W_2 R \beta_2 + u$$

$$u = \lambda W_3 u + \varepsilon$$

$$\varepsilon \approx N(0, \Omega); \Omega_{ii} = h_i(Z\alpha); h_i \geq 0$$

donde:

- $y$  es un vector  $(N,1)$  de observaciones de la variable endógena
- $\rho$  coeficiente autorregresivo espacial que recoge la intensidad de la dependencia entre las observaciones muestrales
- $W_1$  es la matriz de pesos espaciales de la variable endógena
- $W_2$  matriz de pesos espaciales de las variables exógenas espacialmente retardadas
- $X$  es una matriz  $(K_1,N)$  de  $K_1$  variables exógenas y  $N$  observaciones
- $R$  es una matriz  $(K_2,N)$  de  $K_2$  variables exógenas espacialmente retardadas, que pueden o no coincidir con las variables incluidas en  $X$
- $\beta_1$  y  $\beta_2$  son dos vectores  $(K_1,1)$  y  $(K_2,1)$  de parámetros de las variables exógenas y exógenas retardadas espacialmente, respectivamente
- $u$  es la perturbación aleatoria que sigue un proceso AR(1)
- $\lambda$  es el parámetro autorregresivo asociado a  $u$
- $W_3$  es la matriz de pesos espaciales de la variable aleatoria  $u$ .
- $\varepsilon$  vector de perturbaciones aleatorias con una matriz de varianzas y covarianzas diagonal pero heterocedástico, siendo los elementos de la diagonal principal de la matriz de covarianzas  $(\Omega)$  función de  $P+1$  variables exógenas de  $Z$ .

Desde esta expresión general se derivan otras que se ajustan a cada caso concreto en función de la dependencia espacial detectada. Se trata de ir incorporando restricciones a ese modelo general para obtener las expresiones de los modelos.

Según Moreno y Vayá (2000) los modelos de regresión que incorporan dependencia espacial y sus expresiones serían:

**Dependencia espacial sustantiva:**• **De primer orden**

1. Modelo mixto regresivo de regresión espacial, con perturbaciones aleatorias autorregresivas y heteroscedásticas

$$y = \rho W_1 y + X \beta_1 + W_2 R \beta_2 + u$$

$$u = \lambda W_3 u + \varepsilon$$

$$\varepsilon \approx N(0, \Omega); \Omega_{ii} = h_i(Z\alpha); h_i \geq 0.$$

2. Modelo mixto regresivo-autorregresivo espacial de primer orden (modelo del retardo espacial).

$$y = \rho W_1 y + X \beta_1 + u$$

$$u \approx N(0, \sigma^2 I)$$

3. Modelo mixto regresivo-autorregresivo espacial con perturbaciones espaciales autorregresivas

$$y = \rho W_1 y + X \beta_1 + u$$

$$u = \lambda W_2 u + \varepsilon$$

$$\varepsilon \approx N(0, \sigma^2 I)$$

4. Modelo mixto regresivo-regresivo espacial

$$y = \rho W_1 y + X \beta_1 + W_2 R \beta_2 + u$$

$$u \approx N(0, \sigma^2 I)$$

5. Modelo mixto regresivo cruzado-regresivo espacial

$$y = \rho W_1 y + W_2 R \beta_2 + u$$

$$u \approx N(0, \sigma^2 I)$$

6. Modelo mixto regresivo-autorregresivo espacial con perturbaciones espaciales media móvil SARMA (1,1)

$$y = \rho W_1 y + X \beta_1 + u$$

$$u = \theta W_2 \varepsilon + \varepsilon$$

$$\varepsilon \approx N(0, \sigma^2 I)$$

- **De órdenes superiores**

Modelo mixto regresivo-autorregresivo espacial con perturbaciones espaciales media móvil SARMA (p, q)

$$y = \rho_1 W_1 y + \rho_2 W_2 y + \rho_3 W_3 y + \dots + \rho_p W_p y + X \beta_1 + u$$

$$u = \theta_1 W_1 \varepsilon + \theta_2 W_2 \varepsilon + \theta_3 W_3 \varepsilon + \dots + \theta_q W_q \varepsilon + \varepsilon$$

$$\varepsilon \approx N(0, \sigma^2 I)$$

**Dependencia espacial residual:**

**De primer orden**

- Estructura autorregresiva SAR
1. Modelo mixto regresivo de regresión espacial, con perturbaciones aleatorias autorregresivas y heteroscedásticas

$$y = \rho W_1 y + X \beta_1 + W_2 R \beta_2 + u$$

$$u = \lambda W_3 u + \varepsilon$$

$$\varepsilon \approx N(0, \Omega); \Omega_{ii} = h_i(Z\alpha); h_i \geq 0.$$

2. Modelo mixto regresivo-autorregresivo espacial con perturbaciones espaciales autorregresivas

$$y = \rho W_1 y + X \beta_1 + u$$

$$u = \lambda W_2 u + \varepsilon$$

$$\varepsilon \approx N(0, \sigma^2 I)$$

3. Modelo con perturbaciones espaciales autorregresivas de primer orden (modelo del error espacial).

$$y = X \beta_1 + u$$

$$u = \lambda W u + \varepsilon$$

$$\varepsilon \approx N(0, \sigma^2 I)$$

- Estructura media móvil SMA

1. Modelo de medias móviles espacial de orden 1. SMA (1)

$$y = X \beta_1 + u$$

$$u = \theta W_1 \varepsilon + \varepsilon$$

$$\varepsilon \approx N(0, \sigma^2 I)$$

2. Modelo mixto regresivo-autorregresivo espacial con perturbaciones espaciales media móvil SARMA (1,1)

$$y = \rho W_1 y + X \beta_1 + u$$

$$u = \theta W_2 \varepsilon + \varepsilon$$

$$\varepsilon \approx N(0, \sigma^2 I)$$

- Modelo de componentes del error espacial

$$y = X \beta + \varepsilon$$

$$\varepsilon = W v + \psi$$

donde  $v$  y  $\psi$  son dos vectores estocásticos independientes que recogen para cada región  $i$  el shock específico regional ( $\psi$ ) y una combinación lineal de shocks de otras regiones que afectan a  $i$  ( $v$ ).  $W$  es la matriz de pesos espaciales.

### De órdenes superiores

- Estructura autorregresiva SAR

Modelo de regresión lineal con perturbaciones espaciales autorregresivas de orden 2.

$$y = X\beta_1 + u$$

$$u = \lambda_1 W_1 u + \lambda_2 W_2 u + \varepsilon$$

$$\varepsilon \approx N(0, \sigma^2 I)$$

Donde  $W_1$  y  $W_2$  son matrices de contigüidad de primer y segundo orden respectivamente.

- Estructura media móvil SMA

Modelo mixto regresivo-autorregresivo espacial con perturbaciones espaciales media móvil SARMA ( $p, q$ )

$$y = \rho_1 W_1 y + \rho_2 W_2 y + \rho_3 W_3 y + \dots + \rho_p W_p y + X\beta_1 + u$$

$$u = \theta_1 W_1 \varepsilon + \theta_2 W_2 \varepsilon + \theta_3 W_3 \varepsilon + \dots + \theta_q W_q \varepsilon + \varepsilon$$

$$\varepsilon \approx N(0, \sigma^2 I)$$

El modelo de componentes del error anteriormente expuesto recoge en el término de error dos vectores que modelizan externalidades (spillovers) espaciales. Anselin (2001) propone una clasificación de los modelos espaciales de regresión que recojan estas externalidades, distinguiendo entre aquellos que recogen la dependencia global y los que recogen la local. La forma en la que los modelos incorporan los spillovers, es la de

multiplicadores espaciales, utilizando la versión reducida que se obtiene con la transformación del modelo estructural. La clasificación y los modelos serían los siguientes:

1. Modelos de dependencia espacial GLOBAL

Se utilizan cuando una externalidad tiene lugar en la variable  $y$  de la región  $i$  y esto hace que se extienda al resto de regiones, recogiendo en el modelo como un multiplicador espacial. Se produce en modelos que siguen un proceso autorregresivo espacial (SAR) y donde la  $Cov(y_i, y_j) \neq 0$ , al transmitirse el impacto en  $i$  a todas las regiones.

En la siguiente tabla se presentan distintos modelos de dependencia espacial global en función de que las externalidades globales se produzcan en la perturbación aleatoria, las variables explicativas o en ambas a la vez.

**Cuadro 2.2. Modelos de externalidades espaciales globales.**

	FORMA ESTRUCTURAL	FORMA REDUCIDA
u	$y = \lambda Wy + X\beta - \lambda WX\beta + \varepsilon$	$y = X\beta + [I - \lambda W]^{-1}\varepsilon$
X	$y = \rho Wy + X\beta + u - \rho Wu$	$y = [I - \rho W]^{-1}X\beta + u$
Ambos	$y = \rho W_1 y + \lambda W_2 y - \rho \lambda W_1 W_2 y + X\beta - \lambda W_2 X\beta + u - \rho W_1 u$ $y = (\rho + \lambda)Wy - \rho \lambda W^2 y + X\beta - \lambda WX\beta + \varepsilon - \rho W\varepsilon$ $y = \rho Wy + X\beta + u$	$y = [I - \rho W_1]^{-1}X\beta + [I - \lambda W_2]^{-1}u$ $y = [I - \rho W]^{-1}X\beta + [I - \rho W]^{-1}u$

Fuente: Chasco (2003).

2. Modelos de dependencia espacial LOCAL

Cuando el impacto que se produce en la región  $i$  solo tiene influencia en las regiones vecinas, al menos hasta un grado limitado, y no en todas las demás, la dependencia es local. En este caso los modelos propuestos incorporan dependencia espacial en las variables exógenas y/o en el término de perturbación

aleatoria, siguiendo procesos de medias móviles. Se presentan en el siguiente cuadro resumen.

**Cuadro 2.3. Modelos de externalidades espaciales locales.**

	FORMA ESTRUCTURAL	FORMA REDUCIDA
u	$y = X\beta + \varepsilon + \theta W\varepsilon$	$y = X\beta + (I + \theta W)\varepsilon$
X	$y = X\beta_1 + WX\beta_2 + u$ $y = X\beta + \rho WX + u$	
Ambos	$y = X\beta_1 + W_1 X\beta_2 + \varepsilon + \theta W_2 \varepsilon$ $y = X\beta + \rho WX + \varepsilon + \rho W \varepsilon$	$y = X\beta_1 + W_1 X\beta_2 + (I + \theta W_2) \varepsilon$ $y = (I + \rho W)(X\beta + \varepsilon)$

Fuente: Chasco (2003).

La modelización mediante procesos autorregresivos para las externalidades espaciales globales, y de procesos de medias móviles para las externalidades espaciales locales, se puede ilustrar a partir de un modelo SAR (1), en el primer caso, y de un SMA (1) en el segundo, de la forma siguiente:

1. Modelo de regresión espacial con perturbación aleatoria autorregresiva de orden 1

En su forma estructural su expresión es

$$u = \lambda Wu + \varepsilon$$

$$\varepsilon \approx N(0, \sigma^2 I)$$

transformado a su forma reducida se expresa como

$$u = [I - \lambda W]^{-1} \varepsilon$$

siendo  $u = [I - \lambda W]^{-1} \varepsilon$  el multiplicador espacial asociado al proceso autorregresivo.

La matriz de covarianzas sería:

$$\text{Cov} (uu') = E (uu') = E \{ [I - \lambda W]^{-1} \varepsilon \varepsilon' [I - \lambda W]^{-1} \} = \sigma^2 \{ [I - \lambda W]^{-1} [I - \lambda W]^{-1} \}$$

La matriz inversa  $[I - \lambda W]^{-1}$  se puede desarrollar según la expresión de Leontieff del modo siguiente:

$$[I - \lambda W]^{-1} = I + \lambda W + \lambda^2 W^2 + \dots$$

quedando por tanto:

$$\text{Cov} (uu') = I + \lambda W + \lambda W' + \lambda^2 (W^2 + WW' + W^2) + \dots$$

Esta expresión muestra como las sucesivas potencias “expanden” el impacto producido al resto de regiones, ya que la matriz de covarianzas es una matriz completa no nula, recogiendo por tanto los fenómenos de dependencia espacial global. No ocurre lo mismo con procesos de medias móviles, propios de la dependencia local, como se ve a continuación en el caso 2

2. Modelo de regresión espacial con perturbación aleatoria de medias móviles de orden uno.

$$u = \varepsilon + \theta W\varepsilon$$

$$\varepsilon \approx N(0, \sigma^2 I)$$

Esta expresión es ya una forma reducida por lo que no tiene matrices inversas como ocurría en los procesos autorregresivos. La matriz de covarianzas es:

$$\begin{aligned} \text{Cov} (uu') &= E \{ (I + \theta W) \varepsilon \varepsilon' (I + \theta W)' \} = \sigma^2 \{ (I + \theta W) (I + \theta W)' \} = \\ &= \sigma^2 [I + \theta (W + W') + \theta^2 WW'] \end{aligned}$$

los únicos elementos no nulos en esta matriz son aquellos que lo son en  $W$ ,  $W'$  y  $WW'$ , por lo que si, por ejemplo,  $W$  es una matriz de contigüidad de primer orden, solo serán no nulos los elementos, y por tanto correlacionadas las regiones, que tengan relación de vecindad de primer ( $W, W'$ ) y segundo orden ( $WW'$ ). No se expande el efecto al resto de regiones como ocurre en los procesos autorregresivos, por la no existencia de matrices inversas.

Hasta aquí se han recogido los distintos modelos y clasificaciones de los mismos, según distintos autores, para recoger el fenómeno de autocorrelación espacial en modelos de regresión lineal. A continuación se presentan los que recogen e incorporan la heterogeneidad espacial. Anselin (2001) distingue entre los modelos de regresión lineal con heterogeneidad espacial en los que esta se manifiesta como heteroscedasticidad o aquellos en los que se manifiesta como inestabilidad estructural paramétrica.

Según Chasco (2003) “las causas de la existencia de heterogeneidad espacial en un modelo de regresión serían las siguientes:

- *Utilización de datos procedentes de unidades espaciales irregulares, es decir, con diferente área o extensión territorial, como es el caso de las divisiones político-administrativas (países, regiones, provincias, municipios, secciones censales,...).*

- *Tratamiento de unidades geográficas en las que un fenómeno se distribuye de manera desigual en el espacio, sobre todo, cuando se utilizan datos de regiones extremas (centro-periferia, norte-sur, este-oeste) o cuando se trabaja con datos referidos tanto a antiguas áreas metropolitanas como zonas de nuevo asentamiento, provincias urbanas y provincias rurales, secciones censales del centro de una ciudad y secciones del extrarradio, etc. (Anselin, 2001).*

- *Causas de tipo sociológico, como la existencia de diversos gustos o actitudes de la población, o político, cuando en la zona analizada se producen diferentes administraciones o políticas regionales que llevan a respuestas diferentes ante un mismo estímulo (Moreno y Vayá, 2000).*

- *Además, a estas situaciones hay que añadir las causas habituales del problema de heteroscedasticidad en los modelos de regresión lineal: omisión de variables relevantes u otro tipo de especificación errónea del modelo, que producen en el término de la perturbación aleatoria una varianza no constante.*

- *Por último, cabe señalar que algunas causas que provocan la heterogeneidad espacial pueden también originar la aparición de autocorrelación espacial, (especificaciones erróneas o errores de medida, sobre todo), siendo necesaria la contrastación de ambos efectos conjuntamente (Anselin, 1988).”*

## 1. Heterogeneidad espacial en los casos en que esta se manifiesta como heterocedasticidad

En estos casos la  $\text{Var}(u_i) = \sigma_i^2$  donde el subíndice  $i$  de  $\sigma_i^2$  indica que la varianza de la perturbación aleatoria es diferente para cada observación muestral ( $i$ ).

Si en la expresión general anteriormente descrita como más común, que sería la del modelo mixto regresivo de regresión espacial, con perturbaciones aleatorias autorregresivas y heteroscedásticas, eliminamos la parte que recoge la autocorrelación espacial, tendríamos el modelo del error heterocedástico.

La especificación del primero sería:

$$y = \rho W_1 y + X\beta_1 + W_2 R\beta_2 + u$$

$$u = \lambda W_3 u + \varepsilon$$

$$\varepsilon \approx N(0, \Omega); \Omega_{ii} = h_i(Z\alpha); h_i \geq 0$$

eliminando la autocorrelación espacial quedaría el mencionado de error heteroscedástico:

$$y = X\beta + u$$

$$\text{Var}(u_i) = \sigma_i^2$$

$$E(uu') = \Omega$$

Dependiendo de la especificación en este modelo de la varianza no constante se pueden desarrollar y distinguir tres modelos distintos, que serían (Chasco, 2003):

- Modelo de heterocedasticidad aditiva.

La varianza de la perturbación aleatoria sería función lineal de  $Z$  variables heteroscedásticas, que no tienen por qué coincidir con las variables explicativas, aunque podría. Su expresión sería:

$$\text{Var}(u) = Z\gamma$$

Donde  $\text{Var}(u)$  es un vector columna de varianzas de la perturbación aleatoria ( $N \times 1$ )

$Z$  es una matriz ( $N \times P$ ) de variables heterocedásticas (en columnas)

$\gamma$  es el vector de coeficientes.

- Modelo de heteroscedasticidad de coeficientes aleatorios.

$$\text{Var}(u) = Z^2 \gamma$$

En este caso la varianza es función del cuadrado de las variables heterocedásticas.

- Modelo de heteroscedasticidad de grupos (groupwise).

En este caso existen estructuras espaciales en los que cambia la varianza pero esta se mantiene constante dentro del grupo. Se estimaría a partir de los residuos de cada estructura espacial. Se da en casos de relaciones tipo centro-periferia, norte- sur, etc.

## 2. Heterogeneidad espacial en los casos en los que esta se manifiesta como inestabilidad estructural paramétrica

La inestabilidad estructural paramétrica implica que puede darse una variación, tanto en la forma funcional, como en los parámetros de una regresión, dependiendo de la localización geográfica, no siendo así homogéneos en toda la muestra de datos.

Según Moreno y Vayá (2000) la especificación en esta situación sería:

$$y_i = f_i(X_i \beta_i + \varepsilon_i)$$

donde  $i$  se refiere a los  $N$  puntos del espacio geográfico.  $i = 1, 2, \dots, N$ .

$y_i$  es la variable dependiente en la localización  $i$ .

$X_i$  es un vector ( $1 \times K$ ) de  $K$  variables explicativas.

$\beta_i$  es un vector ( $K, 1$ ) de parámetros asociados a las variables explicativas.

$\varepsilon_i$  es la perturbación aleatoria.

$f_i ( )$  expresa la relación funcional que cambia en función de la localización  $i$ .

En esta especificación se produciría el problema de los parámetros incidentales, que consiste en que tendría insuficientes observaciones en la muestra de tamaño  $N$  para poder estimar  $N$  vectores de parámetros  $\beta_i$ .

La inestabilidad paramétrica puede ser discreta o continua, lo que permite clasificar los modelos de la forma siguiente:

### **Modelos con inestabilidad paramétrica continua**

La variación espacial es continua en toda la muestra de datos, es decir se da en todas las localizaciones o variables. Los coeficientes varían de forma continua en el espacio.

- Modelo espacial de parámetros aleatorios, de Hildreth-Houck

$$y_i = X_i \beta_i$$

donde  $X_i$  es un vector fila ( $1 \times K$ ) de variables explicativas

$\beta_i$  es un vector columna ( $K \times 1$ ) de coeficientes para cada observación  $i$ .

En este modelo no se conoce la forma funcional de la variación paramétrica, por lo que el vector  $\beta_i$  se estima como la suma de dos elementos

$$\beta_i = \beta + \mu_i$$

donde  $\beta$  es una constante y  $\mu_i$  es la perturbación aleatoria de distribución normal, con media nula y matriz de covarianzas que puede ser diagonal siempre que no exista correlación entre los términos de error y los coeficientes de regresión del modelo.

Sustituyendo la expresión general quedaría así:

$$y_i = X_i \beta + u_i$$

$$u_i = X_i' \mu_i$$

donde  $E(u_i) = 0$  y  $V(u_i) = X_i' \sum_i X_i$ , es decir la varianza del error aleatorio es función de las variables exógenas.

- Modelo de expansión espacial, de Casetti

Se consideran unas variables de expansión (z) de forma que los coeficientes  $\beta_j$  son función de estas variables. En el caso general se considera una función lineal con m variables de expansión,  $z_1, \dots, z_m$ , de la forma siguiente:

$$\beta_j = \gamma_{0j} + \gamma_{1j} z_1 + \dots + \gamma_{mj} z_m$$

- Modelo de superficie tendencial

Las variables explicativas de este modelo de regresión lineal son los elementos de un polinomio de las coordenadas terrestres de longitud (m) y latitud (p) correspondientes a la localización de una observación i en el espacio ( $m_i, p_i$ ). La forma de un modelo de superficie tendencial de 2º orden sería:

$$y = \alpha + \beta_1 m + \beta_2 p + \beta_3 m^2 + \beta_4 p^2 + \beta_6 m p$$

- Modelo de regresiones ponderadas geográficamente (RPG), de Fotheringham, Charlton y Brundson

Se trata de un modelo estimado por mínimos cuadrados ponderados donde los pesos son función de la distancia entre cada punto y el resto.

$$y_i = X_i \beta_i + u_i$$

donde el vector de parámetros ( $\beta_i$ ) se obtiene a través de un método de estimación de mínimos cuadrados ponderados,

$$\beta_i = [X' W_i X]^{-1} X' W_i y$$

$W_i$  es una matriz diagonal de orden (N x N) que de forma distinta a la que se suele usar con el método de mínimos cuadrados ponderados, es diferente en cada localización i. Solo son elementos no nulos los de la diagonal principal en la que se sitúan las ponderaciones  $w_{ij}$  obtenidas como función de la distancia entre i y el resto.

## Modelos con inestabilidad paramétrica discreta

La variación espacial se plantea entre un número de grupos de tamaño menor a la muestra, evitando así el problema de parámetros incidentales o falta de grados de libertad. Los coeficientes varían entre grupos, permaneciendo constantes dentro de ellos.

- Modelo ANOVA espacial (SANOVA)

Es la aplicación espacial del modelo ANOVA tradicional. Se contrasta si el grado de variación entre una variable y la media de dicha variable en un grupo de datos es significativo o casual.

$$y = \beta_0 + \beta_1 f + u$$

$$u \approx N(0, \sigma^2 I)$$

donde  $y$  es una variable dependiente sometida a diferentes tratamientos espaciales

$f$  es una variable categórica que define los diferentes tratamientos (estructuras)

$\beta_0$  y  $\beta_1$  son coeficientes a estimar

$u$  es la perturbación aleatoria esférica.

- Modelo espacial de regresiones cambiantes (“switching regressions”)

Se estiman tantos valores diferentes como estructuras o regiones cambiantes se hayan definido. Par el caso de que sean dos, siguiendo el criterio de una variable ficticia

$$y_1 = X_1 \beta_1 + u_1 \quad d = 0$$

$$y_2 = X_2 \beta_2 + u_2 \quad d = 1$$

## 2.4.2. ESTIMACIÓN Y CONTRASTES

La estimación de un modelo espacial de regresión lineal suele comenzar con la estimación del modelo básico de regresión lineal, sin efectos espaciales, por el método habitual de mínimos cuadrados ordinarios. Además de realizar los contrastes habituales en modelos sin efectos espaciales, para verificar el cumplimiento de las propiedades deseadas, se realiza, también, un contraste sobre la perturbación aleatoria, para verificar la existencia o no de autocorrelación y heterogeneidad espacial. Si no se puede rechazar la ausencia de estos efectos espaciales, se debe empezar un proceso que incluya, tanto nuevas especificaciones, que recojan apropiadamente el problema espacial, como la utilización de nuevos métodos de estimación, que proporcionen estimadores con buenas propiedades estadísticas.

La estimación por MCO en presencia de autocorrelación espacial hace que se pierdan las propiedades estadísticas deseadas de los estimadores, por lo que no es un método de estimación deseable si no podemos rechazar, desde un primer momento, la ausencia de dependencia y heterogeneidad espaciales.

En el caso de existencia de dependencia espacial residual los estimadores MCO, aunque siguen siendo insesgados, son ineficientes. Si lo que se produce es autocorrelación espacial por retardo de la variable endógena los estimadores MCO serían sesgados e inconsistentes (Moreno y Vayá, 2002).

Sobre los métodos de estimación Pérez Pineda (2006) dice que *“en concreto y de forma general se consideran dos métodos de estimación, el primero y más utilizado es el método de máxima verosimilitud (MV) y el segundo, el método de variables instrumentales (VI). Como procedimiento alternativo, para muestras grandes y para tratar la dependencia espacial simultánea también se pueden usar métodos condicionales como el coding method basado en MCO sobre un subconjunto de observaciones no contiguas... Otros métodos menos usados son: estimación bayesiana (LeSage, 1999) y métodos de estimación robustos basados en la técnica de bootstrap o jacknife (Anselin, 1988)”*.

Aunque a grandes rasgos los métodos de estimación más apropiados serían los mencionados de máxima verosimilitud y variables instrumentales, además de que puede

variar su conveniencia en cada caso concreto, esta va a depender del tipo de relación espacial que se presente. La mayoría de autores recomiendan unos u otros métodos en función de si se da un problema de heterogeneidad espacial o si se da una dependencia espacial sustantiva, es decir modelos de retardo espacial, o si se trata de dependencia espacial residual, donde corresponde hablar de modelos de error espacial, o las posibles combinaciones de los mismos. Además dependiendo de si se está modelizando de una forma u otra será conveniente aplicar distintos contrastes. Estos se pueden agrupar en los que se aplican al modelo básico de regresión lineal estimado por mínimos cuadrados ordinarios, para ver si existe o no heterogeneidad y/o dependencia espacial, y los que se aplican posteriormente, una vez no se acepta la no existencia de autocorrelación espacial, heterogeneidad espacial o ambas a la vez. En este segundo grupo serían también distintos los contrastes a aplicar según se diese el caso de autocorrelación espacial o el caso de heterogeneidad espacial. Además, dentro de estos dos casos, se debe distinguir entre autocorrelación espacial sustantiva y autocorrelación espacial residual, y heterogeneidad espacial por heteroscedasticidad o heterogeneidad espacial por inestabilidad paramétrica, y esta última, a su vez, puede ser discreta o continua. Los distintos modelos posibles en cada caso anterior sirven de base a algunos autores para explicar y recomendar unos u otros contrastes y métodos de estimación.

Según **PÉREZ PINEDA (2006)** *“siguiendo las estructuras en que se puede especificar la autocorrelación en los modelos de regresión (lag o error), existen una serie de estadísticos propuestos para la dependencia espacial sustantiva y residual. Originalmente en Anselin (1999) se sugiere sólo 3 estadísticos: un estadístico global de autocorrelación, Moran I, una para estructuras de retardo LM-LAG y una para estructuras de error LM-ERR. Sin embargo en Vaya y Moreno (2000) se presentan en mayor profundidad estos estadísticos diferenciando entre los de dependencia sustantiva y residual.”*

**Cuadro 2.4. Contrastes de dependencia espacial sustantiva en modelos de regresión**

Contrastes de dependencia espacial sustantiva	Características
<p>1. Test LM-LAG:</p> $LM-LAG = [e' Wy / S^2] / RJ_{\rho-\beta}$	<p>Donde <math>e</math> es el vector de residuos MCO del modelo lineal comentado, <math>Wy</math> el termino espacial retardado, <math>S^2</math> la estimación de la varianza residual del mismo modelo y <math>RJ_{\rho-\beta} = [T1+(WX\beta)'M(WX\beta)/S^2]</math></p> <p>Este test se distribuye como una <math>\chi^2</math> con un grado de libertad. Se ha impuesto <math>\lambda=\alpha\rho=0</math> con <math>H_0= \rho=0</math></p>
<p>2. Test LM-LE</p> $LM-LE = [ (e' Wy / S^2 - e' We / S^2) / (RJ_{\rho-\beta} - T1) ]^2$	<p>Con el mismo significado en los elementos definidos en el estadístico anterior y con <math>T1= Traza (W'W+W^2)</math>.</p> <p>Se distribuye como una <math>\chi^2</math> con un grado de libertad. A diferencia del test anterior, este es robusto ante posibles especificaciones erróneas locales, como la existencia de un término de perturbación correlacionado espacialmente.</p>

Fuente Pérez Pineda (2006) a partir de Vaya y Moreno (2000).

**Cuadro 2.5. Contrastes de dependencia espacial residual en modelos de regresión**

Contrastes de dependencia espacial residual	Características
<p>1. I de Moran</p> $I = (N/S)(e' We / e'e)$	<p>Donde <math>e</math> es el vector de residuos de MCO de un modelo lineal, <math>N</math> el tamaño muestral y <math>S = \sum_i \sum_j w_{ij}</math> (la suma de los elementos <math>w_{ij}</math> de la matriz de pesos). En un contexto asintótico y el supuesto de residuos incorrelacionados, la I de moran estandarizada se distribuye como una normal estándar.</p>
<p>2. Test K-R (Kelejian- Robinson)</p> $K-R = (\gamma'Z'Z\gamma) / (\alpha'\alpha / h_R)$	<p>Expresión derivada de una regresión auxiliar (cuya variable dependiente es <math>C=e_i e_i</math>) en que se usan los productos cruzados de los residuos de las observaciones, potencialmente correlacionadas espacialmente (según la matriz de contactos) y los productos cruzados de las variables explicativas de tales observaciones. Las variables explicativas de la regresión auxiliar <math>Z_n</math>, están formadas por productos cruzados <math>X_i</math> y <math>X_j</math>. <math>\gamma</math> es el vector de coeficientes obtenido de MCO en una regresión de <math>C</math> en <math>Z</math>, y <math>\alpha</math> el vector asociado a los residuos.</p>
<p>3. Test LM-ERR</p> $LM-ERR = [ (e'We / S^2) / T1 ]^2$	<p>Donde <math>e</math> es el vector de residuos de MCO, <math>T1= Traza (W'W+W^2)</math>, <math>S^2</math> la estimación de la varianza residual de dicho modelo. A diferencia de los dos estimadores anteriores este se distribuye como una <math>\chi^2</math> con un grado de libertad.</p>
<p>4. Test LM-EL</p> $LM-EL = [ (e'We / S^2) - T1(Rj_{\rho-\beta})^{-1} (e'Wy / S^2) ]^2 / [ T1 - T1^2 (Rj_{\rho-\beta})^{-1} ]$	<p>Donde <math>M</math> es una matriz idempotente y <math>Rj_{\rho-\beta} = [T1+(WX\beta)'M(WX\beta)/S^2]</math>. <math>e</math> es el vector de residuos MCO del modelo lineal comentado, <math>Wy</math> el termino espacial retardado, <math>S^2</math> la estimación de la varianza residual y <math>T1= Traza (W'W+W^2)</math></p>

Fuente Pérez Pineda (2006) a partir de Vaya y Moreno (2000).

También se plantea “*un contraste que evalúa la presencia simultánea de autocorrelación sustantiva y residual, basado en el Multiplicador de Lagrange, que para un modelo mixto regresivo espacial autorregresivo con perturbaciones espaciales, con estructura SARMA (1,1), contrastando simultáneamente para  $\rho$  y  $\theta$ , bajo  $H_0 = \rho = \theta = 0$ , plantea la siguiente estructura:*

$$\text{SARMA} = [(e'Wy / S^2 - e'We / S^2) / (R_{j\rho-\beta} - T1)]^2 + [(e'We / S^2) / T1]^2$$

*Este test, conocido también como test multidireccional SARMA supone  $W_1 = W_2 = W$ , donde  $e$  es el vector de residuos MCO del modelo bajo  $H_0$  y se distribuye bajo una  $\chi^2$  con dos grados de libertad. Por sus características dicho test es obtenible también como  $\text{SARMA} = \text{LM-ERR} + \text{LM-LE} = \text{LM-LAG} + \text{LM-EL}$ .*

También Pérez Pineda (2006) propone una metodología de estimación que se basaba en el programa informático SpaceStat (al que se puede considerar precursor del actual GeoDa, que lo mejora): “*el procedimiento que seguiremos parte de una primera estimación de MCO con diagnósticos espaciales, de la cual esperamos encontrar autocorrelación espacial, con ello podemos reespecificar nuestro modelo y utilizar otra estructura que recoja mejor la dependencia encontrada. Las principales opciones serán modelos tipo LAG o tipo ERROR, las cuales se suelen tratar por alguno de los siguientes métodos: Máxima Verosimilitud, Variables Instrumentales, o VI robustas, como principales estructuras modelizables, basados en regresiones genéricas como método de estimación estándar, a menos que se requiera otro.*”

En el cuadro 2.6 se resume su propuesta.

Cuadro 2.6.

MODELO	MÉTODO DE ESTIMACIÓN	TEST DE CONTRASTE		
		MEDIDAS DE AJUSTE	HETEROSCEDASTICIDAD	DEPENDENCIA ESPACIAL
MODELO CLÁSICO	MÍNIMOS CUADRADOS ORDINARIOS	<p>- <math>R^2</math></p> <p>- <math>R^2</math> ajustada</p> <p>Además se requiere NORMALIDAD y para su contraste se sugiere el estadístico Jarque-Vera. Para la detección de posible MULTICOLINEALIDAD se propone la Utilización combinada <math>R^2</math> y T-Student así como el número de condición de Belsley.</p>	<p>- ML de Breusch-Pagan</p> <p>- Koenker-Basset</p>	<p>I de Moran, LM-ERR, K-R y LM-LAG, los tres primeros para el caso de dependencia espacial residual y el último para dependencia espacial sustantiva.</p>
RETARDO ESPACIAL	MÁXIMA VEROSIMILITUD	<p><math>R^2</math> no es aplicable por lo que se proponen dos medidas más también llamadas pseudo <math>R^2</math>: la primera el ratio simple de la varianza de valores predichos sobre la varianza de los valores observados de la variable dependiente (como <math>R^2</math>), la segunda es la correlación al cuadrado entre los valores predichos y observados, (como Sq. Corr). Sin embargo las medidas más adecuadas para estos modelos son las basadas en la función de verosimilitud, entre las que se encuentran: el Log likelihood, el criterio de información de Akaike y el criterio de Schwartz. El modelo con el mayor Log Likelihood, o con los AIC o SC más bajos serán los mejores.</p>	<p>- Breusch-Pagan espacial basado en los residuos de la estimación de Máxima Verosimilitud</p> <p>- BP basado en la regresión estándar del modelo</p>	<p>Test LR para el coeficiente autoregresivo espacial <math>\rho</math>. Adicionalmente al test LR para modelos con retardos, se proponen otros dos que son asintóticamente equivalentes, en concreto el test Wald (W) y el LM-LAG para retardos espaciales en el modelo estándar de regresión. Si la especificación de la dependencia espacial de los retardos es la adecuada no deberá haber dependencia en los residuos, para ello se usa el test LM-ERR.</p>

	VARIABLES INSTRUMENTALES		De forma similar al caso anterior (estimación bajo MV) la $R^2$ no es útil y se utilizan los mismos dos estadísticos del coeficiente de determinación señalados: $R^2$ y Sq. Corr.	Se considera similar al método anterior para la heteroscedasticidad y dependencia espacial, en este último con tests LM para este caso.	
ERROR ESPACIAL	MÁXIMA VEROSIMILITUD		Nuevamente la $R^2$ no es de utilidad. Se proponen tres pseudo medidas de $R^2$ , las dos primeras las mismas que en el modelo de retardos, la $R^2$ y el Sq. Corr.. El tercero es una aplicación de ajustes sugeridos por Buse (1973) para modelos espaciales y que se encuentra como $R^2$ (Buse). Adicionalmente se proporcionan tres medidas de ajuste, basadas en el principio de MV: el Logaritmo de la función de verosimilitud (LIK), el AIC y el SC antes comentados, medidas comparables con las de anteriores modelos (el de regresión estándar y el de retardo). Al comparar estas medidas basadas en el <i>criterio de información</i> , se preferirá el modelo de errores sobre el de retardo.	- BP no ajustado  - BP espacialmente ajustado.	Ratio de verosimilitud ( <i>likelihood Ratio</i> ) LR para el coeficiente espacial autorregresivo $\lambda$ .  Se distribuye como una $\chi^2$ con un grado de libertad.
	ERRORES HETEROSCEDÁSTICOS	MÍNIMOS CUADRADOS GENERALIZADOS FACTIBLES  MÉTODOS ITERATIVOS	Similar a los casos anteriores, la $R^2$ no es aplicable y se proponen las mismas tres medidas pseudo $R^2$ , la $R^2$ , la Sq. Corr. y la $R^2$ (Buse) así como las tres medidas basadas en el criterio de información derivadas del principio de MV: el logaritmo de la función de verosimilitud, AIC, y SC. Tales medidas sólo se usan para los estimadores de ML y no para los de FGLS.	- Test de Wald  - LR	Los diagnósticos de dependencia espacial se incluyen sólo para estimaciones de MV. Así el test que se incluye es el basado en Multiplicador de Lagrange para dependencia espacial en error de un modelo heteroscedástico (LM-ERR). El cual es asintótico y se distribuye como una $\chi^2$ con un grado de libertad.

Fuente: Elaboración propia a partir de Pérez Pineda (2006)

En SpaceStat, además de esta, desarrollada para el modelo general, se incluyen otras cuatro metodologías de estimación para casos especiales, que serían:

- MODELO DE TENDENCIA DE SUPERFICIE.
- MODELO DE REGÍMENES ESPACIALES.
- MODELO DE EXPANSIÓN ESPACIAL.
- MODELO ANOVA (ANÁLISIS DE LA VARIANZA).

Según **CHASCO (2003)** *“el análisis confirmatorio espacial puede llevarse a cabo a través de un proceso que tiene por primera fase la especificación del modelo propuesto como un modelo básico de regresión lineal (MBRL) sin efectos espaciales, que se estima por el método de mínimos cuadrados ordinarios (MCO). A continuación, sobre la perturbación aleatoria se aplican una serie de contrastes de autocorrelación y heterogeneidad espacial con el objeto de aceptar o rechazar la hipótesis nula de ausencia de efectos espaciales (aunque también deben contrastarse el resto de hipótesis básicas, comunes al contexto temporal clásico, como la normalidad de la perturbación aleatoria y no multicolinealidad). La aceptación de dicha hipótesis nula supondrá la aceptación del MBRL y sus estimadores MCO, pero el rechazo de la misma requerirá de nuevas especificaciones del modelo (en un proceso de “feedback”) que incluyan convenientemente los efectos espaciales, así como de otros métodos de estimación (máxima-verosimilitud, mínimos cuadrados espaciales en dos etapas, método de los momentos, etc.), que produzcan buenas propiedades en los estimadores.”*

*“Este proceso, propuesto por Anselin (1995 y 1999), tiene su punto de partida en la especificación de un modelo espacial de regresión lineal sin efectos espaciales (MBRL), que se estimará por MCO y sobre el que se contrastarán las hipótesis nulas de ausencia de autocorrelación y heterogeneidad espacial en el mismo. El rechazo de dichas hipótesis conducirá a nuevas especificaciones, como los modelos de dependencia espacial y/o modelos de heterogeneidad espacial, que habrán de ser estimados a través de procedimientos diferentes de los mínimos cuadrados ordinarios y sobre los que habrá que contrastar también la hipótesis nula de ausencia de efectos espaciales.”*

*“En el análisis de datos espaciales, es frecuente encontrar este problema (heteroscedasticidad), sobre todo cuando se utilizan datos procedentes de unidades*

*espaciales irregulares (con áreas diferentes); también se produce cuando hay diferencias regionales sistemáticas en las relaciones del modelo (por ejemplo, estructuras espaciales); o cuando hay una deriva espacial continua en los parámetros de modelo (es decir, expansión espacial), como se vio en el apartado anterior. La presencia de cualquiera de estos efectos espaciales invalidaría cualquier modelo de regresión tradicional que los ignorase. Por tanto, el indicio de heteroscedasticidad puede apuntar a la necesidad de incorporar más explícitamente los efectos espaciales, en forma de estructuras espaciales o expansión espacial de los parámetros.”*

En el primer paso de estimación del MBRL, propone la realización de los tests de heteroscedasticidad siguientes:

- **Test del multiplicador de Lagrange, desarrollado por Breusch y Pagan (1979).**
- **Test de Koenker-Bassett (1982).**
- **Test de White (1980).**

Los siguientes tests son los propuestos para contrastar la existencia o no de dependencia espacial en el MBRL. Los dos primeros se consideran contrastes “ad-hoc” y los siguientes se basan en el principio de máxima verosimilitud.

- **Test I de Moran**

$$I = (N/S_0) (\sum w_{ij} e_i e_j / \sum_{i=1}^n e_i^2) = (N/S) (e'We / e'e)$$

Donde:

- N es el número de observaciones
- $S_0$  es la suma de los pesos espaciales
- e es el vector de residuos MCO

La inferencia se basa en un valor z, estandarizado, que se distribuye asintóticamente como una normal. Es un contraste muy extendido pero tiene importantes problemas ya que puede recoger otro tipo de errores (mala especificación, no normalidad, heteroscedasticidad) e incluso la existencia de dependencia espacial sustantiva, que puede impedir detectar de forma separada la dependencia espacial residual.

- **Test K-R (Kelejian y Robinson (1992)).**

Partimos de:

$$C = Z\gamma + A$$

Donde:

- C está formada por los productos cruzados de los residuos MCO de las observaciones, siendo el término h-ésimo:  $c_h = e_i e_j$
- Z son los productos cruzados de las variables explicativas del MBRL, siendo el término h-ésimo:  $Z_h = X_i X_j$
- $\gamma$ : vector de coeficientes
- A: perturbación aleatoria de la regresión auxiliar.

Siendo  $\alpha$  el vector de residuos resultante la expresión del estadístico K-R sería:

$$K-R = (\gamma' Z' Z \gamma) / (\alpha' \alpha / h_R)$$

donde  $h_R$  es el número total de observaciones del vector auxiliar.

Este estadístico se distribuye según una  $\chi^2$  con K grados de libertad, siendo K el número de variables explicativas de X. No requiere, como hipótesis previa, normalidad en los términos de la perturbación aleatoria, y suele ser útil en grandes bases de datos y menos útil con muestras pequeñas. Al igual que la I de Moran, no presupone ningún esquema de autocorrelación espacial, por lo que no son útiles para determinar si la dependencia espacial es sustantiva o residual.

- **Test LM-ERR**

Basado en el principio de los multiplicadores de Lagrange, fue originalmente propuesto por Burridge (1980).

$$LM-ERR = [e'W'e / s^2]^2 / \text{tr}[W'W + W^2]$$

donde

- $s^2 = e'e / N$  es el estimador MV de la varianza de la perturbación aleatoria

- $e$  es el vector de residuos de la estimación MCO.

Se distribuye como una  $\chi^2$  con un grado de libertad.

Existe una relación entre este test y la I de Moran cuando la Matriz está estandarizada por filas (Moreno y Vayá (2000)) de la siguiente forma:

$$\text{LM-ERR} = [\text{NI}]^2 / \text{tr} [\text{W}'\text{W} + \text{W}^2]$$

Siendo I la I de Moran.

Anselin (1994) propone, además, un nuevo contraste LM-ERR (2) que permite contrastar la existencia de procesos autorregresivos o de medias móviles de orden 2. Su expresión sería:

$$\text{LM-ERR (2)} = [e' W_1 e / s^2]^2 / \text{tr} [W_1' W_1 + W_1^2] + [e' W_2 e / s^2]^2 / \text{tr} [W_2' W_2 + W_2^2]$$

Donde  $W_1$  y  $W_2$  son las matrices de pesos espaciales de orden 1 y 2.

Sigue una distribución  $\chi^2$  con 2 grados de libertad, al contrastar la significación de los dos coeficientes autorregresivos y medias móviles simultáneamente. El contraste LM-ERR (2) podría extenderse a órdenes superiores a 2 como la suma de los respectivos tests unidireccionales, añadiendo los sumandos correspondientes.

- **Test LM-EL**

*“Es un contraste de dependencia espacial de la perturbación aleatoria, similar al test LM-ERR, aunque robusto a una mala especificación local de la dependencia espacial como sería el caso de la existencia de una variable endógena espacialmente retardada (Anselin et al., 1996)” (Chasco, 2003).* La hipótesis nula a contrastar sería, en este caso,  $H_0 (\lambda=0, \theta=0)$ , robusto a  $\rho \neq 0$ ,

$$\text{LM-EL} = [ (e' W e / s^2) - T_1 (R \hat{J}_{\rho-\beta})^{-1} (e' W y / s^2) ]^2 / [T_1 - T_1^2 (R \hat{J}_{\rho-\beta})]$$

donde:

- $(R \hat{J}_{\rho-\beta})^{-1} = [ T_1 + [(WX\beta)' M (WX\beta)] / S^2 ]$
- $T_1 = \text{tr} (W'W + W^2)$

- $W$  es la matriz de pesos espaciales.
- $WX\beta$  es el retardo espacial de los valores procedentes de la estimación MCO de  $y$  sobre  $X$ .

Se distribuye como una  $\chi^2$  con 1 grado de libertad.

- **Test LM-LAG**

Propuesto por Anselin (1988) es para retardos espaciales de la variable endógena, se requiere hipótesis de normalidad y se distribuye como una variable  $\chi^2$  con 1 grado de libertad.

$$\text{LM-LAG} = [e' Wy / s^2]^2 / R \hat{J}_{\rho-\beta}$$

Siendo  $y$  el vector  $(N, 1)$  de observaciones de la variable dependiente

- **Test LM-LE**

Es el contraste de retardos, robusto a la presencia de dependencia local de error espacial.

$$\text{LM-LE} = [(e' W_1 y / s^2) - (e' W_1 e / s^2)]^2 / [R \hat{J}_{\rho-\beta} - T_1]$$

Se distribuye también como una  $\chi^2$  con 1 grado de libertad y ha sido construida bajo el supuesto de coincidencia entre las matrices de pesos espaciales correspondientes a la variable endógena y a la perturbación aleatoria.

- **Test SARMA multidireccional.**

Al igual que los cuatro anteriores se basa en el multiplicador de Lagrange y permite estimar de forma conjunta la existencia de autocorrelación espacial sustantiva y residual. La hipótesis nula sería  $H_0 (\rho=\theta=0)$  ó  $H_0 (\rho=\lambda=0)$ .

$$\text{SARMA} = [[(e' Wy / s^2) - (e' We / s^2)]^2 / [R \hat{J}_{\rho-\beta} - T_1]] + [(e' We / s^2)^2 / T_1]$$

Se distribuye como una  $\chi^2$  con 2 grados de libertad.

- **Test LM-LAG<sub>λ</sub>**

Se utiliza para contrastar la presencia de un retardo espacial en la variable endógena de un modelo en el que la perturbación aleatoria presenta un esquema autorregresivo espacial.

$$LM_p^A = [\hat{u}' B' B W_1 y]^2 / (H_p - H_{vp} \hat{v}ar[\hat{\theta}] H'_{vp})$$

Donde:

- $\hat{u}$  es el vector de residuos de la estimación máximo verosímil del modelo, obtenido bajo la hipótesis nula de no dependencia espacial sustantiva, aunque con errores espacialmente autorregresivos.
- $y = X\beta + (I - \lambda W_2)^{-1} u$
- $v = [\beta', \lambda, \sigma^2]$
- $B = I - \lambda W_2$
- $H_p = \text{tr } W_1^2 + \text{tr } [BW_1 B^{-1}]' [BW_1 B^{-1}] + (1/\sigma^2) [BW_1 X\beta]' [BW_1 X\beta^{-1}]$
- $H_{vp} = [ \text{fila 1 } (1/\sigma^2) [(BX)' BW_1 X\beta \text{ fila 2 } \text{tr}[ W_2 B^{-1}]' BW_1 B^{-1} + \text{tr } W_2 W_1 B^{-1} \text{ fila 3 } 0]$
- $\hat{v}ar[\hat{\theta}]$  es la matriz de varianzas estimadas del vector de parámetros  $\theta$  en el modelo estimado bajo la hipótesis nula.

Se distribuye como una  $\chi^2$  con 1 grado de libertad.

*“El rechazo de la hipótesis nula de ausencia de autocorrelación y heterogeneidad espacial en el MBRL, supone una re-especificación del mismo, a través de modelos que incorporen estos efectos en la forma funcional.” (Chasco, 2003).*

La autora distingue entre modelos de dependencia espacial y modelos de heterogeneidad espacial. En el cuadro siguiente (2.7) se resume su propuesta incluyendo los métodos de estimación y los estadísticos recomendados en cada caso.

Cuadro 2.7.

MODELOS		METODOS DE ESTIMACIÓN		TEST- ESTADÍSTICOS
DEPENDENCIA ESPACIAL	RETARDO ESPACIAL	MV		<p>“El método MV dispone de un menor número de contrastes sobre los efectos espaciales que el método MCO, siendo los más utilizados el test del multiplicador de Lagrange (LM) y el test del cociente de verosimilitud (LR) que, por ser asintóticos, pueden dar lugar a conclusiones inconsistentes en el caso de muestras pequeñas. El contraste de heteroscedasticidad en el modelo del retardo espacial puede realizarse a través del test de Breusch-Pagan, basado en los residuos de la estimación MV. Aunque de forma estricta este contraste es incorrecto, dado que ignora la presencia de dependencia espacial en el modelo, en la práctica suele utilizarse habitualmente sin muchos problemas. En cuanto a los contrastes de dependencia espacial, el test LR del cociente de verosimilitud sobre el coeficiente autorregresivo (<math>\rho</math>) se calcula como el doble de la diferencia entre el logaritmo de verosimilitud del modelo de retardos espaciales y el logaritmo de verosimilitud en el MBRL, con el mismo conjunto de variables explicativas (para <math>\rho = 0</math>). Este estadístico se distribuye como una variable <math>\chi^2</math> con 1 grado de libertad. A pesar de que tanto el test LR, como el contraste de Wald (test t asintótico) y LMLAG son todos equivalentes asintóticamente, tienden a ofrecer resultados distintos en muestras finitas. En la mayoría de las circunstancias, el orden de dichos estadísticos en términos de su magnitud es la siguiente: <math>W \geq LR \geq LM</math>. Si el modelo de retardos espaciales especificado la especificación correcta, entonces no debería quedar en los residuos de la regresión ningún tipo de dependencia espacial. El test del multiplicador de Lagrange para la autocorrelación espacial de los errores LM-ERR propuesto por Anselin (1988) contrasta esta cuestión.”</p>
		VI. No se exige la hipótesis de normalidad en la distribución de la perturbación aleatoria	MÍNIMOS CUADRADOS BIETÁPICOS	<p>“Dado que se trata de métodos robustos, no se ofrece ningún tipo de contraste”.</p>

			BOOTSTRAP	“Dado que se trata de métodos robustos, no se ofrece ningún tipo de contraste”.
	ERROR ESPACIAL	Si el coeficiente fuera conocido MCG y en caso contrario MV.		“El test de heteroscedasticidad de Breush-Pagan (del multiplicador de Lagrange) y el contraste del cociente de verosimilitud (LR) para el coeficiente espacial autorregresivo ( $\lambda$ ). Además, existe también un test LR y otro de Wald para la hipótesis del factor común, que es un contraste sobre la consistencia interna de la especificación del error espacial”.
HETEROGENEIDAD ESPACIAL	ERROR HETEROCEDÁSTICO	-MÍNIMOS CUADRADOS GENERALIZADOS FACTIBLES  - MV		“El contraste de heteroscedasticidad en este modelo, puede realizarse a través del test de Wald (W) y el cociente de verosimilitud (LR). En cuanto a la dependencia espacial, este efecto puede ser estimado a través de los contrastes LM-ERR (dependencia residual) y LM-LAG (dependencia sustantiva).”
	INESTABILIDAD PARAMÉTRICA CONTINUA	MODELO DE SUPERFICIE TENDENCIAL	“Este modelo, expresado únicamente como función de las coordenadas, puede ser estimado por MCO. Pero, en este tipo de especificación, podrían implementarse también otras formulaciones, como el modelo del retardo espacial, error espacial o error heteroscedástico, debiendo emplearse, en estos casos, los métodos de estimación adecuados anteriormente expuestos”.	
		MODELO DE PARÁMETROS ALEATORIOS	MCO MCO-robustos	Test F de inestabilidad  W de Wald
	REGRESIONES CAMBIANTES	MCO (otros, según especificación)		Chow espacial

Fuente: Elaboración propia a partir de Chasco (2003)

La propia autora presenta un cuadro resumen de todos los métodos de estimación y contrastes que propone.

**Cuadro 2.8. Métodos de estimación y contrastes de efectos espaciales**

MODELO	MÉTODO DE ESTIMACIÓN	CONTRASTES DE DEPENDENCIA ESPACIAL		CONTRASTES HETEROSCEDASTICIDAD
		RETARDO ESPACIAL	RESIDUOS	
MBRL	MCO	LM-LAG	I de Moran Kelejia Robinson (K-R)	Breusch-Pagan (BP) Koenker-Bassett(KB)
		LM-LE	LM-ERR LM-EL	
		SARMA LM-LAG $\lambda$		
Modelo del retardo espacial	MV	W de Wald	LM-ERR	Breusch-Pagan (BP)
	VI-MC2E	LR		
	VI-BOOTSTRAP	LM-LAG		
Modelo del error espacial	MCG	W de Wald	LM-ERR	Breusch-Pagan (BP)
	MV	LR		
Modelo del error heteroscedástico	MCO-robustos MCGF	LM-LAG	LM-ERR	W de Wald
	MV			LR
Modelo de superficie tendencial	MCO (otros, según especificación)			
Modelo de parámetros aleatorios	MCO			Test F de inestabilidad
	MCO-robustos			W de Wald
Modelo de regresiones cambiantes	MCO (otros, según especificación)			Chow espacial

Fuente: Chasco (2003).

### **Anselin y Rey (2014)**

A las clasificaciones sobre test y métodos de estimación “clásicos” anteriormente descritos conviene añadir la propuesta de Anselin y Rey (2014), más actualizada ya que incluye las técnicas en las que se basan algunos de los softwares y programas informáticos más recientes y avanzados, que además son los que vamos a utilizar para esta tesis en el estudio del caso práctico que desarrollamos en los siguientes apartados.

Su propuesta de métodos de estimación, aunque pensado para los programas de código abierto que desarrolla GeoDaCenter, “sucesores” de SpaceStat (GeoDa, GeoDaSpace y PySal) puede ser utilizado como guía metodológica para poder ser desarrollada por otros programas, concretamente Matlab, R y Stata.

Los métodos se basan en la secuencia de un proceso de decisión que comienza con una estimación clásica por MCO. A los residuos se les aplican los test de I de Moran, LM LAG, LM ERROR, robust LM LAG, robust LM ERROR y test SARMA. Inicialmente se consideran los resultados de significatividad para LM LAG y LM ERROR. Si ninguno de los dos es significativo aceptamos el modelo estimado por MCO.

Si lo es LM LAG y no LM ERROR se estima un modelo de retardo espacial. Hay dos métodos para estimarlo. Uno sería el de máxima verosimilitud (ML), recomendado en caso de existir normalidad en los errores. En caso de no existir normalidad en los errores se debe estimar aplicando el método de los momentos generalizados (GMM), decidiendo si se incluyen o no regresores endógenos. De no incluirse se utilizan como variable instrumental e instrumentos los retardos espaciales de la endógena y de las exógenas, respectivamente. Si se incluyen se utilizarían junto con estas. El método sería mínimos cuadrados bietápicos espaciales (S2SLS) y se puede aplicar con el método de estimadores robustos de White.

Si por el contrario es significativo LM ERROR y no lo es LM LAG, se estima un modelo de error espacial. Se podría estimar por máxima verosimilitud si los errores son normales. Si no es el caso se estimaría por momentos generalizados (GM) o por el método de los momentos generalizados (GMM), siendo este último más recomendable por ser más completo y robusto a heteroscedasticidad. También está la opción de incluir o no regresores endógenos. Si no se incluyen el primer paso de estimación se

realiza con una transformación que equivale a utilizar mínimos cuadrados factibles y el método entonces sería mínimos cuadrados espacialmente ponderados (SWLS). Si se incluyen regresores endógenos el primer paso es una estimación por mínimos cuadrados bietápicos. El método sería en este caso mínimos cuadrados bietápicos espacialmente ponderados (SW2SLS). Se puede aplicar el método de estimación propuesto por Kelejian y Prucha (2010) robusto a heteroscedasticidad y al igual que en los modelos clásicos no espaciales se podía aplicar el método que calcula estimadores robustos a heteroscedasticidad de White y/o a heteroscedasticidad y autocorrelación espacial HAC.

Si ambos test LM ERROR y LM LAG son significativos, tendríamos que estudiar los robustos y estimar retardo o error en función del robusto que si lo sea. Si ambos lo son el más significativo es el que indica el tipo de modelo espacial a estimar. Aunque el test SARMA es menos tenido en cuenta, también se puede estimar un modelo mixto del error y el retardo, ambos autorregresivos por el método GM o GMM con la opción nuevamente de aplicar el procedimiento robusto a heteroscedasticidad de Kelejian y Prucha (2010). El método para este modelo es como el del error, mínimos cuadrados bietápicos espacialmente ponderados (GSW2SLS) pero para el modelo combo.

En caso de existir heterogeneidad espacial se tiene como opción aplicar el método de regresiones cambiantes o regímenes espaciales (spatial régimes), que dividiría el espacio en distintas zonas o regímenes para modelizar la inestabilidad paramétrica continua, si es el caso.

#### **2.4.5. PROGRAMAS Y PAQUÉTES INFORMÁTICOS.**

La mayoría de paquetes estadísticos tradicionales van teniendo en cuenta el análisis espacial e incorporando herramientas que permiten su análisis estadístico. De todas formas cada vez existen más paquetes informáticos específicos para la aplicación de técnicas espaciales. Algunos de ellos serían los que se consideran en Fisher y Getis (2010):

- ArcGIS

- SAS
  
- R
  
- GeoDa
  
- STARS
  
- Space-Time Intelligence System Software for the Analysis of Complex Systems
  
- Geostatistical Software
  
- GeoSurveillance: GIS-based Exploratory Spatial Analysis Tools for Monitoring Spatial Patterns and Clusters
  
- Web-based Analytical Tools for the Exploration of Spatial Data
  
- PySAL: A Python Library of Spatial Analytical Methods

Algunos de los anteriores paquetes informáticos se desarrollan por el GeoDa Center for Geospatial Analysis and Computation. Este centro surge de la School of Geographical Sciences and Urban Planning y sucede al Spatial Analysis Laboratory (SAL) fundado por Luc Anselin. El GeoDa Center, entre otras cosas, desarrolla métodos y técnicas para el análisis geoespacial y para la modelización de procesos espaciales. Utiliza paquetes informáticos de código abierto, accesibles desde su página web a toda la comunidad científica. La lista de software disponible del GeoDa Center y sus asociados (R-Geo) en su página es la siguiente, que figura en el cuadro 2.9:

Cuadro 2.9.

PAQUETE INFORMÁTICO	FUNCIONALIDADES	PLATAFORMA Y CÓDIGO	LENGUAJE
Geoda	Introduction to Spatial Data Analysis: EDA, ESDA & ML Spatial Regression	Cross platform and open source	C++
PySAL	Open Source Library for Spatial Analysis: Weights, computational geometry, ESDA, spatial econometrics, clustering and spatial dynamics	Cross platform and open source	Python
GeoDaSpace	Spatial econometrics (lag and error, endogenous variables, HAC, robust standard errors, spatial regimes)	Cross platform code in pysal 1.3+	Python
GWR	Stewart Fotheringham's Geographically Weighted Regression software	W Vista, 7or8 64bit & XP and later 32bit, closed source	.NET and Visual C++ 2012
R-Geo	Spatial weights, spatial econometrics, geostatistics, point pattern analysis	Cross platform, open source	R
CAST	Crime Analytics for Space-Time: Space-time cluster and other methods for point and polygon data	Cross-platform, code in PySAL	Python, C++
ArcGIS PySAL TOOLS	Spatial weights and regression PySAL tool for ArcGIS. Additional PySAL functionality will be integrated with ArcGIS 11.	Windows, ArcGIS 10.1 (+ latest service pack); code from PySAL	Python
GeoDanet	Spatial point pattern analysis on networks	Cross-platform, code in PySAL 1.4-1.6	Python
DynTM: Scalable Tile Map Web Service	Dynamic choropleth mapping	Cross platform, open source	Python
STARS	Space-Time Analysis of Regional Systems: EDA & ESDA (part of PySAL)	Cross platform, open source	Python
Pweights: Weights Transformations	Dynamic Manipulation of Weights Using Web Services	Cross platform, open source	Python
Java Tools for Spatial Data Analysis	EB & SR smoothing, excess risk maps, global Moran scatterplot and spatial reference distribution	Cross platform, open source	Java
Web-Based ESDA Tools	EB & SR smoothing, excess risk and other maps, global Moran scatterplot	Cross platform, closed source	Java

Fuente: Elaboración propia a partir de la información disponible en <https://geodacenter.asu.edu/>

## GeoDa.

Busca ser una herramienta que permita un análisis sencillo, desde una simple visualización inicial de mapas, siguiendo con un análisis exploratorio espacial, y terminando con la modelización de una regresión espacial. Se basa en un entorno de ventanas que simplifica su uso y no requiere de programación como otros paquetes informáticos existentes (por ejemplo R). En muchos aspectos GeoDa es una reinención de SpaceStat, que ahora está bastante anticuado y con limitaciones de rendimiento (Fisher y Getis, 2010). Es uno de los paquetes que forma parte de la librería PySal. En el siguiente cuadro (2.10) se muestran las funcionalidades del programa, que son básicamente seis.

**Cuadro 2.10. GeoDa functionality overview**

Category	Functions
Spatial data	Data input from shape file (point, polygon). Data input from text (to point or polygon shape). Data output to text (data or shape file). Create grid polygon shape file from text input, Centroid computation. Thiessen polygons.
Data transformation	Variable transformation (log, exp, etc.), Queries, dummy variables (regime variables). Variable algebra (addition, multiplication, etc.). Spatial lag variable construction. Rate calculation and rate smoothing. Data table join.
Mapping	Generic quantile choropleth map. Standard deviational map. Percentile map. Outlier map (box map). Circular cartogram. Map movie. Conditional maps. Smoothed rate map (EB, spatial smoother). Excess rate map (standardized mortality rate, SMR).
EDA	Histogram. Box plot. Scatter plot. Parallel coordinate plot. Three-dimensional scatter plot. Conditional plot (histogram, box plot, scatter plot).
Spatial autocorrelation	Spatial weights creation (Rook, Queen, distance, $k$ -nearest). Higher order spatial weights. Spatial weights characteristics (connectedness histogram). Moran scatterplot with inference. Bivariate Moran scatterplot with inference. Moran scatterplot for rates (EB standardization). Local Moran significance map. Local Moran cluster map. Bivariate local Moran. Local Moran for rates (EB standardization).
Spatial regression	OLS with diagnostics (e.g., LM test, Moran's $J$ ). Maximum likelihood spatial lag model. Maximum likelihood spatial error model. Predicted value map. Residual map.

Fuente: Anselin, Syabri y Kho (2010)

GeoDa es una herramienta muy útil para el AEDE, y es el programa que usaremos para la realización de esta fase en nuestro caso de estudio. Aunque no son todas, a

continuación explicaremos algunas de las técnicas gráficas que podrían ser consideradas como apropiadas para el mismo.

### **Análisis univariante.**

#### ***Histograma de frecuencias***

Se calcula el histograma de frecuencias de las variables para 8 clasificaciones (es el número que viene por defecto en GeoDa).

#### ***Mapa de deciles.***

Se utiliza sobre un mapa un color distinto para cada grupo de observaciones.

#### ***Diagrama de caja.***

Parte del cálculo de los cuartiles y la mediana de una variable, así como de las cotas superior e inferior, que se obtienen, a su vez, como el producto de los valores del tercer y primer cuartil por 1.5 veces el rango intercuartílico (un criterio más estricto sería multiplicar por 3 en vez de 1.5). Se consideran como valores atípicos aquéllos situados por encima o por debajo de dichas cotas.

#### ***Mapa de caja.***

Resulta de trasladar al mapa, en colores para las distintas áreas geográficas, los resultados del diagrama de caja.

#### ***Diagrama de dispersión de Moran.***

Se trata de un diagrama de dispersión que representa en el eje X la variable previamente estandarizada y en el eje Y se representa el retardo espacial de dicha variable estandarizada, siendo este el promedio ponderado de los valores que adopta una variable en el subconjunto de observaciones vecinas. La pendiente de la recta de regresión es el valor del estadístico I de Moran, y cuanto mayor sea esta, mayor será el grado de autocorrelación espacial de la variable. Se divide el tipo de asociación espacial en cuatro categorías: dos para autocorrelación espacial positiva (valores altos de una variable rodeados de valores altos o valores bajos rodeados de valores bajos) y dos para

autocorrelación espacial negativa (valores altos rodeados por valores bajos, y viceversa). Además, ya que los valores de la variable se encuentran estandarizados, es posible analizar tanto la asociación espacial global (pendiente de la recta de regresión) como la asociación espacial local. Así los valores en el diagrama de dispersión de Moran situados a más de dos unidades del origen pueden considerarse como puntos atípicos.

El programa permite calcular el pseudo p valor resultante de un proceso de aleatorización con permutaciones de los valores de la variable para contrastar la hipótesis de existencia o no de autocorrelación espacial. Se puede escoger el número de permutaciones del proceso (99, 199, 499, 999, o más).

### ***Mapa de desviación típica***

Se agrupan los valores en intervalos, que se acotan en función de un número determinado de unidades de desviación típica con respecto a la media. GeoDa lo hace en 6 intervalos determinados por 1, 2 y 3 unidades de la desviación típica.

### ***Mapa de percentiles.***

Es un caso del mapa de cuantiles, pero con un tipo de categorías que permite resaltar valores atípicos. GeoDa realiza seis grupos con los percentiles  $<1$ ,  $[1,10)$ ,  $[10,50)$ ,  $[50,90)$ ,  $[90,99)$ ,  $>99$ .

### ***Cartograma.***

El cartograma es un mapa que representa con un círculo cada unidad geográfica y la superficie de dicho círculo es proporcional al valor de la variable en dicha unidad geográfica. En GeoDa se usan colores distintos para cuatro cuantiles y otros dos más para resaltar valores atípicos (rojo para altos y azul para bajos). Los atípicos se determinan con el mismo criterio que en el gráfico de caja.

### ***Mapas LISA.***

Los gráficos LISA incluidos en GeoDa se basan en el estadístico I de Moran de asociación local y en los estadísticos G y G\* de Getis y Ord.

El mapa de significación muestra los concejos que presentan autocorrelación, una vez se ha realizado el contraste correspondiente, utilizando distintos colores para los distintos niveles de significatividad. El método de contraste utilizado se basa en el cálculo del pseudo valor de p tras una aleatorización con 99 permutaciones (se podrían escoger 199, 499, 999, o más) de los valores de la variable. Para la G local también utiliza además del anterior el cálculo del p valor asumiendo la normalidad de la distribución.

En el mapa de clusters se identifican mediante distintos colores cuatro tipos de concejos en los que la autocorrelación espacial local es significativa, municipios con valores de la variable altos rodeados de municipios con valor de la variable también altos, municipios con valores de la variable bajos rodeados de municipios con valor de la variable bajo, municipios con valores de la variable altos rodeados de municipios con valor de la variable bajo y por último municipios con valores de la variable bajos rodeados de municipios con valores de la variable altos. Las dos primeras situaciones se corresponden con autocorrelación espacial local positiva mientras las dos últimas ponen de manifiesto autocorrelación espacial local negativa.

### ***Mapa dinámico.***

Consiste en la iluminación ordenada sobre el mapa de las distintas unidades geográficas según su valor de una variable, de menor a mayor.

### ***Gráfico de dispersión en 3D***

Realiza movimientos de rotación, traslación y acercamiento/alejamiento (“zooming”). También las habituales funciones de enlace y cepillado (“linking” y “brushing”).

### **Análisis multivariante.**

#### ***Diagrama de dispersión***

Se trata de un análisis bivariante. Representa en los ejes la distribución de las dos variables estudiadas. GeoDa ajusta la recta de regresión por mínimos cuadrados ordinarios y adjunta el valor de la pendiente de la misma. Este gráfico se puede también calcular para los valores estandarizados de las variables, de forma que la pendiente de la

recta de regresión se corresponda con el coeficiente de correlación de Pearson. Analizando los cuatro cuadrantes del diagrama de dispersión se muestran las localizaciones con valores superiores o inferiores a la media en ambas variables o también los valores superiores a la media en una variable e inferiores a la media en otra.

### ***Gráfico de coordenadas paralelas***

Las variables se representan como líneas paralelas, en las que los valores de las variables son las coordenadas. Para cada observación se unen los puntos correspondientes en cada línea, de forma que el resultado final es un conjunto de tantas líneas como observaciones. Es útil para detectar clusters o agrupaciones en función de grupos con pendientes o formas similares entre si y distintas al resto.

### ***Gráficos condicionales.***

Se basan en el estudio de una variable condicionada a otras dos que se dividen en grupos. GeoDa hace tres grupos por variable y para cada grupo se crean otros tres gráficos de la variable condicionada, por lo que se manejan nueve gráficos en total. Permite diseñar cuatro tipos de gráficos condicionales: mapas, diagramas de caja, histogramas condicionales y diagramas de dispersión condicionales.

### ***Diagrama de dispersión de Moran bivalente.***

Se trata de un diagrama de dispersión de Moran en el que se representa en el eje vertical Y el retardo espacial de la variable que se desea explicar y en el eje X la variable explicativa. Las dos deben haber sido estandarizadas. Mide por tanto la relación entre una variable y el promedio de los valores vecinos a cada observación de la otra.

### ***Mapas LISA bivariantes.***

Son los mapas LISA bivalente que se basan en el estadístico de Moran. Son los mismos que en el caso univariante (mapa de significatividad y mapa de clusters), pero en el caso bivariante el estadístico LISA tiene en cuenta para cada unidad geográfica, los valores de una variable y el retardo espacial de otra variable diferente (en vez del retardo espacial de la propia variable como ocurre en el caso univariante).

## GeodaSpace

Es otro de los paquetes que forma parte de PySAL, escrito en Python y de código abierto. Puede ser utilizado con los sistemas operativos Windows, Macintosh y Linux. Se basa en una interfaz amigable para el usuario, como su predecesor, GeoDa, y a grandes rasgos se puede decir que así como el anterior se centra más en el análisis exploratorio espacial, con muchas funcionalidades de gráficos y mapas, GeoDaSpace se centra más en el análisis confirmatorio, es decir en la estimación de modelos. Los métodos de estimación que permite aplicar serían mínimos cuadrados ordinarios (MCO) y mínimos cuadrados bivariados (MC2E) para modelos no espaciales. Mínimos cuadrados espaciales bivariados (S2SLS) para el modelo del retardo espacial, con y sin regresores endógenos. Mínimos cuadrados espacialmente ponderados (SWLS) y mínimos cuadrados bivariados espacialmente ponderados (SW2SLS) para el modelo del error espacial (con y sin regresores endógenos). También se puede estimar un modelo mixto de retardo y error espacial, ambos autorregresivos, denominado combo, que también se estima por mínimos cuadrados bivariados espacialmente retardados (GSW2SLS) con o sin regresores endógenos. Para los modelos anteriores se dispone de dos opciones, una estimar usando el método de los momentos generalizados (GMM), y otra estimar por el de los momentos generalizados (GM). Además los modelos lag y error también se pueden estimar por máxima verosimilitud (como en GeoDa). Se presenta la posibilidad de incluir varios métodos de estimación robustos a heteroscedasticidad y autocorrelación (White, HAC, y K-P). Regímenes espaciales para modelar modelos de regresiones cambiantes es otra capacidad de GeoDaSpace. A continuación se resumen las funcionalidades del programa, que también incluyen el tratamiento de los datos y la creación de matrices espaciales (cuadro 2.11).

**Cuadro 2.11. GEODASPACE FUNCTIONALITY OVERVIEW**

<b>Category</b>	<b>Functions</b>
Spatial data	<ul style="list-style-type: none"> <li>. Data input from shape file (point, polygon)</li> <li>. Data input from dBase (dbf), Comma Separated Values (csv), model specification files (mdl) or results text files (txt).</li> <li>. Data output to model specification files (mdl) or results text files (txt).</li> </ul>
Spatial weight matrices	<ul style="list-style-type: none"> <li>. Spatial weights creation (contiguity rook and queen, binary distance bands, inverse distance, k-nearest neighbors and kernel weights), higher order spatial weights.</li> <li>. Spatial weights properties (editor, point-map viewer, connectedness histogram, transformations).</li> <li>. Spatial weights reading/writing from/to other extensions: ArcGIS (dbf, swm, txt), dat, gal, geoBUGS, gwt, kwt, MatLab, MatrixMarket, STATA text files.</li> </ul>
Spatial regression	<ul style="list-style-type: none"> <li>. OLS with diagnostics (e.g., Jarque-Bera, Breusch-Pagan, LM tests, Moran's I, etc.)</li> <li>. 2SLS for non-spatial endogeneity with diagnostics (Anselin-Kelejian test)</li> <li>. Spatial 2SLS+GMM spatial lag, spatial error and spatial lag+error models with HAC and spatial HAC (Kelejian-Prucha) estimation</li> <li>. Spatial regimes for basic and spatial models</li> <li>. Output text files for predicted values and residuals, variance-covariance matrix and complete model specification</li> </ul>

Fuente: Chasco (2013)

## STARS

Es un paquete de código abierto diseñado para el análisis dinámico de datos espaciales, es decir permite contemplar conjuntamente las dimensiones temporal y espacial.

Se basa en módulos que permiten ser utilizados de diferentes formas por los usuarios, bien enlazándolos entre sí, pudiendo usar gráficos en un entorno de usuario amigable, o bien utilizándolos como biblioteca en otros programas que utilicen scripts con el lenguaje de programación Python con el que ha sido implementado.

Los dos módulos básicos son el geocomputacional y el de visualización. A continuación se muestran las funcionalidades que incluye cada uno (cuadros 2.12 y 2.13).

**Cuadro 2.12. Geocomputational methods contained in STARS**

Category	Description
Descriptive statistics	Distribution and summary measures for variables by cross-section, time period, or pooled
Exploratory spatial data analysis	Various methods specifically designed to analyze spatial dependence. Global and local versions of Moran's $I$ , Geary's $c$ and the $G$ statistic are provided
Inequality	Techniques that quantify and decompose inequality over time and space. Includes classic and spatial Gini coefficients as well as Theil decomposition
Mobility	Recent advances in internal mobility dynamics are presented through the $\tau$ and $\theta$ statistics
Markov analysis	Transitional dynamics of distributional attributes are examined through the use of classic Markov and spatial Markov techniques

Fuente: Rey y Janikas en Fisher y Getis (2010)

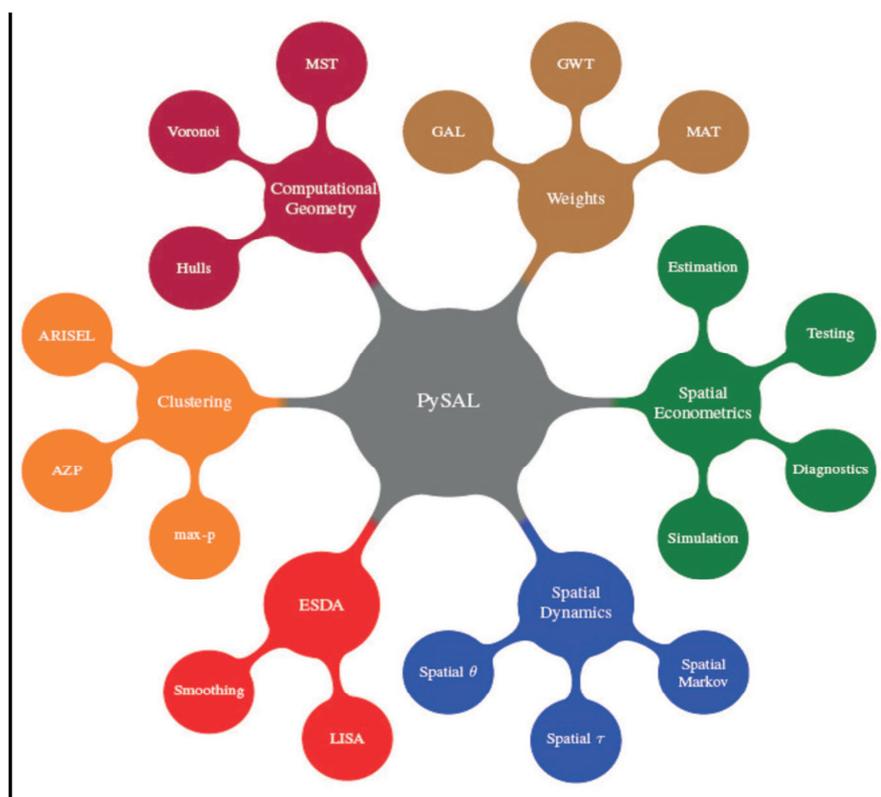
**Cuadro 2.13. Visualization capabilities in STARS**

Category	Description
Map	A variety of sequential, categorical and user-defined choropleth maps
Scatter plot	A basic two-dimensional view, the scatter plot can be used to analyze cross-sectional, time period or bivariate correspondence in X-Y space
Conditional scatter plot	Extends the traditional scatter plot to three dimensions by conditioning the color of the data points by the level of a third variable
Parallel coordinate plot	Allows the user to view multivariate relationships over space and time
Time series plot	Plots the evolution of a variable for a given spatial unit
Time path plot	Demonstrates the co-movement of a variable for two spatial units over time
Histogram	Creates a basic partitioning of a variable into respective bins.
Density	Contains empirical kernel density estimation for the analysis of dispersion, modality, and skewness
Box plots	Another distributional view with an added focus on quantiles and outliers

Fuente: Rey y Janikas en Fisher y Getis (2010)

## PySAL

Es una biblioteca de código abierto, para el análisis espacial, escrita en el lenguaje de programación Python. Combina el desarrollo de funcionalidades de GeoDa, GeoDaSpace y STARS, entre otras las del diagnóstico de efectos espaciales y la estimación de modelos espaciales. La utilización de módulos que desarrollan las necesidades de otros programas permite integrar esfuerzos y avances paralelos que se integran posteriormente en otros proyectos. La utilización de Python, al igual que una parte importante de la comunidad científica permite integrar rápidamente avances de otros campos, como por ejemplo el desarrollo de sistemas y herramientas GIS de código abierto que están teniendo importantes avances recientes. En cuanto al software de análisis de datos espaciales tanto en el caso de las herramientas que usan un interfaz gráfico en un entorno amigable, como las que usan lenguaje de programación, pueden aprovechar funcionalidades de PySAL. Sus funcionalidades se pueden agrupar en seis grupos tal como muestran el gráfico siguiente y el cuadro 2.14.



Fuente: <https://geodacenter.asu.edu/projects/pysal>

**Cuadro 2.14. PySAL functionality by component**

<b>Component</b>	<b>Capabilities</b>
	<b>Data analytic functions</b>
File input-output	Read and write common spatial data formats
Map calculations	Map algebra
Computational geometry	Geometric summaries of spatial patterns
Spatial weights	Efficient construction/manipulation of spatial weights matrices
Rate Smoothing	Spatial and non-spatial smoothing of rate data
	<b>ESDA and modeling functions</b>
Spatial autocorrelation	Local and global spatial autocorrelation
Space-time correlation	Spatial and temporal correlation measures
Markov and mobility	Spatial Markov and distributional dynamics
Regionalization	Spatially constrained clustering
Spatial regression	Classic spatial econometric methods
Spatial panel regression	Spatial methods for panel data

Fuente: Rey y Anselin en Fisher y Getis (2010)

## R PROJECT

R es un lenguaje y entorno de programación para análisis estadístico y gráfico. Es uno de los más utilizados en estadística junto con S, del que deriva. Es un proyecto de software libre, resultado de implementar S con el sistema operativo GNU, tras la integración del Proyecto R en el Proyecto GNU, también de software libre. GNU es un sistema operativo totalmente compatible con UNIX (sistema operativo no libre) pero distinto al mismo.

R además de libre es programable y permite el lenguaje matricial. Funciona por línea de comandos y puede ocurrir que algunas funciones no se encuentren disponibles, ya que su funcionamiento se basa en el uso de paquetes. Cada paquete es una colección de funciones que permiten la realización de determinadas tareas que previamente debe instalarse. Para el área de econometría espacial, R dispone de diversos paquetes que se complementan.

## ArcGIS

La tecnología GIS permite la organización, manipulación, análisis y visualización de datos espaciales. A las funcionalidades tradicionales de los sistemas de información geográfica (SIG) que permiten, con la información geográficamente referenciada, realizar consultas, editar datos, mapas, analizar y presentar esa información, se le han ido incorporando herramientas que mejoran esas funcionalidades. Entre ellas están las que permiten el análisis y estudio de los datos mediante técnicas de estadística espacial. Las herramientas estadísticas de la versión ArcGIS 9.3 se pueden agrupar en cuatro grupos: medición de distribuciones geográficas, análisis de patrones, cartografía de clusters y modelización de relaciones espaciales (Scott y Janikas, 2010). También permite el desarrollo de herramientas personalizadas y su combinación con otros programas, como PySAL, para aumentar sus capacidades.

## QGIS

Se trata de un sistema de información geográfica libre y de código abierto, con licencia GNU. Es un proyecto oficial de Open Source Geospatial Foundation (OSGeo). Actualmente funciona en la mayoría de plataformas Unix, Windows y OS X. Pretende hacer ampliamente accesible el software SIG, lo que le ha hecho enormemente popular en la actualidad y muy utilizado. Al igual que otros sistemas de información geográfica (SIG) permite el tratamiento de información geográficamente referenciada. Maneja formatos ráster y vectoriales a través de la bibliotecas GDAL y OGR, así como bases de datos. QGIS está desarrollado en C++, lo que le hace más amigable, y acepta complementos tanto en C++ como en Python. Entre sus puntos fuertes destaca la facilidad de interconexión con otras bases de datos geoespaciales y la posibilidad de trabajo en cualquiera de los sistemas operativos con similar forma de funcionamiento (GNU/Linux, BSD, Unix, Mac OSX, Windows y Android).

## CAPÍTULO 3. ENFOQUE METODOLÓGICO, VARIABLES Y FUENTES DE INFORMACIÓN ESTADÍSTICA.

---

### 3.1. POSIBLES ENFOQUES METODOLÓGICOS Y REVISIÓN DE LA LITERATURA RECIENTE.

La literatura relacionada con el objeto de estudio de esta tesis es amplia y su repaso susceptible de múltiples enfoques, ya que abarca ámbitos diversos. Por un lado la política de desarrollo rural de la Unión Europea, en la comunidad autónoma de referencia, se basa, fundamentalmente, en las subvenciones que incentivan la inversión en bienes de capital, que se supone que activan la economía. Al tratarse de subvenciones sectoriales, pero que se aplican en un ámbito territorial relativamente pequeño, su impacto sobre la economía se ve necesariamente abocado a un enfoque regional y probablemente a requerir del tratamiento de los datos con técnicas de econometría espacial. Las ayudas al desarrollo rural se dirigen en principio a agentes económicos del sector primario, pero también incluyen enfoques territoriales en los que, una vez que un territorio se considera rural, se pueden destinar las ayudas concedidas a cualquier otro sector, para favorecer así la diversificación de la economía del territorio y crear un tejido económico y social que frene el abandono, tanto del sector, como de la zona rural. Es por esto que la temática que se aborda está relacionada con enfoques teóricos propios de la teoría del crecimiento y de la convergencia, de la economía regional, de la geografía económica (incluyendo la nueva geografía económica), del estudio de impactos de medidas de política económica, etc.

Los enfoques teóricos, tanto conceptuales como metodológicos, que se pueden aplicar a este tema de estudio son múltiples, pero todos tienen en común que tratan de cuantificar o medir la eficacia y los resultados de la inversión pública sobre diversos aspectos económicos de territorios concretos. Se debe aplicar siempre un enfoque regional, territorial, geográfico o espacial, independientemente de que el ámbito de estudio sean países, regiones, municipios, individuos o cualquier otra unidad de actuación.

La literatura sobre la relación entre inversión pública y actividad económica es susceptible de ser clasificada de diversas formas.

Según Cervelló y Segura (2010) “*existen dos campos básicos en los que se centran los estudios de carácter socioeconómico en los procesos de intervención: a) la dotación y provisión de bienes y servicios públicos, y b) la gestión de externalidades (Guellec y Ralle, 1995).*”

Los enfoques metodológicos para estudiar las relaciones entre inversión pública y actividad económica, considerada esta en sentido amplio, han sido clasificados por varios autores. Algunos, como por ejemplo Boscá, Escribá y Murguá (2011), repasan la literatura existente. Estos autores consideran que “*la literatura ha utilizado diferentes aproximaciones al estudio empírico de los efectos del capital público sobre la actividad privada: el enfoque de funciones de producción, el enfoque dual tanto a través de funciones de coste como de beneficios, modelos vectoriales autorregresivos (VAR), modelos de crecimiento y modelos de equilibrio general. El enfoque más usual ha sido el de funciones de producción, aunque cada vez más frecuentemente se observa la utilización del enfoque dual*”.

Según Díaz y Martínez (2005) los distintos enfoques que lo relacionan serían los siguientes:

- “*Quizás el planteamiento más efímero haya sido el del **Enfoque del Potencial del Desarrollo Regional (EPDR)**, que tuvo su punto álgido en el llamado Informe Biehl (1986) y con extensiones en el tiempo hacia atrás (Biehl, 1979) y hacia delante (Biehl, 1988,1991). Desde la perspectiva del EPDR, las regiones disponen de un nivel potencial de desarrollo económico que se sería el alcanzado si todos sus recursos se utilizan de forma óptima, tanto desde un punto de vista individual como en términos de su combinación conjunta. Los recursos con los que habitualmente trabajan este tipo de estudios suelen ser el capital humano y productivo privado, la localización geográfica del territorio, el grado de aglomeración de los agentes económicos, las características de su población y estructura sectorial, y las infraestructuras públicas. Definido el nivel de renta per cápita máximo que podría alcanzar una región, el análisis del EPDR compara éste con el nivel efectivo, a fin de identificar aquellos factores que suponen un lastre en la consecución del nivel potencial; en esta comparación se presta especial atención al papel desempeñado por el capital público, con la derivación de evidentes implicaciones normativas.*”

- “Posiblemente el enfoque del que más trabajos se han derivado, en parte por la sencillez de su implementación empírica, en parte por la versatilidad de su especificación a distintas metodologías econométricas, es el de la **función de producción con capital público**. Las variantes que este planteamiento presenta difieren entre sí, no en el fundamento teórico, sino en el procedimiento econométrico, que estima la importancia de los factores productivos sobre el output. A partir de una sencilla especificación de la función de producción con capital privado y trabajo, la literatura ha añadido capital público para evaluar el efecto de las infraestructuras sobre la actividad económica. La forma funcional elegida en la inmensa mayoría de los casos es la conocida Cobb-Douglas, que además ha visto cómo se ampliaban sus argumentos para incluir capital humano y tecnológico, progreso técnico, diversas externalidades, etc.

*Realmente en este caso no debiera hablarse de un enfoque teórico per se, sino más bien de las distintas aproximaciones econométricas que la estimación de funciones de producción con capital público ha generado.”*

- En cierta medida, como una extensión de los análisis anteriores, aparece una tercera aproximación al fenómeno del capital público y su efecto sobre la actividad económica: el denominado **enfoque dual**. Éste parte de la estimación de sistemas de ecuaciones formados por funciones de costes o beneficios y demandas de factores de producción privados. Bajo semejante planteamiento no solo se tiene en cuenta el comportamiento optimizador de las empresas sino que permite el estudio de factores adicionales implicados en la discusión sobre los efectos del capital público en la actividad económica, como son las elasticidades de la demanda de factores a las infraestructuras, la incorporación de una mayor flexibilidad tecnológica a la hora de relacionar output e inputs o la posibilidad de contrastar si los factores de producción (cuasi) fijos se encuentran en sus valores de equilibrio.

- “Una cuarta aproximación es la llevada a cabo desde postulados propios de **economía urbana** (Martín y Rogers; 1995; Alonso, 2001). Como se destaca en Caramés y Lago (2002), en estos modelos los equilibrios estáticos que se alcanzan están determinados por la confluencia de tres tipos de fuerzas: i) rendimientos a escala crecientes, que favorecen la concentración de factores productivos; ii) el tamaño de los mercados, que propicia un acercamiento de los productores a los consumidores; y iii)

*la existencia de costes de transporte, por los que la producción tiende a dispersarse. Dados estos puntos, las políticas públicas pueden ejercer sus efectos a través de dos vías: por un lado, con políticas de rentas susceptibles de modificar el tamaño de los mercados y, por otra parte, con provisión de infraestructuras que afectan a los costes de transporte. Este último punto supone una innovación respecto al grueso de la literatura anterior dado que admite la posibilidad de que la política regional afecte negativamente a la convergencia al favorecer la aglomeración de actividades productivas vía reducción en los costes de transporte. Esta línea de trabajo es eminentemente teórica. El rigor analítico de los mismos impide un inmediato contraste empírico, y a lo más que estos trabajos llegan es a ofrecer simulaciones numéricas sobre los resultados del modelo en base a parámetros definidos al margen de la realidad. No obstante, dichas simulaciones proporcionan conclusiones interesantes acerca del efecto de las infraestructuras sobre la localización espacial de la actividad, una circunstancia ésta que debe ser considerada por los agentes de política económica.”*

*- “Desde otra perspectiva distinta a la de la economía urbana, surgen también a lo largo de la década de los noventa aportaciones novedosas que estudian los efectos económicos de la provisión de infraestructuras empleando técnicas de **econometría espacial** (Rephann e Isserman, 1994; Kopp, 1995; Moreno et al., 1997; Kelejian y Robinson, 1997; Boarnet, 1998; Gómez de Antonio, 2001), **análisis de frontera** (Maudos et al., 1998; Pedraja et al., 1999; Delgado y Álvarez, 2003; Salinas, 2004) o **economía del transporte** (Ginés de Rus, 2001; Chandra y Thompson, 2000; Shirley y Winston, 2004). En general, todas ellas ofrecen evidencia a favor del efecto positivo del capital público sobre la actividad económica, aunque alguna de estas aportaciones matiza considerablemente dicho impacto.”*

Esta, quizá sea la clasificación más completa de los distintos enfoques aplicados. El más oportuno en este caso, de los considerados en la clasificación anterior, es el enfoque que utiliza técnicas de econometría espacial, ya que se ajusta más al objeto de estudio planteado, ha sido menos utilizado por otros autores, y dada la naturaleza de los datos y el ámbito de aplicación, que se ciñen a una escala microterritorial, se requiere un tratamiento que tenga en cuenta los posibles efectos espaciales para que los resultados sean relevantes.

La econometría espacial tiene una serie de planteamientos teóricos con ciertos problemas en lo referente, entre otras cosas, a la endogeneidad de las variables explicativas. Suele ocurrir en los casos en los que se ven correlacionadas las exógenas con el retardo de la endógena, en los casos de modelos SAR. Entre otras técnicas esto se puede corregir con la aplicación del método de los momentos generalizados. Como este, existen otra serie de problemas a los que la econometría ha dado solución, pero también otros en que esto no ha sido posible. No parece conveniente estimar modelos sin una base teórica, pero tampoco lo parece extender sin más, al contexto espacial, ampliaciones de modelos teóricos existentes, como extensiones de los mismos, sin considerar todos los problemas empíricos que pueden presentar. Pinkse y Slade (2009) repasan estos posibles problemas y soluciones a la vez que proponen el desarrollo de la econometría espacial aprovechando las soluciones propuestas por los avances recientes y adaptando los modelos teóricos a las mismas, y no al revés, para dar así solución directa a los problemas que se puedan presentar. Nos parece procedente y lo más adecuado al caso que nos ocupa, utilizar las técnicas y herramientas disponibles para evitar precisamente esos problemas, fundamentalmente el de endogeneidad, siempre respetando la base y sentido teórico de la relación a contrastar.

González Torres (2013) realiza un estudio del impacto del turismo rural en las comarcas turísticas de Castilla y León, cuantificando la relación entre el desarrollo de dichas comarcas rurales y esta actividad. De forma indirecta, analiza la eficacia de programas de desarrollo rural como LEADER y PRODER, ya que menciona su influencia en el turismo rural. Estos programas, iniciativas o metodología actúan sobre otras variables y sectores además del turístico, y a su vez sobre el sector turístico se aplican otras políticas además de las mencionadas, por lo que el impacto sectorial que se estudia es indirecto con respecto a la política de desarrollo rural en su conjunto. Realiza un análisis espacial después de determinar previamente las comarcas turísticas existentes, y en dicho análisis utiliza modelos espaciales una vez que ha detectado la presencia tanto de autocorrelación como de heterogeneidad espaciales. Plantea distintos análisis de correlación para contrastar la relación entre las variables objeto de estudio y con capacidad explicativa, y en base a los resultados confirma la selección previa, acorde a la teoría económica, y plantea modelos de regresión que recojen los problemas espaciales detectados.

*“Para determinar las variables a incorporar en los distintos modelos hemos calculado los correspondientes coeficientes de correlación y posteriormente hemos realizado un contraste de hipótesis para determinar si dicha correlación lineal es significativa.*

*Teniendo en cuenta la correlación existente entre las variables hemos planteado una colección de modelos que reflejen la relación existente entre las variables en términos absolutos, y por otro lado una serie de modelos para las variables con valores en términos relativos. Los modelos construidos pretenden recoger la posible influencia del turismo rural en la fijación de población, en el envejecimiento de esta o en las variables económicas. A la hora de especificarlos hemos tenido en cuenta la correlación entre la variable dependiente y la variable turística correspondiente que junto a otras se incluye como variable explicativa.”* (González Torres, 2013).

El enfoque metodológico que vamos a aplicar en este caso está basado en la econometría espacial. Este enfoque considera las relaciones teóricas previas y acepta las que se confirman desde el punto de vista econométrico, adaptando los modelos a las soluciones que proporcionan los avances recientes en econometría espacial. Se intenta así evitar los posibles problemas empíricos que se pudiesen presentar y que a posteriori probablemente tendrían difícil solución. Creemos que es la metodología que mejor se adapta al ámbito de estudio, dada la disponibilidad de datos y la complejidad de relaciones posibles que se pueden considerar entre las múltiples variables comarcales y municipales sobre las que se pretende actuar desde la política de desarrollo rural de la Unión Europea.

### **3.2. EL PROBLEMA DEL NIVEL DE AGREGACIÓN DE LOS DATOS ESPACIALES.**

El análisis a niveles territoriales de desagregación a escala muy pequeña condiciona la elección de las variables y el ámbito de estudio, tanto por su escasa disponibilidad como por la posible pérdida de validez de teorías previamente contrastadas que funcionan a niveles agregados pero que pueden no ser válidas a un nivel microterritorial.

*“En muchos fenómenos sociales, como la economía, las principales teorías se han elaborado en ámbitos territoriales superiores (países o regiones) no siendo siempre directamente aplicables en una escala microterritorial, cercana al individuo. Efectivamente, el citado sesgo de la agregación consiste precisamente en una pérdida de información (o desvirtuación de la misma) que se produce en el proceso de agregación de datos más o menos individuales para su expresión en una escala superior. Por eso, como ya se ha demostrado, algunos estadísticos básicos cambian de tamaño e incluso de signo, pudiendo no ser válidas las teorías propuestas para ámbitos agregados”. (Chasco, 2003)*

Cuando trabajamos con datos a un nivel territorial muy pequeño encontramos dificultades para obtenerlos y también dificultades para que se mantengan las relaciones existentes en distintos niveles de agregación, por dos problemas fundamentalmente, el de la inferencia ecológica (King, 1997) y el de la unidad espacial modificable (Arbia, 1989).

El denominado problema de unidad espacial modificable, que se da cuando se producen cambios o desviaciones en las mediciones estadísticas al cambiar el área o la unidad espacial elegida, se puede dividir en dos, según Alañón (2002), el problema de escala y el problema de agregación.

El problema de escala surge cuando se impone de forma exógena una división preexistente, como puede ser una división político administrativa. Al darse cambios según la escala espacial utilizada se debe determinar cuál es la más conveniente para cada caso de estudio. Chasco (2003), *“así, por ejemplo, la agregación de datos microterritoriales (individuos, secciones, municipios) a una escala geográfica superior (provincias, regiones) tiene, al menos, las siguientes consecuencias sobre los estadísticos básicos de una variable medida en el ámbito agregado:*

- a) Disminución de la varianza.*
- b) Disminución del grado de autocorrelación espacial.*
- c) Aumento del coeficiente de correlación lineal entre dos variables.*
- d) Cambios en la estacionariedad espacial.”*

El problema de agregación surge cuando las variaciones se producen por la forma de las observaciones y no por su escala. Es decir dependiendo de cómo se agrupen las unidades menores cambian los resultados. En la escala se pierde información al agregar y en la agregación no se pierde, pero esta es distinta dependiendo de que la forma escogida sea una u otra.

El problema de la inferencia ecológica deriva y está relacionado con el problema de la unidad de área modificable. Se produce cuando se infieren datos de un grupo a las partes que lo forman, de ahí el término ecológica. Las partes serían unidades no modificables, que son indivisibles, por ejemplo individuos y el grupo la suma de los mismos, por ejemplo grupo de individuos. Inferir relaciones o datos a los individuos a partir de las que se dan en el grupo puede ocasionar graves errores de medición.

Todos estos problemas van a condicionar el estudio ya que la disponibilidad de datos y las fuentes de información se ven condicionados por los problemas anteriormente descritos.

### **3.3. DEMARCACIÓN TERRITORIAL DE LA UNIÓN EUROPEA CON FINES ESTADÍSTICOS.**

La realización de un estudio sobre fondos europeos destinados a la política regional va a requerir que se utilice en el análisis estadístico de las regiones el mismo criterio territorial que se emplea en la redistribución de estos fondos. Además también se verá condicionado por la disponibilidad estadística de los datos que seguirá el mismo criterio de distribución territorial. La Unión Europea utiliza una nomenclatura común de unidades territoriales estadísticas denominadas NUTS, derivado de las siglas en francés de Nomenclature des Unités Territoriales Statistiques.

El artículo 2 del **REGLAMENTO (CE) No 1059/2003 DEL PARLAMENTO EUROPEO Y DEL CONSEJO de 26 de mayo de 2003 por el que se establece una nomenclatura común de unidades territoriales estadísticas (NUTS)**, dice textualmente:

“1. La nomenclatura NUTS subdivide el territorio económico de los Estados miembros, tal y como se define en la Decisión 91/450/CEE, en unidades territoriales. Asigna un código y un nombre específico a cada unidad territorial.

2. La nomenclatura NUTS es jerárquica. Subdivide cada Estado miembro en unidades territoriales del nivel NUTS 1, que a su vez se subdividen en unidades territoriales del nivel NUTS 2, subdivididas por último en unidades territoriales del nivel NUTS 3.

3. No obstante, una unidad territorial concreta podrá estar clasificada en varios niveles NUTS.

.....

5. En cada Estado miembro podrá haber más niveles de detalle jerarquizado, establecidos por el Estado miembro, que subdividan el nivel NUTS 3.....”

y el artículo 3 dice:

“1. Las unidades administrativas existentes en los Estados miembros constituirán el primer criterio utilizado para definir las unidades territoriales. Con este fin, la expresión «unidad administrativa» significará una zona geográfica con una autoridad administrativa que esté facultada para tomar decisiones administrativas o políticas para dicha zona, dentro del marco jurídico e institucional del Estado miembro.

2. Para establecer el nivel NUTS en que habrá de clasificarse una clase dada de unidades administrativas de un Estado miembro, el tamaño medio de dicha clase de unidades administrativas del Estado miembro deberá hallarse dentro de los siguientes umbrales de población:

**Cuadro 3.1.**

NIVEL	MÍNIMO	MÁXIMO
NUTS 1	3 MILLONES	7 MILLONES
NUTS 2	800000	3 MILLONES
NUTS 3	150000	800000

*Si la población de todo un Estado miembro está por debajo del umbral mínimo de un nivel NUTS dado, el Estado miembro completo será una unidad territorial NUTS para dicho nivel.”*

Los datos y el criterio de reparto, es decir la distribución territorial o espacial sigue un criterio que aprovecha las unidades administrativas existentes y que fija el nivel de agregación con un criterio poblacional.

El artículo 4 dice:

*“1. En un plazo de seis meses siguientes a partir de la entrada en vigor del presente Reglamento, la Comisión publicará los componentes de cada unidad territorial del nivel NUTS 3 en términos de las unidades administrativas menores enumeradas en el anexo III comunicadas a la Comisión por los Estados miembros.”*

En España las unidades administrativas menores que figuran en el anexo III son los municipios. El Principado de Asturias por tanto es una región de nivel NUTS 2 y nivel NUTS 3 ya que es una comunidad autónoma uniprovincial. Las unidades territoriales menores serían los municipios.

### **3.4. VARIABLES CONSIDERADAS Y FUENTES DE INFORMACIÓN ESTADÍSTICA.**

#### **3.4.1. FUENTES DE INFORMACIÓN ESTADÍSTICA.**

El estudio planteado pretende estudiar el comportamiento de variables sobre las que incide, o pretende tener influencia, la inversión pública destinada a políticas de desarrollo rural en Asturias. El Principado de Asturias es una comunidad autónoma uniprovincial, que tiene transferidas competencias en materia de política agraria y se encarga de realizar los planes de desarrollo rural para esta región de forma independiente a otras comunidades autónomas. Sus características culturales y territoriales, entre otras la de tener una sola provincia y un territorio no muy extenso, en comparación con otras comunidades autónomas, condiciona la disponibilidad de datos

para su estudio, ya que se necesita un nivel de agregación en unidades territoriales muy pequeñas.

El enfoque de econometría espacial requiere tener en cuenta las unidades territoriales en las que se divide el territorio a estudiar, en este caso Asturias, que es la unidad territorial que se ve afectada por la misma política agraria dependiente del Gobierno del Principado de Asturias. La división administrativa inferior sería la municipal, y esta es la principal unidad territorial de la que se dispone de datos estadísticos. En el Estatuto de Autonomía de Principado de Asturias se divide su territorio en concejos y comarcas (formadas por varios concejos), aunque la división comarcal no ha sido todavía desarrollada. Si que existen unas Directrices Regionales de Ordenación del Territorio (Decreto 11/91, de 24 de enero por el que se aprueban las Directrices Regionales de Ordenación del Territorio de Asturias. BOPA núm. 45 de 23 de febrero de 1991) en la que se establece una distribución comarcal a efectos de planificación territorial que se utiliza a efectos estadísticos, entre otros fines. También existe una distribución de zonas agroganaderas que recoge la comarcalización propuesta por la Consejería de Medio Rural y Pesca para la explotación estadística de los resultados de las diferentes Campañas de Saneamiento Ganadero y para la difusión de la información relativa a la cuota láctea. Ambas distribuciones son usadas por SADEI en la obtención de datos estadísticos.

La consideración de comarcas a efectos económicos tiene su justificación en la existencia de ciertas similitudes geográficas, culturales, de prestación de servicios, etc, que dan cierta lógica a unas unidades territoriales y/o administrativas superiores a la municipal. Además se debe tener en cuenta que la distribución municipal en Asturias comprende muchos municipios de tamaño muy reducido. En política agraria un ejemplo claro sería la creación de grupos de acción local para gestionar estrategias comarcales de desarrollo rural bajo la metodología LEADER.

La disponibilidad de datos comarcales es aun más escasa, pero en algunos casos se pueden obtener datos para las comarcas por la agregación de los datos municipales, ya que las comarcas están formadas por municipios.

A continuación se detallan las variables disponibles consideradas de las que hay datos a nivel municipal.

## INE

Las variables disponibles con datos municipales relevantes para nuestro estudio pertenecen a temática **agraria**, concretamente datos derivados del Censo Agrario, a **demografía**, con datos del Padrón y del Censo de Población y Viviendas. De este último también se dispone de datos municipales de **vivienda**. Sobre **mercado laboral** se dispone de la Estadística de Afiliación de Trabajadores a SS y de la **Estadística del Movimiento Laboral Registrado**. En este caso, aunque lo proporciona el INE la fuente es externa ya que este organismo, al igual que otros como SADEI, proporciona variables de otras fuentes estadísticas que disponen de información más específica.

## SADEI.

En el caso de SADEI existe más información estadística disponible a niveles municipales o comarcales que en el caso de otros organismos oficiales (como puede ser el INE), ya que se centra en el territorio específico de Asturias.

Se proporcionan datos de más de 200 variables a estos niveles de agregación, de las que muchas, evidentemente, son irrelevantes para nuestro caso de estudio. Muchos de los datos disponibles proporcionan información sobre las mismas variables a considerar en los modelos, por lo que conviene seleccionar la más apropiada en cada caso.

Entre las más importantes se encuentran las variables referentes a la **población**, que incluyen datos padronales, datos censales, de vivienda, movimiento natural de la población, movimientos migratorios y nomenclátor poblacional. Otro grupo importante sería el referente a **mercado de trabajo** del que se dispone de datos municipales y comarcales de empleo, de paro registrado, de contratación laboral y de expedientes de regulación de empleo. Sobre **magnitudes económicas** se presentan datos de valor añadido bruto y de renta de los hogares. Se dispone de muchos datos específicos del **sector primario**, algunos obtenidos del censo agrario, que fundamentalmente aportan información sobre superficie, explotaciones, titulares, cabaña ganadera y sector forestal. Para la **actividad empresarial** solo se tienen datos de licencias de IAE y áreas industriales. Hay información municipal sobre **construcción y urbanismo** que proporcionan datos sobre licencias de obra, edificios, transacciones inmobiliarias, número y superficie de parcelas y datos catastrales de las mismas. De **infraestructuras**

y **transportes** solo se dispone del parque de vehículos. Es amplia la información estadística de **turismo** sobre establecimientos y plazas. Del **sistema financiero** solo se dispone del número de oficinas de entidades de crédito por concejos y también proporciona con respecto a **administraciones públicas** el presupuesto de ingresos y gastos de los ayuntamientos. Muchas variables municipales y comarcales referentes a la **enseñanza** sobre centros, alumnos y profesores. Sobre **servicios sanitarios** se dispone de la infraestructura hospitalaria, plazas y personal. También se dispone de datos geográficos sobre **territorio**, fundamentalmente superficie y alguna variable medioambiental.

El resto de organismos o fuentes de información estadística son los que proporcionan datos específicos de su competencia y funciones (como Consejería de Agroganadería, Consejería de Empleo, Seguridad Social, Gobierno Central a través de los ministerios correspondientes, etc), En muchos casos, los institutos oficiales de estadística, y concretamente SADEI, en el caso de Asturias, publican los resultados proporcionados por estas fuentes.

### 3.4.2. VARIABLES CONSIDERADAS.

#### INDICADORES COMUNES DE CONTEXTO

Existe un Marco Común de Seguimiento y Evaluación a efectos de evaluar los objetivos y grado de ejecución de la PAC, y por tanto de los distintos Programas de Desarrollo Rural (PDR) que ejecutan los estados miembros de la UE, y que se cofinancian con fondos europeos (actualmente FEADER y anteriormente FEOGA-O). Se dan directrices sobre cómo realizar la cuantificación de dichos programas (PDR), que en el caso de España, al estar transferidas las competencias en materia de política agraria a las comunidades autónomas, existe un PDR para cada una. Incluyen una serie de indicadores que se utilizan tanto para el seguimiento y evaluación de los programas, como para el diagnóstico socioeconómico de las regiones en el planteamiento de estrategias futuras, cuando se opta a los nuevos fondos de cada periodo de programación. La disponibilidad de datos estadísticos de muchos de estos indicadores se da en EUROSTAT hasta nivel NUTS 2 y 3, o sea para el conjunto de Asturias. Para

nuestro estudio necesitamos mayor desagregación y muchos de estos indicadores no están disponibles. De todas formas los vamos a usar de referencia para que, en los casos en los que sea posible, se incluyan en el planteamiento de los modelos, ya que al estar incluidos en el diseño y los objetivos que plantean estos PDR se hace necesaria su consideración, al menos en parte. Muchos de esos indicadores son de contenido socioeconómico, y totalmente relevantes para nuestro estudio económico (producción, renta, población, infraestructuras, datos sectoriales de actividad agraria, etc).

El punto 2 del artículo 14 del REGLAMENTO DE EJECUCIÓN (UE) No 808/2014 DE LA COMISIÓN de 17 de julio de 2014 por el que se establecen disposiciones de aplicación del Reglamento (UE) no 1305/2013 del Parlamento Europeo y del Consejo, relativo a la ayuda al desarrollo rural a través del Fondo Europeo Agrícola de Desarrollo Rural (Feader) textualmente dice: *“En el anexo IV figura el conjunto común de indicadores de contexto, resultados y productividad de la política de desarrollo rural. Dicho anexo también recoge los indicadores que deben utilizarse para el establecimiento de objetivos cuantificados en relación con los ámbitos de interés del desarrollo rural”*.

Estos son los indicadores planteados para el próximo periodo de programación 2014-2020, ya que se recogen en las normas relativas al FEADER correspondiente a ese periodo, y se usan en el diagnóstico territorial para las nuevas estrategias y próximo PDR. En cada periodo de programación se establecen normas de seguimiento y evaluación, por lo que hay variaciones entre las que se usan en cada periodo. Este estudio no pretende analizar exhaustivamente los resultados de los periodos anteriores en términos de los indicadores planteados en cada periodo de programación, que además, aunque los hay comunes, pueden variar y adaptarse a cada territorio, sino los resultados posibles de la política agraria cofinanciada con fondos europeos sobre variables económicas relevantes desde el punto de vista de la teoría económica y teniendo en cuenta el problema espacial en su estimación. Por tanto, se considerarán en la elección de variables aquellos indicadores relevantes para los gestores de los fondos, pero que sean relevantes también desde el punto de vista del análisis de la teoría económica y de la econometría espacial. Esto hace que para incorporar algunos de estos aspectos como variables de nuestros modelos, podamos usar como base los indicadores

actualmente planteados para este periodo de programación, ya que son una referencia de la que seleccionaremos algunas variables que sean relevantes para este estudio.

De los tipos de indicadores que son los de contexto, los de resultados, los de productividad y los de objetivos, nos parecen relevantes para su consideración en este estudio los de contexto, ya que los de resultados, productividad y objetivos cuantifican fundamentalmente aspectos de ejecución de los programas. Los de contexto en cambio estudian variables socioeconómicas del territorio y permiten analizar su evolución para realizar un diagnóstico del mismo. La lista propuesta que los describe los divide en indicadores socioeconómicos, del 1 al 12, indicadores sectoriales, del 13 al 30, e indicadores medioambientales del 31 al 45. Serían los que figuran en el cuadro 3.2:

**CUADRO 3.2. Indicadores de contexto.**

<b>INDICADORES SOCIOECONÓMICOS</b>	<b>INDICADORES SECTORIALES</b>	<b>INDICADORES MEDIOAMBIENTALES</b>
C1. Población	C13. Empleo por actividad económica	C31. Cubierta terrestre
C2. Estructura de edades	C14. Productividad laboral en el sector agrario	C32. Zonas desfavorecidas
C3. Territorio	C15. Productividad laboral en el sector forestal	C33. Intensidad agrícola
C4. Densidad de población	C16. Productividad laboral en la industria alimentaria	C34. Zonas «Natura 2000»
C5. Tasa de empleo	C17. Explotaciones agrícolas	C35. Índice de aves ligadas a medios agrícolas (FBI)
C6. Tasa de empleo por cuenta propia	C18. Superficie agrícola	C36. Estado de conservación de los hábitats agrícolas (pastos)
C7. Tasa de desempleo	C19. Superficie agrícola en el marco de la agricultura ecológica	C37. Agricultura con alto valor natural
C8. PIB per cápita	C20. Tierra de regadío	C38. Bosques protegidos
C9. Tasa de pobreza	C21. Unidades de ganado mayor	C39. Extracción de agua en la agricultura
C10. Estructura de la economía	C22. Mano de obra agrícola	C40. Calidad del agua
C11. Estructura del empleo	C23. Estructura de edades de los agricultores	C41. Materia orgánica del suelo en tierras de cultivo
C12. Productividad laboral por sector económico	C24. Formación agrícola de los agricultores	C42. Erosión del suelo por la acción del agua
	C25. Renta de los factores agrícolas	C43. Producción de energía renovable procedente de la agricultura y la silvicultura
	C26. Renta empresarial agraria	C44. Consumo de energía en la agricultura, la silvicultura y la industria alimentaria
	C27. Productividad total de los factores en la agricultura	C45. Emisiones de gases procedentes de la agricultura
	C28. Formación bruta de capital fijo en la agricultura	
	C29. Bosques y otras superficies forestales	
	C30. Infraestructura turística	

Bajo nuestro criterio, las variables o temas que recogen se pueden agrupar o sintetizar en:

- Población. (C1, C2, C4).
- Territorio. Sería una variable que recoge datos geográficos de superficie. (C3).
- Empleo. (C5, C6, C11, C13).
- Paro (C7).
- Producción y renta (C8, C9, C10).

- Productividad laboral (C12).
- Sector primario. Datos específicos del sector agrario, explotaciones, rentas, empleo, productividad, etc. (C14 a C29).
- Sector turístico (C30).
- Medioambiente (C31 a C45).

En el anterior periodo de programación el “MANUAL SOBRE EL MARCO COMÚN DE SEGUIMIENTO Y EVALUACIÓN” (2006) define los indicadores de dicho periodo:

*“Indicadores de medios y recursos. Se refieren al presupuesto o a otros recursos asignados a cada nivel de la asistencia. Los indicadores de contribuciones financieras se utilizan para realizar un seguimiento de los avances desde el punto de vista del compromiso y el pago (anual) de los fondos disponibles para toda operación, medida o programa en relación con sus costes subvencionables. Ejemplo: gasto por medida declarado a la Comisión.*

***Indicadores de ejecución.** Miden las actividades realizadas directamente dentro de los programas. Estas actividades son el primer paso para la consecución de los objetivos operativos de la intervención y se miden en unidades físicas o monetarias. Ejemplo: número de sesiones de formación organizadas, número de explotaciones que reciben ayudas a la inversión, volumen total de inversión.*

***Indicadores de resultados.** Miden los efectos directos e inmediatos de la intervención. Proporcionan información sobre los cambios de comportamiento, capacidad o actuación de los beneficiarios directos y se miden en términos físicos o monetarios. Ejemplo: número bruto de puestos de trabajo creados, resultados satisfactorios de la formación.*

***Indicadores de repercusiones.** Se refieren a las ventajas del programa más allá de los efectos inmediatos en su beneficiarios directos, no sólo desde el punto de vista de la intervención, sino también y de manera más general en el ámbito del programa. Estos indicadores están vinculados con objetivos del programa más amplios. En general se expresan en términos «netos», lo que supone descontar efectos que no pueden*

*atribuirse a la intervención (por ejemplo, recuento doble, peso muerto) y tener en cuenta los efectos indirectos (desplazamiento y multiplicadores). Ejemplo: aumento del empleo en las zonas rurales, mayor productividad en el sector agrario, mayor producción de energía renovable.*

**Indicadores de base** *Los indicadores de base se utilizan en el análisis SWOT y para la definición de la estrategia de los programas. Pertenecen a dos categorías:*

- *Indicadores de base relacionados con los objetivos. Están vinculados directamente con objetivos del programa más amplios. Se utilizan para efectuar el análisis SWOT en relación con los objetivos determinados en el Reglamento. También se utilizan como base (o referencia) con respecto a la cual se evalúa la repercusión de los programas. Los indicadores de base reflejan la situación al comienzo del periodo de programación y la tendencia a lo largo del tiempo. La evaluación de la repercusión debe reflejar esa parte del cambio a lo largo del tiempo que puede atribuirse al programa una vez que se han tenido en cuenta la tendencia de base y otros factores.*
- *Indicadores de base relacionados con el contexto. Proporcionan información sobre aspectos pertinentes de las tendencias contextuales generales que pueden influir en la actuación del programa. Estos indicadores tienen, pues, una doble finalidad: i) contribuir a determinar los puntos fuertes y las deficiencias de la región y ii) ayudar a interpretar las repercusiones logradas en el programa a la luz de las tendencias económicas, sociales, estructurales o medioambientales generales.”*

En este caso serían de referencia los indicadores de base y los de repercusiones fundamentalmente. Aunque utilicemos de referencia los anteriormente descritos, correspondientes al próximo periodo de programación, se puede comprobar que aunque se observe alguna discrepancia entre periodos en la clasificación o planteamiento de indicadores, siempre se persigue actuar, entre otras, sobre determinadas variables del territorio de naturaleza económica cuya medición es pertinente para valorar el mayor o menor éxito de estos programas. La medición directa se realiza en las correspondientes evaluaciones, utilizando estos indicadores, pero no se realiza un estudio espacial o territorial basado en la teoría económica y con técnicas de econometría espacial.

Para estos aspectos, que son referencia de la política de desarrollo rural de la UE, y que deducimos de sus métodos de medición, vamos a buscar variables estadísticas que los definan o resuman, siempre que tengan coherencia desde el punto de vista de la teoría económica, que estén disponibles al nivel de desagregación conveniente para aplicar el enfoque y las técnicas de la econometría espacial, y siempre también que permitan complementarse con otras variables que la teoría económica considere relevantes.

## VARIABLES

Cuadro 3.3. Variables utilizadas.

GRUPO DE VARIABLES	VARIABLES	DESCRIPCIÓN	UNIDAD	FUENTE
<b>Inversión pública.</b>	<b>IPDR13.07</b>	Suma de la cuantía total de dinero público destinado a pagos a beneficiarios finales de ayudas públicas cofinanciadas por FEADER dentro del PDR 2007-2013 del Gobierno del Principado de Asturias.	Euros	Consejería de Agroganadería y Recursos Autóctonos.
	<b>I1PIL13.07</b>	Suma de la cuantía total de dinero público destinado a pagos a beneficiarios finales de ayudas públicas correspondientes al primer pilar de la PAC (financiación FEAGA) del Gobierno del Principado de Asturias en el periodo 2007-2013.	Euros	Consejería de Agroganadería y Recursos Autóctonos.
	<b>IPAC13.07</b>	Suma de la cuantía total de dinero público destinado a pagos a beneficiarios finales de ayudas públicas correspondientes al primer y segundo pilar de la PAC (financiación FEADER y FEAGA) del Gobierno del Principado de Asturias en el periodo 2007-2013.	Euros	Consejería de Agroganadería y Recursos Autóctonos.
<b>Población.</b>	<b>POBPAD14.</b>	Población según Padrón Municipal de Habitantes en 2014	Habitantes	INE. Padrón Municipal de Habitantes. Elaborado por SADEI.
	<b>VPOBPAD14.07</b>	Variación de la población para el periodo 2007-2013 según Padrón Municipal de Habitantes.	Habitantes	Elaboración propia a partir de datos del INE. Padrón Municipal de Habitantes

Empleo y paro.	<b>CONTRAT13</b>	Contratos realizados en 2013	Contratos	SADEI
	<b>PAR14</b>	Paro registrado en 2014	Personas	Servicios Públicos de Empleo. Estatal (SPEE) y Autonómico (SEPEPA).
	<b>VPAR14.07</b>	Variación del paro registrado en 2014 con respecto al registrado en 2007	Personas	Elaboración propia a partir de datos de los Servicios Públicos de Empleo Estatal (SPEE) y Autonómico (SEPEPA)
	<b>EMP13</b>	Empleo en 2013	Empleos	SADEI
	<b>VEMP13.07</b>	Variación del empleo en el periodo 2007-2013	Empleos	Elaboración propia a partir de datos de SADEI
	<b>EMPAGRI13</b>	Empleo en el sector agricultura y pesca en 2013	Empleos	SADEI
	<b>VEMPAGRI13.07</b>	Variación del empleo en el sector agricultura y pesca en el periodo 2007-2013	Empleos	Elaboración propia a partir de datos de SADEI.
	<b>ASS14</b>	Afiliaciones a la Seguridad Social en diciembre de 2014	Personas	Servicio Público de Empleo
Producción y renta	<b>RNET10</b>	Renta disponible neta en 2010	Miles de euros	SADEI. La renta de los municipios
	<b>RFAMNET10</b>	Renta familiar disponible ajustada neta por habitante en 2010	Euros	SADEI. La renta de los municipios
	<b>VAB10</b>	Valor añadido bruto a precios básicos en 2010	Miles euros	SADEI. La renta de los municipios
	<b>VABAGRI10</b>	Valor añadido bruto a precios básicos en el sector agricultura y pesca en 2010	Miles euros	SADEI. La renta de los municipios
Estructura económica	<b>SUPTECHIND12</b>	Superficie total de techo industrial en 2012	Metros cuadrados	Cámaras Oficiales de Comercio, Industria y Navegación de Oviedo, Gijón y Avilés.
	<b>LIAE13</b>	Número de licencias del IAE en 2013	Licencias	Cámaras Oficiales de Comercio, Industria

				y Navegación de Oviedo, Gijón y Avilés.
	<b>LIAEAGRI13</b>	Número de licencias del IAE en agricultura en 2013	Licencias	Cámaras Oficiales de Comercio, Industria y Navegación de Oviedo, Gijón y Avilés.
	<b>LICOBRI13</b>	Licencias de obra concedidas en 2013	Licencias	SADEI. Ministerio de Fomento
	<b>VEHIC13</b>	Parque de vehículos en 2013	Vehículos	Ministerio del Interior. Dirección General de Tráfico.
<b>Sector primario</b>	<b>EXPLOT14</b>	Número de explotaciones en 2014	Explotaciones	Consejería de Medio Rural y Pesca. Elaborado por SADEI
	<b>ANIM14</b>	Número de animales en 2014	Animales	Consejería de Medio Rural y Pesca. Elaborado por SADEI
	<b>CORTMAD12</b>	Cortas de madera en 2012	Metros cúbicos	Consejería de Medio Rural y Pesca. Elaborado por SADEI
<b>Turismo</b>	<b>REST13</b>	Restaurantes en 2013	Restaurantes	Principado de Asturias. Registro de empresas y actividades turísticas. Elaborado por SADEI.
	<b>PLAZTUR13</b>	Plazas en establecimientos turísticos en 2013	Plazas de alojamiento	Principado de Asturias. Registro de empresas y actividades turísticas. Elaborado por SADEI.

## CAPÍTULO 4. ESTIMACIÓN Y MODELO ECONÓMÉRICO.

---

### 4.1. ANÁLISIS EXPLORATORIO DE DATOS ESPACIALES (AEDE).

Vamos a analizar la posible existencia de autocorrelación espacial de cada variable, tanto con técnicas gráficas como con contrastes estadísticos, para contrastar la existencia de autocorrelación espacial global así como la existencia de autocorrelación espacial local. También se estudia en este capítulo la posible presencia de heterogeneidad espacial, ya sea asociada o no a la existencia de autocorrelación espacial. Se distingue entre análisis univariante, para cada variable, y el multivariante para hacer un análisis conjunto de dos o más variables, centrándonos, para el caso que nos ocupa, en el primero de ellos.

Se aplican, a las variables seleccionadas, las técnicas gráficas del análisis exploratorio espacial que proporciona el programa GeoDa. Concretamente se estudia para el análisis univariante el histograma de frecuencias, mapa de deciles y mapa de desviación típica. También el diagrama de dispersión de Moran. Para la representación de las observaciones atípicas que se producen a nivel local se calculan los gráficos de caja (diagrama y mapa), mapa de percentiles, cartograma y mapas LISA de dependencia espacial local. Vamos a aplicar el enfoque reticular (lattice) dado el carácter discreto de las unidades administrativas existentes (concejos) y de la información estadística disponible que se basa en valores de dichas unidades espaciales. Se plantea una matriz de contigüidad de tipo reina.

#### 4.1.1. ANÁLISIS UNIVARIANTE.

Como son muchas variables, vamos a presentar el análisis univariante comentando los resultados de las mismas por grupos, utilizando los grupos temáticos de variables que se expusieron previamente en el cuadro 4.3. En el desarrollo de cada bloque se detalla para cada prueba los resultados independientes de las distintas variables, salvo cuando todas tengan un mismo comportamiento.

## Inversión Pública.

### IPDR13.07, IIPIL13.07, IPAC13.07.

**Histograma de frecuencias.** La mayoría de concejos (más de 50) se sitúan en el primer intervalo, para las tres variables. La frecuencia va disminuyendo bruscamente para el segundo y tercer intervalos. Las distribuciones muestran forma de cola a la derecha en los tres casos. (Ver gráficos 1,2 y 3 del anexo 1)

### Mapa de deciles.

Los concejos rurales más grandes y con más actividad agraria de las zonas oriental y occidental se colorean con los tonos más oscuros, que corresponden con las inversiones más altas. Los concejos rurales más pequeños y los más urbanos de la zona centro se muestran con los colores más claros, de menor inversión. (Ver mapas 1, 2 y 3 del anexo 1).

### Diagrama de caja y mapa de caja.

Los **diagramas y mapas de caja**, para las tres variables consideradas, no reflejan outliers por debajo de la cota inferior. Si se obtienen, en cambio, por encima de la cota superior, 5 concejos para IPDR13.07 (Llanes en el oriente y Valdés, Tineo, Salas y Cangas del Narcea en occidente) y tres para IIPIL13.07 y para IPAC13.07 (Valdés, Tineo Cangas del Narcea), todos ellos concejos extensos y con un importante sector primario. (Ver gráficos 4,5 y 6 y mapas 4, 5 y 6 del anexo 1).

### Diagrama de dispersión de Moran.

El índice de Moran tiene un valor en los tres casos superior a 0.25 y es mayor a su valor esperado ( $E(I) = -0.0130$ ), lo que sería indicativo de autocorrelación espacial positiva. Se realiza un test de permutaciones (con 999 permutaciones), obteniendo una distribución con un p-valor inferior a 0.002. Al repetir la prueba salen distintos valores siempre inferiores, por lo que se rechaza la hipótesis nula de no existencia de autocorrelación espacial, con un nivel de significatividad del 99%. Esta prueba indica la posible existencia de autocorrelación espacial en las tres variables de inversión pública considerada. (Ver gráficos 7, 8 y 9 del anexo 1).

Si eliminamos los concejos que se consideran atípicos según el criterio del mapa de caja, vemos que se incrementa la pendiente de la recta de regresión para IPDR13.07, aunque no demasiado, y prácticamente nada en el caso de las otras dos variables consideradas. (Ver gráficos 10, 11 y 12 del anexo 1).

### **Mapa de desviación típica.**

Tres concejos correspondientes a los concejos de Valdés, Tineo y Cangas del Narcea se sitúan en el intervalo superior de tres desviaciones típicas por encima de la media (color rojo en los mapas). La mayoría de concejos se sitúa en el intervalo de una desviación típica por debajo de la media (color azul en los mapas). (Ver mapas 7, 8 y 9 del anexo 1).

### **Mapa de percentiles.**

Un concejo en el primer y último centil en cada variable.

IPDR13.07. Centil 1, Muros de Nalón. Centil 100, Cangas del Narcea.

IIPIL13.07. Centil 1, Ribera de Arriba. Centil 100, Tineo.

IPAC14.07. Centil 1, Muros de Nalón. Centil 100, Tineo.

(Ver mapas 10, 11 y 12 del anexo 1).

### **Cartograma.**

Se representan en círculos, categorizados por colores y tamaños, los cuatro intervalos, así como los valores extremos ( en rojo los mas altos y en azul los mas bajos), que coinciden con los ya descritos en el mapa y diagrama de caja, todos sobre la cota superior. (Ver mapas 13, 14 y 15 del anexo 1).

### **Mapas LISA.**

Para los mapas de clusters basados en el índice local de Moran, calculados con 999 permutaciones en el proceso de aleatorización, y una matriz de pesos espaciales con criterio de contigüidad de tipo queen, se obtienen fundamentalmente dos clusters, uno, el más grande, en la zona occidental con valores altos, rodeados de valores altos, zona

---

de fuerte presencia del sector primario. Y otro cluster de valores significativos que rechazan la hipótesis nula de no existencia de autocorrelación espacial, se da en la zona centro, con valores bajos rodeados de valores bajos. Si se calculan tomando como base la G de Getis y Ord los resultados son similares también para las tres variables. (Ver mapas 16, 17, 18, 19, 20 y 21 del anexo 1).

## Población

### POBPAD14

#### **Histograma de frecuencias.**

Más del 90% (72 concejos) en el primer intervalo que va de 166 habitantes a 34600). La media es 13612,3 y la mediana 2459.5. (Ver gráfico 13 del anexo 1).

#### **Mapa de deciles.**

En el decil más alto, con el color más oscuro, se muestran los 8 concejos con más de 17000 habitantes, en la zona central de Asturias, que son los más urbanizados e industrializados (Oviedo, Gijón, Avilés, Langreo, Mieres, Siero, Castrillón y San Martín del Rey Aurelio). Los más bajos, que se corresponden con el decil mas claro en el mapa, están, la mayor parte, en el ala occidental y tienen valores por debajo de los 498 habitantes (Yernes y Tameza, Santo Adriano, Pesoz, Villanueva de Oscos, Illano, San Martín de Oscos, Santa Eulalia de Oscos y San Tirso de Abres). (Ver mapa 22 del anexo 1).

#### **Diagrama de caja y mapa de caja.**

Considera valores extremos, por encima de la cota superior, los de los seis concejos con las ciudades más pobladas (Gijón, Oviedo, Avilés, Siero, Langreo y Mieres) y una importante tradición industrial. No considera ninguno por debajo de la cota inferior. El resto se incluyen mayoritariamente en el segundo y tercer cuartil (20 y 20), dejando 19 y 13 para el primer y cuarto cuartil respectivamente. (Ver gráfico 14 y mapa 15 del anexo 1).

### **Diagrama de dispersión de Moran.**

Al igual que con el método anterior se consideran atípicos los seis concejos ya descritos. Los valores se concentran dando muestras de posible autocorrelación espacial, aunque no siguen claramente la forma de la recta ajustada y el valor de  $I$  se acerque más a cero y esté alejado de 1. El índice de Moran es mayor a su valor esperado ( $E(I) = -0.0130$ ), lo que sería indicativo de autocorrelación espacial positiva. Si realizamos un proceso de aleatorización para contrastar la significatividad de  $Z(I)$  (con 999 permutaciones), obtenemos una distribución con un p-valor superior a 0.05. Al repetir la prueba salen distintos valores siempre superiores a 0.06, incluso cercanos a 0.1, por lo que no se puede rechazar la hipótesis nula de no existencia de autocorrelación espacial con un nivel de significatividad del 95% y si que se rechazaría al 90%. Cambios en la matriz de pesos espaciales cambian el valor de  $I$  pero no las conclusiones ni resultados obtenidos. (Ver gráfico 15 del anexo 1).

Si eliminamos los valores atípicos, según el criterio del diagrama de caja, aumenta la pendiente de la recta de regresión del diagrama de dispersión, lo que hace suponer una elevada correlación espacial. Pero al realizar un proceso de aleatorización con 999 permutaciones y repetirlo varias veces, siempre nos sale un resultado con un p valor superior a 0.05, por lo que no rechazaríamos la hipótesis nula de no existencia de autocorrelación espacial. Si que la rechazaríamos al 90% ya que siempre da un p valor menor de 0.1. (Ver gráfico 16 del anexo 1).

### **Mapa de desviación típica.**

En el intervalo más alto (tres veces la desviación típica) se encuentran dos concejos (Gijón y Oviedo), uno en el segundo (Avilés), 11 en el tercero y los 64 restantes en el primero, que se encuentra una unidad de desviación típica por debajo de la media. Ninguno está a dos o tres unidades por debajo de la media. (Ver mapa 23 del anexo 1).

### **Mapa de percentiles.**

Dos son los concejos con valores extremos, Yernes y Tameza en el percentil 1 y Gijón en el 100. (Ver mapa 24 del anexo 1).

## **Cartograma.**

Seis son los concejos con valores extremos en el intervalo más alto (Oviedo, Gijón, Avilés, Langreo, Mieres y Siero), no encontrándose ninguno en el más bajo. (Ver mapa 25 del anexo 1).

## **Mapas LISA.**

*Basados en I Local de Moran*

El mapa LISA considera la existencia de autocorrelación espacial local significativa en 11 concejos Seis al 95% y 5 al 99%. El procedimiento se basa en un test de aleatorización con 999 permutaciones. Los concejos son Taramundi y Vegadeo en el ala occidental, y Corvera de Asturias, Ribera de Arriba, Langreo, Noreña (estos al 95%), San Martín del Rey Aurelio, Siero, Sariego, Llanera y Carreño.

El mapa de clusters los clasifica como bajo-bajo Taramundi y Vegadeo, alto-bajo San Martín del Rey Aurelio, bajo-alto Noreña, Ribera de Arriba, Carreño y Sariego y alto-alto los cuatro restantes, Siero, Langreo, Llanera y Corvera de Asturias. (Ver mapa 26 del anexo 1).

*Basados en G Local de Getis y Ord.*

$G_i$

Considera significativos también 11 concejos si seguimos el procedimiento de aleatorización con 999 permutaciones para el cálculo del pseudo p-valor. Se estandariza por filas la matriz de pesos espaciales. Estos concejos serían los mismos que en el caso de la I de Moran salvo dos, Ribera de Arriba y Noreña. Sí que son significativos con este estadístico Boal y Villaviciosa. La significatividad es al 95% salvo en Siero, Llanera y San Martín del Rey Aurelio que es al 99%.

En cuanto al mapa de clusters considera de valor bajo Taramundi, Vegadeo, San Martín del Rey Aurelio y Boal y de valores altos el resto. Si se realiza el test de aleatorización con 999 permutaciones solo son significativos al 95% Llanera y Carreño formando un cluster de valores altos. (Ver mapa 27 del anexo 1).

$G_i^*$ 

Con este estadístico se obtienen los mismos resultados en los mapas. Si se asume la aleatorización con 999 permutaciones solo Llanera es significativo. (Ver mapa 28 del anexo 1).

Ante la sensibilidad de los estadísticos a la matriz de pesos espaciales se prueba con distintas matrices obteniendo resultados que consideran nuevos concejos como significativos, y viceversa, es decir presenta cambios, pero en todos los casos se da la característica común de considerar a grandes rasgos dos clusters, uno de valores altos en la zona centro, donde se concentran los valores atípicamente altos, y otro de valores bajos en la zona occidental, en el entorno de los concejos con valores muy bajos.

#### **VPOBPAD14.07**

**Histograma de frecuencias.** La variación se acumula en el tercer intervalo (más del 80%) que va de perder más de 1000 habitantes a ganar menos de 200). (Ver gráfico 17 del anexo 1).

#### **Mapa de deciles.**

Se registran las variaciones más altas correspondientes al décimo decil, con valores positivos, en ocho concejos. Llanes en el oriente y los otros siete en la zona central más poblada que incluyen Oviedo, Gijón y limítrofes. Los intervalos más oscuros, correspondientes a los valores más altos, predominan por tanto, en la zona centro y las alas. Son mas frecuentes los tonos oscuros en concejos pequeños que han perdido menos población por su menor tamaño y mayor estabilidad. Las mayores pérdidas corresponden a concejos grandes de las alas, sobre todo occidental y a la zona centro de interior (cuencas mineras). (Ver mapa 29 del anexo 1).

#### **Diagrama y mapa de caja.**

Considera atípicos por encima de la cota superior cinco concejos de la zona centro, coincidentes con los dos más poblados y otros tres contiguos a ellos (Oviedo, Gijón, Corvera de Asturias, Siero y Llanera). Por debajo de la cota inferior nueve concejos,

que coinciden con algunos de los más industrializados, en proceso de reconversión y también algunos grandes de occidente de carácter más rural (Avilés, Aller, Lena, Langreo, Mieres, San Martín del Rey Aurelio, Cangas del Narcea, Valdés y Tineo). (Ver gráfico 18 y mapa 30 del anexo 1).

### **Diagrama de dispersión de Moran.**

El índice de Moran tiene un valor de 0.0482384, y aunque cercano a cero y alejado de 1, es mayor a su valor esperado ( $E(I) = -0.0130$ ), lo que sería indicativo de autocorrelación espacial positiva. Se realiza un test de permutaciones (con 999 permutaciones), obteniendo una distribución con un p-valor superior a 0.05. Al repetir la prueba salen distintos valores siempre superiores a 0.1, por lo que no se rechaza la hipótesis nula de no existencia de autocorrelación espacial con un nivel de significatividad del 95%. (Ver gráfico 19 del anexo 1).

Al eliminar los 14 concejos considerados outliers según el criterio del diagrama de caja (9 inferiores y 5 superiores) la poca pendiente de la recta del diagrama de dispersión de Moran aumenta mostrando una importante dependencia espacial, aunque el proceso de aleatorización con 999 permutaciones no rechaza la hipótesis nula de no existencia de autocorrelación espacial. (Ver gráfico 20 del anexo 1).

### **Mapa de desviación típica**

Dos concejos (Oviedo y Siero) figuran en el intervalo donde se supera la media en tres desviaciones estándar, y otros dos (Langreo y Mieres) en el que está tres desviaciones estándar por debajo de la media. Los cuatro concejos son limítrofes. Mieres y Langreo son los concejos más poblados de las cuencas mineras del Caudal y del Nalón, respectivamente. (Ver mapa 31 del anexo 1).

### **Mapa de percentiles.**

Son en este caso Oviedo en el percentil 100 y Mieres en el 1 los valores extremos de la muestra. (Ver mapa 32 del anexo 1).

### **Cartograma.**

Nueve concejos en el intervalo más bajo y 5 en el más alto son los outliers considerados, los mismos que en el diagrama y mapa de caja ya que se usa el mismo criterio. (Ver mapa 33 del anexo 1).

### **Mapas LISA.**

Con una matriz tipo reina y un test de 999 permutaciones para contrastar la significatividad de la I de Moran, se obtienen clusters de autocorrelación espacial local en la zona centro de la región, tal como se observa en el mapa de clusters, que colorea los concejos estadísticamente significativos para los que se detecta existencia de autocorrelación espacial local. (Ver mapa 34 del anexo 1).

Para la G y G\* local de Getis y Ord, los resultados son similares, ya que coinciden todos los concejos. La diferencia con la G\* es sobre el tipo de autocorrelación, ya que esta considera valores altos aquellos que la G considera bajos y viceversa. (Ver mapas 35 y 36 del anexo 1).

## **Empleo y paro.**

### **CONTRAT13, PAR14, EMP13, EMPAGRI13 y ASS14.**

Las cinco variables presentan **histogramas de frecuencias** (ver gráficos 21, 22, 23, 24 y 25 del anexo 1) con la mayor parte de concejos concentrados en el primer intervalo de valores más bajos. En el **mapa de deciles** (ver mapas 37, 38, 39, 40 y 41 del anexo 1) se corresponden, con los valores más altos, en tonos más oscuros, concejos de la zona central, más dinámica, en las variables ASS14, CONTRAT13, PAR14 y EMP13. Cambia en el empleo agrario, EMPAGRI13, donde no se reflejan algunos concejos del área central, más urbanos, y si figuran en los deciles altos concejos de occidente, más rurales. En el empleo total EMP13, también figuran concejos de las alas, quizá por la aportación del empleo agrario, y en la zona oriental Llanes, concejo con amplia oferta turística, está también en deciles altos.

Los **diagramas y mapas de caja** (ver gráficos 26, 27, 28, 29 y 30 y mapas 42, 43, 44, 45 y 46 del anexo 1) muestran concejos con valores atípicos por encima cotas superiores en las cinco variables, pero no por debajo de las cotas inferiores. Para la variable CONTRAT13 son ocho los concejos (Avilés, Gijón, Langreo, Llanera, Llanes, Mieres, Siero y Oviedo). Para PAR14 y EMP13 son siete (los mismos menos Llanes). Cambia EMPAGRI13, donde se consideran atípicos cuatro concejos, tres de occidente, Tineo, Valdés, y Cangas de Narcea, con más peso agrario en su economía, y Gijón que tiene alto empleo agrario en su zona rural, a pesar de tener la zona urbana más poblada de Asturias. En las afiliaciones a la seguridad social son 10 concejos limítrofes de la zona centro los outliers (Avilés, Carreño, Castrillón, Corvera, Gijón, Langreo, Llanera, Mieres, Siero y Oviedo).

Los **diagramas de dispersión de Moran** (ver gráficos 31, 32, 33, 34 y 35 del anexo 1) muestran índices y comportamientos parecidos en las variables CONTRAT13 y PAR14. Es algo superior la pendiente de la recta en la variable EMP13 y más aún en EMPAGRI13. Los valores de I son superiores al valor esperado en los cinco casos. El ajuste gráfico a la nube de puntos no da evidencias de autocorrelación espacial, salvo en el caso de EMPAGRI13, donde sí sigue un patrón algo más claro. El índice de Moran es mayor a su valor esperado en todos los casos ( $E(I) = -0.0130$ ), lo que sería indicativo de autocorrelación espacial positiva. Cambian los resultados si realizamos un proceso de aleatorización para contrastar la significatividad de Z (I) (con 999 permutaciones). Obtenemos una distribución con un p-valor superior a 0.05 para CONTRAT13 y PAR14, y al repetir la prueba salen distintos valores siempre superiores a 0.06, incluso cercanos a 1, por lo que no se rechazaría la hipótesis nula de no existencia de autocorrelación espacial con un nivel de significatividad del 95% y se rechazaría al 90%. Para ASS14 y EMP13 ocurre lo mismo en la mayoría de las pruebas, pero en algunas sí que presenta valores inferiores a 0.05, por lo que aunque mayoritariamente no se rechazaría la no existencia de autocorrelación espacial, existen ciertas dudas. En el caso de EMPAGRI13, sí que se rechaza la hipótesis nula y se acepta la existencia de autocorrelación espacial, ya que se obtienen p-valores inferiores a 0.01, por lo que el nivel de significatividad es incluso más alto (99%).

Si eliminamos los valores atípicos aumentan las pendientes de la recta de regresión del diagrama de dispersión de Moran, salvo para EMPAGRI13, aunque la disminución es

tan pequeña que se queda prácticamente igual. El aumento es considerable en el resto, especialmente en CONTRAT13, PAR14 y EMP13, lo que indica que los valores atípicos pueden estar distorsionando y ocultando la alta dependencia espacial. (Ver gráficos 36, 37, 38, 39 y 40).

Los **mapas de desviación típica** (ver mapas 47, 48, 49, 50 y 51 del anexo 1) no presentan concejos situados en los dos intervalos más bajos. Oviedo y Gijón están en el más alto y la mayor parte de concejos están en el tercer intervalo, es decir en una medida de la desviación típica por debajo de la media. Ocho concejos (nueve en el caso de CONTRAT13), de la zona centro y limítrofes con Oviedo y Gijón, están en el cuarto y quinto intervalo (junto con Llanes en el caso de CONTRAT14). Cambia el mapa de EMPAGRI13, en el que los concejos del intervalo más alto son Tineo, Valdés y Cangas del Narcea, y aunque 51 se encuentran en el tercer intervalo, el resto que se encuentran en el cuarto y quinto se reparten por toda la geografía y no solo se concentran en el área central.

Los **mapas de percentiles** (ver mapas 52, 53, 54, 55 y 56 del anexo 1) muestran un concejo en el percentil 1 y otro en el 100 en todas las variables. En el caso de paro y contrataciones son el concejo de Yernes y Tameza y el concejo de Gijón. Para el empleo total y las afiliaciones a la seguridad social son Yernes y Tameza y Oviedo. Para el empleo agrario Noreña y Tineo.

**Cartogramas.** (Ver mapas 57, 58, 59, 60 y 61 del anexo 1). Tampoco presentan outliers inferiores. Si que los presentan superiores, que coinciden con concejos más industrializados o urbanizados de la zona central, salvo en el caso del empleo del sector primario que presenta cuatro outliers en Tineo, Valdés, Cangas del Narcea y Gijón.

### **Mapas LISA.**

El índice local de Moran detecta clusters en el centro y occidente de la región para las afiliaciones a la seguridad social, el paro, el empleo total y los contratos realizados. El de occidente de valores bajos rodeados de valores también bajos (Low-Low). El de la zona centro es de valores altos-altos y también con algún concejo de valores bajos rodeado de valores altos. Para EMPAGRI13 se detecta autocorrelación local en concejos grandes de occidente con valores altos-altos y alguno bajo-alto. También algún

concejo pequeño de occidente y zona centro presenta dependencia local con valores bajo-bajo. (Ver mapas 62, 63, 64, 65 y 66 del anexo 1).

Si calculamos los índices locales de autocorrelación espacial  $G$  y  $G^*$  se obtienen resultados similares usando la misma matriz de pesos espaciales de tipo queen, ya que con la diferencia de algún concejo limítrofe significativo, se localizan los grupos con autocorrelación espacial local en las mismas zonas de la región. (Ver mapas 67, 68, 69, 70, 71, 72, 73, 74, 75 y 76 del anexo 1).

### **VPAR14.07, VEMP13.07 y VEMPAGRI13.07**

#### **Histograma de frecuencias**

En las tres variables más de 70 concejos se concentran en uno de los ocho intervalos. En el caso del empleo total y del agrario se concentran en el último intervalo (desde perder 2.210 empleos hasta crear 196) y penúltimo (desde perder 74 hasta perder 2), respectivamente. Son variables fundamentalmente negativas. En cuanto a la variable de variación del paro, fundamentalmente con valores positivos, la mayoría de concejos están en el tramo que va de perder cuatro parados a ganar 1660. (Ver gráficos 41, 42 y 43 del anexo 1)

#### **Mapa de deciles.**

Se observan mayores valores en la variación del paro en las zonas del área central, así como los menores valores en la variación del empleo (mayores en valor absoluto), es decir la zona central presenta las mayores oscilaciones en las variaciones de empleo y paro. En el empleo agrario se observa mayor dispersión en las oscilaciones, aunque sus valores negativos más bajos se dan en algunos concejos de la zona occidental y algunos también de la zona centro. (Ver mapas 77, 78 y 79 del anexo 1)

#### **Diagrama de caja y mapa de caja.**

Seis concejos son outliers, por debajo de la cota inferior, en la variación del empleo agrario, Gijón y Oviedo, que son a su vez los dos más poblados, junto con Valdés, Tineo, Salas y Cangas del Narcea, en occidente y con gran peso ganadero en la región.

Para la variación del empleo total son ocho los aún más bajos que la cota inferior y corresponden con concejos limítrofes del área central y con zonas más urbanas e industriales. Para la variación del paro son seis los outliers superiores y también en la zona más poblada y urbana. (Ver gráficos 44, 45, 46 y mapas 80, 81 y 82 del anexo 1).

### **Diagrama de dispersión de Moran.**

La I de Moran presenta valores bajos para las tres variables, especialmente para las variaciones de paro y empleo total. Es algo superior en la variación del empleo agrario. En los tres casos es superior al valor esperado. Los test de permutaciones realizados con 999 permutaciones consideran que no se puede rechazar la hipótesis nula de no existencia de autocorrelación espacial, a un nivel de significatividad del 95% en el caso de VPAR13.07 y de VEMP14.07. Si que rechazamos la hipótesis nula en el caso de VEMPAGRI14.07 ya que da un p-valor siempre inferior a 0.05. (Ver gráficos 47, 48 y 49 del anexo 1). Para las tres variables aumenta la pendiente de la recta del diagrama de dispersión de Moran al eliminar los valores atípicos según el criterio del diagrama de caja. Podrían estar distorsionando los resultados del estadístico I de Moran. (Ver gráficos 50, 51 y 52 del anexo 1).

### **Mapa de desviación típica**

Mapas con outliers en concejos del área central para VPAR13.07 y de VEMP14.07. Para VEMPAGRI14.07 cinco outliers, tres en occidente (Valdés, Salas y Tineo), junto con Oviedo y Gijón. (Ver mapas 83, 84 y 85 del anexo 1).

### **Mapa de percentiles.**

Un concejo en el primer y último centil en cada variable. (Ver mapas 86, 87 y 88 del anexo 1).

VPAR13.07. Centil 1, Degaña. Centil 100, Gijón.

VEMP14.07. Centil 1, Oviedo. Centil 100, Belmonte de Miranda.

VEMPAGRI14.07. Centil 1, Gijón. Centil 100, Lena.

## Mapas LISA.

Si consideramos la I local de Moran, y el test de permutaciones de 999, obtenemos unos resultados que consideran significativos dos grupos de concejos, uno en la zona central y otro en la occidental. Estos concejos no son los mismos en las tres variables, pero siguen ese patrón geográfico. La diferencia más importante se observa en el comportamiento de VEMPAGRI14.07 que considera representativos en occidente a los concejos más grandes en territorio y en peso ganadero, mientras que las otras dos variables consideran significativos en occidente a concejos muy pequeños y con menor peso en la economía regional. (Ver mapas 89, 90 y 91 del anexo 1).

Para la G y G\* locales de Getis y Ord, se obtienen resultados similares en cuanto a las zonas en las que se concentran los clusters de autocorrelación espacial local. (Ver mapas 92, 93, 94, 95, 96 y 97 del anexo 1).

## Producción y renta.

### RNET10, RFAMNET10, VAB10, VABAGRI10

Los **histogramas de frecuencias** reflejan la concentración en el primer intervalo en todas las variables salvo en RFAMNET10, que muestra una distribución más repartida, acumulándose las frecuencias más altas en los cuatro últimos intervalos. (Ver gráficos 53, 54, 55 y 56 del anexo 1).

Los **mapas de deciles** muestran la concentración de los valores más altos en el área central más poblada, salvo en el VAB agrario que también alcanza valores altos en algunos concejos de occidente, con más tradición ganadera, más poblados y de mayor superficie que sus vecinos. Los más bajos suelen estar en concejos de interior, con menor población y superficie. (Ver mapas 98, 99, 100 y 101 del anexo 1).

Los **diagramas y los mapas de** caja recogen un solo outlier, situado por debajo de la cota inferior en RFAMNET10, y se corresponde con Yernes y Tameza. Para el resto de variables los outliers son más altos que a la cota superior, y se corresponden con Tineo y Valdés para VABAGRI10. Para la renta neta y el valor añadido bruto los outliers son

los concejos con las áreas más urbanas de la zona centro. Lo mismo se observa en los correspondientes **cartogramas**, (que utilizan el mismo criterio para los valores atípicos). (Ver gráficos 57, 58, 59, 60 y mapas 102, 103, 104, 105, 106, 107, 108 y 109 del anexo 1).

Los **diagramas de dispersión de Moran** para las cuatro variables son positivos y mayores que E (I). Los test de permutación realizados sugieren rechazar la hipótesis nula de no existencia de autocorrelación espacial con una significatividad del 99% en los casos de RFAMNET10 y VABAGRI10. En el caso de RNET10 no podemos rechazar la hipótesis nula de no existencia de autocorrelación espacial con una significatividad del 95%. Para VAB10 la sucesiva realización de distintos test de permutaciones (con 999 permutaciones), arroja distintos resultados del p valor, en unos casos con valores ligeramente superiores a 0.05 y en otros casos con valores ligeramente inferiores, por lo que presenta dudas a la hora de considerar la existencia o no de autocorrelación espacial. (Ver gráficos 61, 62, 63 y 64 del anexo 1).

Al eliminar los valores extremos según el criterio del diagrama de caja dos variables, RNET y VAB10, muestran una pendiente mucho más elevada de la recta de regresión en el diagrama de dispersión de Moran. Para las otras dos variables, RFAMNET10 y VABAGRI10, aumentan pero muy poco, ya que solo presentan uno y dos concejos respectivamente como valores atípicos. (Ver gráficos 65, 66, 67 y 68 del anexo 1).

**Mapa de desviación típica.** En el primer intervalo, tres desviaciones típicas por debajo de la media están Yernes y Tameza y Cabranes, para la variable renta familiar. Para las variables de renta neta y valor añadido bruto están Oviedo y Gijón por encima de la media, con las dos ciudades más pobladas de Asturias. Para el VAB agrario Valdés y Tineo están en el último intervalo. (Ver mapas 110, 111, 112 y 113 del anexo 1).

**Mapa de percentiles.** Yernes y Tameza en el percentil 1 para las cuatro variables. Oviedo en el 100 para la renta familiar y el VAB, Gijón para la renta neta y Tineo para el VAB del sector primario. (Ver mapas 114, 115, 116 y 117 del anexo 1).

Los **mapas LISA** basados en la I local de Moran consideran la existencia de autocorrelación espacial local significativa en dos grupos de concejos para las variables RTANET10 y VAB10. Uno en la zona centro, para concejos cercanos o limítrofes a

Oviedo y Gijón y otra en occidente, en algunos de los concejos más pequeños de interior. Para RTAFAMNET10 ocurre algo parecido pero se extiende a más concejos sobre todo de la zona central. Para VABAGRI10 se detecta un cluster en la zona occidental, fundamentalmente en los concejos más grandes. Lo mismo ocurre utilizando la G y la G\* de Getis y Ord. (Ver mapas 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128 y 129 del anexo 1).

## Estructura económica

**SUPTECHIND12, LIAE13, LIAEAGRI13, LICOBRI13, VEHIC13.**

### Histograma de frecuencias.

Se observa la misma tendencia en las cinco variables, al igual que en la mayoría de las estudiadas hasta el momento, de acumulación de la mayor parte de los valores en el primer intervalo. Su distribución se aleja mucho de la esperada en poblaciones normales. (Ver gráficos 69, 70, 71, 72 y 73 del anexo 1).

### Mapa de deciles.

En los cinco mapas se observa, en términos generales, con independencia de las peculiaridades de cada variable, una concentración de los colores más oscuros, correspondientes a los valores más altos, en la zona central, en los concejos más dinámicos y urbanizados, acompañados por Llanes en el oriente y seguidos de los concejos más grandes y agrarios de occidente y zonas industriales de las cuencas mineras. (Ver mapas 130, 131, 132, 133 y 134 del anexo 1).

### Diagrama y mapa de caja.

No existen outliers inferiores aplicando el criterio de 1.5 veces el rango intercuartílico. Si existen outliers superiores, cinco o más concejos, dependiendo de la variable. El reparto se observa en los mapas y box plot incluidos en el anexo 1 (gráficos 74, 75, 76, 77 y 78). Las zonas con valores más altos coinciden con los concejos más dinámicos tal

como se describe en el apartado anterior, correspondiente a los mapas de deciles. (Ver mapas 135, 136, 137, 138 y 139 del anexo 1).

### **Diagrama de dispersión de Moran.**

Los valores de la I de Moran son positivos y aunque muy bajos son mayores al valor esperado  $E(I)$ , lo que puede sugerir existencia de autocorrelación espacial. Los test de permutaciones realizados (con 999) y una matriz tipo queen dan los siguientes resultados. Para SUPTECHIND12 y LICOBR13 no rechazan la hipótesis nula de no existencia de autocorrelación espacial al 95%. Para LIAE13 y VEHIC13 rechazan casi siempre, aunque en algunas ocasiones se obtiene un p valor algo superior a 0.05, por lo que se presentan ciertas dudas. Para LIAEAGRI13 rechaza la hipótesis nula de no existencia de autocorrelación espacial al 95%. (Ver gráficos 79, 80, 81, 82 y 83 del anexo 1).

Si eliminamos los outliers del diagrama de caja no aumenta la pendiente de la recta en el caso de SUPTECHIND12 y lo hace en poca cuantía para LICOBR13. Donde se observan aumentos importantes, detectando así el efecto distorsionador de los valores extremos, al disminuir los valores de la pendiente, es para las otras tres variables (LIAE13, LIAEAGRI13 y VEHIC13). (Ver gráficos 84, 85, 86, 87 y 88 del anexo 1).

### **Mapa de desviación típica**

Ningún concejo se encuentra en los dos intervalos más bajos (dos y tres desviaciones típicas por debajo de la media) para ninguna de las cinco variables. La mayor parte se encuentra en el intervalo de una desviación típica por debajo de la media. En el intervalo más alto y en el segundo más alto se encuentran pocos concejos (entre 4 y 11, según la variable), que pertenecen a la zona centro, salvo excepciones como Llanes y Cangas del Narcea en el sexto intervalo, Colunga en el quinto para las licencias de obra, o Tineo en el sexto intervalo para las licencias de IAE del sector agrario. (Ver mapas 140, 141, 142, 143 y 144 del anexo 1).

### **Mapa de percentiles.**

En los mapas del anexo se observan los valores extremos que proporcionan los mapas de percentiles. Corresponde a Gijón el valor superior en al caso de licencias de IAE

totales y agrarias, así como en parque de vehículos. Cangas del narcea lo sería para licencias de obra y Avilés para la superficie de techo industrial. Como outliers bajos se sitúan en el primer percentil Yernes y Tameza en vehículos y licencias de IAE totales. Ningún concejo se sitúa en este percentil para las otras tres variables. (Ver mapas 145, 146, 147, 148 y 149 del anexo 1).

### **Cartograma.**

En el anexo 1 se observan los cartogramas y sus criterios de clasificación con los mismos outliers que los del diagrama de caja por usar el mismo método de cálculo. (Ver mapas 150, 151, 152, 153 y 154 del anexo 1).

### **Mapas LISA.**

Para los mapas de clusters basados en el índice local de Moran, calculados con 999 permutaciones y una matriz de pesos espaciales con criterio de contigüidad de tipo queen, en cuatro variables (todas excepto SUPTECHIND12) se obtienen dos clusters, uno de valores alto-alto y bajo-alto en los concejos limítrofes a Gijón y otro en el suroccidente en la zona de los Oscos. Además para LICOBR13 también se observa significatividad en algunos concejos de interior cercanos a Cangas del Narcea. Para SUPTECHIND12 se observa solo un cluster en la zona central en Oviedo, Siero, Las Regueras, Corvera y Gozón, y ninguno de valores bajos. (Ver mapas 155, 156, 157, 158 y 159 del anexo 1).

Pocas variaciones en los resultados si usamos la G y a G\* de Getis y Ord en vez de la I local de Moran. Los cambios serían que en SUPTECHIND12 considera significativos muchos más concejos en el grupo de bajos, sobre todo en la zona del oriente, considerando altos algunos concejos del centro y el resto considerados bajos en ambas alas de la región. Las otras cuatro variables coinciden mayoritariamente en los resultados, con la salvedad de la inclusión de Llanes en LICIAGRI13 y de Morcín en LICOBR13. (Ver mapas 160, 161, 162, 163, 164, 165, 166, 167, 168 y 169 del anexo 1).

## Sector primario

### EXPLOT14, ANIM14, CORTMAD12

Los **histogramas de frecuencias** de las tres variables presentan la mayor concentración de valores en el primer intervalo para las tres variables, seguido del segundo intervalo y dejando solo unos pocos valores para los intervalos más altos. (Ver gráficos 89, 90 y 91 del anexo 1).

Los **mapas de deciles** muestran los valores más altos en los concejos con mayor tradición agraria y forestal, fundamentalmente en los concejos más extensos de la costa oriental y suroccidente. Los más bajos se dan en los concejos más pequeños del interior de las alas, tanto de oriente como de occidente. (Ver mapas 170, 171 y 172 del anexo 1).

Los **diagramas y mapas de caja** no muestran outliers por debajo de la cota inferior. Por encima de la cota superior muestran 5 concejos para el número de explotaciones, correspondientes a concejos de interior de occidente y zona centro, 4 para la cabaña ganadera que se corresponden con Llanes en el oriente y Valdés, Tineo y Cangas del Narcea en occidente, y 5 para las cortas de madera del sector forestal, que se corresponden con concejos de costa de las zonas centro y oriente, donde proliferan especies alóctonas de ciclo corto como el eucalipto, y concejos del interior de occidente con predominancia de especies autóctonas. (Ver gráficos 92, 93, 94 y mapas 173, 174 y 175 del anexo 1)

Los **diagramas de dispersión de Moran** para las tres variables son positivos y mayores que E (I). Los test de permutación realizados sugieren rechazar la hipótesis nula de no existencia de autocorrelación espacial con una significatividad del 99% en los casos de EXPLOT14 y ANIM14 y del 95% para CORTMAD12. (Ver gráficos 95, 96 y 97 del anexo 1). Si eliminamos los valores atípicos casi no se producen variaciones en la pendiente de la recta, que aumenta levemente para EXPLOT14 y ANIM14 y lo hace considerablemente para CORTMAD12. (Ver gráficos 98, 99 y 100 del anexo 1).

El **mapa de desviación típica** presenta una distribución en las tres variables en la que la mayor parte de concejos se sitúan en el intervalo de una desviación típica por debajo de la media (más de 50 en los tres casos), no contando prácticamente con ningún valor en los dos intervalos inferiores (solo uno en la variable EXPLOT14). Tres concejos en el

intervalo superior para CORTMAD12 y ANIM14 y cinco para EXPLOT14. (Ver mapas 176, 177 y 178 del anexo 1).

Los **mapas de percentiles** muestran como valores extremos en el primer y último percentil a Muros de Nalón y Siero, respectivamente, para EXPLOT14, a Muros de Nalón y Tineo para ANIM14 y a ninguno en el primer percentil y Villaviciosa en el 100 para CORTMAD12. (Ver mapas 179, 180 y 181 del anexo 1).

Los **cartogramas** muestran con círculos los resultados ya comentados en el mapa de caja, ya que aplica el mismo criterio para los outliers. (Ver mapas 182, 183 y 184 del anexo 1).

### **Mapas LISA.**

El mapa LISA basado en el índice local de Moran para la variable EXPLOT14 muestra la presencia de autocorrelación espacial local en varios concejos, concentrándose fundamentalmente en dos grupos, uno en la zona central con valores altos rodeados de valores altos y también algunos concejos de valores bajos con valores altos a su alrededor, y otro grupo en el surcoccidente de varios concejos pequeños con valores bajos rodeados de valores también bajos.

Para ANIM14 el principal cluster de valores “alto-alto” está en los concejos más grandes de occidente. También otros dos se incluyen en el grupo de valores “bajo-alto”, uno junto a los del anterior grupo y otro en la zona interior oriental. Como “bajo-bajo” encontramos cuatro concejos, también de pequeñas dimensiones, dos occidentales y dos en la costa central. Para CORTMAD12 vemos cluster “bajo-bajo” en el interior del centro oriente y “alto-alto” junto con dos “bajo-alto” en la zona del centro oriente. (Ver mapas 185, 186 y 187 del anexo 1). Resultados similares en cuanto a su distribución en el espacio obtenemos si usamos la G y la G\* de Getis y Ord. (Ver mapas 188, 189, 190, 191, 192 y 193 del anexo 1).

## Turismo

### REST13, PLAZTUR13

El **histograma de frecuencias** muestra una concentración de la mayor parte de los valores en el primer intervalo, que va desde 5 a 1630 plazas de alojamiento (66 concejos), en el caso de las plazas de alojamientos turísticos, y de 1 a 280 (71 concejos) en el caso de los restaurantes. En el último intervalo solo figura un concejo en ambos casos, Llanes para las plazas de alojamiento, a mucha distancia del resto y Gijón para los restaurantes, solo seguida en el siguiente intervalo por Oviedo. (Ver gráficos 101 y 102 del anexo 1).

Los **mapas de deciles** muestran los valores más altos en tonos más oscuros en concejos de costa, junto con Oviedo y Cangas de Onís en el interior, para el caso de los alojamientos, y una clara concentración en la zona centro de la región, junto con Llanes, para REST13. (Ver mapas 194 y 195 del anexo 1).

El **diagrama y mapa de caja** consideran valores atípicos por encima de la cota superior a Llanes, Oviedo, Gijón, Cangas de Onís, Cabrales, Cudillero, Valdés, Ribadesella y Villaviciosa para PLAZTUR13, concejos costeros junto con tres de interior, la capital del principado y dos con territorio en el Parque Nacional de Picos de Europa. Para REST13 son siete los concejos, Oviedo, Gijón, Avilés, Langreo, Mieres, Siero y Llanes. Salvo el último, se corresponden con los más urbanizados de la región. (Ver gráficos 103, 104 y mapas 196 y 197 del anexo 1).

El **diagrama de dispersión de Moran** se muestra en el anexo 1. En ambos casos los valores de  $I$  son positivos y mayores que  $E(I)$ . Los test de permutación realizados sugieren rechazar la hipótesis nula de no existencia de autocorrelación espacial con una significatividad del 98% en el caso de PLAZTUR13. Lo contrario ocurre en el caso de REST13, donde para los test de permutaciones realizados el  $p$  valor obtenido es mayor a 0.05 en todos los casos, lo que no permite rechazar la hipótesis nula. (Ver gráficos 105 y 106 del anexo 1). Eliminando valores atípicos aumentan las pendientes de las rectas de regresión del diagrama de dispersión de Moran, obteniéndose resultados propios de alta correlación espacial al anularse el efecto distorsionador de los outliers. (Ver gráficos 107 y 108 del anexo 1).

### **Mapa de desviación típica.**

Ningún concejo en los dos intervalos más bajos. Oviedo y Gijón en número de restaurantes se sitúan en el intervalo tres veces la desviación estándar sobre la media. En el caso de plazas de alojamiento también Llanes está en ese intervalo. (Ver mapas 198 y 199 del anexo 1).

### **Mapa de percentiles.**

Llanes en el 100 y Yernes y Tameza en el 1, son los valores extremos de la muestra. (Ver mapas 200 y 201 del anexo 1).

### **Cartograma.**

Se muestran los cuatro intervalos y como valores extremos los mismos que los ya descritos en el mapa y diagrama de caja, todos sobre la cota superior. (Ver mapas 202 y 203 del anexo 1).

### **Mapas LISA.**

Los **mapas LISA** basados en la I local de Moran consideran la existencia de autocorrelación espacial significativa en dos grupos de concejos en la variable REST13. En la zona central, limitando con Oviedo y Gijón para valores “alto-alto” y “bajo-alto” y en el ala occidental para valores bajos rodeados de valores bajos. En la variable de plazas de alojamiento los valores altos rodeados de valores altos y los bajos rodeados de valores altos se concentran en el oriente de Asturias. Otros dos clusters “bajo-bajo” se consideran como significativos con autocorrelación espacial local, uno en la zona central interior, y otro en el occidente, también de interior. Prácticamente los mismos resultados si utilizamos la G y la G\* de Getis y Ord. (Ver mapas 204, 205, 206, 207 208 y 209 del anexo 1).

### 4.1.2. RESULTADOS DEL AEDE.

Para las variables de inversión pública cofinanciada o financiada por FEADER y FEAGA (IPDR13.07, I1PIL13.07 e IPAC13.07), los principales resultados del AEDE univariante reflejan la existencia de autocorrelación espacial. También se detecta la presencia de outliers superiores, principalmente en la zona de occidente, en los concejos más extensos y con mayor peso del sector primario.

En los tres casos el índice I de Moran, es positivo, en torno a 0.3, y es significativo al realizar distintas pruebas de aleatorización para simular su distribución. Al eliminar los considerados outliers, según el criterio del diagrama de caja, se mantiene la autocorrelación y su significatividad.

Con respecto a la autocorrelación espacial local, los test LISA tanto basados en la I de Moran, como en la G de Getis y Ord son significativos en varios concejos, resultando un cluster importante de valores altos en occidente, así como también en zonas con importante actividad del sector primario y concejos colindantes. También son significativos pero con valores bajos algunos concejos de la zona centro. Esto puede ser indicativo de heterogeneidad espacial y puede sugerir inestabilidad paramétrica, si consideramos la diferencia existente en Asturias entre el centro y las alas, así como la que pueda también existir entre la costa y el interior.

Para las que vamos a considerar variables dependientes en los modelos (VPOBPAD14.07, VEMP13.07, VEMPAGRI13.07 y LIAEAGRI13), no sería significativo el test de Moran al 95% en el caso de VEMP13.07 y LIAEAGRI13 (aunque sí al 90%). No es tampoco significativo para VPOBPAD14.07 y en cambio, sí lo es para VEMPAGRI13.07. Salvo para esta última variable, los valores de la I de Moran son inferiores a 0.1. La dependencia espacial parece clara solo en este último caso, no siendo significativa ni relevante para la variación de la población y presentando dudas en las otras dos variables de variación del empleo total y licencias agrarias. Para la existencia de autocorrelación local existen evidencias claras al analizar los resultados de los test, tanto los basados en la I de Moran como los de la  $G_i$  y  $G_i^*$  de Getis y Ord. Para las cuatro variables se observa un cluster en la zona centro de valores altos el caso de empresas agrarias. En los otros tres casos se observan otros clusters en distintos lugares, concretamente en la zona central interior para la población, ala occidental para el empleo y occidente para el empleo agrario. Hay una importante evidencia de patrones

espaciales de dependencia local concentrada en distintas zonas de Asturias, con la característica, común a las cuatro variables, de concentrarse en los cuatro casos en el centro de la región.

Con respecto al comportamiento del resto de variables, así como para todo el conjunto, se observa, de forma general, que la dependencia espacial no es alta ni significativa para las variables no agrarias (salvo para la renta familiar neta, en que si es significativa al 95%) y si lo sería para las variables propias del sector primario. Para las variables no agrarias, con autocorrelación espacial no significativa, podemos distinguir dos grupos, uno en el que los resultados no son especialmente concluyentes, debido a que aunque no son significativas al 95% si lo son al 90% y otro grupo que sería el de las variables para las que el test de Moran no es significativo con un amplio margen, ya que tampoco lo sería al 90%. En muchos casos, al eliminar los valores extremos, aumenta en alto grado la dependencia espacial. Esto aconseja estimar los modelos con y sin valores extremos, para comparar ambos resultados. Además se deben utilizar modelos espaciales en los casos en que se confirme la existencia de autocorrelación espacial, ya que las variables agrarias, que son las que presentan dependencia espacial en el análisis univariante realizado, serán siempre consideradas en el planteamiento inicial de los modelos (especialmente las de inversión).

Lo que si se da en todas las variables es una importante presencia de clusters con autocorrelación espacial local. Casi siempre, salvo en tres variables agrarias (EMPAGRI13, VABAGRI10 y ANIM14), está localizada en la zona centro. Esto sugiere considerar la posible existencia de heterogeneidad espacial y plantear modelos que corrijan este problema, de confirmarse su existencia. El problema de la concentración de la actividad en la zona centro y la existencia de distintas dinámicas en algunas zonas del territorio asturiano son cuestiones a considerar en este estudio para su contraste, y los resultados del análisis exploratorio van en esa línea.

En el cuadro 4.1 se resumen, para todas las variables consideradas, los resultados del I de Moran y su significatividad, así como la existencia o no de autocorrelación local describiendo las zonas de mayores concentraciones de concejos con autocorrelación espacial local significativa.

Cuadro 4.1.

VARIABLE	I MORAN	SIGNIFICATIVO AL 95%	TEST LISA SIGNIFICATIVO (95%) EN ALGÚN CONCEJO-ZONAS*
IPDR13.07	0.347453	SI	SI-OCCIDENTE-CENTRO
I1PIL13.07	0.263426	SI	SI-OCCIDENTE-CENTRO
IPAC13.07	0.31059	SI	SI-OCCIDENTE-CENTRO
POBPAD14	0.720753	NO(al 90% si)	SI-CENTRO-ALA OCCIDENTAL
VPOBPAD14.07	0.0482384	NO	SI-CENTRO-CENTRO INTERIOR
CONTRAT13	0.0771281	NO (al 90% si)	SI-CENTRO-ALA OCCIDENTAL
PAR14	0.0742348	NO (al 90% si)	SI-CENTRO-ALA OCCIDENTAL
VPAR14.07	0.0830172	NO (al 90% si)	SI-CENTRO-ALA OCCIDENTAL
EMP13	0.103275	NO SIEMPRE(al 90% si)	SI-CENTRO-ALA OCCIDENTAL
VEMP13.07	0.0501124	NO (al 90% si)	SI-CENTRO-ALA OCCIDENTAL
EMPAGRI13	0.215014	SI	SI-OCCIDENTE
VEMPAGRI13.07	0.154703	SI	SI-CENTRO-OCCIDENTE
ASS14	0.0931376	NO SIEMPRE(al 90% si)	SI-CENTRO-ALA OCCIDENTAL
RNET10	0.0709382	NO (al 90% si)	SI-CENTRO-ALA OCCIDENTAL
RFAMNET10	0.343659	SI	SI-CENTRO-ALA OCCIDENTAL
VAB10	0.113502	NO SIEMPRE(al 90% si)	SI-CENTRO-ALA OCCIDENTAL
VABAGRI10	0.296729	SI	SI-OCCIDENTE
SUPTECHIND12	0.0556332	NO	SI-CENTRO-ORIENTE-ALA OCCIDENTAL
LIAE13	0.0993931	NO SIEMPRE (al 90% si)	SI-CENTRO-ALA OCCIDENTAL
LIAEAGRI13	0.0829393	NO (al 90% si)	SI-CENTRO
LICOBR13	0.037748	NO	SI-CENTRO-OCCIDENT-ALA OCCIDENTA
VEHIC13	0.0988298	NO SIEMPRE (al 90% si)	SI-CENTRO-ALA OCCIDENTAL
EXPLOT14	0.209182	SI	SI-CENTRO INT-CENTRO-ALA OCCIDEN
ANIM14	0.194155	SI	SI-OCCIDENTE
CORTMAD12	0.120027	SI (NO SIEMPRE, al 90% si)	SI-CENTRO INTERIOR-CENTRO
REST13	0.0704182	NO ( al 90% si)	SI-CENTRO-ALA OCCIDENTAL
PLAZTUR13	0.180412	SI	SI-ORIENTE-CENTRO INT-ALA OCCIDEN

Fuente: Elaboración propia.

\* Las zonas descritas son aquellas en las que varios concejos agrupados son significativos, sin considerar como zona concejos aislados con autocorrelación local significativa.

## 4.2. ANÁLISIS CONFIRMATORIO

Se analiza en este apartado la relación existente entre las distintas variables consideradas. Calculamos la correlación entre variables y contrastamos su significación (análisis conjunto de las variables). La información obtenida nos sirve para construir los modelos explicativos, teniendo en cuenta a su vez la autocorrelación espacial y la heterogeneidad espacial en caso de que existan.

De las variables consideradas hacemos un análisis de la correlación entre ellas, para ayudarnos a plantear distintos modelos en base a estos resultados. Se calculan los índices de correlación de Pearson ( $r$ ). Para ello se hace previamente un test de normalidad, ya que se requiere que las variables sean normales para garantizar la validez de los contrastes de significatividad de la  $r$  de Pearson. Se aplica el test de Shapiro-Wilks. Para la mayoría de variables se rechaza la hipótesis nula de que se distribuyen normalmente. Por ello se calculan también como alternativas a la  $r$  de Pearson,  $\rho$  (rho) de Spearman y la  $\tau$  (tau) de Kendall, para los que los contrastes se pueden aplicar sin la necesidad de asumir la hipótesis de normalidad de las variables.

Las variables más importantes sobre las que se quiere incidir con las políticas rurales objeto de estudio son la población y el empleo, por tanto algunas de las variables de esta naturaleza, son las que se consideran como dependientes, cada una en su modelo correspondiente, a la vez que se consideran como independientes el resto de variables que pueden mantener una relación de causalidad desde el punto de vista teórico, junto con la inversión o gasto público destinado a financiar el PDR de Asturias para este periodo de programación. Esta es la variable clave. Su relevancia a la hora de explicar variables poblacionales y laborales indicará su capacidad de contribución para alcanzar los objetivos teóricos. Para comparar resultados también se va a considerar la inversión de pagos directos del primer pilar de la PAC, así como la suma de pagos totales para el conjunto de la PAC.

Se excluyen las variables que pudiendo tener una causalidad sobre la dependiente presentan altos índices de correlación entre ellas (próxima a 1) para evitar problemas de multicolinealidad (Novales, 1993).

En el cuadro 4.2 se muestran las variables inicialmente consideradas. Son aquellas que presentan correlación estadísticamente significativa con la variable dependiente y presentan una posible causalidad teórica desde el punto de vista de las políticas de desarrollo rural planteadas y de la teoría económica. Entre las que cumplen estas dos condiciones se han eliminado aquellas que presentan una correlación próxima a uno para evitar la multicolinealidad. Algunas de estas que hayan sido eliminadas a priori pueden ser consideradas e incluidas más adelante al analizar los resultados de los posibles modelos iniciales si se observa que su inclusión o sustitución por otra inicialmente considerada mejora las propiedades del mismo.

Se calculan modelos considerando dos escenarios, el primero incluye a todos los concejos y el segundo excluye a los seis concejos más poblados (por encima de los 40000 habitantes) con carácter más urbano e industrial, con valores extremos en la mayoría de variables. Ya hemos visto en el AEDE los outliers considerados, con distintos criterios, que evidentemente no se atribuyen a los mismos concejos en cada variable, pero en términos generales y sobre todo en variables relativas a la población, el empleo y la actividad económica casi siempre están estos seis concejos muy por encima del resto. Desde el punto de vista econométrico conviene eliminar la posible influencia de valores extremos que puedan ocultar o cambiar propiedades de la muestra. Además, dada la naturaleza rural de estos fondos, consideramos adecuado, también desde el punto de vista teórico, realizar el análisis sin estos valores extremos en la muestra, para ver qué ocurre con el comportamiento de los modelos sin su posible distorsión sobre los mismos. En Asturias todos los concejos tienen una parte rural dedicada al sector primario, con baja densidad de población, pequeños núcleos poblacionales dispersos, con tradiciones y cultura agraria, etc. De hecho las variables agrarias consideradas (incluidas las de inversión) también muestran valores elevados en los concejos con ciudades y con importantes estructuras urbanas, pero también es evidente que en ciertas zonas, sobre todo del área central, este peso en relación a actividades más urbanas e industriales hace que los valores considerados estén condicionados por mecanismos ajenos a las políticas agrarias de la Unión Europea. A continuación, en el cuadro 4.2 se muestran las distintas variables consideradas inicialmente al utilizar los criterios anteriormente expuestos para, a partir de ellas, ir construyendo los distintos modelos con sus variaciones según los distintos escenarios considerados.

Cuadro 4.2.

GRUPO DE MODELOS	VARIABLE DEPENDIENTE	VARIABLES INDEPENDIENTES INICIALMENTE CONSIDERADAS
MODELO 1	VPOBPAD14.07	IPDR13.07, I1PIL13.07, IPAC13.07, VEMP13.07, VEMPAGRI13.07, VABAGRI10, LIAEAGRI13, VAB10, EXPLOT14, ANIM14.
MODELO 2	VEMP13.07	IPDR13.07, I1PIL13.07, IPAC13.07, VPOBPAD14.07, EMPAGRI13, VEMPAGRI13, RFAMNET10, VAB10, VABAGRI10, SUPTECHIND12, LIAEAGRI13, LICOBRI13, PLAZTUR13.
MODELO 3	VEMPAGRI13.07	IPDR13.07, I1PIL13.07, IPAC13.07, ASS14, VPOBPAD14.07, VEMP13.07, RFAMNET10, VABAGRI10, SUPTECHIND12, LIAEAGRI13, LICOBRI13, EXPLOT14, ANIM14, CORTMAD12, PLAZTUR13.
MODELO 4	LIAEAGRI13	IPDR13.07, I1PIL13.07, IPAC13.07, VPOBPAD14.07, EMPAGRI13, VEMPAGRI13.07, RNET10, RFAMNET10, VABAGRI10, SUPTECHIND12, LICOBRI13, CORTMAD13, REST13, PLAZTUR13.

#### 4.2.1. AUTOCORRELACIÓN ESPACIAL.

Para la estimación se utiliza en un primer paso el método habitual de mínimos cuadrados ordinarios, sin tener en cuenta el problema espacial. Sobre esta estimación inicial se realizan, además de los contrastes habituales del modelo “clásico”, los test específicos sobre los residuos que puedan indicar la presencia o no de autocorrelación espacial.

Dentro de los cálculos y contrastes no espaciales se considera el  $R^2$  y el  $R^2$  ajustado, la suma de cuadrados de los residuos y el sigma cuadrado, considerando para su cálculo tanto los grados de libertad como el total de variables. También se considera el valor del estadístico F de significatividad conjunta del modelo y el p valor asociado, considerando como hipótesis nula que los coeficientes de las variables explicativas son cero. Además se considera el logaritmo de la verosimilitud y los valores del criterio de

información de Akaike y el criterio de Schwarz, así como la desviación estándar de los coeficientes del modelo junto con el valor del estadístico  $t$  para la significación individual de los mismos, bajo la hipótesis nula de valor cero, y el  $p$  valor asociado. Para la multicolinealidad se considera el número de condición de multicolinealidad. El test Jarque Bera para la normalidad de los residuos. Para la heteroscedasticidad los test de Breush Pagan (BP) y Koenker Basset (KB), así como el test robusto de White.

Los test específicos para la autocorrelación espacial sobre los residuos de la estimación MCO, son el  $I$  de Moran, y los basados en los multiplicadores de Lagrange, concretamente el LM-LAG, el LM-ERR, el LM-LAG robusto, el LM-ERR robusto y el LM-SARMA.

Estos test nos van a dar información sobre la existencia o no de autocorrelación espacial en el modelo y sobre la forma de incorporación de la misma, si es que está presente, según se trate de autocorrelación espacial sustantiva, residual o ambas. Si los test no constatan la existencia de ningún tipo de autocorrelación espacial, plantearemos el modelo tal como está y resolveremos los posibles problemas que hubiese con las técnicas habituales, es decir aceptamos el modelo estimado por MCO. En caso contrario nos llevaría al siguiente paso que es especificar un nuevo modelo que si la considere, estimado por el método de máxima verosimilitud si aceptamos la normalidad en los errores y por variables instrumentales en caso contrario. Para ello se estudian los test más importantes que serían el LM-LAG y el LM-ERR (Anselin, 2005). Si solo uno de ellos es significativo, escogemos el modelo del error espacial o el del retardo espacial según proceda. Si ambos son significativos, es cuando se debe estudiar los test robustos (ya que en caso contrario pueden no ser apropiados). Si uno de ellos es significativo escogemos ese criterio para plantear el modelo del error o del retardo correspondiente. Si ambos vuelven a serlo tomaremos como criterio el de la significatividad más alta, aunque revisando posibles problemas de mala especificación del modelo y/o de las matrices de pesos espaciales.

Los nuevos modelos espaciales estimados por MV que incorporan la problemática de la dependencia espacial proporcionan también los resultados del  $R^2$ , la suma de cuadrados de los residuos y el sigma cuadrado, el logaritmo de la verosimilitud y los valores del criterio de información de Akaike y el criterio de Schwarz. Además para la heteroscedasticidad se considera el test de Breusch y Pagan y el test del cociente de

verosimilitud para la dependencia espacial del nuevo modelo. Para la validez del test se requiere homocedasticidad, por lo que también se consideran, junto con el valor de este último estadístico del cociente de verosimilitud (LR), los valores de los estadísticos de Wald (W, que sería el cuadrado del valor estandarizado del coeficiente del retardo espacial) y el del multiplicador de Lagrange del error (LM-ERR), ya que según Anselin (2005) se debe cumplir que  $W > LR > LM$ . Si esto no se mantiene, sería indicativo de problemas de especificación del modelo, que pueden ser debidos a términos de error no normales, relación no lineal entre las explicativas y la explicada, o mala elección de variables explicativas, entre otras posibles causas.

Se calcula con los valores previstos, los errores previstos y los residuos del modelo espacialmente retardado la I de Moran. Para los residuos se deben obtener valores cercanos a 0, como resultado de la eliminación de la dependencia espacial en los mismos, al considerar su correcta especificación en el modelo. En comparación con el valor de la I de Moran para el error previsto, que no considera la inclusión de las variables espacialmente retardadas, debe ser menor (Anselin, 2005). Si esto no es así sería indicativo de no haber solucionado el problema de autocorrelación espacial.

Si con lo anterior no obtenemos modelos y estimaciones con buenas propiedades o en el caso de que no podamos aceptar la distribución normal de los residuos, usaremos como alternativa al método de máxima verosimilitud métodos basados en el principio de variables instrumentales como serían método de mínimos cuadrados bietápicos (MC2E) y momentos generalizados (GM) o método de los momentos generalizados (GMM). Se aplican en el caso de correlación entre alguna de las variables explicativas y el término de error (Anselin y Rey, 2014) También en el caso de no normalidad de los residuos es uno de los métodos recomendados (Chasco, 2013), ya que si no se cumple dicha condición, no son correctos los test de significatividad y pueden dar resultados erróneos. Usaremos el programa GeodaSpace.

Con este método vamos a obtener además de la media y varianza de la variable dependiente, el pseudo  $R^2$ , así como la significación individual de los coeficientes del modelo. Para el diagnóstico de los efectos espaciales de los modelos de retardo espacial se aplica el test de Anselin y Kelejian (1997).

Si se pone de manifiesto la existencia de dependencia espacial los métodos alternativos a MV aplicados y los procesos y resultados obtenidos son los propuestos por Anselin y Rey (2014), que proporciona el programa GeodaSpace:

1. Mínimos cuadrados bietápicos espaciales (spatial two-stage least squares, S2SLS) para estimar el modelo del retardo espacial (modelo lag), si así lo sugiriesen los test calculados previamente sobre los errores de los modelos no espaciales, con la posibilidad de incluir o no regresores endógenos. Se basa en variables instrumentales y en el método de los momentos generalizados. Al igual que con el método de mínimos cuadrados bietápicos aplicado al modelo no espacial, se calcula la media y varianza de la variable dependiente, el pseudo  $R^2$ , así como la significación individual de los coeficientes del modelo. Para el diagnóstico de los efectos espaciales se aplica el test de Anselin-Kelejian.

2. Para el modelo del error espacial se tienen dos opciones que serían utilizar mínimos cuadrados espacialmente ponderados (spatial weighted least squares, SWLS) o mínimos cuadrados bietápicos espacialmente ponderados (spatial weighted two-stage least squares, SW2SLS). En el segundo caso se requiere la inclusión de regresores endógenos. Se puede estimar tanto por el método de los momentos generalizados como por momentos generalizados (GMM o GM). El primer paso requiere una estimación de los coeficientes del modelo sobre una transformación espacial de las variables, para obtener  $\lambda$ . En el caso de SWLS sería por mínimos cuadrados generalizados factibles (feasible generalized least squares, FGLS) y en el caso de usar variables instrumentales, por considerar la existencia de regresores endógenos (aplicando el método de SW2SLS), se usa MC2E. Se aplican a variables espacialmente ponderadas y se obtiene un estimador de  $\lambda$  consistente, pero todavía no eficiente. En el segundo paso se obtiene un estimador de  $\lambda$  eficiente (Anselin, Amaral y Arribas-Bel, 2012) y se calculan los demás estimadores. También se puede aplicar adicionalmente un procedimiento que calcula estimadores robustos a heteroscedasticidad y autocorrelación (KP-HET) (Kelejian y Prucha, 2010). Este lo consideraremos también cuando proceda para asegurar la validez de los modelos.

3. Para el modelo mixto que combina el modelo del retardo con el del error espacial (autorregresivo), que se denomina combo, se utiliza el método de mínimos cuadrados

bietápicos generalizados espacialmente ponderados (generalized spatial weighted two-stage least squares, GSW2SLS). También es un proceso iterativo que aplica filtros espaciales a las variables y que partiendo de MC2E estima en tres pasos estimadores eficientes y consistentes aplicando el método de los momentos generalizados (Anselin, 2011).

Tanto en el método de MCO como en el de MC2E para estimar los modelos no espaciales, así como para S2SLS para calcular el modelo del retardo espacial, vamos a considerar y calcular cuando proceda, además de los habituales, los estimadores robustos a heteroscedasticidad de White, así como los que utilizan el método desarrollado por Kelejian y Prucha (2007) que son robustos a la heteroscedasticidad y a la autocorrelación espacial. Cuando usemos el método de SWLS, SW2SLS o GSW2SLS para calcular, tanto el modelo del error espacial, como el mixto (lag+error), consideraremos y calcularemos, cuando proceda, los estimadores robustos a heteroscedasticidad y autocorrelación propuestos por Kelejian y Prucha (2010).

### *Modelo 1*

El primer modelo estimado por MCO con todas las variables inicialmente consideradas presenta un  $R^2$  de 0.5543 y un  $R^2$  ajustado de 0.5027. El test de significatividad conjunta del modelo resulta favorable y se rechaza la hipótesis nula de coeficientes iguales a 0. Como problemas se presentan la alta multicolinealidad, con un número de condición cercano a 30, los errores no son normales, según los resultados del test de Jarque Bera y también salen significativos los tres test de heteroscedasticidad realizados, que rechazan la hipótesis nula de residuos homocedásticos. No resultan significativos los test de dependencia espacial, así como la significatividad individual de las explicativas IPDR13.07, VAB10, VABAGRI10 y VEMPAGRI13.07. Si calculamos los coeficientes robustos a heteroscedasticidad de White a las anteriores habría que añadir EXPLOT14 y VEMP13.07, pero si resulta significativa IPDR13.07. Al eliminar las variables no significativas en ambos casos mejora la multicolinealidad con un número de condición de 11.522. Mejoran levemente los valores del criterio de información de Akaike y Schwarz, y empeora el logaritmo de la verosimilitud. Los coeficientes son

significativos, salvo el robusto de White para EXPLOT14. El signo de los coeficientes es negativo en todas las variables salvo en ANIM14.

Al eliminar alguna de las variables que tenía alta correlación con otras explicativas probamos a introducir algunas de las descartadas inicialmente. Se obtienen mejores resultados al introducir la variable RNET10 y al eliminar las no significativas, quedando como explicativas esta junto a VEMP13.07 e IPDR13.07, aunque siguen teniendo todas las variables signo negativo en contra de lo esperado. Mejoran los valores del criterio de información de Akaike y Schwarz, y también el logaritmo de la verosimilitud. El número de condición de multicolinealidad es de 11.524, y los resultados de los test de heteroscedasticidad de BP y KB no rechazan la hipótesis nula de homocedasticidad, aunque si el test robusto de White. Salen significativos el test de LM-ERR y no el LM-LAG, lo que sugiere estimar un modelo del error espacial. Al persistir la significatividad del Jarque Bera asumiendo la no normalidad de los errores no es aconsejable el método de máxima verosimilitud y si el de variables instrumentales, concretamente usaremos el método de los momentos generalizados. (Anselin y Rey, 2014)

El modelo clásico estimado por MC2E, incluyendo como instrumental VEMP13.07 e instrumentos VAB10 y VPAR14.07, considera significativas todas las variables. También si calculamos los errores estándar tanto por el método robusto de White, como por el método robusto a heteroscedasticidad y autocorrelación espacial (HAC). Considera significativa la dependencia espacial del test de Anselin y Kelejian al 99%. Vamos a probar con el modelo de error tal como sugerían los resultados de la estimación por MCO.

El mejor modelo de los considerados es:

$$\begin{aligned} \text{VPOBPAD14.07} &= -51.921744 - 0.00003 \text{ IPDR13.07} - 0.0063363 \text{ RNET10} - 1.4724344 \text{ VEMP13.07} + \\ &U \\ U &= 0.5011952 \text{ WU} + \varepsilon \end{aligned}$$

Se estima por SWLS (Anselin y Rey, 2014) por el método robusto a heteroscedasticidad (KP-HET), ya que en el paso previo de estimación por MCO, aunque no eran significativos los test de KB y BP, si lo era el robusto de White. Al rechazar también la hipótesis nula de normalidad de los errores se escoge este método para estimar el modelo del error espacial, tal como sugiere la significatividad del test LM-ERR, junto a

la no significatividad del LM-LAG. Salen significativos todos los coeficientes, incluido el retardo espacial. La matriz de pesos espaciales es de contigüidad tipo queen. Da mejores resultados este modelo que el que incluye también VEMP13.07 como instrumental y como instrumentos VAB10 y VPAR14.07, estimado por mínimos cuadrados bietápicos espacialmente ponderados (SW2SLS). En los modelos que suponen homocedasticidad todos los coeficientes son significativos salvo IPDR13.07 y el término constante.

Para la variable I1PIL13.07, que recoge los fondos públicos destinados al primer pilar de la PAC, y para IPAC13.07, que recoge el total del primer y segundo pilares, tras un proceso similar al anteriormente descrito los modelos con mejores propiedades, de todos los considerados, son los siguientes modelos de error espacial:

$$\text{VPOBPAD14.07} = - 50.618485 - 0.0000187 \text{ I1PIL13.07} - 0.0063309 \text{ RNET10} - 1.4736904 \text{ VEMP13.07} + U$$

$$U = 0.5097277 \text{ WU} + \varepsilon$$

$$\text{VPOBPAD14.07} = - 44.1628291 - 0.0000124 \text{ IPAC13.07} - 0.00633 \text{ RNET10} - 1.4728552 \text{ VEMP13.07} + U$$

$$U = 0.5059038 \text{ WU} + \varepsilon$$

En los tres modelos se obtienen coeficientes con valores en los test de significación individual altamente significativos. Los tres coeficientes de las variables explicativas tienen signo negativo. En el caso de RNET10 se justifica por la mayor pérdida de población en los concejos con mayor nivel de renta, ya que estos concejos coinciden con los más grandes, en un contexto general de declive poblacional y económico. Para VEMP13.07 las variaciones en el número de empleos de mayor pérdida o ganancia de los mismos, han generado mayores o menores valores en las variaciones de la población respectivamente (es decir mayores ganancias o menores pérdidas y viceversa), y se han producido menores variaciones en aquellos en los que el empleo ha sido más estable. Mayor número de empleos creados no ha contrarrestado la pérdida poblacional. Los concejos con mayor caída del empleo, que son los más urbanos han tenido mejores resultados en la variación de la población por lo que posteriormente, al eliminarlos de la muestra, comprobaremos si cambia el signo al eliminar su efecto. Y por último con las

variables de inversión se obtiene una correlación también negativa que refleja que la inversión pública en políticas de desarrollo rural ha tenido un efecto negativo, o al menos no lo ha contrarrestado lo suficiente, con respecto a las variaciones de la población en términos absolutos, a mayor inversión menor variación de la población, es decir mayor pérdida. La influencia negativa es menor en el conjunto de inversión total IPAC13.07, que en la inversión del primer pilar IIPIL13.07 y estas menores que la del segundo pilar IPDR13.07, que presenta los peores resultados.

Vamos a estimar este modelo para la muestra sin outliers, intentando eliminar su posible efecto sobre el total de la muestra, ya que pueden condicionar los resultados anteriores y desvirtuar las conclusiones obtenidas, sobre todo porque los seis concejos eliminados son los que tienen la estructura urbana más importante de toda la región y estas zonas urbanas no son en principio beneficiarios de las ayudas a zonas rurales. Se eliminan de la muestra Oviedo, Gijón, Avilés, Mieres, Langreo y Siero (todos con población superior a los 40000 habitantes).

En la primera estimación del modelo clásico por mínimos cuadrados ordinarios se obtienen resultados significativos para el test de verificación conjunta. El  $R^2$  y el  $R^2$  corregido son ambos superiores a 0.5. También salen significativos los test de normalidad de los errores y de heteroscedasticidad por lo que ambos problemas se encuentran presentes en el modelo. La multicolinealidad no parece un problema muy grave, ya que el número de condición de multicolinealidad es inferior a 20 (17.561). Se eliminan las variables que no resultan significativas individualmente y se prueba cual de las no consideradas inicialmente, por muy alta correlación con otras posibles explicativas, funciona mejor. Se sustituye VAB10 por CONTRAT13 al obtener mejores resultados. Quedarían como explicativas VEMP13.07, IPDR13.07 y CONTRAT13. Las tres son muy significativas individualmente tanto con los estimadores habituales como con los robustos a heteroscedasticidad de White, como con los robustos a heteroscedasticidad y autocorrelación espacial conjuntamente, HAC (Anselin y Rey, 2014) utilizando una matriz de distancia de Kernel basada en la distancia euclídea y considerando 5 vecinos. Al eliminar las variables no significativas disminuye de forma importante el número de condición de multicolinealidad quedando en 3.335 y también mejoran los valores del criterio de información de Akaike y Schwarz, así como el logaritmo de la verosimilitud. Aumentan el  $R^2$  y el  $R^2$  corregido pasando a ser superiores a 0.6. Los test de dependencia espacial no indican la presencia de este

---

problema en el modelo. Persiste la heteroscedasticidad y la no normalidad de errores. Esta puede invalidar los resultados de los test planteados. Se estima el modelo por mínimos cuadrados bietápicos como método alternativo. Se considera como independiente VEMP, como instrumental CONTRAT13 y como instrumentos LIAE13 y RNET10. Siguen siendo significativas las variables individualmente. El pseudo  $R^2$  da un valor de 0.6267 y el test de dependencia espacial de Anselin-Kelejian no es significativo por lo que no rechazamos la hipótesis nula de no existencia de autocorrelación espacial, en la misma línea que en el método de mínimos cuadrados ordinarios. Además se calculan también los errores estándar para contrastar la significatividad individual de los coeficientes de las variables por los métodos HAC y robusto de White y también son altamente significativos. El modelo quedaría:

$$\text{VPOBPAD14.07} = -46.1994561 + 0.1801377 \text{ CONTRAT13} - 0.0000317 \text{ IPDR13.07} + 1.0761226 \text{ VEMP13.07} + U$$

Los coeficientes del modelo son positivos para CONTRAT13 y VEMP13.07, lo que implica que para valores altos y positivos de estas variables mayor variación de la población en términos positivos. Para IPDR13.07 el coeficiente es muy bajo y negativo, al igual que ocurría en el modelo calculado para todos los concejos. Por tanto no se ha conseguido el objetivo de dinamizar la población con las inversiones realizadas en desarrollo rural, sino más bien lo contrario, y a mayores inversiones valores más negativos de variación de la población. Lo que se observa es que sin valores los extremos considerados no aparece el fenómeno de dependencia espacial, aunque sí el de heteroscedasticidad que se abordará en el apartado siguiente.

Para I1PIL13.07 el modelo estimado con mejores propiedades de los considerados es:

$$\text{VPOBPAD14.07} = 25.5898472 - 0.0000177 \text{ I1PIL13.07} + 0.0014268 \text{ VAB10} + 0.7705851 \text{ VEMP13.07} + 0.3986629 \text{ W VPOBPAD14.07} + U$$

Es un modelo de retardo espacial que ha sido estimado por el método de S2SLS, utilizando como instrumental el retardo de la endógena (W VPOBPAD14.07) y como instrumentos los retardos espaciales de las explicativas (W I1PIL13.07, W VAB10 y W VEMP13.07). Todas las variables son significativas individualmente y se elimina la dependencia espacial ya que no sale significativo el test de Anselin-Kelejian que recoge la existencia de autocorrelación espacial en los residuos. La matriz de pesos espaciales utilizada es de contigüidad de tipo queen.

Con el total de fondos públicos para ambos pilares IPAC13.07 el modelo estimado por MC2E, utilizando para el análisis de la dependencia espacial una matriz de contigüidad tipo rook, no requiere la utilización de modelo de retardo espacial, error espacial o combinación de ambos, al no salir significativos los test de dependencia espacial, tanto en la estimación inicial por MCO, como en la segunda por el método de MC2E. Al no obtener residuos normales se utiliza este método, obteniendo el siguiente modelo como el mejor de los considerados:

$$\text{VPOBPAD14.07} = -46.7288451 + 0.185674 \text{ CONTRAT13} + 1.0601361 \text{ VEMP13.07} - 0.0000130 \text{ IPAC13.07} + U$$

Las variables CONTRAT13 y VEMP13.07 tienen coeficientes significativos individualmente y con signo positivo, lo que significa que valores altos y positivos contribuyen a obtener valores de variación de la población en el periodo también altos y positivos. Ocurre lo contrario con IPAC13.07, que tiene un valor negativo y muy pequeño, lo que indica una contribución negativa donde a mayor inversión peor variación, lo que indica que no se consiguen los objetivos perseguidos por parte de estas políticas en lo referente a la población. Los otros modelos considerados y descartados por tener peores propiedades que este, también coinciden en la significatividad del coeficiente y su valor siempre es de signo negativo.

La comparativa entre los modelos con y sin outliers muestra que las variables que recogen los fondos públicos destinadas a políticas de desarrollo rural tienen una relación negativa con la variación de la población. Así variables como la variación del empleo y el número de contratos ven cambiado su signo al eliminar de la muestra las zonas más urbanizadas y con valores más extremos que recogen comportamientos anómalos con respecto al resto. No sucede lo mismo con la inversión pública en política agraria. Se observa además que al eliminar los outliers no se detecta la presencia de dependencia espacial que si existía con el total de la muestra.

## *Modelo 2*

Se hace una primera estimación por MCO, que resulta con una alta multicolinealidad y varios coeficientes no significativos individualmente, incluida la variable IPDR13.07. No se da normalidad en los errores. Existe dependencia espacial. El  $R^2$  y  $R^2$  corregido

son cercanos a 1. Al eliminar las variables no significativas (salvo IPDR13.07) baja la multicolinealidad hasta un número de condición inferior a 3, los test dan resultados de significatividad positiva tanto de forma conjunta como individual, salvo para IPDR13.07. Los test de Breush y Pagan (BP) y Koenker Basset (KB) no rechazan la hipótesis nula de homocedasticidad, pero si el test robusto de White. Al eliminar IPDR13.07 pasa a ser no significativo el coeficiente de PLAZTUR13, por lo que la eliminamos también. Siguen siendo significativos los test de dependencia espacial. Como explicativas están SUPTECHIND12 y VAB10. Como no se cumple el supuesto de normalidad en los errores estimamos el modelo no espacial por MC2E y llegamos a las mismas variables como relevantes tras analizar los resultados de las alternativas, quedando como mejor opción la que considera como explicativa SUPTECHIND12, como instrumental VAB10 y como instrumento RNET10. El test de Anselin-Kelejian sale significativo, lo que indica que rechazamos la hipótesis nula de no existencia de autocorrelación espacial. Al igual que en la estimación inicial por MCO se confirma que existe dependencia espacial en los residuos por lo que conviene plantear modelos espaciales que lo recojan.

Al no tener normalidad en los residuos no conviene estimar por máxima verosimilitud, sino por el método de los momentos generalizados (GMM). El mejor modelo entre los considerados es un modelo de error espacial estimado por mínimos cuadrados espacialmente ponderados, SWLS (Anselin y Rey, 2014), bajo el supuesto de heteroscedasticidad que utiliza el procedimiento KP-HET (ya que el test robusto de White había salido significativo). La matriz de pesos espaciales es de contigüidad tipo rook.

$$\text{VEMP13.07} = 69.2725521 + 0.0535363 \text{ SUPTECHIND12} - 0.0038189 \text{ VAB10} + U$$

$$U = 0.4561540 \text{ WU} + \varepsilon$$

No sale significativa la variable IPDR13.07, por lo que no se considera relevante su aportación a la creación de empleo. La superficie de techo industrial si que es significativa y positiva. El valor añadido bruto es también significativo pero negativo, ya que en los concejos con el VAB más alto la pérdida de empleos ha sido más acusada.

Si consideramos la variable IIPIL13.07 la realización del proceso de estimación también nos lleva a un modelo que no considera significativa la inversión correspondiente al primer pilar de la PAC.

Para la variable IPAC13.07 ocurre lo mismo y presenta el modelo el mismo comportamiento que con la variable anterior.

Los resultados para la muestra que excluye los concejos con valores extremos muestran cambios con respecto a la muestra completa. La diferencia fundamental es que no existe dependencia espacial en ninguno de los tres casos considerados, es decir para las tres variables que recogen la aportación de fondos públicos (IPDR13.07, IIPIL13.07 e IPAC13.07). También que se obtienen como favorables dos tipos de modelos, con las mismas variables explicativas en los tres casos, uno con solo dos variables explicativas y otro con cinco.

Los modelos de dos variables, estimados por MC2E, con matriz de contigüidad tipo queen, considerando como instrumental VPOBPAD14.07 y como instrumentos CONTRAT13 y RNET10, son:

Con IPDR13.07, con IIPIL13.07 y con IPAC13.07 que son no significativas individualmente:

$$\text{VEMP13.07} = -75.1834109 - 0.0547171 \text{ SUPTECHIND12} + 0.5285516 \text{ VPOBPAD14.07} + U$$

El coeficiente de SUPTECHIND12 es negativo, porque al estar las variables en términos absolutos se ha registrado una mayor caída de empleo en los concejos que previamente tenían más infraestructuras industriales. Con VPOBPAD14.07 ocurre lo contrario, como era de esperar, a mayor variación en la población mayor variación en el empleo y en el mismo sentido positivo.

Los modelos de cinco variables, estimados por MC2E, con VPOBPAD14.07 como variable instrumental y CONTRAT13 y RNET10 como instrumentos, son:

Con IPDR13.07

$$\text{VEMP13.07} = -34.3746686 + 0.0000174 \text{ IPDR13.07} + 0.3571715 \text{ VPOBPAD14.07} - 0.0389411 \text{ SUPTECHIND12} - 3.0007389 \text{ LIAEAGRI13} - 8.1685940 \text{ LICOBRI3} + U$$

Con IIPIL13.07

$$\text{VEMP13.07} = -25.6131836 + 0.0000195 \text{ IIPIL13.07} + 0.3535183 \text{ VPOBPAD14.07} - 0.0377047 \text{ SUPTECHIND12} - 6.0591866 \text{ LIAEAGRI13} - 8.0011108 \text{ LICOBRI13} + U$$

Con IPAC13.07

$$\text{VEMP13.07} = -31.2142767 + 0.0000102 \text{ IPAC13.07} + 0.3566159 \text{ VPOBPAD14.07} - 0.0379217 \text{ SUPTECHIND12} - 4.7190061 \text{ LIAEAGRI13} - 8.2877397 \text{ LICOBRI13} + U$$

En el primer bloque de modelos, con dos variables explicativas, la aportación de fondos públicos no es significativa pero si en el segundo bloque. Además en este caso con signo positivo, al contrario de lo ocurrido con todos los demás modelos considerados. Al comparar las características de ambos se obtienen resultados más favorables, por MCO, en cuanto a los criterios de información de Akaike y Schwarz, así como log likelihood, en los modelos del bloque con más variables. También tienen un  $R^2$  y  $R^2$  ajustado más alto, pero los modelos del primer bloque, con menos variables, no tienen heteroscedasticidad, lo que en principio mejoraría a los otros. En el caso de IPDR13.07 los coeficientes son altamente significativos individualmente, pero los coeficientes robustos calculados por el método de White no son significativos al 95% si estimamos por MCO, y tienen un p valor de 0.04987 si estimamos por MC2E. Quizá todo esto permita considerar como mejores los modelos con menos variables, ya que es posible que la heteroscedasticidad, unido a la no normalidad de los errores, estén dando resultados anómalos en los modelos con más variables. La mayoría de modelos considerados en este apartado, correspondiente al tratamiento de la dependencia espacial, presentan heteroscedasticidad, probablemente originada por la existencia de heterogeneidad espacial, según los indicios detectados en el AEDE que entre otras cosas manifestaban la existencia de clusters. Esta se tratará posteriormente con las técnicas correspondientes y se considerarán ambos grupos de modelos, para que cuando se haya corregido este problema, podamos aclarar el signo y la significatividad de los coeficientes de la inversión pública. De momento se aceptan como favorables los modelos que no muestran heteroscedasticidad y que no consideran significativas las variables de inversión pública, ya que es posible que al eliminar la heteroscedasticidad el signo sea negativo tal como ocurre en el resto de modelos, concretamente en los de dos variables, en los que no son significativos los test de White, B-P y K-B.

### *Modelo 3*

Se repite el proceso habitual que comienza con una primera estimación por MCO, con las variables preseleccionadas para la muestra completa, que incluye todos los concejos asturianos. De todos los modelos considerados se selecciona el que considera significativas las variables IPDR13.07, VEMP13.07, VABAGRI10 y ASS14, con una multicolinealidad aceptable (número de condición de 10.22), y no significativos los test de KB y robusto de White de heteroscedasticidad, aunque si significativo el de BP. En cuanto a la dependencia espacial es significativo el test LM-LAG al 95%. Si se estima por MC2E, ya que los errores no podemos aceptar que sean normales, al salir significativo el test de Jarque-Bera en la estimación inicial por MCO, el test de A-K no es significativo. Si lo son, de forma individual, las mismas variables descritas. De todas formas se plantea el modelo espacial autorregresivo (modelo tipo lag), resultando significativo el retardo de la endógena y resultando no significativa la existencia de dependencia espacial en el término de error. El mejor modelo de los considerados es

$$\text{VEMPAGRI13.07} = 17.4046370 + 0.3425330 \text{ WVEMPAGRI13.07} - 0.0000048 \text{ IPDR13.07} - 0.0059773 \text{ VABAGRI10} + 0.0045867 \text{ ASS14} + 0.0399541 \text{ VEMP13.07} + U$$

Se estima por el método de los momentos generalizados con regresores endógenos, considerando como instrumental el retardo de la endógena y como instrumentos los retardos de las exógenas. La variable IPDR13.07 es significativa pero de signo negativo, lo que indica un efecto inverso con respecto a la variación del empleo agrario experimentada en el periodo. Lo mismo ocurre con el valor añadido bruto del sector primario, siendo indicativo de una mayor caída o menor subida del empleo agrario en los concejos con el VABAGRI10 más alto. Son también significativas pero con relación directa la variación del empleo total y las afiliaciones a la seguridad social, siendo por tanto positiva su relación.

Si consideramos las inversiones destinadas al primer pilar, en vez de las destinadas a desarrollo rural, las variables explicativas serían las mismas, y con los mismos signos, salvo la que recoge la inversión pública, en este caso IPIL13.07 que no es significativa. El modelo también estimado por el método de los momentos generalizados, es también un modelo de retardo espacial, que no considera relevante la aportación de la inversión del primer pilar de la PAC.

$$\text{VEMPAGRI13.07} = 14.6848274 + 0.3960755 \text{ WVEMPAGRI13.07} - 0.0086241 \text{ VABAGRI10} + 0.0046429 \text{ ASS14} + 0.0387496 \text{ VEMP13.07} + U$$

Sí se considera significativa, y de signo negativo, la aportación total que incluye ambos pilares, resultando como seleccionado el siguiente modelo de retardo espacial.

$$\text{VEMPAGRI13.07} = 17.2211445 + 0.3638805 \text{ W VEMPAGRI13.07} - 0.0000021 \text{ IPAC13.07} - 0.0048905 \text{ VABAGRI10} + 0.0045703 \text{ ASS14} + 0.0399868 \text{ VEMP13.07} + U$$

Ante la posible influencia de los valores extremos, coincidentes con los concejos con núcleos más urbanos, estimamos los modelos para la muestra sin los valores de los concejos de Oviedo, Gijón, Avilés, Langreo, Mieres y Siero.

El modelo seleccionado con mejores propiedades se estima por MC2E, al salir significativo el test Jarque Bera de no normalidad de los errores. No se detecta dependencia espacial, ni en los test aplicados al modelo básico, estimado por MCO, ni el AK que se aplica a los residuos del modelo estimado por MC2E. Con la muestra sin valores extremos desaparece este problema. Son significativas con signo negativo tanto VABAGRI10, lo que muestra peores resultados en la variación del empleo en los concejos con mayor VAB agrario, como IPDR13.07, lo que también indica que con la muestra limitada a concejos menos urbanos también se detecta una relación negativa entre la variación del empleo agrario y la aportación de fondos públicos para desarrollo rural. El modelo es:

$$\text{VEMPAGRI13.07} = -3.3895721 - 0.0000052 \text{ IPDR13.07} - 0.0073425 \text{ VABAGRI10} + U$$

La inversión pública en el primer pilar (IIPIL13.07) no resulta significativa, y si lo serían junto con el VAB agrario (como antes con signo negativo), las licencias de actividad agrícola, con signo positivo, y el número de explotaciones con signo negativo. Esto sugiere que los concejos con más explotaciones tienen peores datos de variación de empleo agrario, perdiéndose más empleo en las que más explotaciones ganaderas tienen en 2014, quizás debido a que en los concejos que más dedicación ganadera tienen en 2014, más empleos por cierre de explotaciones se han perdido durante el periodo. Ocurre lo contrario con el número de licencias de actividad en agricultura en 2013, que tiene una relación directa y positiva con la variación en el empleo, ya que en aquellos concejos con más actividad agraria en 2013, no solo ganadera, ha sido positiva la variación del empleo en el periodo considerado. No sale significativa la existencia de

dependencia espacial. El modelo estimado por MC2E, con VABAGRI10 como variable instrumental y RNET10 y VEHIC13 como instrumentos, utilizando una matriz de contigüidad tipo rook sería:

$$\text{VEMPAGRI13.07} = 7.1450823 + 3.326321 \text{LIAEAGRI13} - 0.0231744 \text{VABAGRI10} - 0.1013536 \text{EXPLOT14} + U$$

Para el caso de la inversión total en política agraria (IPAC13.07), tampoco se detecta dependencia espacial y el modelo estimado por MC2E con IPAC como instrumental y ANIM14 y LIAEAGRI13 como instrumentos sería:

$$\text{VEMPAGRI13.07} = -5.3051705 - 0.0000022 \text{IPAC13.07} - 0.0063626 \text{VABAGRI10} + U$$

Se contempla como en el caso de IPDR13.07 una relación negativa y significativa de IPAC13.07, lo que indica la no consecución de aportación positiva a la variación del empleo agrario. Lo mismo ocurre con un alto valor añadido agrario en 2010.

#### *Modelo 4*

Para el modelo 4, con LIAEAGRI13 como variable endógena y la muestra completa (es decir para la totalidad de los concejos de Asturias), tras una primera estimación del modelo clásico por MCO y una segunda por MC2E (al presentar no normalidad en los errores) se llega a que el modelo más adecuado es el que considera significativas como exógenas a VPAR14.07 y a IPDR13.07, ambas con signo positivo. No salen significativos los test de dependencia espacial, por lo que no procede la formulación de modelos espaciales. La estimación de los modelos definitivos por MC2E, considerando como instrumental IPDR13.07, y como instrumentos ANIM14 y CORTMAD13, con matriz de contigüidad tipo rook, sería:

$$\text{LIAEAGRI13} = -3.4574905 + 0.0000055 \text{IPDR13.07} + 0.0202045 \text{VPAR14.07} + U$$

Para las inversiones del primer pilar y para las del total de la PAC (I1PIL13.07 e IPAC13.07) el proceso es similar y los modelos son:

$$\text{LIAEAGRI13} = -0.7657033 + 0.0000030 \text{I1PIL13.07} + 0.0191573 \text{VPAR14.07} + U$$

$$\text{LIAEAGRI13} = -1.7044260 + 0.0000019 \text{ IPAC13.07} + 0.0195225 \text{ VPAR14.07} + U$$

En estos tres casos, para la muestra completa salen significativas y con signo positivo las variables de inversión pública en políticas agrarias, y podemos constatar que según estos modelos han tenido una contribución positiva en el número de licencias agrarias en 2013. Lo mismo ocurre con la variación del paro, lo que es indicativo de que un mayor aumento del paro total presenta una relación positiva y significativa con el número de licencias en el sector agrícola.

Para la muestra que no considera los seis concejos más urbanos y que tienen valores extremos en la mayoría de variables consideradas, los resultados son los siguientes.

Para los fondos destinados al primer pilar de la PAC, en el primer paso se estima por MCO resultando significativas VABAGRI10 e IPDR13.07. Esta última presenta signo negativo. Se detecta dependencia espacial, que se recoge en el modelo espacial definitivo, en este caso un modelo mixto de retardo y error espacial, estimado por el método de los momentos generalizados, concretamente por mínimos cuadrados bietápicos espacial, considerando además la existencia de heteroscedasticidad por lo que se usa el método propuesto por Kelejian y Prucha (2010) para obtener estimadores robustos a heteroscedasticidad y autocorrelación espacial. En este modelo, que se muestra a continuación, no es significativa la variable de inversión pública.

$$\text{LIAEAGRI13} = 15.0028449 - 0.6386360 \text{ WLIAEAGRI13} + 0.0048738 \text{ VABAGRI10} + U$$

$$U = 0.5418809 \text{ WU} + \varepsilon$$

Para los casos de I1PIL13.07 y de IPAC13.07 no se detecta dependencia espacial, las variables son significativas y positivas. Los modelos estimados por MC2E, con matrices de contigüidad tipo rook, considerando la variable de inversión pública correspondiente en cada caso como instrumental y ANIM14 y CORTMAD13 en ambos casos como instrumentos serían:

$$\text{LIAEAGRI13} = -0.9493081 + 0.0000027 \text{ I1PIL13.07} + 0.0300004 \text{ VPAR14.07} + U$$

$$\text{LIAEAGRI13} = -1.9831207 + 0.0000017 \text{ IPAC13.07} + 0.0314792 \text{ VPAR14.07} + U$$

## 4.2.2. HETEROGENEIDAD ESPACIAL.

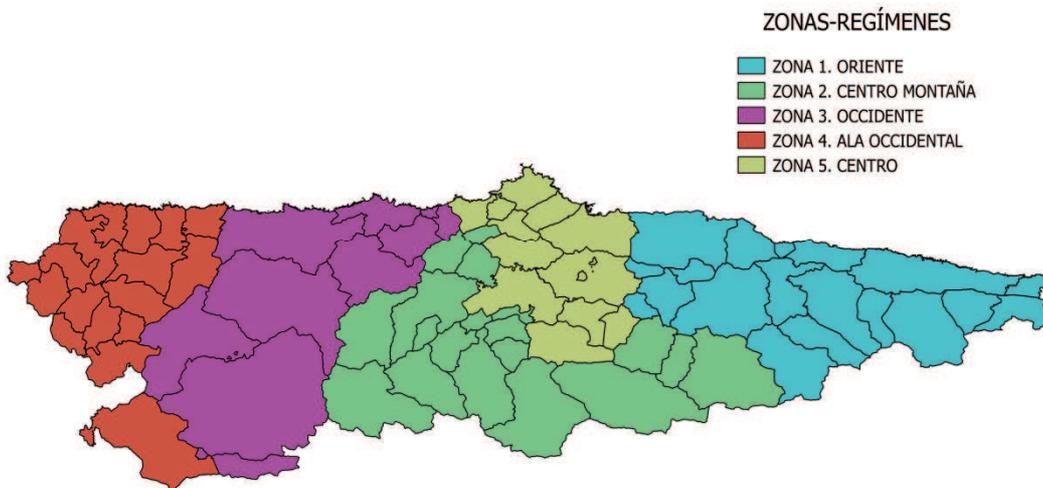
En el apartado anterior se ha estudiado la presencia de autocorrelación espacial en los modelos propuestos, utilizando modelos espaciales, si era procedente, para corregir su efecto. En la fase del AEDE hemos detectado además indicios de heterogeneidad espacial, sobre todo en los mapas de clusters que evidenciaban la existencia de inestabilidad estructural. Además casi todos los modelos analizados en el capítulo 5.2.1 presentan heteroscedasticidad.

Una de las alternativas para modelizar la heterogeneidad espacial en un modelo de regresión son las regresiones espaciales cambiantes o “switching regressions”. Se considera que la inestabilidad estructural es discreta y estima tantos valores para los coeficientes de una regresión como estructuras o regímenes espaciales se establezcan en la muestra total de observaciones.

En nuestro caso hemos propuesto cinco grupos o regímenes basándonos en los indicios detectados en el AEDE, así como en la existencia de comarcas y grupos de desarrollo rural que gestionan una parte del PDR, concretamente el eje LEADER. A estos grupos se les exige una coherencia y cohesión territorial y tienen una trayectoria continuada desde principios de los años 90. Son 11 en Asturias y sirven de base, junto con con criterios económicos y de actividad, así como con los resultados de los clusters del AEDE y también la división administrativa existente en el Principado de Asturias, para realizar los siguientes cinco regímenes espaciales:

1. **Zona Oriente.** Zona rural que incluye los grupos de acción local del Oriente de Asturias y la Comarca de la Sidra. Tiene importante actividad en el sector primario, baja actividad industrial, y su economía se basa fundamentalmente en el sector servicios. Grandes recursos turísticos.
2. **Zona Centro Montaña.** Zona rural dependiente y cercana a zonas industriales y urbanas del centro. Se corresponde con los GAL de Alto Nalón, Montaña Central y Camín Real de la Mesa.
3. **Zona Occidente.** Importante sector primario, con tradición y grandes recursos agrícolas y ganaderos. Grupos de acción local de Ese-Entrecabos, Alto Narcea-Muniellos y Bajo Nalón.

4. **Zona Ala Occidental.** Zona rural con muchos concejos pequeños e importantes dificultades orográficas. Incluye los GAL Navia-Porcía y Oscos-Eo
5. **Zona Centro.** Tiene un carácter más urbano e industrial que el resto, ya que incluye las zonas más pobladas, las ciudades y los núcleos más urbanos. También tiene las zonas con más tradición industrial. Incluye también zonas rurales de baja densidad de población y actividad del sector primario ya que la conformación territorial de Asturias integra zona rural y sector primario junto con zonas urbanas. Incluye las zonas excluidas de los grupos de acción local (es decir zona urbana) y los concejos de Carreño y Gozón, que forman el grupo de acción local del Cabo Peñas (ADICAP).



## MODELOS “SPATIAL REGIMES”

### *Modelo 1*

#### **Con IPDR13.07**

En la primera estimación por MCO, con una matriz tipo queen, se obtienen resultados del test de Chow significativos, tanto para el contraste global, así como para el individual de las variables RNET10 y VEMP13. No son significativos al 95% los test para la dependencia espacial conjunta, aunque si lo sería al 90% el test LM-SARMA. Individualmente no son significativos los test de dependencia espacial en los regímenes 1, 3 y 4. Lo sería, pero con un p valor de 0.487 el SARMA del 5 (y no el resto de este régimen) y el test LM-ERROR y el test SARMA del régimen 2. Los valores del número de condición de multicolinealidad son aceptables en todos los regímenes, ya que son inferiores a 10 en los cuatro primeros y solo superior, con 15.842, en el 5. La significatividad conjunta del modelo, con el estadístico F, permite rechazar la hipótesis nula de coeficientes iguales y nulos en las 5 zonas. El  $R^2$  y el  $R^2$  ajustado son bastante elevados (en el régimen 3 incluso cercanos a 1), salvo en la zona 1, en que son próximos a 0.5 y 0.4, respectivamente. El resultado del test de Jarque Bera no rechaza la normalidad de los errores en los regímenes 1, 3 y 5. No ocurre lo mismo en el 2 ni en el 4. Los tres test de heteroscedasticidad planteados (KB, BP y White) permiten no rechazar la hipótesis nula de homocedasticidad en los casos de los grupos 3, 4 y 5 pero sí se rechaza en los grupos 1 y 2. La significatividad individual de los coeficientes de las variables de inversión pública (IPDR13.07) permite incluirlos en las zonas 3 y 4, con signo negativo, y no son significativos en el resto, es decir 1, 2 y 5. Si observamos los resultados al aplicar el método de White, que calcula errores estándar robustos a heteroscedasticidad, se obtienen resultados muy similares, con la diferencia de que en el grupo 2 también sale significativo (aunque con 0.0426) el coeficiente de IPDR13.07, que también es negativo. Dado que en el modelo que no considera regímenes se presentaba dependencia espacial, y que en algún grupo los test presentaban dudas, e incluso eran significativos, hemos planteado el modelo de error espacial para comparar los resultados. Estos son satisfactorios ya que tanto por GMM como por máxima verosimilitud (ML o MV) en varios regímenes el retardo del error es significativo. El

que obtiene mejores resultados es el que estima por MV, y es por tanto el seleccionado, ya que en la mayoría de regímenes se podía asumir la normalidad de los errores (en tres de cinco). El test de Chow, en este caso, es muy significativo ya que lo es para todas las variables individualmente, así como de forma global. IPDR13.07 no es significativa en los grupos 1 y 5 y sí lo es en el resto con coeficiente negativo. Los regímenes del modelo del error espacial estimado por máxima verosimilitud quedarían así:

#### MAXIMUM LIKELIHOOD SPATIAL ERROR

$$1\_VPOBPAD14.07 = 0.0021483 \ 1\_RNET10 + 0.8099345 \ 1\_VEMP13.07 + 1\_U$$

$$2\_VPOBPAD14.07 = 61.3159469 - 0.0000447 \ 2\_IPDR13.07 + 0.8300463 \ 2\_VEMP13.07 + 2\_U$$

$$3\_VPOBPAD14.07 = -0.0000276 \ 3\_IPDR13.07 - 0.0032366 \ 3\_RNET10 + 3\_U$$

$$4\_VPOBPAD14.07 = -0.0000429 \ 4\_IPDR13.07 + 4\_U$$

$$5\_VPOBPAD14.07 = -0.0078894 \ 5\_RNET10 - 1.7935678 \ 5\_VEMP13.07 + 5\_U$$

$$2\_U = -0.9999960 \ 2\_W\_U + 2\_ε$$

$$3\_U = 0.5370677 \ 3\_W\_U + 3\_ε$$

$$4\_U = -0.7771899 \ 4\_W\_U + 4\_ε$$

$$5\_U = 0.6715758 \ 5\_W\_U + 5\_ε$$

Es procedente, a la vista de los resultados, la utilización de regímenes espaciales, ya que disminuye significativamente la heteroscedasticidad, el test de Chow es significativo tanto globalmente como individualmente (ver cuadro 4.3 del anexo 2) y se corrigen tanto autocorrelación como heterogeneidad espaciales. La variable de inversión pública es significativa en las zonas 3 y 4 pero con signo negativo y no es significativa en los otros tres, indicando nula o mala influencia en la variación de la población en el periodo considerado, según el modelo planteado. No se consigue el objetivo perseguido de dinamización de la población, al menos para el total de la región. Más adelante comprobaremos los resultados sin valores extremos en la muestra, que excluyen ciudades y zonas más urbanas. El resto de variables cambian de signo, cuando son significativas, dependiendo de la zona, lo que explica las distintas dinámicas

económicas en las zonas consideradas, debido probablemente a su heterogeneidad, lo que se manifiesta en la inestabilidad paramétrica de tipo discreto modelizada.

### Con IIPIL13.07

Para la variable IIPIL, en la estimación inicial por MCO, resultan significativos los test de dependencia espacial global, por lo que procede estimar modelos espaciales. No se detectan problemas de multicolinealidad y el test de Chow de inestabilidad estructural es significativo, tanto el global como los individuales, salvo para IPDR13.07, lo que indica la procedencia de los grupos propuestos. También mejora, con respecto al modelo anterior, que no tenía en cuenta los regímenes espaciales, la heteroscedasticidad y la normalidad de los errores. El modelo seleccionado, que mejores propiedades presenta y mejor explica sería un modelo mixto de retardo y error espaciales (combo) estimado por mínimos cuadrados bietápicos espacialmente ponderados (GSW2SLS) por el método de Kelejian y Prucha robusto a heteroscedasticidad.

#### SPATIALLY WEIGHTED TWO STAGE LEAST SQUARES (HET)

$$\begin{aligned}
 1\_VPOBPAD14.07 &= 0.0030257 \ 1\_RNET10 + 0.7432703 \ 1\_VEMP13.07 + 1\_U \\
 2\_VPOBPAD14.07 &= 82.5992392 \ -0.0000368 \ 2\_IIPIL13.07 + 0.8065020 \ 2\_VEMP13.07 + \\
 &0.2263630 \ 2\_W\_VPOBPAD14.07 + 2\_U \\
 3\_VPOBPAD14.07 &= 315.6154255 \ -0.0059739 \ 3\_RNET10 + 0.3325693 \ 3\_W\_VPOBPAD14.07 + 3\_U \\
 4\_VPOBPAD14.07 &= 0.0027779 \ 4\_RNET10 + 1.1254552 \ 4\_VEMP1307 + 4\_U \\
 5\_VPOBPAD14.07 &= -0.0074534 \ 5\_RNET10 \ -1.6902686 \ 5\_VEMP1307 + 5\_U \\
 2\_U &= -1 \ 2\_W\_U + 2\_ε \\
 5\_U &= 0.7581094 \ 5\_W\_U + 5\_ε
 \end{aligned}$$

El test de Chow es significativo tanto globalmente como individualmente para todas las variables, por lo que se confirma la inestabilidad estructural paramétrica de las zonas propuestas (ver cuadro 4.4 del anexo 2).

La variable IIPIL13.07 solo es significativa en la zona 2, y con signo negativo, por lo que al igual que en el caso de IPDR13.07, tampoco está cumpliendo con las expectativas previstas. El resto de variables cambia de signo de unos grupos a otros por lo que también muestran las distintas dinámicas, concretamente cambia de los grupos 1 al 4, en que tienen signo positivo y muestran una correlación significativa y positiva con la variación de la población, con respecto al grupo 5, mas urbano y en el que los coeficientes de la renta disponible y variación del empleo tienen signo negativo, probablemente por la influencia de los grandes centros urbanos y la polarización centro-alias que se da en la región.

### Con IPAC13.07

Tras un proceso similar en el que en la primera estimación por MCO mejora mucho la heterocedasticidad, la normalidad de los errores, la multicolinealidad es baja, la significatividad conjunta e individual de las variables es positiva y los test de dependencia espacial apuntan a un modelo error o mixto, se selecciona como mas apropiado el siguiente modelo mixto de retardo y error espacial (combo) estimado por mínimos cuadrados bietápicos espacialmente ponderados (GSW2SLS) por el método de Kelejian y Prucha robusto a heteroscedasticidad.

#### SPATIALLY WEIGHTED TWO STAGE LEAST SQUARES (HET) COMBO

$$1\_VPOBPAD14.07 = 0.002604 \ 1\_RNET10 + 0.7325867 \ 1\_VEMP13.07 + 1\_U$$

$$2\_VPOBPAD14.07 = 79.6585366 \ -0.0000196 \ 2\_IPAC13.07 + \ 0.8023719 \ 2\_VEMP13.07 + 0.2051332 \ 2\_W\_VPOBPAD14.07 + 2\_U$$

$$3\_VPOBPAD14.07 = 273.1372289 \ -0.0054954 \ 3\_RNET10 + 0.2659463 \ 3\_W\_VPOBPAD14.07 + 3\_U$$

$$4\_VPOBPAD14.07 = -0.0000167 \ 4\_IPAC13.07 + 0.0027978 \ 4\_RNET10 + 0.9615930 \ 4\_VEMP13.07 + 4\_U$$

$$5\_VPOBPAD14.07 = -0.0074060 \ 5\_RNET10 \ -1.6795816 \ 5\_VEMP \ 13.07 + 5\_U$$

$$1\_U = -0.7931191 \ 2\_W\_U + 1\_ \varepsilon$$

$$2\_U = -1 \ 3\_W\_U + 2\_ \varepsilon$$

$$5\_U = 0.7511778 \ 5\_W\_U + 5\_ \varepsilon$$

El test de Chow es significativo para todas las variables individualmente y también globalmente (ver cuadro 4.5 del anexo 2), al igual que ocurría con IPIL13.07 y se repite también el mismo comportamiento en el signo de las variables, resultando negativo en el caso de IPAC13.07, para el segundo y cuarto regímenes, en los que son significativas, ya que en el resto no lo son, y positivo para RNET10 y VEMP13.07 en todas las zonas salvo en la 5, zona central más “urbana”.

### *Modelo 1. MUESTRA SIN OUTLIERS*

Tal como hicimos anteriormente, en los modelos que no consideraban la inestabilidad paramétrica, vamos a estimar de nuevo los modelos con la muestra que elimina aquellos concejos con valores más altos en la mayoría de variables consideradas, que coinciden con los concejos con núcleos más urbanos.

#### **Con IPDR13.07**

Los test de dependencia espacial sobre los residuos de la primera estimación por MCO son significativos en el caso del LM-ERROR y LM-SARMA. El resto de test realizados indican una importante mejora en la heteroscedasticidad y normalidad de errores con respecto al modelo estimado sin considerar regímenes espaciales. Los test de significación conjunta e individual, multicolinealidad y  $R^2$ , siguen en valores correctos. Para recoger el fenómeno de dependencia espacial surgido al estimar por regímenes espaciales se estiman distintos modelos espaciales escogiéndose, por sus mejores propiedades, el siguiente modelo del error espacial, estimado por máxima verosimilitud.

#### MAXIMUM LIKELIHOOD SPATIAL ERROR

$$1\_VPOBPAD14.07 = 0.1099710 \ 1\_CONTRAT13 + 0.7922932 \ 1\_VEMP13.07 + 1\_U$$

$$2\_VPOBPAD14.07 = -0.0000499 \ 2\_IPDR13.07 + 1.0924876 \ 2\_VEMP13.07 + 2\_U$$

$$3\_VPOBPAD14.07 = -0.0000360 \ 3\_IPDR13.07 + 0.5268054 \ 3\_VEMP13.07 + 3\_U$$

$$4\_VPOBPAD14.07 = -0.0000381 \ 4\_IPDR13.07 + 0.6009630 \ 4\_VEMP13.07 + 4\_U$$

$$5\_VPOBPAD14.07 = -0.0005178 \ 5\_IPDR13.07 - 0.4842678 \ 5\_CONTRAT13 + 2.4252931$$

$$5\_VEMP13.07 + 5\_U$$

$$2_U = -0.9999960 \ 2_{W_U} + 2_{\varepsilon}$$

$$3_U = -0.6852236 \ 3_{W_U} + 3_{\varepsilon}$$

$$4_U = -0.8121820 \ 4_{W_U} + 4_{\varepsilon}$$

$$5_U = -0.796171 \ 5_{W_U} + 5_{\varepsilon}$$

La significatividad del test de Chow global e individual para todas las variables, salvo para el coeficiente autorregresivo espacial del error, indican la validez del método (ver cuadro 4.6 del anexo 2). Los coeficientes significativos de la variable IPDR13.07, en cuatro de las cinco zonas, vuelven a indicar, según este modelo, la falta de eficacia del gasto público para favorecer la variación positiva de la población, tal como se pretendía. Sí que presenta capacidad explicativa, y con correlación positiva, la variación del empleo en los cinco regímenes, como era de esperar. CONTRAT13 solo lo tiene en las zonas 1 y 5, con signo positivo y negativo respectivamente, poniendo de manifiesto de nuevo la distinta estructura económica en la zona centro, no solo en los coeficientes estimados sino también en el sentido en el que explican las mismas variables.

### Con I1PIL13.07

Con la variable I1PIL13.07 no se presentan problemas de dependencia espacial al calcular los test LM-ERROR ni LM-LAG para los residuos de la estimación por MCO. El modelo que mejor explica es el estimado por mínimos cuadrados bietápicos calculando los errores estándar por el método de White robusto a heteroscedasticidad.

#### TWO STAGE LEAST SQUARES ESTIMATION (White Standard Errors)

$$1\_VPOBPAD14.07 = 0.0028343 \ 1\_VAB10 + 0.7709403 \ 1\_VEMP13.07 + 1\_U$$

$$2\_VPOBPAD14.07 = -0.0000388 \ 2\_I1PIL13.07 + 2\_U$$

$$3\_VPOBPAD14.07 = -0.0053878 \ 3\_VAB10 + 3\_U$$

$$4\_VPOBPAD14.07 = -63.6381616 + 0.0011069 \ 4\_VAB10 + 1.0946051 \ 4\_VEMP13.07 + 4\_U$$

Variable instrumental VAB10

Variable instrumento RNET10

El test de Chow (ver cuadro 4.7 del anexo 2) resulta significativo globalmente e individualmente para VEMP13.07 y VAB10. No lo sería para I1PIL13.07, pero es que esta solo es significativa en la zona 2. Tiene signo negativo, lo que indica el efecto contrario al esperado, unido a la no significatividad en el resto de zonas del territorio asturiano. Es de resaltar la no significatividad de ninguna variable en la zona 5, ya que al quitar los outliers de la muestra, que son todos de esa zona, se pone de manifiesto la distinta dinámica frente al resto de zonas. Para las zonas dos y cuatro son significativas las mismas variables con el mismo signo (positivo ambas), y no lo son para el resto, es decir también siguen distinta dinámica para explicar la variación de la población las zonas 2 y 3 cuando se considera I1PIL13.07 como variable explicativa.

### Con IPAC13.07

Con la inversión total en fondos de ambos pilares de la PAC si que es conveniente utilizar modelos espaciales. El modelo que mejor explica es un modelo mixto del retardo y del error estimado por mínimos cuadrados bietápicos espacialmente ponderados, utilizando la propuesta de Kelejian y Prucha (2010), GSW2SLS (HET):

#### SPATIALLY WEIGHTED TWO STAGE LEAST SQUARES (HET) COMBO

$$\begin{aligned}
 1\_VPOBPAD14.07 &= 0.1270776 \ 1\_CONTRAT13 + 0.7745191 \ 1\_VEMP13.07 + 1\_U \\
 2\_VPOBPAD14.07 &= 85.7884117 - 0.0000195 \ 2\_IPAC13.07 + 0.810316 \ 2\_VEMP + 0.2581287 \\
 & \ 2\_W\_VPOBPAD14.07 + 2\_U \\
 3\_VPOBPAD13.07 &= -0.000006 \ 3\_IPAC13.07 - 0.2873099 \ 3\_CONTRAT13 + 0.6484948 \\
 & \ 3\_VEMP13.07 + 0.3039132 \ 3\_W\_VPOBPAD14.07 + 3\_U \\
 4\_VPOBPAD13.07 &= -0.0000115 \ 4\_IPAC13.07 + 0.1523662 \ 4\_CONTRAT13 + 0.9780488 \\
 & \ 4\_VEMP13.07 + 4\_U \\
 5\_VPOBPAD13.07 &= 0.3033120 \ 5\_CONTRAT13 + 1.8807645 \ 5\_VEMP13.07 + 5\_U \\
 1\_U &= -0.8371299 \ 1\_W\_U + 1\_ε \\
 2\_U &= -1 \ 2\_W\_U + 2\_ε
 \end{aligned}$$

Nuevamente IPAC13.07 tiene signo negativo en las zonas en las que es significativa. El modelo pone de manifiesto la inestabilidad paramétrica entre las distintas zonas de estudio, aunque el test de Chow no es significativo para VEMP13.07, ni para el retardo espacial de la endógena, ni para el retardo del error, si lo es para el resto de variables así como globalmente (ver cuadro 4.8 del anexo 2). Ha mejorado la heteroscedasticidad la normalidad de errores y se han corregido con respecto al modelo sin regímenes espaciales tanto la heterogeneidad como la dependencia espacial.

## *Modelo 2*

### **Con IPDR13.07**

En la primera estimación por MCO no son significativos los test LM-LAG ni LM-ERROR. El test de Jarque-Bera no permite rechazar la hipótesis nula de normalidad en los errores en tres de los cinco regímenes, por lo que se estima también por mínimos cuadrados bietápicos (2SLS) dando mejores resultados la estimación robusta a heteroscedasticidad de White para los errores estándar.

#### TWO STAGE LEAST SQUARES ESTIMATION (White Standard Errors)

$$2\_VEMP13.07 = -0.0075663 \ 2\_VAB10 + 2\_U$$

$$3\_VEMP13.07 = -0.0028832 \ 3\_VAB10 + 3\_U$$

$$4\_VEMP13.07 = -0.0000214 \ 4\_IPDR \ 13.07 - 0.0015567 \ 4\_VAB10 + 4\_U$$

$$5\_VEMP13.07 = 0.0459874 \ 5\_SUPTECHIND12 - 0.0039001 \ 5\_VAB10 + 5\_U$$

IPDR13.07 solo es significativa, y con un coeficiente negativo en la zona 4, que es el ala occidental, y no son significativas las variables consideradas en la zona 1, el ala oriental. No parece que la inversión en desarrollo rural sea estadísticamente relevante con los modelos considerados. Sí que es conveniente la modelización de la inestabilidad paramétrica discreta planteada ya que los test de Chow son significativos globalmente y para VAB10 (ver cuadro 4.9 del anexo 2). No lo son para IPDR13.07 ni para SUPTECHIND12, pero es que solo son significativas en un régimen cada una.

SUPTECHIND12 lo es como era de esperar en la zona central, que se corresponde con la más industrializada.

### Con IIPIL13.07 y con IPAC13.07

Ninguna de las dos variables es significativa, por lo que se eliminan del modelo. El que mejores propiedades y poder explicativo presenta es el estimado por mínimos cuadrados espacialmente ponderados, robusto a heteroscedasticidad, SWLS (HET). Los coeficientes estimados para la variable VAB10 tienen signo negativo en los cuatro regímenes en los que es significativa ya que la caída en el empleo ha sido mayor en los concejos con niveles de producción más altos en términos de valor añadido, ya que son los que más empleos han perdido al ser los más grandes y estar medida dicha variación en términos absolutos. La superficie industrial es explicativa y relevante solo para la zona central y con mayor grado de industrialización. En la zona 1, es decir el oriente de Asturias, no son significativas las variables de este modelo. El test de Chow de inestabilidad estructural, es significativo en todos los casos, individual y globalmente (ver cuadro 4.10 del anexo 2).

#### SPATIALLY WEIGHTED LEAST SQUARES ESTIMATION (HET)

$$2\_VEMP13.07 = -0.0063645 \ 2\_VAB10 + 2\_U$$

$$3\_VEMP13.07 = -0.0033377 \ 3\_VAB10 + 3\_U$$

$$4\_VEMP13.07 = -50.8592605 - 0.0015251 \ 4\_VAB10 + 4\_U$$

$$5\_VEMP13.07 = -0.0038666 \ 5\_VAB10 + 0.0451281 \ 5\_SUPTECHIND12 + 5\_U$$

$$5\_U = 0.5354364 \ 5\_W\_U + \varepsilon$$

### *Modelo 2. MUESTRA SIN OUTLIERS.*

Para la muestra sin outliers no son significativas las variables de inversión pública. No son significativos los test de dependencia espacial en los residuos de la estimación por MCO. El modelo que mejor explica es el que se estima por MCO con el método de White para estimadores robustos a heteroscedasticidad.

ORDINARY LEAST SQUARES ESTIMATION.

$$1\_VEMP13.07 = -71.7605795 - 0.0019843 \ 1\_VAB10 + 1\_U$$

$$2\_VEMP13.07 = 76.5891837 - 0.0065535 \ 2\_VAB10 + 2\_U$$

$$3\_VEMP13.07 = -0.0029201 \ 3\_VAB10 + 3\_U$$

$$4\_VEMP13.07 = -54.8293463 - 0.0015335 \ 4\_VAB10 + 4\_U$$

Al aplicar regímenes espaciales la variable que mejor explica pasa a ser VAB10. El modelo de cinco variables propuesto en el apartado 4.2.1, que presentaba dudas con respecto al finalmente seleccionado en los modelos estimados sin considerar inestabilidad estructural, tiene peores propiedades que los demás modelos, por lo que se excluye definitivamente. Tendríamos por tanto o bien el considerado inicialmente, que no presentaba heteroscedasticidad y no se le aplican regímenes espaciales (además el test de Chow en ese caso no es significativo individualmente para ninguna variable y lo es con un p valor de 0.041 globalmente) o este que si los considera procedentes. El test de Chow es significativo global e individualmente (ver cuadro 4.11 del anexo 2). En ambos casos las variables de inversión pública no son significativas, lo que refuerza el resultado que indica no haber tenido una relevancia en el proceso de animación del empleo a nivel regional, tampoco en la muestra sin concejos con importantes núcleos urbanos.

***Modelo 3*****Con IPDR13.07**

El modelo clásico por MCO muestra una corrección importante de la heteroscedasticidad que se había manifestado en el modelo sin regímenes espaciales. También mejora la normalidad de los errores en los cinco grupos o zonas. La dependencia espacial es significativa para el test SARMA, y el modelo seleccionado corrige la posible existencia de dependencia espacial y es el que mejor explica.

SPATIALLY WEIGHTED TWO STAGE LEAST SQUARES (HET) (combo)

$$\begin{aligned}
1\_VEMPAGRI13.07 &= -0.0099193 \ 1\_VABAGRI10 + 0.0540769 \ 1\_VEMP13.07 + 1\_U \\
2\_VEMPAGRI13 &= -0.0333285 \ 2\_ASS14 + 2\_U \\
3\_VEMPAGRI13 &= -0.0068628 \ 3\_VABAGRI10 + 0.28774393\_VEMP13.07 + 3\_U \\
4\_VEMPAGRI13 &= -0.0000041 \ 4\_IPDR -0.0078199 \ 4\_ASS14 -0.0031760 \ 4\_VABAGRI10 + \\
&0.0812608 \ 4\_VEMP13.07 + 0.4143796 \ 4\_W\_VEMPAGRI13.07 + 4\_U \\
5\_VEMPAGRI13 &= 48.1110126 + 0.0048394 \ 5\_ASS14 -0.0051956 \ 5\_VABAGRI10 + 0.0448996 \\
&5\_VEMP13.07 + 0.6925558 \ 5\_W\_VEMPAGRI13.07 + 5\_U \\
3\_U &= -1 \ 3\_W\_U + 3\_ε \\
4\_U &= -1 \ 4\_W\_U + 4\_ε
\end{aligned}$$

La variable IPDR13.07 es significativa solo en la zona del ala occidental y con signo negativo, no resultando significativa en el resto de zonas. Tampoco parece haber tenido un efecto positivo en la variación del empleo agrario. El resto de variables significativas tienen los signos esperados, el empleo de todos los sectores positivo, el valor añadido bruto agrario negativo, ya que la caída del empleo en términos absolutos es mayor en los concejos con más actividad del sector primario y las afiliaciones a la seguridad social han sido significativas con signo positivo en la zona central, con más actividad en todos los sectores, y con signo negativo en la zona centro montaña donde las afiliaciones en otros sectores determinan mayor caída del empleo agrario y los problemas de relevo generacional son mayores por la cercanía e influencia del área central más urbana e industrial. El test de Chow de inestabilidad paramétrica resulta significativo globalmente y para todas las variables excepto VABAGRI13 (ver cuadro 4.12 del anexo 2).

Para las variables IIPIL13.07 e IPAC13.07 se escogen modelos estimados por el mismo método anterior, con similar comportamiento y propiedades. La novedad en ambos casos es que estas variables que o no son significativas, o lo son con signo negativo, si que han tenido, como excepción, una correlación positiva y significativa en el régimen 5, es decir la zona centro mas “urbanizada”. Esto quizás sea debido a que algunos concejos de la zona centro pueden presentar explotaciones más competitivas y

profesionalizadas, con mejores condiciones geográficas, con mayor capacidad de creación de empleo y quizás dentro de un entorno más dinámico. En los modelos estimados con la muestra sin outliers veremos si se repite este resultado. Los modelos en este caso son los siguientes (ver resultados del test de Chow en cuadros 4.13 y 4.14 del anexo 2).

### Con I1PIL13.07

#### SPATIALLY WEIGHTED TWO STAGE LEAST SQUARES (HET) (combo)

$$1\_VEMPAGRI13.07 = 70.4917491 - 0.0000059 \ 1\_I1PIL13.07 + 0.0411829 \ 1\_VEMP13.07 + 1.1809658 \ 1\_W\_VEMPAGRI13.07 + 1\_U$$

$$2\_VEMPAGRI13.07 = -23.9612825 - 0.0318945 \ 2\_ASS14 + 2\_U$$

$$3\_VEMPAGRI13.07 = 75.9091984 - 0.0065429 \ 3\_VABAGRI10 + 0.2123068 \ 3\_VEMP13.07 + 0.4603924 \ 3\_W\_VEMPAGRI13.07 + 3\_U$$

$$4\_VEMPAGRI13.07 = 19.4088497 - 0.0000060 \ 4\_I1PIL13.07 + 0.2431089 \ 4\_VEMP13.07 + 0.7264384 \ 4\_W\_VEMPAGRI13.07 + 4\_U$$

$$5\_VEMPAGRI13.07 = 50.0323599 + 0.0052878 \ 5\_ASS14 + 0.0000086 \ 5\_I1PIL13.07 - 0.0083190 \ 5\_VABAGRI10 + 0.0458596 \ 5\_VEMP13.07 + 0.6960644 \ 5\_W\_VEMPAGRI13.07 + 5\_U$$

$$1\_U = -1 \ 1\_W\_U + 1\_ε$$

$$3\_U = -1 \ 3\_W\_U + 3\_ε$$

$$4\_U = -1 \ 4\_W\_U + 4\_ε$$

**Con IPAC13.07**SPATIALLY WEIGHTED TWO STAGE LEAST SQUARES (HET) (combo)

$$1\_VEMPAGRI13.07 = 52.9442885 - 0.0065103 \ 1\_VABAGRI10 + 0.0525881 \ 1\_VEMP13.07 + 0.9184412 \ 1\_W\_VEMPAGRI13.07 + 1\_U$$

$$2\_VEMPAGRI13.07 = -0.0344798 \ 2\_ASS14 + 2\_U$$

$$3\_VEMPAGRI13.07 = 69.2068828 - 0.0061819 \ 3\_VABAGRI10 + 0.2480401 \ 3\_VEMP13.07 + 3\_U$$

$$4\_VEMPAGRI13.07 = 15.9150007 - 0.0000033 \ 4\_IPAC13.07 + 0.1483714 \ 4\_VEMP13.07 + 0.5559430 \ 4\_W\_VEMPAGRI13.07 + 4\_U$$

$$5\_VEMPAGRI13.07 = 49.9153565 + 0.0000064 \ 5\_IPAC13.07 + 0.0051642 \ 5\_ASS14 - 0.0076737 \ 5\_VABAGRI10 + 0.0455797 \ 5\_VEMP13.07 + 0.7021797 \ 5\_W\_VEMPAGRI13.07 + 5\_U$$

$$1\_U = -0.8724065 \ 1\_W\_U + 1\_ \varepsilon$$

$$3\_U = -1 \ 3\_W\_U + 3\_ \varepsilon$$

$$4\_U = -1 \ 4\_W\_U + 4\_ \varepsilon$$

***Modelo 3. MUESTRA SIN OUTLIERS.*****Con IPDR13.07**

El modelo seleccionado es un modelo de error espacial. Aunque al estimar por MCO no son significativos los test LM LAG ni LM ERROR, si que sale significativo LM SARMA (aunque solo con un p valor de 0.045 al usar una matriz de contigüidad de tipo queen). Al probar modelos espaciales no son significativos los retardos espaciales, salvo con el modelo error estimado por máxima verosimilitud. Al comprobar los resultados del diagrama de dispersión de Moran para los valores previstos y para los residuos se observa que se elimina con el modelo espacial parte de la poca autocorrelación existente al incluir el “filtrado espacial”, por lo que el modelo del error mejora al modelo estimado por MCO en lo referente a la dependencia espacial, además también mejora en el resto de propiedades y por tanto se selecciona como el más apropiado.

MAXIMUM LIKELIHOOD SPATIAL ERROR

$$1\_VEMPAGRI13.07 = -0.0116488 1\_VABAGRI10 + 1\_U$$

$$2\_VEMPAGRI13.07 = -0.0152456 2\_VABAGRI10 + 2\_U$$

$$3\_VEMPAGRI13.07 = -0.0152456 3\_VABAGRI10 + 3\_U$$

$$4\_VEMPAGRI13.07 = -0.0000055 4\_IPDR13.07 - 0.0087859 4\_VABAGRI13.07 + 4\_U$$

$$5\_VEMPAGRI13.07 = -0.0000955 5\_IPDR13.07 + 0.0196215 5\_VABAGRI10 + 5\_U$$

$$3\_U = -0.7681237 3\_W\_U + 3\_ε$$

$$5\_U = 0.8916655 5\_W\_U + 5\_ε$$

El test de Chow es altamente significativo en todos los casos (global e individual) salvo para el término constante (ver cuadro 4.15 del anexo 2). IPDR13.07 es significativa pero con signo negativo en los grupos 4 y 5. VABAGRI13 es significativa en todos, pero en el grupo 5 con signo contrario, lo que refleja distinto comportamiento entre centro y resto de zonas, incluso al eliminar los valores extremos que podrían distorsionar la muestra y utilizando solo zonas no urbanas.

**Con IIPIL13.07**

IIPIL13.07 no era significativa en el modelo anterior pero considerando inestabilidad estructural paramétrica es significativo en el cuarto de los cinco regímenes propuestos, para un modelo de error espacial que se estima por el método de MV, ya que previamente, en la primera estimación por MCO se detecta dependencia espacial y se hace aconsejable modelizar este fenómeno. Al considerar regímenes espaciales mejora mucho la heteroscedasticidad ya que se corrige el fenómeno de heterogeneidad. Los test de Chow son significativos globalmente y de forma individual para todas las variables excepto IIPIL13.07, lo que demuestra la conveniencia de este método para corregir la heterogeneidad espacial.

MAXIMUM LIKELIHOOD SPATIAL ERROR

$$\begin{aligned}
1\_VEMPAGRI13 &= -9.9650488 -1.5399566 1\_LIAEAGRI13 + 1\_U \\
2\_VEMPAGRI13 &= -0.1314019 2\_EXPLOTT14 + 2.6460996 2\_LIAEAGRI13 -0.0140190 \\
&2\_VABAGRI10 + 2\_U \\
3\_VEMPAGRI13 &= 48.6341874 -0.2357847 3\_EXPLOTT14 + 1.8477336 3\_LIAEAGRI13 -0.0184555 \\
&3\_VABAGRI10 + 3\_U \\
4\_VEMPAGRI13 &= 0.000005 4\_I1PIL -0.1740887 4\_EXPLOTT14 + 0.5285194 4\_LIAEAGRI13 - \\
&0.0122596 4\_VABAGRI10 + 4\_U \\
2\_U &= -0.8972995 2\_W\_U + 2\_ε \\
3\_U &= -0.999996 3\_W\_U + 3\_ε
\end{aligned}$$

El test de Chow (ver cuadro 4.16 del anexo 2) no es significativo para I1PIL13.07, ya que solo es significativa en la zona 4 del ala occidental. No lo es en el resto pero en esta zona si lo ha sido y además con signo positivo. Aunque es un resultado positivo solo lo ha sido en un régimen, no siéndolo en el resto del territorio asturiano. En el grupo 5 de la zona central no es significativa ninguna variable. Se vuelva a obtener un resultado distinto al resto del territorio.

**Con IPAC13.07**

La variable IPAC13.07 no es significativa en ningún régimen. El mejor modelo es el estimado por MCO, donde el coeficiente de la variable VABAGRI13.07 es significativo y negativo en las zonas 1 a 4 y no lo es en la zona 5 más urbana. El modelo sería el siguiente, utilizando una matriz de contigüidad rook.

ORDINARY LEAST SQUARES ESTIMATION

$$\begin{aligned}
1\_VEMPAGRI13 &= -0.0109148 1\_VABAGRI13 + 1\_U \\
2\_VEMPAGRI13 &= -0.0148525 2\_VABAGRI13 + 2\_U \\
3\_VEMPAGRI13 &= -0.0113191 3\_VABAGRI13 + 3\_U \\
4\_VEMPAGRI13 &= -0.0096212 4\_VABAGRI13 + 4\_U
\end{aligned}$$

## Modelo 4

### Con IPDR13.07

El modelo planteado solo contempla significativo IPDR13.07 en el régimen 5, es decir en la zona central, y no en el resto. Es un modelo no espacial estimado por MCO, ya que no son significativos los test de dependencia espacial aplicados a los errores. Los errores estándar se calculan con el método de White, robusto a heteroscedasticidad. Si no se aplicase este método IPDR13.04 no sería significativa en ninguna zona. En la zona 3 no es significativa ninguna variable. La explicativa VPAR14.07 tiene coeficientes estimados con signo positivo, lo que indica que una variación en el aumento del paro para el periodo explicaría un aumento de las licencias agrarias en 2013. Esto quizás se deba a la consideración de nuevas empresas agrarias como alternativa laboral para algunos colectivos de parados. La aplicación de este método de regresiones cambiantes hace que mejore la normalidad de los errores y aunque no se elimina, si se observa una mejora de la heteroscedasticidad con respecto a la estimación inicial que no consideraba variación en los parámetros.

#### ORDINARY LEAST SQUARES ESTIMATION (White Standard Errors)

$$1\_LIAEAGRI13 = 0.0897293 \ 1\_VPAR14.07 + 1\_U$$

$$2\_LIAGEAGRI13 = 0.0710647 \ 2\_VPAR14.07 + 2\_U$$

$$4\_LIAEAGRI13 = 0.0811421 \ 4\_VPAR14.07 + 4\_U$$

$$5\_LIAEAGRI13 = 0.0000177 \ 5\_IPDR13.07 + 0.0183195 \ 5\_VPAR14.07 + 5\_U$$

Ver cuadro 4.17 del anexo 2 para resultados del test de Chow.

### Con IIPIL13.07

Si que se detecta en este caso dependencia espacial, por lo que se estiman modelos espaciales. El que mejor explica es un modelo de error espacial estimado por máxima verosimilitud, ya que la normalidad de errores se corrige en regresiones cambiantes en casi todos los grupos, por lo que es procedente el método. Se utiliza una matriz de contigüidad rook. El test de Chow de inestabilidad estructural es significativo en todos

los casos salvo para el término constante (ver cuadro 4.18 del anexo 2). IIPIL13.07 es significativo en todas las zonas salvo en el ala oriental.

MAXIMUM LIKELIHOOD SPATIAL ERROR.

$$\begin{aligned}
 1\_LIAEAGRI13 &= 0.0659928 \quad 1\_VPAR14.07 + 1\_U \\
 2\_LIAEAGRI13 &= 0.0000013 \quad 2\_IIPIL13.07 + 0.0543021 \quad 2\_VPAR14.07 + 2\_U \\
 3\_LIAEAGRI13 &= 0.0000025 \quad 3\_IIPIL13.07 + 3\_U \\
 4\_LIAEAGRI13 &= 0.0000031 \quad 4\_IIPIL13.07 + 4\_U \\
 5\_LIAEAGRI13 &= 0.0000053 \quad 5\_IIPIL13.07 + 0.0185595 \quad 5\_VPAR14.07 + 5\_U \\
 1\_U &= 0.5053028 \quad 1\_W\_U + 1\_ \varepsilon \\
 2\_U &= 0.4869712 \quad 2\_W\_U + 4\_ \varepsilon \\
 4\_U &= 0.6807980 \quad 4\_W\_U + 4\_ \varepsilon
 \end{aligned}$$

**Con IPAC13.07**

Al igual que ocurría con IPDR13.07, no se detecta dependencia espacial en el modelo estimado por MCO, con una matriz de contigüidad queen. El test de Chow es significativo en todos los casos salvo para el término constante (ver cuadro 4.19 del anexo 2). Mejora la heteroscedasticidad y la normalidad de los errores. En este caso VPAR14.07 es significativa en todas las zonas y con signo positivo. IPAC13.07 es significativa en tres de los cinco grupos, que se corresponden con centro, occidente y ala occidental.

ORDINARY LEAST SQUARES ESTIMATION

$$\begin{aligned}
 1\_LIAEAGRI13 &= 0.0820761 \quad 1\_VPAR14.07 + 1\_U \\
 2\_LIAEAGRI13 &= 0.0677411 \quad 2\_VPAR14.07 + 2\_U \\
 3\_LIAEAGRI13 &= 0.0000016 \quad 3\_IPAC13.07 + 0.0059459 \quad 3\_VPAR14.07 + 3\_U \\
 4\_LIAEAGRI13 &= 0.0000017 \quad 4\_IPAC13.07 + 0.0357804 \quad 4\_VPAR14.07 + 4\_U \\
 5\_LIAEAGRI13 &= 0.0000042 \quad 5\_IPAC13.07 + 0.0185159 \quad 5\_VPAR14.07 + 5\_U
 \end{aligned}$$

**Modelo 4. MUESTRA SIN OUTLIERS.****Con IPDR13.07**

Al eliminar los concejos mas urbanos de la muestra se aconseja la modelización de la autocorrelación espacial con un modelo de error espacial por el método es máxima verosimilitud. La variable de inversión es significativa en el ala oriental y la zona central. En el primer caso con signo negativo y positivo en el segundo. El desarrollo rural con enfoque territorial y multisectorial tiene mejores resultados en la zona más dependiente o interrelacionada con otros sectores y áreas urbanas, asi como efecto negativo en el ala oriental. En el resto de zonas tampoco es significativo por lo que no se consiguen los resultados deseados.

MAXIMUM LIKELIHOOD SPATIAL ERROR

$$1\_LIAEAGRI13 = -0.0000021 \ 1\_IPDR13.07 + 0.007709 \ 1\_VABAGRI10 + 1\_U$$

$$2\_LIAEAGRI13 = 0.0079599 \ 2\_VABAGRI10 + 2\_U$$

$$3\_LIAEAGRI13 = 0.0046078 \ 3\_VABAGRI10 + 3\_U$$

$$4\_LIAEAGRI13 = 0.0033927 \ 4\_VABAGRI10 + 4\_U$$

$$5\_LIAEAGRI13 = 12.4056966 + 0.0000146 \ 5\_IPDR13.07 + 5\_U$$

$$2\_U = 0.5427764 \ 2\_W\_U + 4\_ε$$

$$4\_U = 0.6965370 \ 4\_W\_U + 4\_ε$$

$$5\_U = -0.8967208 \ 5\_W\_U + 5\_ε$$

Ver cuadro 4.20 del anexo 2 para resultados del test de Chow.

**Con IIPIL13.07**

El mismo tipo de modelo es el más adecuado para las inversiones del primer pilar de la PAC. La diferencia es que en este caso, salvo en el oriente, si son significativos y con signo positivo los coeficientes de IIPIL13.07, lo que indica resultados en línea con lo esperado en la mayor parte del territorio asturiano, al menos en lo referente a fomentar las nuevas altas en actividades agrarias en el último año del periodo de programación.

MAXIMUM LIKELIHOOD SPATIAL ERROR.

$$\begin{aligned}
1\_LIAEAGRI13 &= 0.0659928 \quad 1\_VPAR14.07 + 1\_U \\
2\_LIAEAGRI13 &= 0.0000013 \quad 2\_I1PIL13.07 + 0.0543021 \quad 2\_VPAR14.07 + 2\_U \\
3\_LIAEAGRI13 &= 0.0000025 \quad 3\_I1PIL13.07 + 3\_U \\
4\_LIAEAGRI13 &= 0.0000031 \quad 4\_I1PIL13.07 + 4\_U \\
5\_LIAEAGRI13 &= 0.0000027 \quad 5\_I1PIL13.07 + 0.0217554 \quad 5\_VPAR14.07 + 5\_U \\
1\_U &= 0.5053028 \quad 1\_W\_U + 1\_ \varepsilon \\
2\_U &= 0.4869712 \quad 2\_W\_U + 2\_ \varepsilon \\
4\_U &= 0.6807980 \quad 4\_W\_U + 4\_ \varepsilon \\
5\_U &= -0.9446677 \quad 5\_W\_U + 5\_ \varepsilon
\end{aligned}$$

Ver cuadro 4.21 del anexo 2 para resultados del test de Chow.

**Con IPAC13.07**

Tres son las zonas en los que IPAC13.07 resulta significativa y con signo positivo. Serían ala occidental, occidente y centro. El modelo es no espacial y estimado por MCO, con una matriz queen. Buenos resultados para el test de Chow (ver cuadro 4.22 del anexo 2), y mejoran mucho los resultados de heteroscedasticidad con respecto al modelo sin inestabilidad estructural inicial.

ORDINARY LEAST SQUARES ESTIMATION.

$$\begin{aligned}
1\_LIAEAGRI13 &= 0.0820761 \quad 1\_VPAR14.07 + 1\_U \\
2\_LIAEAGRI13 &= 0.0677411 \quad 2\_VPAR14.07 + 2\_U \\
3\_LIAEAGRI13 &= 0.0000016 \quad 3\_IPAC13.07 + 3\_U \\
4\_LIAEAGRI13 &= 0.0000017 \quad 4\_IPAC13.07 + 4\_U \\
5\_LIAEAGRI13 &= 0.0000024 \quad 5\_IPAC13.07 + 5\_U
\end{aligned}$$

### 4.2.3. RESULTADOS DEL ANÁLISIS CONFIRMATORIO.

Para la contribución por parte de la inversión o gasto público con fondos agrarios a la variación de la población en el periodo 2007-2013, medido en los modelos que consideran dependiente la variable VPOBPAD14.07, los resultados son negativos en cuanto a la significatividad y el signo de los coeficientes de dicha variable. En los modelos que no consideran inestabilidad estructural paramétrica y que corrigen solo la autocorrelación espacial, las tres variables de inversión son significativas y tienen signo negativo. Si eliminamos los concejos con valores extremos en la mayoría de las variables consideradas, que coinciden con las zonas con núcleos más urbanos, más industrializadas y menos “rurales” (son los seis concejos con más de 40000 habitantes, todos en el área central) tanto la significatividad como el signo negativo indican los mismos resultados. El dinero público destinado a políticas tanto rurales, con enfoque territorial y multisectorial, como las exclusivamente agrarias con enfoque sectorial, así como también la agregación total y conjunta de ambas, no ha tenido relación positiva con la variación de la población, es decir no se detecta una significatividad que indique que una mayor inversión se ha traducido en variaciones positivas o menos negativas de la población, sino todo lo contrario, a mayor inversión mayor pérdida. No se ha logrado por tanto ni frenar ni corregir el éxodo poblacional, ni para el conjunto de Asturias ni para la muestra que prescinde de las zonas más urbanas.

Con la muestra completa se detecta dependencia espacial, por lo que se requiere en los tres casos utilizar modelos de error espacial, autorregresivos, para tratar la autocorrelación espacial global, que así se elimina. Al utilizar la muestra sin los seis concejos considerados outliers, los modelos para IPDR13.07 e IPAC13.07, no presentan autocorrelación espacial y se presentan modelos no espaciales. Para el caso de IIPIL13.07 si que se debe utilizar como más conveniente un modelo autorregresivo de retardo espacial. Entre ambas muestras, además de la dependencia espacial, se presentan cambios en las variables explicativas que tienen coeficientes significativos, así como en su signo. Para la muestra completa en los tres modelos son VEMP13.07 y RNET10, ambas con signo negativo. Esto probablemente se deba que en las zonas más dinámicas en cuanto a habitantes, se haya atraído población aun con variaciones negativas de empleo y rentas bajas. En cambio al eliminar los outliers poblacionales de la muestra la variación del empleo pasa a ser positiva. Además son significativas, también con signo

positivo, el número de contratos y el VAB en 2010, lo que ya es más acorde a lo esperado, aumentos en el empleo se concretan en aumentos en población.

Los seis modelos presentan heteroscedasticidad, y ante los indicios de heterogeneidad espacial detectados en el AEDE se plantean modelos que puedan corregir los posibles efectos negativos que conlleva. Se plantean seis modelos de regresiones cambiantes en los que se corrige la heteroscedasticidad, a la vez que se obtienen resultados positivos para los correspondientes test de Chow de inestabilidad paramétrica, lo que avala la conveniencia de modelizar con este método para las cinco zonas planteadas, que serían “Oriente”, “Centro Montaña”, “Occidente”, “Ala Occidental” y “Centro”. De los resultados podemos obtener la significatividad y signo de las variables por zonas. Con respecto a las de inversión pública, todas son significativas en alguna de las zonas o regímenes, pero nunca en todas, tanto en la muestra completa como en la que excluye concejos, y en todos los casos son coeficientes con signo negativo, lo que refrenda los resultados de ineficiencia obtenidos anteriormente. Es importante el comportamiento de la variable VEMP13.07, ya que en las zonas en las que es significativa tiene signo positivo salvo en el régimen 5, es decir la zona central, más urbana, en la que el signo es negativo. La variación del empleo no contribuye a la variación de la población ya que caídas importantes del empleo no han ido acompañadas de caídas similares en población, sino todo lo contrario ya que incluso hay concejos en los que la población ha aumentado. La zona centro, fundamentalmente la que no coincide con las cuencas mineras, sumidas recientemente en una crisis industrial y poblacional, son captadores de población con respecto al resto.

Cuando la variable dependiente ha sido VEMP13.07, modelo 2, ninguna de las tres variables de inversión es significativa, lo que viene a señalar poca o nula eficacia en la aportación de las inversiones y gastos de la PAC para influir en la variación del empleo. El modelo considerado es un modelo de error espacial que corrige la dependencia espacial detectada. Se consideran explicativas las variables de valor añadido bruto en 2010 y la superficie de techo industrial en 2012, con signos negativo y positivo respectivamente. En los concejos con más valor añadido la pérdida de empleos ha sido mayor y los concejos con mayores dotaciones industriales han sido los que mejores datos de variación en el empleo han tenido.

Para la estimación por el método de regímenes espaciales, se constata la conveniencia del método de estimación y se corrige mucho la heteroscedasticidad. No son significativas ni IIPIL13.07, ni IPAC13.07. Sí lo es IPDR13.07, pero solo en un régimen, el correspondiente a la zona 4, correspondiente con el ala occidental, y además con signo negativo, lo que en conjunto supone una pobre aportación en discordancia con los objetivos planteados.

Para la muestra sin outliers se plantean dudas sobre la elección de un modelo de dos variables explicativas, que no considera significativas ninguna de las tres variables de inversión o de otros tres con cinco variables cada uno, que si las considera significativas y positivas. Se acepta provisionalmente el primer modelo que no presenta heteroscedasticidad, ante la posibilidad que los otros resultados no sean válidos por su efecto. Posteriormente se comprueba al aplicar regímenes espaciales la conveniencia del primer modelo de dos variables, tal como se preveía, ya que se anula la heteroscedasticidad y cambia las propiedades y el sentido de los resultados de los modelos de cinco variables, que se rechazan. No son significativas ninguna de las tres variables de inversión pública y no se detecta dependencia espacial. Se repiten los malos resultados para la aportación positiva a la variación del empleo ya que de todos los modelos contrastados ninguno presenta coeficientes en las variables de inversión positivos y significativos.

En el caso de considerar endógena VEMPAGRI13.07, para comprobar la aportación de la inversión o gasto público en política agraria en Asturias al empleo sectorial, ya que, al ser fondos agrarios, quizás se hayan obtenido resultados positivos solo en el empleo del sector agrario, se plantean nuevos modelos para las dos muestras y las tres variables de inversión con fondos PAC. En ambas muestras, con y sin outliers IIPIL13.07 no es significativo mientras que IPDR13.07 e IPAC13.07 son significativas y negativas, con lo que parece que tampoco se consigue dinamizar el empleo agrario en Asturias. El valor añadido del sector agrario es significativo y negativo en los seis modelos por lo que valores altos en esta variable supusieron mayores pérdidas de empleo agrario. Con la muestra completa se constataba la necesidad de modelizar la autocorrelación espacial a través de un modelo autorregresivo de retardo espacial y al eliminar los outliers esta desaparece y se modeliza por mínimos cuadrados bietápicos. Los valores extremos de

los concejos más urbanos causaban dependencia espacial aunque esto no afectaba al sentido de los resultados de los fondos de la PAC.

Al aplicar regresiones cambiantes para los tres modelos con la muestra completa se plantean modelos espaciales mixtos autorregresivos por el método de los momentos generalizados propuesto por Kelejian y Prucha (2010) robusto a heteroscedasticidad. Los resultados para IPDR13.07 no son buenos ya que solo es significativo en el ala oriental pero con signo negativo. Tampoco son buenos para las otras dos variables IIPIL13.07 e IPAC13.07, ya que en las zonas 1 a 4 no son significativas o lo son con signo negativo. Cambia en la zona centro donde si que son ambas significativas y con signo positivo. Es en la zona más dinámica e industrial y en su área de influencia donde parece haber tenido una pequeña aportación positiva. Al comprobar lo que ocurre con la muestra sin los valores atípicos, que puedan desvirtuar los resultados, y centrándonos más en la zona rural, vemos que IPDR13.07 sigue con malos resultados siendo significativa solo en dos grupos y con signo negativo. Son malos para IPAC13.07 que no es significativa, por lo que sin contar con la zona centro al completo, esta variable no es relevante. Para IIPIL13.07 se produce un cambio importante, ya que pasa a ser significativa solo en el ala occidental y con signo positivo. Ya no lo es en la zona centro. Es necesario considerar que al eliminar estos concejos también se elimina junto con los urbanos, todos los datos de sus zonas rurales, que son muchas, y periurbanas, que tienen un alto grado de actividad del sector primario, por lo que este cambio se puede deber tanto al efecto negativo de los valores atípicos, como a considerar la zona cinco solo parcialmente. Independientemente de esta circunstancia es evidente que se manifiesta nuevamente la distinta dinámica entre el centro y el resto de zonas, que se detecta en el cambio de signo de variables explicativas (como VABAGRI10), la necesidad de las regresiones cambiantes, que corrigen la heteroscedasticidad, y la influencia económica del área central, incluso en variables sectoriales propias del sector primario. También parece clara la poca relevancia que muestran los modelos sobre la aportación beneficiosa de los fondos públicos a la variación del empleo agrario, ya que en la mayoría de casos no es significativa o lo es de forma negativa, salvo en dos zonas puntuales en las que el primer pilar puede haber tenido mejor resultado, que se anula al considerar la muestra restringida o no, es decir sin o con outliers.

Cuando las nuevas licencias de actividad en 2013, LIAEAGRI13, ha sido la variable explicada, considerando explicativas, entre otras, la variables de inversión pública con fondos agrarios, los resultados cambian con respecto a los tres grupos de modelos anteriores, ya que ahora sí que son significativas y positivas IPDR13.07, I1PIL13.07 e IPAC13.07 cuando planteamos modelos con una única estructura para toda Asturias. Solo IPDR13.07 no es significativa cuando utilizamos la muestra restringida sin los concejos con los seis núcleos urbanos más poblados. Todos son modelos no espaciales salvo un modelo mixto que es el que considera no significativa IPDR13.07. Es un resultado que refleja la mayor eficacia de los fondos sectoriales en las licencias de empresas agrarias en 2013. Es relevante la significatividad positiva de la variación en número de parados ya que puede ser debido a que un aumento de paro hace que ciertos colectivos de parados vean como una salida laboral la actividad agraria, lo que significaría favorecer la “vuelta al campo”.

De todas formas este resultado favorable tiene matices, ya que al estimar considerando inestabilidad paramétrica discreta en las cinco zonas anteriormente descritas vemos que los resultados no son los mismos en todas las zonas. Para IPDR13.07 solo es positivo y significativo con la muestra completa en la zona central, y con la muestra restringida lo es también en la zona central, aunque es negativo y significativo en el oriente. Es llamativo que en la zona más urbana y en su área de influencia es donde mejores resultados obtiene, siendo bastante negativos en el resto de zonas. Para I1PIL13.07 sí que se obtienen valores positivos y significativos, para ambas muestras, en las zonas 2, 3, 4 y 5, y para IPAC13.07, es decir la inversión total, en las zonas 3, 4, y 5. Es una aportación positiva a la creación de empresas agrarias, que quizás no se haya visto reflejada en los datos de empleo agrario por la más que probable baja capacidad de creación de empleo agrario de las explotaciones asturianas y su baja productividad, debido, entre otras posibles causas, a razones orográficas y estructurales.

---

## CAPÍTULO 5. CONCLUSIONES Y DISCUSIÓN.

---

### 5.1. DISCUSIÓN.

Sobre los resultados del análisis confirmatorio se pueden plantear distintos escenarios e hipótesis que se presentan a discusión, ya que se ven afectados por distintos factores, propios de una cadena con múltiples agentes participantes, tanto en el diseño como en la ejecución de los planes de desarrollo rural y de la política agraria comunitaria en su conjunto.

La nula o negativa aportación detectada con estos modelos, de los fondos de los dos pilares de la PAC a las variables de población y empleo, claves en los objetivos perseguidos por el desarrollo rural, así como la pobre aportación del segundo pilar a las empresas agrarias, fundamentalmente en zonas urbanas o bajo su influencia, unido a la aportación positiva del primer pilar al número de empresas agrarias en casi todo el conjunto regional, debería servir de apoyo para replantear tanto el diseño como la ejecución de las políticas de desarrollo del mundo rural asturiano. Se plantean a discusión una serie de reflexiones a modo de conclusión y como posibles futuras líneas de investigación, una vez se han sentado las bases fundamentales con este estudio:

1. La aportación positiva y relevante, según los modelos planteados, solo se ha producido por parte de los fondos del primer pilar de la PAC, que es el destinado a ayudas directas, sobre el número de explotaciones en 2013 (la única aportación al empleo agrario en la zona 5 desaparece y pasa a la zona 4 al considerar la muestra sin outliers, por lo que además de marginal no es concluyente). Este resultado con respecto a esta variable dependiente es lógico, ya que son pagos directos a inversiones y gastos de las explotaciones agrícolas y ganaderas. Parece evidente que las aportaciones económicas directas deben mantener una relación positiva con las explotaciones y ser mayores donde más explotaciones se concentren. También parece evidente que aportaciones directas para sufragar sus gastos e inversiones contribuyen positivamente a su mantenimiento y desarrollo. Pero si esas aportaciones fuesen eficientes, ya que se realizan en un marco que persigue actuar sobre variables socioeconómicas como la población y el empleo, tendrían que traducirse en mejoras en dichas variables. Esto no

es así, ya que los resultados obtenidos o no lo consideran relevante o incluso, en múltiples casos, se consigue el efecto contrario, salvo en la excepción del ala occidental en que da resultados positivos en la variación de empleo agrario para la muestra sin outliers y en el centro para la muestra con outliers (resultado contradictorio y por tanto no concluyente).

2. El resto de inversiones, concretamente las que son objeto directo de estudio, que son las del PDR asturiano, es decir las del segundo pilar, no han tenido más que una aportación marginal para el número de explotaciones, dando resultados positivos solo en el área central más urbanizada, no dando tampoco buenos resultados en otras zonas. Además para población y empleo los resultados han sido también o nulos o negativos.

3. Los resultados en conjunto no son buenos, según estos modelos, ya que parecen los propios de un plan de “subsidio” rural, que mantiene explotaciones y actividades incapaces de frenar la caída poblacional y crear empleo. Sin poner en tela de juicio su conveniencia, ya que no sabemos los resultados de un escenario alternativo sin las mismas, no parecen propias de un plan de desarrollo como su nombre indica. Quizá su utilidad, diseño y ejecución sean los propios de un subsidio a zonas y actividades marginales para que perduren en el tiempo, se alivien y tengan una fácil transición hasta su desaparición, como pueda pasar en otros sectores industriales con dificultades de reconversión. Creemos que de ser así, y sin entrar a juzgar su conveniencia ni los beneficios sociales que pueda tener, que pueden ser muchos, se debería llamar por su nombre y adecuar los objetivos a sus capacidades y naturaleza. No sería entonces, a nuestro juicio, un fracaso socioeconómico, sino una acción solidaria. Pensamos, de todas formas, que se debería intentar reorientar el diseño y ejecución de verdaderas políticas de desarrollo que aprovechen potencialidades que actualmente no son aprovechadas, ya que nos consta que existen agentes sociales dispuestos a aprovecharlas. También opinamos que, si estos planes se diseñan como un subsidio por las administraciones, que lo condicionan desde el principio como tal, no puede tener sino resultados negativos en muchos casos, como se constata empíricamente con los modelos presentados.

4. La predominancia del primer pilar sobre el segundo, que pone más énfasis en los aspectos socioeconómicos en su conjunto y en la búsqueda de alternativas, quizás esté condicionando y produciendo estos resultados, ya que aunque se lleva mucho tiempo

intentando enfocar toda la PAC hacia el desarrollo territorial, esto sigue sin llevarse a cabo.

5. El control de la política agraria, al intentar cohesionar y homogeneizar económicamente los territorios, se diseña con una base y unas medidas comunes que muchas veces no tiene en cuenta las peculiaridades y diversidad de cada uno de ellos. Los regímenes espaciales han demostrado que una región de nivel NUTS II y NUTS III como Asturias tiene distintas realidades económicas y distintas estructuras en sus dinámicas. Aunque Asturias desarrolle su propio PDR este se debe ceñir al Reglamento FEADER. Además el PDR asturiano no contempla las distintas zonas con sus peculiaridades, o al menos no con criterios económicos, sino más bien administrativos, ya que la única división que hace es a efectos de grupos de acción local, que son 11, y más bien se ajustan a mancomunidades de concejos creadas con criterios políticos y administrativos. Además los fondos, que se reparten con criterios territoriales a estos grupos, son muy escasos dentro del propio PDR (en torno al 10%). Parece apropiado diseñar distintas acciones y medidas para cada zona, es decir utilizar un enfoque territorial, pero determinando estas zonas en función de criterios económicos y sociales, y extender esto a todo el PDR, eso sí, manteniendo las estructuras administrativas para ejecutar y poner en práctica dichas acciones, como instrumento de cercanía al ciudadano.

6. Dentro de la estructura en cinco zonas para el territorio asturiano, que resulta conveniente, tal como se deduce de los resultados obtenidos, también se constata una distinta dinámica en la zona central de Asturias con respecto al resto, que puede indicar la existencia de un modelo centro-periferia. Esta circunstancia también puede ser interpretada como la dicotomía rural-urbano, que en Asturias es especialmente interesante por la importancia de la cultura rural y actividad agrícola y ganadera en los municipios con grandes núcleos urbanos. Esto puede dar pie a futuras líneas de investigación.

7. Las políticas de desarrollo rural, así como el conjunto de ayudas de la PAC en Asturias, deben ser rediseñadas y ejecutadas con importantes cambios, para que consigan acercarse a la consecución de los objetivos previstos en términos de población y empleo. Pasarían a ser así verdaderos planes de desarrollo, pero es necesario que en su diseño participen agentes comprometidos con el medio rural que lo vean como la

fuelle de recursos que es, la base económica y cultural sobre la que se asientan otros sectores y sobre la que se asienta también la idea misma de la UE, y no como una actividad marginal, desarrollada en zonas también marginales, que deba ser compensada o mantenida económicamente para evitar problemas sociales.

Planteamos a continuación una serie de cuestiones, a modo de reflexión, que nos surgen tras el análisis de los resultados obtenidos. No todas son cuestiones claras ni afirmaciones que se deduzcan de forma directa, en su totalidad, de los resultados del estudio, pero sí que tienen siempre una relación evidente con los mismos, ya que muchas son cuestiones o críticas que se pueden hacer y se hacen a la PAC en general y a su aplicación en Asturias, en concreto, sobre la falta de creación de empleo y despoblamiento de la zona rural. Se presentan para su discusión y se comenta el posible enlace o relación con nuestras conclusiones. También analizaremos algunas de las cuestiones habituales en foros de discusión sobre el mundo rural y especialmente sobre las políticas de desarrollo rural y la ejecución de fondos agrarios en Asturias, que pueden ser objeto de debate al contrastarse con los resultados de este estudio.

➤ **Enfoque territorial vs enfoque sectorial. ¿Son compatibles? ¿En qué proporción o peso de cada uno en la totalidad de la PAC?**

La clara inestabilidad estructural paramétrica de tipo discreto detectada, muestra una clara división en la estructura económica en cuanto a las variables objeto de estudio entre las cinco zonas propuestas, así como entre la zona centro, más urbana, y el resto. Parece conveniente por tanto que sea tenido en cuenta el enfoque territorial, que aunque parece que será el que definitivamente acabe adoptando la PAC, sigue siendo el menos dotado en cuanto a inversiones. Sería conveniente adaptar las políticas agrarias a los territorios y sus necesidades y que para ello se sigan criterios económicos frente a los criterios administrativos o políticos. No se obtienen resultados mejores por parte de las inversiones de un pilar con respecto al otro, ni tampoco considerando su aportación conjuntamente, por lo que en principio no se puede afirmar ni desmentir, en base a los resultados obtenidos, la existencia de indicios de incompatibilidad de ambos pilares, ni de la conveniencia clara de uno frente al otro en términos de rendimiento en cuanto a las variables consideradas, salvo la conveniencia de la ya contrastada disparidad territorial.

➤ **Conveniencia de la demarcación territorial existente.**

La actual demarcación territorial o bien no considera diferencias entre territorios, aplicando la misma política para toda Asturias, o bien considera en otros casos fronteras administrativas existentes, sin tener, por tanto, en cuenta criterios de la evolución de la actividad económica y de la realidad social de los territorios. La significatividad de la existencia de inestabilidad estructural pone de manifiesto la posibilidad de replantearse la actual demarcación territorial. La primera y fundamental distinción sería la de la zona central frente al resto, y la más específica y conveniente, en nuestra opinión y basándonos en los resultados de este estudio, es la que considera las cinco comarcas propuestas, con criterios económicos comunes en cada una, frente al resto.

➤ **Adecuación de objetivos. ¿Son alcanzables los actuales?**

El apartado 1 del artículo 4 del REGLAMENTO (CE) no 1698/2005 DEL CONSEJO de 20 de septiembre de 2005 relativo a la ayuda al desarrollo rural a través del Fondo Europeo Agrícola de Desarrollo Rural (FEADER), textualmente dice:

*1. La ayuda al desarrollo rural contribuirá a alcanzar los siguientes objetivos:*

- a) aumentar la competitividad de la agricultura y la silvicultura mediante la ayuda a la reestructuración, el desarrollo y la innovación;*
- b) mejorar el medio ambiente y el medio rural mediante ayudas a la gestión de las tierras;*
- c) mejorar la calidad de vida en las zonas rurales y fomentar la diversificación de la actividad económica.*

Si analizamos detenidamente la descripción de objetivos del FEADER observamos que son bastante generalistas ya que se basan en competitividad, medio ambiente, calidad de vida y diversificación. El planteamiento es muy genérico a nuestro juicio, lo que permite considerar el vaso medio lleno o medio vacío. Necesitamos seguir analizando tanto los indicadores como los documentos específicos para cada operación y documentos estratégicos concretos para encontrar afirmaciones como la que se extrae del PDR 2007-2013 de Asturias:

*“La aplicación del PDR FEADER en el Principado de Asturias para el próximo período de programación 2007-2013 desplegará indudables efectos positivos de impacto sobre la renta y el empleo de la región”*. Podemos darnos cuenta que tanto el empleo como la población son indicadores fundamentales para medir el rendimiento de los programas, así como su aportación a la consecución de los objetivos generales del reglamento FEADER, al menos desde el enfoque regional.

Si analizamos los resultados de este estudio, que plantea la medición de la evolución de variables clave, tanto en el diagnóstico económico territorial, como en la aportación fundamental de las mismas a los objetivos de competitividad del sector primario, mejora de la calidad de vida y diversificación, no parece que la pobre aportación de las inversiones públicas a la población, y sobre todo al nivel empleo, contribuya a la consecución de un sector primario competitivo, diversificado ni con un alto nivel de vida. Todo esto sin perjuicio, ya que no es lo que se puede concluir de los resultados obtenidos, de que se pueda estar contribuyendo a mantener o corregir ciertas carencias en estos ámbitos.

- **¿Existe una línea clara en la consecución de objetivos de población y empleo o la supuesta acción multidisciplinar y participativa ejecuta medidas contradictorias y plantea distintas acciones que se anulan o contrarrestan?**

La ambigüedad y generalidad de los objetivos previstos propone y desarrolla estrategias y planes de acción con muchos frentes, lo que a veces los hace más versátiles y enriquecedores, pero también más complejos y contradictorios. La consideración de la población y el empleo, junto con las empresas agrarias, como algunas de las variables sobre las que actuar para obtener resultados positivos puede suscitar críticas por la posibilidad de no conseguir alcanzar los efectos deseados sobre las variables fundamentales. Los resultados obtenidos en este estudio sobre cuatro variables fundamentales no son compatibles con la existencia de una línea clara de actuación que consiga resultados visibles en estos ámbitos. Creemos conveniente concretar, reducir y adecuar más los objetivos a las verdaderas posibilidades y alcance de estos fondos públicos.

➤ **¿Actúan las administraciones de forma coordinada?**

La coordinación entre las distintas administraciones de estas políticas a nivel europeo, nacional, regional e incluso comarcal, como en el caso del enfoque LEADER, requiere una buena coordinación entre las mismas. Sin entrar a valorar aspectos de carácter administrativo o burocrático sería conveniente que el grado de condicionamiento existente desde que se aprueban los reglamentos de los fondos europeos, hasta que llega al beneficiario final, no fuese un obstáculo que impidiese la actuación en una sola dirección con la suficiente flexibilidad para considerar las peculiaridades y necesidades de los beneficiarios finales maximizando así su rendimiento. Los resultados negativos en la significatividad entre gasto público y población y empleo unido a la falta de consideración administrativa de zonas con distinta estructura económica no parecen respaldar la correcta coordinación de los distintos niveles de la administración pública.

➤ **¿Sirven para dinamizar o sirven para fomentar y mantener la marginalidad de actividades, colectivos y territorios?**

Los malos resultados en población y empleo, así como los pobres resultados en número de empresas de IPDR13.07 contrastando en este caso con la mejora de la dinámica de IIPIL13.07, puede indicar un posible escenario de mantenimiento de actividades y territorios en crisis permanente, que apuntalan sobre todo la actividad, pero no la ponen en los niveles deseados ni se obtienen de la misma los efectos esperados en cuanto a población y empleo. Los resultados son los propios de una política de compensación, que sin entrar a valorar su conveniencia o utilidad, siguen sin cambiar de momento las dinámicas negativas. La alta financiación del primer pilar en detrimento del segundo impide valorar mejor escenarios alternativos, aunque la significatividad negativa en muchos de los modelos y territorios no favorece la hipótesis de que el segundo pilar no de buenos resultados solamente por falta de financiación.

➤ **¿Son auténticas políticas económicas de desarrollo o son políticas a secas que se determinan bajo acuerdos y criterios políticos que se acuerdan en el marco de negociaciones con estados, regiones, colectivos y grupos de interés o presión a los que se destinan inversiones públicas con fines contradictorios por la heterogeneidad y distintos intereses de los mismos?**

Aunque este sea un aspecto que no se puede deducir de los resultados obtenidos, ya que el planteamiento es más generalista sin entrar a valorar estos detalles, es posible que el efecto directo solo sobre las explotaciones por parte de las ayudas directas, sin tener repercusión en el empleo, pueda estar relacionado con la falta de expansión de los efectos positivos esperados en este ámbito. Si las posibles mejoras sobre territorios o agentes económicos deben ser compensadas por otras contradictorias que se destinan a otros territorios o agentes, se puede estar frenando su posible efecto multiplicador. Además la adopción de criterios territoriales en el destino de los fondos con fronteras administrativas, y no con criterios económicos, puede tener su origen en que se utilizan como instrumentos de negociación política, y terminar derivando en el desajuste territorial, al considerar territorios como homogéneos sin serlo realmente.

- **¿Entorpece la capacidad de maniobra de los agentes económicos la excesiva burocratización y control administrativo que encasilla actividades y limita su expansión y crecimiento al ajustar la capacidad a la inversión concedida y no al revés?**

El destino de los fondos a inversiones en empresas y rentas, de forma poco flexible, que en muchos casos obliga a los beneficiarios de subvenciones a adaptarse a las exigencias de las mismas (sobre todo en el caso de mercados muy controlados como los del sector primario), puede ser una de las causas de la poca capacidad de expansión en aquello que no sea nuevas altas por parte de las ayudas directas, ya que con este estudio no se aprecia un efecto multiplicador con respecto a nuestras variables dependientes.

- **¿Existe una relación equilibrada entre las zonas rurales y urbanas en términos económicos?**

A la vista de la diferencias que muestran los resultados de este estudio entre la zona central y el resto, no parece que ambos modelos territoriales se complementen y convivan integrados, al menos en el caso que nos ocupa, tal como se podía esperar, ya que las zonas urbanas siempre han sido núcleos de actividad que proveen de servicios y productos a zonas rurales, que con otro modelo de gestión del espacio proveen a su vez de múltiples productos y servicios (además de alimentos) a las zonas urbanas. Esta relación se supone integrada ya que sería conveniente que formasen una unidad

socioeconómica que contenga ambos modelos y se enriquezca de su relación, pero con ciudadanos y agentes sociales con las mismas oportunidades y con actividades homogéneas y equiparables, aunque adaptadas a cada espacio. No creemos que sea adecuado ni sostenible una relación campo ciudad, en la que la primera parte sea considerada marginal y deba ser subsidiada o mantenida por la otra.

Como habíamos comentado al inicio de este apartado las reflexiones planteadas son propuestas de debate y sirven de guía para futuros proyectos de investigación, ya que la complejidad del tema así como la diversidad de factores y agentes implicados requieren estudios más concretos, una vez se han sentado las bases de partida con el que nos ocupa.

## 5.2. CONCLUSIONES.

Tras una primera introducción y repaso tanto a la PAC como a la historia de la econometría espacial y su contexto, la ciencia regional, hemos detallado las principales técnicas econométricas para la estimación y contraste de modelos y variables con datos de naturaleza espacial. Todo esto, acompañado de un análisis de las principales herramientas informáticas disponibles, ha servido para concluir la necesidad, en nuestro caso, de la utilización de un análisis espacial para el estudio que nos ocupa, que pasamos a dividir en dos grandes bloques, el análisis exploratorio y el análisis confirmatorio.

Previo a la ejecución de los mismos, hemos realizado y justificado la elección de las variables, que son de ámbito microterritorial, valorando y describiendo los posibles problemas que podían presentar, como el de unidad de área modificable y el de inferencia ecológica. Esto unido a la demarcación territorial con fines estadísticos de la UE, así como a sus recomendaciones de indicadores comunes de contexto a efectos de analizar territorios rurales, ha servido para escoger las variables a utilizar siempre condicionados por la disponibilidad estadística en niveles microterritoriales. El repaso de literatura reciente y los aspectos anteriormente comentados han servido para justificar y ayudar a decantarnos por una metodología acorde a las necesidades del

estudio y a los condicionamientos de las políticas puestas en marcha, que junto con la teoría económica, han determinado el sentido y estructura de los modelos. Dichas políticas no se sustentan de forma directa en modelos propios de la teoría macroeconómica, lo que nos habría dado un modelo previo, sino en la intervención de la administración pública con la aportación de fondos destinados a la iniciativa privada para mejorar datos socioeconómicos, en principio y como más importantes, la población y el empleo. Por esto hemos decidido la utilización de las técnicas econométricas que puedan aprovechar los recursos de que disponíamos y que se ajustasen al sentido de esos planes de política económica. Se concretaron en una colección de modelos, divididos en cuatro grandes grupos, coincidentes con las cuatro variables explicativas de interés, dos generales y dos específicas del sector primario, que nos han permitido valorar la aportación de las inversiones del primer y segundo pilar de la PAC, en Asturias, en el periodo de programación 2007-2013. El resultado final han sido 44 modelos, que son variantes de los 4 principales, y que han permitido contrastar el comportamiento que las variables implicadas en las políticas de dinamización implementadas, que debemos tener en cuenta que ya nos vienen condicionadas por la UE. De los cuatro bloques principales, que coinciden con las cuatro variables explicativas, tres son las variaciones que han tenido lugar para la población, para el nivel de empleo total y para el nivel de empleo del sector agrario, medidas las tres en términos absolutos y definidas como la diferencia entre el nivel correspondiente al año de final de periodo y el correspondiente al del comienzo, intentando así captar la dinámica del proceso, que es plurianual y que se ejecuta de forma continua y/o periódica durante esos años. Se persigue evitar así la posible relación causal entre población y empleo en la determinación de la inversión pública, en contra de lo que se estipula por la teoría. Las variaciones que no se han producido no pueden servir de criterio para el destino de fondos públicos, aunque a veces si lo puedan ser datos fijos del pasado, pero de cuantía mínima y dentro de elecciones multicriterio, que nunca incluyen variaciones. De todas formas eliminamos así esa posibilidad. Una cuarta variable sería el número de altas en la actividad agraria en último año del periodo de programación, que aunque no recoge su variación sí que ofrece un valor sobre la actividad empresarial a fin de periodo. Para las tres primeras variables los resultados han sido contrarios a lo esperado, resultando incluso negativo su efecto sobre las variables dependientes. Para la última han sido datos más en línea con lo esperado, pero

solo en zonas de influencia urbana, en el caso de los fondos del segundo pilar PAC, y extendido a más zonas en el caso de los fondos del primer pilar.

El AEDE dio unos resultados que apuntaban a la heterogeneidad espacial, posteriormente corregida con regímenes espaciales, y a la dependencia espacial en algunos casos, tal como apuntaba el análisis univariante (no en todos) lo que se concretó en modelos espaciales autorregresivos tanto de tipo error, lag, como mixtos. En cada caso se han aplicado los métodos correspondientes de estimación. Las técnicas espaciales han sido imprescindibles para modelizar con estos datos espaciales de corte transversal.

Los principales resultados del AEDE, realizado desde el enfoque univariante a todas las variables seleccionadas, ha mostrado una clara dependencia espacial en todas las variables sectoriales propias del sector primario, sin excepción. De las que no son propias del sector primario, solo la renta familiar disponible mostró valores significativos de presencia de autocorrelación espacial, aunque ha resultado ser una variable no significativa y nada relevante para el análisis confirmatorio. El resto de variables no han sido significativas para la existencia de dependencia espacial, con un grado de confianza del 95%, aunque parte de ellas si lo eran al 90%, incluso al realizar los test de permutaciones y repetir las pruebas, en algún caso si se obtuvieron algunos p valor inferiores a 0.05. Otra cuestión importante es el aumento de la dependencia espacial mostrada en el diagrama de dispersión de Moran al eliminar los outliers, ya que en muchos casos se incrementa. La conclusión principal por tanto del análisis exploratorio univariante con respecto a la autocorrelación espacial es que existe para todas las variables del sector primario, incluyendo especialmente a las de inversión pública de la PAC y que en el resto de variables, aunque presentan dudas en muchos casos, no parece afectar demasiado con la muestra completa, pero si se ve afectada en un número importante de casos al eliminar los valores atípicos. Debe ser considerada por tanto la autocorrelación espacial ya que el AEDE indica, en mayor o menor grado, su presencia, y debe ser considerada también la posible distorsión de los valores atípicos, lo que nos hizo presentar dos escenarios, uno con la muestra completa y otro con la muestra sin los concejos más urbanizados e industrializados para poder así comparar los resultados.

Con respecto a la heterogeneidad espacial, también concluimos, al analizar los resultados de los test de dependencia espacial local, la presencia de clusters en casi todas las variables. La zona central ha resultado tener, en todos los casos, varios concejos con valores significativos tanto en los test basados en la I de Moran, como en los basados en la  $G_i$  y  $G_i^*$  de Getis y Ord. Otras zonas también han acompañado a esta en casi todas las variables, lo que sugiere la posible presencia de distintas dinámicas económicas que ha sido tenida en cuenta en el posterior análisis confirmatorio.

El planteamiento para la estimación de los modelos ha sido estudiar la aportación de las inversiones o gasto público cofinanciado o financiado en su totalidad por fondos europeos destinados a la PAC, en Asturias, durante el periodo 2007-2013, planteando una colección de modelos en los que se tuviesen en cuenta los resultados del AEDE previo. Por ello se plantearon como explicativas algunas de las variables, en nuestra opinión, más relevantes sobre las que se deben obtener resultados positivos si estas políticas europeas de desarrollo rural cumplen, al menos en parte, su cometido. Estas serían las referentes a población y empleo. Se utilizan dos variables de conjunto, variación del periodo en población según el padrón municipal y también variación en el periodo del nivel de empleo y otras dos variables sectoriales específicas del sector primario, variación del empleo agrario en el periodo y nivel de empresas agroganaderas en el año 2013. Todas las variables se consideran en términos absolutos, por lo que nos pareció procedente considerar en al menos en tres de los cuatro casos la variación en el periodo de programación, para recoger su efecto sobre el cambio y así evitar posibles inconsistencias en la causalidad teórica de algunas de las variables explicativas, fundamentalmente en las referentes a las aportaciones de fondos públicos. El resultado ha sido que de las variaciones de muestra con y sin outliers, y la consideración de tres variables de fondos públicos, que se corresponden con primer pilar, segundo pilar y la suma total de ambos, han salido 24 casos o modelos a contrastar como variantes de los cuatro grupos o modelos iniciales. Como también se han considerado en dos bloques los modelos que corrigen la dependencia espacial y posteriormente sobre ellos se ha tratado en un segundo paso la heterogeneidad espacial se han considerado un total de 24 escenarios, de los que hemos obtenido conclusiones.

Se ha tratado la dependencia y la heterogeneidad espaciales por los métodos más convenientes en función de cada caso. Cuando se ha detectado la presencia de autocorrelación espacial el método fundamental ha sido el de momentos generalizados,

ya que así se ha evitado el problema habitual de endogeneidad de las variables explicativas. Para la heterogeneidad espacial se han considerado cinco zonas o regímenes espaciales ya que se ha detectado un problema de inestabilidad paramétrica discreta. Se ha contrastado la conveniencia de este método para las zonas propuestas.

Las principales conclusiones con los modelos propuestos serían:

1. Con respecto a las variables no sectoriales (VPOBPAD14.07 y VEMP13.07) la **NO EXISTENCIA** de evidencia que apoye una relación significativa y positiva de los fondos públicos agrarios a la mejora poblacional, ni tampoco a la mejora en el nivel de empleo en términos absolutos, ya que o bien no se considera significativa o incluso en muchos casos es significativa y negativa, lo que sería indicativo de un efecto o relación antagónico entre ambas. Se da para las inversiones de los dos pilares y para la suma total de ambas.

2. Con respecto a las variables sectoriales (VEMPAGRI13.07 y LIAEAGRI13) se distinguen dos comportamientos distintos. Para la variación en el periodo de programación considerado del nivel de empleo agrario en términos absolutos, las conclusiones son casi las mismas que para las variables no sectoriales, ya que también se obtienen resultados negativos o no significativos. La única diferencia se da para la zona del ala occidental donde si es significativa y positiva la aportación del primer pilar en la muestra sin outliers así como para la zona central pero en la muestra con outliers. Es decir tampoco se ha contribuido prácticamente a la mejora del nivel de empleo sectorial. Para el número de licencias de actividad agrarias si se obtienen resultados positivos, aunque con matices, ya que aunque en los modelos iniciales, que no consideran regímenes espaciales, sí que son significativas las tres variables de inversión pública, al considerar modelos de inestabilidad estructural para el caso de IPDR13.07 solo lo sería en la zona centro con la muestra completa y en el oriente con la muestra retringida, para IIPIL13.07 lo sería en 4 zonas y para total recogido por IPAC13.07 en 3 zonas. Es decir solo el primer pilar ha contribuido positivamente en la mayor parte, aunque no en todo el territorio asturiano.

3. La combinación de los resultados con respecto a las cuatro variables dependientes **NO MUESTRA LA CONSECUCCIÓN DE MEJORAS** propias de un plan de

desarrollo territorial que actúe positivamente sobre dos de los principales indicadores socioeconómicos de las zonas rurales, a efectos de conseguir su continuidad e integración en un sistema económico y territorial, en condiciones de peso e importancia dentro del mismo similares a las que tradicionalmente ha tenido.

4. Existe una evidente **DIFERENCIA TERRITORIAL** en el comportamiento económico de las variables y las relaciones entre ellas, en función de las cinco zonas seleccionadas, detectándose además un comportamiento distinto entre la zona central y las otras cuatro.

Las anteriores conclusiones han servido para plantear ciertas cuestiones a debate y discusión, centrándose las mismas en intentar buscar posibles causas y relaciones entre algunas de las críticas habituales a la PAC, que suelen ser objeto de discrepancia por parte de distintas corrientes de opinión o pensamiento, tanto por técnicos como por los propios agentes socioeconómicos de las zonas rurales. Se han planteado como reflexiones apuntando su posible relación con los resultados de este estudio, pero sin considerar que sean concluyentes en un sentido u otro, ya que simplemente abren nuevas vías de futuras investigaciones. Son aspectos a considerar en la redacción de planes futuros ya que existe margen de actuación y cambio para dichos planes que permitan mejorar sus resultados. Un resumen de las más relevantes sería:

- Se necesitan cambios en la implementación de estos programas ya que no se obtienen beneficios contrastables en materia de empleo y población.
- Parecen políticas de subsidio y compensación y no verdaderas políticas dinamizadoras.
- No se consideran criterios económicos para distribuir por zonas el territorio e implementar planes según estas necesidades.
- La excesiva rigidez, la falta de coordinación entre administraciones y la preponderancia del enfoque sectorial sobre el territorial son cuestiones que convendría fuesen revisadas, ya que pueden convertirse en herramientas de cambio de cara a probar nuevas estrategias que mejoren los actuales resultados.
- Las relaciones y la dualidad rural-urbano o campo-ciudad deben afrontarse desde la igualdad e integración de dos modelos complementarios de gestión económica y territorial que forman parte del mismo conjunto.

Nos gustaría proponer la consideración de un modelo económico en red con importantes y distintos núcleos urbanos de concentración de actividad y población, en varios grados de importancia, pero en el que dicha red se asiente sobre un territorio humanizado, gestionado y con actividad económica, que forme parte de un contexto global y que permita un alto grado de movilidad e integración entre campo y ciudad, frente al modelo en red que una puntos urbanos mediante vías de comunicación sin considerar, o al menos con el mismo grado de importancia, la gestión del territorio sobre el que se asiente dicha red.

Después de todo lo expuesto nos planteamos una serie de futuras líneas de investigación que nos permita profundizar en estos temas. Algunas propuestas de nuestro interés serían:

- Análisis dinámico y espacial, con modelos espacio temporales.
- Estudio de escenarios alternativos con planes de desarrollo e inversiones distintas para comparar resultados.
- Estudio de las capacidades, recursos y potencialidades para poder diseñar otros planes y escenarios alternativos. Innovación, medioambiente y cambio climático son objetivos transversales del próximo periodo de programación en los que el medio rural tiene campo de crecimiento y juega un papel fundamental. Las fuentes de energía renovables son claves para el futuro económico de las zonas rurales.
- Estudio de los problemas concretos en la ejecución y diseño de la política agraria comunitaria para poder proponer las correcciones oportunas. Problemas burocráticos y administrativos y falta de coordinación entre agentes implicados, así como poca integración de los aspectos puramente económicos con los culturales y humanísticos, son algunos de los aspectos de interés.
- Estimación y contraste de modelos de economía regional, fundamentalmente de relaciones centro periferia, rural-urbano, etc.
- Estimar modelos que puedan recoger el comportamiento de los fondos públicos para el logro de los objetivos de tipo energético, medioambiental y paisajístico.



*“Cumpliendo con mi oficio  
piedra con piedra, pluma a pluma,  
pasa el invierno y deja  
sitios abandonados  
habitaciones muertas:  
yo trabajo y trabajo,  
debo substituir  
tantos olvidos,  
llenar de pan las tinieblas  
fundar otra vez la esperanza.”*

Pablo Neruda.



---

## BIBLIOGRAFÍA.

---

- “*REGLAMENTO DE EJECUCIÓN (UE) No 808/2014 DE LA COMISIÓN de 17 de julio de 2014 por el que se establecen disposiciones de aplicación del Reglamento (UE) no 1305/2013 del Parlamento Europeo y del Consejo, relativo a la ayuda al desarrollo rural a través del Fondo Europeo Agrícola de Desarrollo Rural (Feader)*”.
  
- “*MANUAL SOBRE EL MARCO COMÚN DE SEGUIMIENTO Y EVALUACIÓN*”. Documento de orientación. Septiembre de 2006. Dirección General de Agricultura y Desarrollo Rural.
  
- ALAÑÓN PARDO, A. (2002). “*ESTIMACIÓN DEL VALOR AÑADIDO PERCÁPITA DE LOS MUNICIPIOS ESPAÑOLES EN 1991 MEDIANTE TÉCNICAS DE ECONOMETRÍA ESPACIAL*”. EKONOMIAZ, nº 51, tercer cuatrimestre 2002.
  
- ALONSO, O. (2001). “*METROPOLITAN AREAS AND PUBLIC INFRASTRUCTURE*”. Investigaciones Económicas, vol. XXV(1); pp. 139-169.
  
- ANSELIN, L. (1980). “*ESTIMATION METHODS FOR SPATIAL AUTOREGRESSIVE STRUCTURES*”. Regional Science Dissertation and Monograph Series (Ithaca, NY).
  
- ANSELIN, L. (1988). “*SPATIAL ECONOMETRICS: METHODS AND MODELS*”. Kluwer Academic Publishers.
  
- ANSELIN, L. (1994). “*TESTING FOR SPATIAL DEPENDENCE IN LINEAR REGRESSION MODELS: A REVIEW*”. Regional Research Institute Research Paper. West Virginia University, Morgantown.
  
- ANSELIN, L. (1998). “*EXPLORATORY SPATIAL DATA ANALYSIS IN A GEOCOMPUTATIONAL ENVIRONMENT*”. Actas de la Conferencia de GeoComputación 1998, Bristol (UK); pp. 17-19.
  
- ANSELIN, L. (2001). “*SPATIAL EFFECTS IN ECONOMETRIC PRACTICE IN ENVIRONMENTAL AND RESOURCE ECONOMICS*”. American Journal of Agricultural Economics, 83 (3); pp. 705-710

- ANSELIN, L. (2001). *"SPATIAL ECONOMETRICS"*. En "A COMPANION TO THEORETICAL ECONOMETRICS", ed. Baltagi, Oxford: Basil Blackwell; pp. 310-330.
- ANSELIN, L. (2005). *"EXPLORING SPATIAL DATA WITH GEODA: A WORKBOOK"*. Spatial Analysis Laboratory. Department of Geography University of Illinois. Center for Spatially Integrated Social Science. Manual. Copyright 2004-2005 Luc Anselin.
- ANSELIN, L. y R. FLORAX (1995). *"NEW DIRECTIONS IN SPATIAL ECONOMETRICS: INTRODUCTION"*. En "NEW DIRECTIONS IN SPATIAL ECONOMETRICS." L. Anselin y J.G.M. Florax editores. Springer; pp. 3-18.
- ANSELIN, L. y H.H. KELEJIAN (1997). *"TESTING FOR SPATIAL ERROR AUTOCORRELATION IN THE PRESENCE OF ENDOGENOUS REGRESSOR"*. Int. Regional Science. Rev. 20 pp.153-182.
- ANSELIN, L. y R. FLORAX (eds.) (2000). *"ADVANCES IN SPATIAL CONOMETRICS"*. Heidelberg: Springer-Verlag.
- ANSELIN, L., R. FLORAX y S. REY (2004). *"ADVANCES IN SPATIAL ECONOMETRICS. METHODOLOGY, TOOLS AND APPLICATIONS."* Springer, Berlin Heidelberg (Germany).
- ANSELIN, L., A. BERA, R. FLORAX y M. YOON (1996). *"SIMPLE DIAGNOSTIC TESTS FOR SPATIAL DEPENDENCE."* Regional Science and Urban Economics, 26; pp. 77-104.
- ANSELIN, L., I. SYABRI y Y. KHO (2010). *"GEODA: AN INTRODUCTION TO SPATIAL DATA ANALYSIS"*. En FISHER, M. y A. GETIS (2010). *"HANDBOOK OF APPLIED SPATIAL ANALYSIS. SOFTWARE TOOLS, METHODS AND APPLICATIONS."* Springer Heidelberg Dordrecht London New York.
- ANSELIN, L. *"GMM ESTIMATION OF SPATIAL ERROR AUTOCORRELATION WITH AND WITHOUT HETEROSKEDASTICITY"*. Nota GeoDa Center, Arizona State University.

- 
- ANSELIN, L., P.V. AMARAL y D. ARRIBAS-BEL (2012). *“TECHNICAL ASPECTS OF IMPLEMENTING GMM ESTIMATION OF THE SPATIAL ERROR MODEL IN PYSAL AND GEODASPACE”*, Working Paper 2/12 .GeoDa Center, Arizona State University.
- ANSELIN, L. y S. REY (2014). *“MODERN SPATIAL ECONOMETRICS IN PRACTICE”*. GeoDa Press LLC. Chicago.
- ARBIA, G. (1989). *“SPATIAL DATA CONFIGURATION IN STATISTICAL ANALYSIS OF REGIONAL ECONOMICS AND RELATED PROBLEMS.”* Dordrecht: Kluwer.
- ARBIA, G. (2006). *“SPATIAL ECONOMETRICS: STATISTICAL FOUNDATIONS AND APPLICATIONS TO REGIONAL CONVERGENCE (ADVANCES IN SPATIAL SCIENCE)”*. Springer-Verlag, Berlin (Germany).
- ASSUNÇÃO, R.M. y E.A. REIS (1999). *“A NEW PROPOSAL TO ADJUST MORAN’S I FOR POPULATION DENSITY.”* STATISTICS IN MEDICINE Statist. Med. 18; pp. 2147–2162.
- BIEHL, D. (1979). *“LOS DETERMINANTES DE LAS DIFERENCIAS REGIONALES Y EL PAPEL DE LA HACIENDA PÚBLICA”*, en MUSGRAVE, R. A. (1984). *“ASPECTOS REGIONALES DE LA POLÍTICA FISCAL”*. Instituto de Estudios Fiscales, Madrid.
- BIEHL, D. (1986). *“THE CONTRIBUTION OF INFRASTRUCTURE TO REGIONAL DEVELOPMENT,”* Final Report of the Infrastructure Study Group, Commission of the European Communities. Luxemburgo.
- BIEHL, D. (1988). *“LAS INFRAESTRUCTURAS Y EL DESARROLLO REGIONAL”*. *Papeles de Economía Española*, 35; pp. 293-310.
- BIEHL, D. (1991). *“THE ROLE OF INFRASTRUCTURE IN REGIONAL DEVELOPMENT”*, en VICKERMAN, R. *“INFRASTRUCTURE AND REGIONAL DEVELOPMENT”*. Pion. Londres.
- BOARNET, M. G. (1998). *“SPILLOVERS AND LOCATIONAL EFFECTS OF PUBLIC INFRASTRUCTURE”*. *Journal of Regional Science*, 38; pp. 381-400.

- BODSON, P. y D. PEETERS (1975). "*ESTIMATION OF THE COEFFICIENTS OF A LINEAR REGRESSION IN THE PRESENCE OF SPATIAL AUTOCORRELATION. AN APPLICATION TO A BELGIAN LABOUR-DEMAND FUNCTION*". Environment and Planning A, 7; pp. 455-472.
  
- BOSCA, J.E., J. ESCRIBÁ, y M.J. MURGUÍ (2011). "*LA EFECTIVIDAD DE LA INVERSIÓN EN INFRAESTRUCTURAS PÚBLICAS: UNA PANORÁMICA PARA LA ECONOMÍA ESPAÑOLA Y SUS REGIONES*". Investigaciones Regionales, 20; pp 195 a 217.
  
- BURRIDGE, P. (1980). "*ON THE CLIFF-ORD TEST FOR SPATIAL AUTOCORRELATION*". Journal of the Royal Statistic Society B, 42; pp. 107-8.
  
- CARAMÉS, L. y S. LAGO (2002). "*LOS EFECTOS DEL GASTO PÚBLICO SOBRE EL CRECIMIENTO ECONÓMICO: EL CASO DE LAS COMUNIDADES AUTÓNOMAS*". Instituto de Estudios Económicos. Fundación Pedro Barrié de la Maza. La Coruña.
  
- CASE, A., H. ROSEN y J. HINES (1993). "*BUDGET SPILLOVERS AND FISCAL POLICY INTERDEPENDENCE: EVIDENCE FROM THE STATES*". Journal of Public Economics, 52; pp. 285-307.
  
- CASSETTI, E. (1972). "*GENERATING MODELS BY THE EXPANSION METHOD: APPLICATIONS TO GEOGRAPHICAL RESEARCH*". Geographical Analysis, 4; pp. 81-91.
  
- CHANDRA, A. y E. THOMPSON. (2000). "*DOES PUBLIC INFRASTRUCTURE AFFECT ECONOMIC ACTIVITY? EVIDENCE FROM THE RURAL INTERSTATE HIGHWAY SYSTEM*". Regional Science and Urban Economics, 30; pp. 457-490.
  
- CHASCO YRIGOYEN, C. (2003), "*ECONOMETRÍA ESPACIAL APLICADA A LA PREDICCIÓN-EXTRAPOLACIÓN DE DATOS MICROTERRITORIALES*". Tesis doctoral. Comunidad de Madrid. Consejería de Economía e Innovación Tecnológica.
  
- CHASCO YRIGOYEN, C. (2010). Curso "*ANÁLISIS ECONÓMICO GEOESPACIAL*". Universidad Autónoma de Madrid. Material didáctico del Curso.

- 
- CHASCO YRIGOYEN, C. (2013). *"GEODASPACE: A RESOURCE FOR TEACHING SPATIAL REGRESSION MODELS"* Departamento de Economía Aplicada. Universidad Autónoma de Madrid
- CERVELLÓ ROYO, R. y B. SEGURA GARCÍA DEL RÍO (2010). *"UN MODELO PARA EVALUAR Y OPTIMIZAR EL IMPACTO ESPACIAL EN LAS INVERSIONES EN REGENERACIÓN URBANA"*. Investigaciones Regionales, 17; pp 125-137.
- CLIFF, A. y J. ORD (1981). *"SPATIAL PROCESSES, MODELS AND APPLICATIONS"*. London: Pion.
- COLINO SUEIRAS, J. y J.M. MARTÍNEZ PAZ (2005). *"EL DESARROLLO RURAL: SEGUNDO PILAR DE LA PAC"*. Departamento de Economía Aplicada. Universidad de Murcia. En GARCÍA DELGADO, J.L. y M.J. GARCÍA GRANDE (2005). *"POLÍTICA AGRARIA COMÚN: BALANCE Y PERSPECTIVAS"*. Nº 34. Colección de Estudios Económicos de la Fundación La Caixa.
- DACEY, M.F. (1969). *"SIMILARITIES IN THE AREAL DISTRIBUTIONS OF HOUSES IN JAPAN AND PUERTO RICO"*. Area, 3; pp. 35-37.
- DE RUS, G. (2001). *"INFRAESTRUCTURAS: ¿QUÉ PODEMOS DECIR LOS ECONOMISTAS, EN LA INVESTIGACIÓN ECONÓMICA EN ESPAÑA: 1990-2000. UNA DÉCADA DE CAMBIOS"*. IVIE.
- DELGADO, M. J. y I. ÁLVAREZ. (2003). *"EFICIENCIA TÉCNICA Y CONVERGENCIA EN LOS SECTORES PRODUCTIVOS REGIONALES"*. Investigaciones Regionales, 3; pp. 115-126.
- DÍAZ ROLDÁN, C. y D. MARTÍNEZ LÓPEZ (2005). *"INVERSIÓN PÚBLICA Y CRECIMIENTO ECONÓMICO. UNA REVISIÓN CRÍTICA CON PROPUESTA DE FUTURO"*. Documento de Trabajo. Serie Economía E2005/10. Fundación Centro de Estudios Andaluces. Consejería de la Presidencia. Junta de Andalucía.

- FISCHER, M. y A. GETIS (1997). "*ADVANCES IN SPATIAL ANALYSIS*". EN "*RECENT DEVELOPMENTS IN SPATIAL ANALYSIS*". Eds. M. Fischer y A. Getis, Springer-Verlag, Berlín; pp. 1-14.
  
- FISHER, M. y A. GETIS (2010). "*HANDBOOK OF APPLIED SPATIAL ANALYSIS. SOFTWARE TOOLS, METHODS AND APPLICATIONS.*" Springer Heidelberg Dordrecht London New York.
  
- FOTHERINGHAM, A., M. CHARLTON y C. BRUNSDON (1997). "*MEASURING SPATIAL VARIATIONS IN RELATIONSHIPS WITH GEOGRAPHICALLY WEIGHTED REGRESSION*". En FISCHER, M. Y A. GETIS (1997). "*RECENT DEVELOPMENT IN SPATIAL ANALYSIS*". Eds. Springer-Verlag, Berlín; pp. 60-82.
  
- FUJITA, M., P. KRUGMAN y A.J. VENEABLES (1999). "*THE SPATIAL ECONOMY: CITIES, REGIONS AND INTERNATIONAL TRADE*". MIT Press. Cambridge MA.
  
- FUJITA, M. y P. KRUGMAN (2004). "*LA NUEVA GEOGRAFÍA ECONÓMICA: PASADO, PRESENTE Y FUTURO*". Investigaciones Regionales, primavera, número 004. Asociación Española de Ciencia Regional. Alcalá de Henares. España; pp. 177-206.
  
- GARCÍA DELGADO, J.L. Y M.J. GARCÍA GRANDE (2005). "*NACIMIENTO Y DESARROLLO DE UNA IDEA: DE LA CONFERENCIA DE STRESSA EN 1958 A LA REFORMA MACSHARRY EN 1992.*" En GARCÍA DELGADO, J.L. y M.J. GARCÍA GRANDE (2005). "*POLÍTICA AGRARIA COMÚN: BALANCE Y PERSPECTIVAS*". Nº 34. Colección de Estudios Económicos de la Fundación La Caixa.
  
- GARCÍA GRANDE, M.J. (2005). "*EL ÚLTIMO DECENIO: APLICACIÓN Y CONSECUENCIAS DE LAS REFORMAS DE LA PAC.*" En GARCÍA DELGADO, J.L. y M.J. GARCÍA GRANDE (2005). "*POLÍTICA AGRARIA COMÚN: BALANCE Y PERSPECTIVAS*". Nº 34. Colección de Estudios Económicos de la Fundación La Caixa.
  
- GEARY, R. (1954). "*THE CONTIGUITY RATIO AND STATISTICAL MAPPING*". The Incorporated Statistician, 5; pp. 115-45.

- 
- GÓMEZ DE ANTONIO, M. (2001). "*EVALUACIÓN DEL IMPACTO DEL STOCK DE CAPITAL PÚBLICO EN EL CRECIMIENTO DE LA RENTA PER CÁPITA DE LAS PROVINCIAS ESPAÑOLAS (1981-1991) MEDIANTE EL EMPLEO DE TÉCNICAS ECONOMETRICAS DE CARÁCTER ESPACIAL*". Colección Investigaciones, Instituto de Estudios Fiscales, Madrid.
- GONZÁLEZ TORRES, A. (2013). "*TURISMO RURAL EN CASTILLA Y LEÓN: DELIMITACIÓN COMARCAL Y ANÁLISIS ESPACIAL*." Tesis Doctoral. Departamento de Economía Aplicada. Facultad de CC Económicas y Empresariales. Universidad de Valladolid.
- GUELLEC, D., y P. RALLE (1995). "*LES NOUVELLES THÉORIES DE LA CROISSANCE*". La Decouverte. París.
- ISARD, W. (1956). "*LOCATION AND SPACE ECONOMY*". MIT Press, Cambridge.
- ISARD, W. (1960). "*METHODS OF REGIONAL ANALYSIS AN INTRODUCTION TO REGIONAL SCIENCE*". Cambridge. Technology Press of the Massachusetts Institute of Technology y Wiley, New York.
- ISARD, W. (2001). "*THE FUTURE OF REGIONAL SCIENCE: REMARKS PROMPTED BY PROFESSORS ALONSO AND TEITZ*". *International Regional Science Review*, 23:3, July 2001; pp. 414-421.
- JOHNSTON, J. (1984). "*ECONOMETRIC METHODS*." 3ª edición, McGraw-Hill.
- KELEJIAN, H. y D. ROBINSON (1992). "SPATIAL AUTOCORRELATION: A NEW COMPUTATIONALLY SIMPLE TEST WITH AN APPLICATION TO PER CAPITA COUNTY POLICY EXPENDITURES". *Regional Science and Urban Economics*, 22; pp. 317-31.
- KELEJIAN, H. H. y D. P. ROBINSON (1997). "*INFRASTRUCTURE PRODUCTIVITY ESTIMATION AND ITS UNDERLYING ECONOMETRIC SPECIFICATIONS: A SENSITIVITY ANALYSIS*". *Papers in Regional Science*, 76; pp. 115-131.

- KELEJIAN, H. H. e I.R. PRUCHA (2007). "*HAC ESTIMATION IN A SPATIAL FRAMEWORK.*" *J. Econometrics* 140.pp. 131-154.
- KELEJIAN, H. H. e I.R. PRUCHA (2010). "*SPECIFICATION AND ESTIMATION OF SPATIAL AUTOREGRESSIVE MODELS WITH AUTOREGRESSIVE AND HETEROSKEDASTIC DISTURBANCES.*" *J. Econometrics* 157, pp. 53-67.
- KING, G. (1997). "*A SOLUTION TO THE ECOLOGICAL INFERENCE PROBLEM. RECONSTRUCTING INDIVIDUAL BEHAVIOR FROM AGGREGATE DATA*". Princeton. Princeton University Press.
- KMENTA, J. (1971). "*ELEMENTS OF ECONOMETRICS*". New York: MacMillan.
- KOPP, A. (1995). "*PUBLIC INVESTMENT-KEY TO EAST GERMAN GROWTH?*". 51 Congreso del International Institute of Public Finance. Lisboa. Agosto de 1995.
- KRUGMAN, P. (1991). "*INCREASING RETURNS AND ECONOMIC GEOGRAPHY*". *Journal of Political Economy*, 99. pp. 483-499.
- KRUGMAN, P. (1995). "*DESARROLLO, GEOGRAFÍA Y TEORÍA ECONÓMICA*". Antoni Bosch editor, Barcelona España.
- KRUGMAN, P. (1998). "*WHAT'S NEW ABOUT THE NEW ECONOMIC GEOGRAPHY?*". *Oxford Review of Economic Policy*, Vol. 14, No.2, pp. 7-17.
- KRUGMAN, P. y A.J. VENEABLES (1995). "*GLOBALIZATION AND THE INEQUALITY OF NATIONS*". *Quarterly Journals of Economics*, 110(4); pp. 857-880.
- LE SAGE, J. (1999). "*SPATIAL ECONOMETRICS*". Department of Economics, University of Toledo.
- LE SAGE, J.P. y R.K. PACE (2009). "*INTRODUCTION TO SPATIAL ECONOMETRICS*". Chapman & Hall/CRC.
- MARTIN, P. y C. A. ROGERS. (1995). "*INDUSTRIAL LOCATION AND PUBLIC INFRASTRUCTURE*". *Journal of International Economics*. 39; pp. 335-351.

- 
- MAUDOS, J., J. M. PASTOR y L. SERRANO. (1998). "*CONVERGENCIA EN LAS REGIONES ESPAÑOLAS: CAMBIO TÉCNICO, EFICIENCIA Y PRODUCTIVIDAD*". *Revista Española de Economía*, 15 (2); pp. 235-264.
- MEYER, J.R. (1966). "*REGIONAL ECONOMICS: A SURVEY*". En "*SURVEYS OF ECONOMIC THEORY, VOL. II*". American Economic Association and Royal Economic Society. Pag 6.
- MORAN, P. (1948). "THE INTERPRETATION OF STATISTICAL MAPS". *Journal of the Royal Statistical Society B*, vol. 10; pp. 243-251.
- MORENO, R. y E. VAYÁ (2000). "*TÉCNICAS ECONOMETRÍAS PARA EL TRATAMIENTO DE DATOS ESPACIALES: LA ECONOMETRÍA ESPACIAL*". Edicions Universitat de Barcelona. Colecció UB 44. Manuals.
- MORENO, R., M. ARTÍS, E. LÓPEZ-BARO y J. SURIÑACH (1997). "*EVIDENCE ON THE COMPLEX LINK BETWEEN INFRASTRUCTURE AND REGIONAL GROWTH*". *International Journal of Development Planning Literature*. 12 (1-2); pp. 81-108.
- MUR, J. y A. ANGULO (2008). "*PUBLICACIONES DE LA ASIGNATURA MODELOS REGIONALES. TEMA 2: PROCESOS ESTOCÁSTICOS ESPACIALES*". Departamento de Análisis Económico. Facultad de CC Económicas y Empresariales. Universidad de Zaragoza.
- NOVALES, A. (1993). "*ECONOMETRÍA*". Ed. McGraw-Hill. Madrid.
- PAELINCK, J.H.P. y L.H. KLAASEN (1979). "*SPATIAL ECONOMETRICS*". Saxon House.
- PEDRAJA, F., J. RAMAJO y J. SALINAS. (1999). "*EFICIENCIA PRODUCTIVA DEL SECTOR INDUSTRIAL ESPAÑOL: UN ANÁLISIS ESPACIAL Y SECTORIAL*". *Papeles de Economía Española*, 80; pp. 51-68.
- PÉREZ PINEDA, J.A. (2005). "*CRECIMIENTO Y DESEQUILIBRIOS REGIONALES: UN MODELO ESPACIAL PARA MÉXICO*". Tesis doctoral. Universidad Complutense de Madrid.
-

- PÉREZ PINEDA, J.A. (2006). "*ECONOMETRÍA ESPACIAL Y CIENCIA REGIONAL*". Investigación Económica. Vol LXV. Número 258; pp 129-160.
  
- PERROUX, F. (1950). "*ECONOMIC SPACE: THEORY AND APPLICATIONS.*" The Quarterly Journal of Economics, Vol. LXIV; pp. 89-104.
  
- PINKSE, J. y M.E. SLADE (2009). "*THE FUTURE OF SPATIAL ECONOMETRICS*". Prepared for the 50th anniversary symposium of the Journal of Regional Science held at the New York Federal Reserve. Center for Auctions, Procurements, and Competition Policy, Department of Economics, The Pennsylvania State University, 608 Kern Graduate Building, University Park 16802
  
- REPHANN, T. J. y A. ISSERMAN (1994). "*NEW HIGHWAYS AS ECONOMIC DEVELOPMENT TOOLS: AN EVALUATION USING QUASI-EXPERIMENTAL MATCHING METHODS*". Regional Science and Urban Economics. 24; pp. 723-751.
  
- REY S.J. y L. ANSELIN (2010). "*PYSAL: A PYTHON LIBRARY OF SPATIAL ANALYTICAL METHODS*". En FISHER, M. y A. GETIS (2010). "*HANDBOOK OF APPLIED SPATIAL ANALYSIS. SOFTWARE TOOLS, METHODS AND APPLICATIONS.*" Springer Heidelberg Dordrecht London New York.
  
- REY S.J. y M.V. JANIKAS (2010). "*STARS: SPACE-TIME ANALYSIS OF REGIONAL SYSTEMS*". En FISHER, M. y A. GETIS (2010). "*HANDBOOK OF APPLIED SPATIAL ANALYSIS. SOFTWARE TOOLS, METHODS AND APPLICATIONS.*" Springer Heidelberg Dordrecht London New York.
  
- RICHARDSON, H.W. (1973). "*ECONOMÍA REGIONAL: TEORÍA DE LA LOCALIZACIÓN, ESTRUCTURAS URBANAS Y CRECIMIENTO REGIONAL*". Editorial Vincens-Vives, Barcelona. España.
  
- SAEZ I ZAFRA, M. y C. SAURINA CANALS (2007). "*ESTADÍSTICA Y EPIDEMIOLOGÍA ESPACIAL*" Volumen 17 de UdG Publicacions. Documenta Universitaria.

- SALINAS, M. M. (2004): "*PUBLIC INFRASTRUCTURE AND PRIVATE PRODUCTIVITY IN THE SPANISH REGIONS*". Journal of Policy Modelling, 26 (1); pp. 47-64.
  
- SCOTT, L.M. y M.V. JANIKAS (2010). "*Spatial Statistics in ArcGIS*". En FISHER, M. y A. GETIS (2010). "*HANDBOOK OF APPLIED SPATIAL ANALYSIS. SOFTWARE TOOLS, METHODS AND APPLICATIONS.*" Springer Heidelberg Dordrecht London New York.
  
- SHIRLEY, C. y C. WINSTON (2004). "*FIRM INVENTORY BEHAVIOR AND THE RETURNS FROM HIGHWAY INFRASTRUCTURE INVESTMENT*". Journal of Urban Economics 55 (2); pp. 398- 415.



# ANEXOS

## ANEXO 1 - AEDE

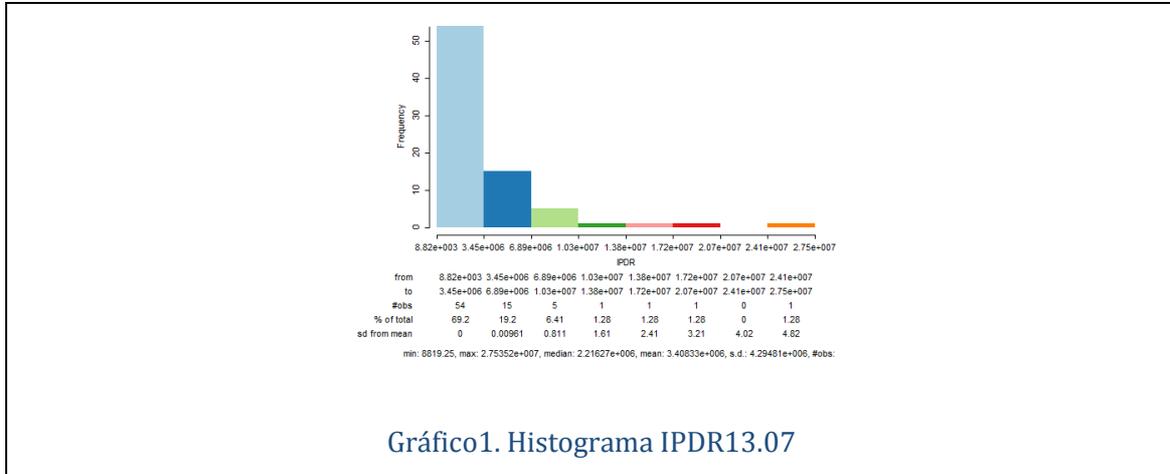


Gráfico1. Histograma IPDR13.07

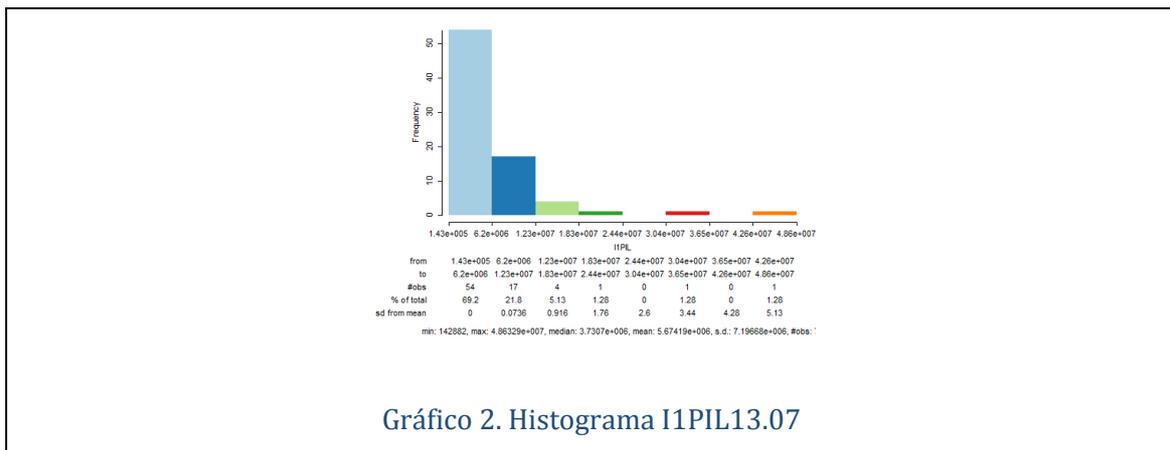


Gráfico 2. Histograma I1PIL13.07

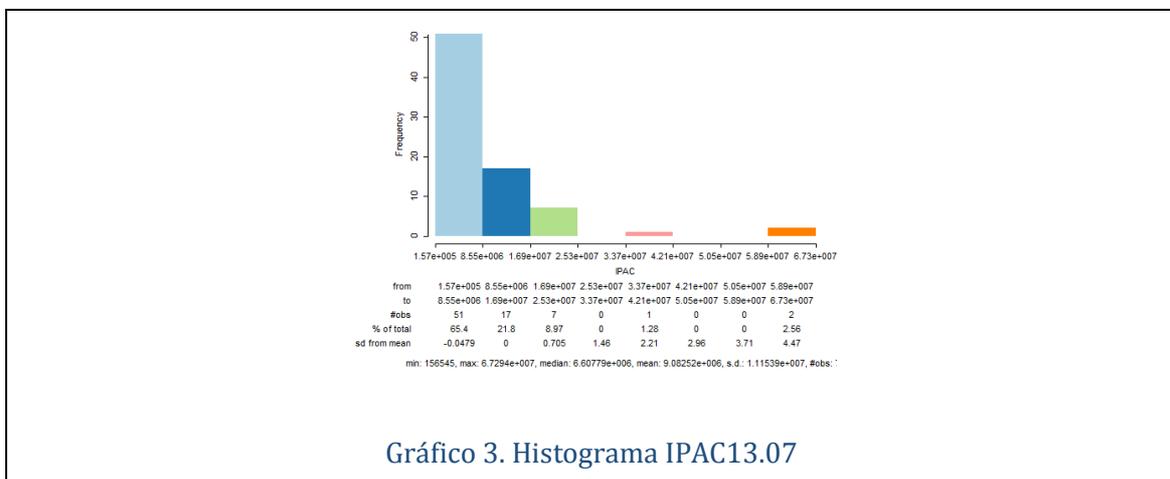
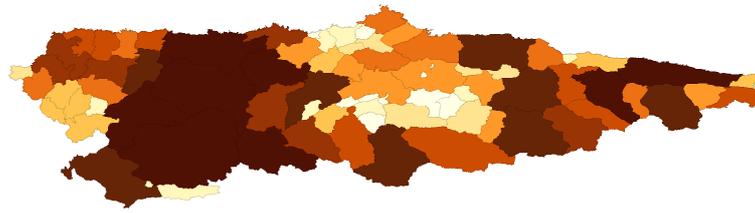
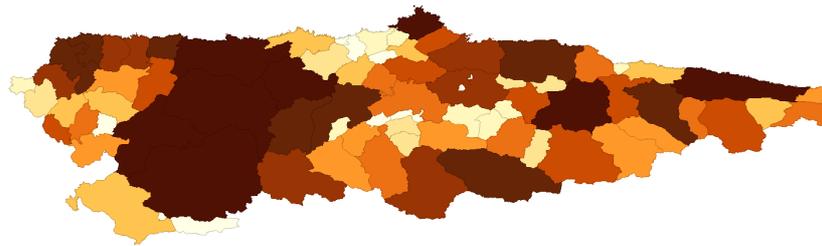


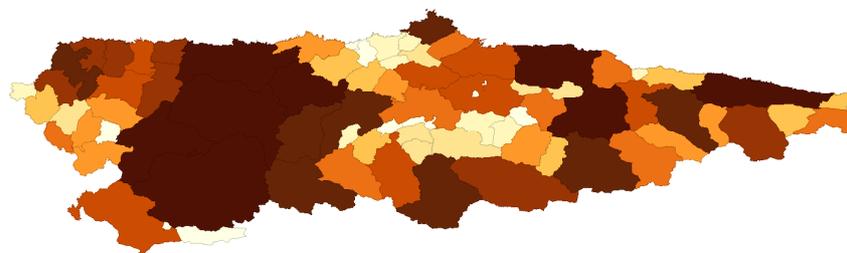
Gráfico 3. Histograma IPAC13.07



Mapa 1. Mapa de deciles IPDR13.07



Mapa 2. Mapa de deciles I1PIL13.07



Mapa 3. Mapa de deciles IPAC13.07

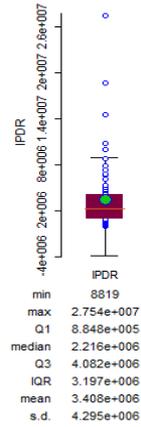


Gráfico 4. Diagrama de caja IPDR13.07

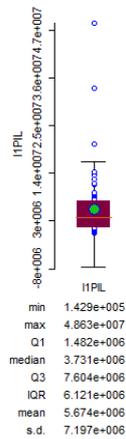


Gráfico 5. Diagrama de caja I1PIL13.07

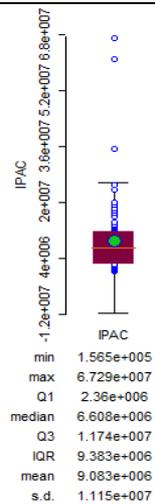
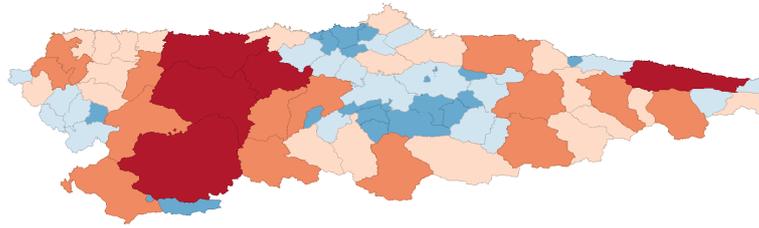
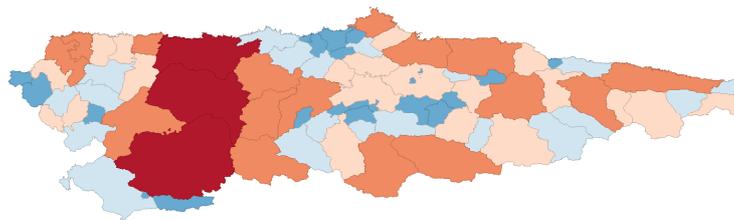


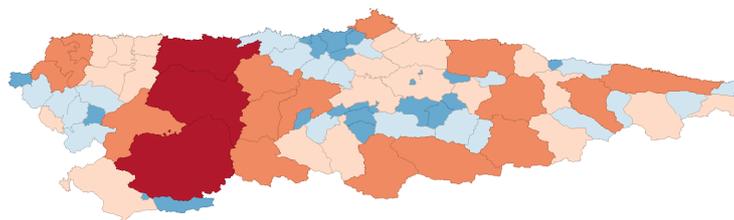
Gráfico 6. Diagrama de caja IPAC13.07



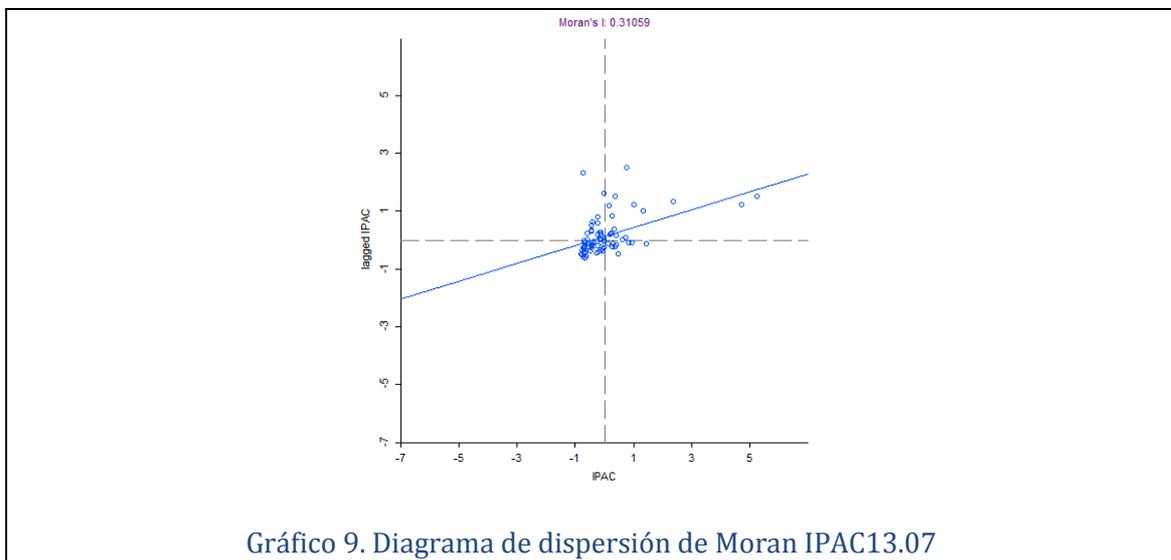
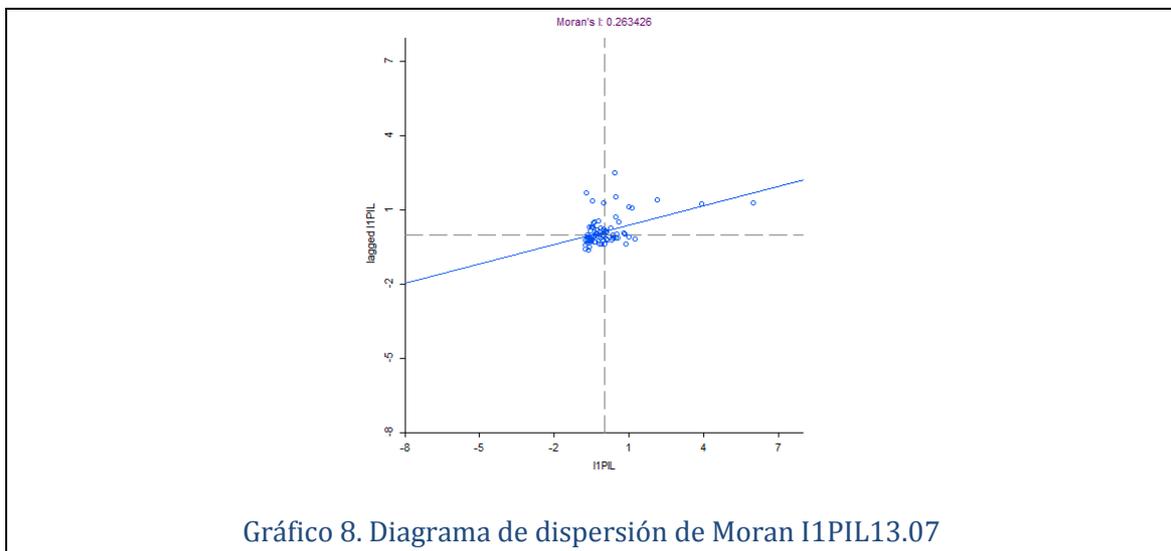
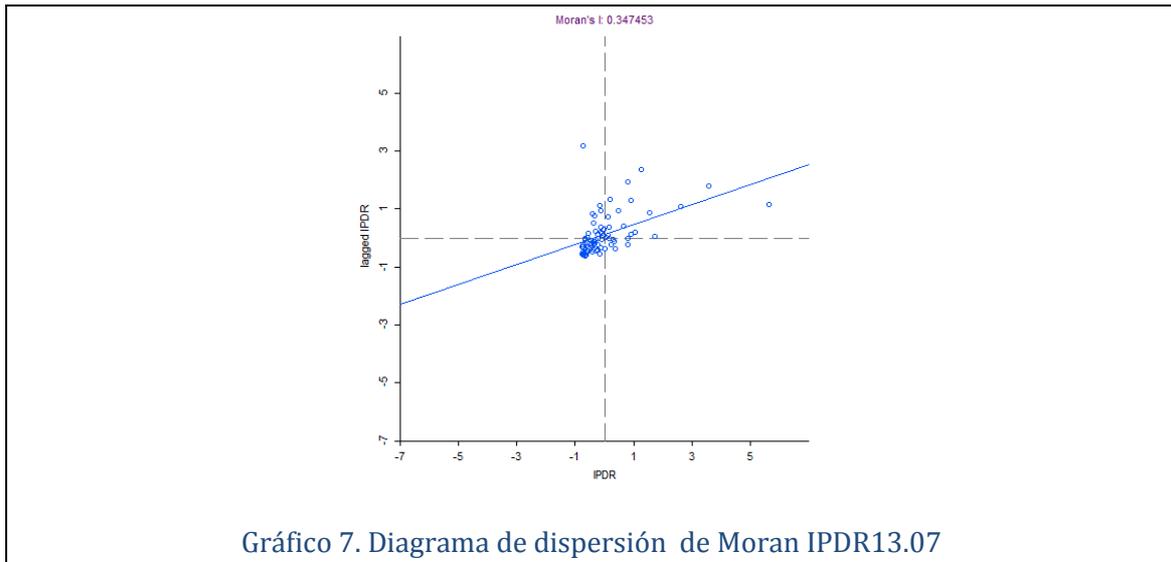
Mapa 4. Mapa de caja IPDR13.07



Mapa 5. Mapa de caja I1PIL13.07



Mapa 6. Mapa de caja IPAC13.07



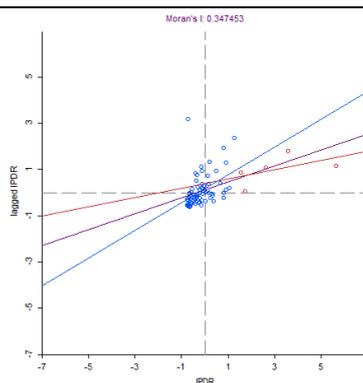


Gráfico 10. Diagrama de dispersión de Moran IPDR13.07 con las rectas de regresión que eliminan valores atípicos

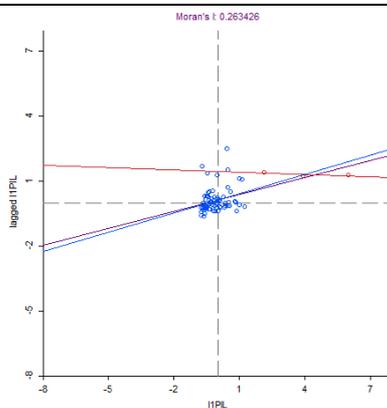


Gráfico 11. Diagrama de dispersión de Moran I1PIL13.07 con las rectas de regresión que eliminan valores atípicos

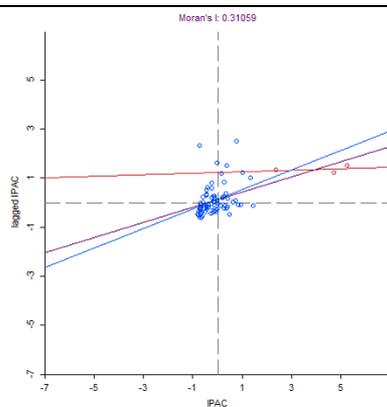
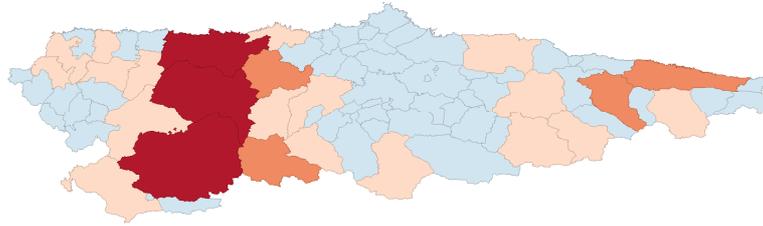
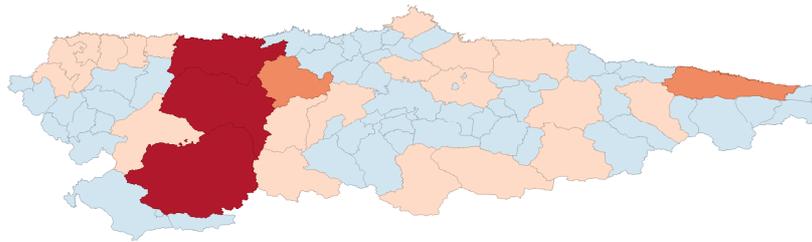


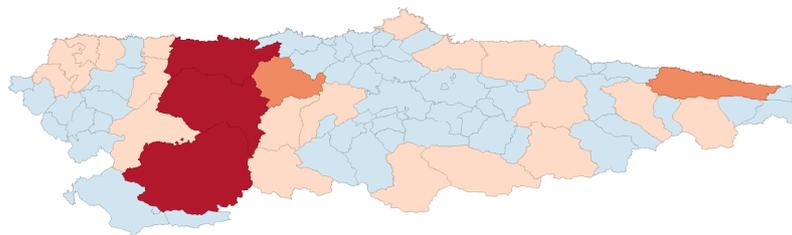
Gráfico 12. Diagrama de dispersión de Moran IPAC13.07 con las rectas de regresión que eliminan valores atípicos



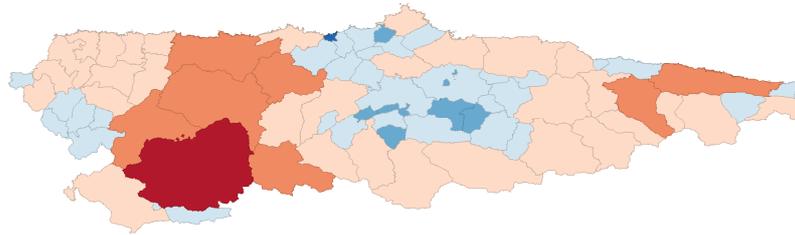
Mapa 7. Mapa de desviación típica IPDR13.07



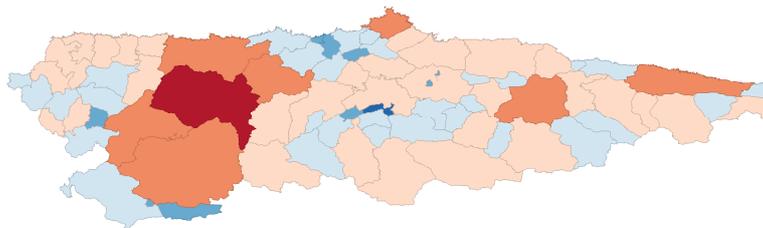
Mapa 8. Mapa de desviación típica I1PIL13.07



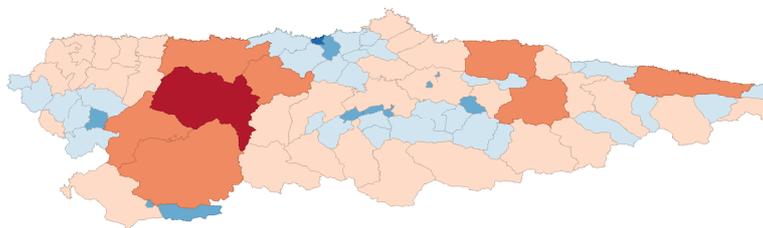
Mapa 9. Mapa de desviación típica IPAC13.07



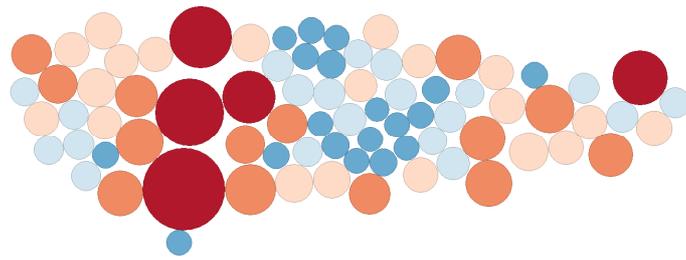
Mapa 10. Mapa percentiles IPDR13.07



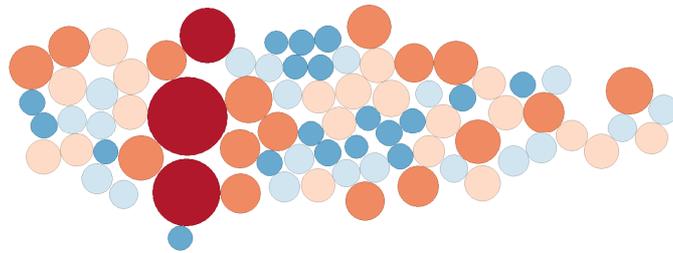
Mapa 11. Mapa percentiles I1PIL13.07



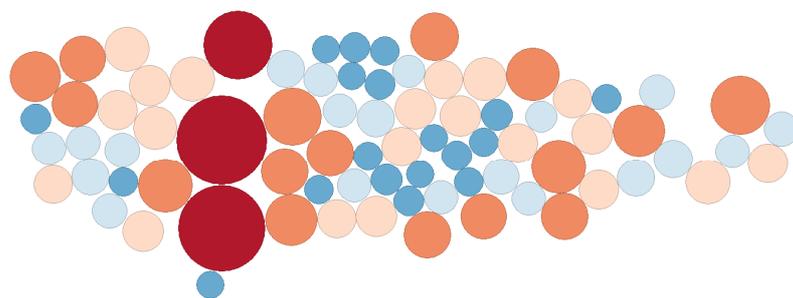
Mapa 12. Mapa percentiles IPAC13.07



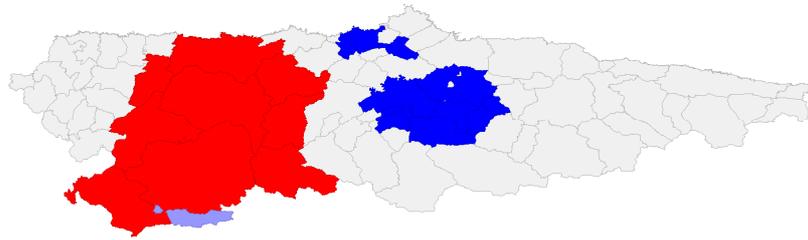
Mapa 13. Cartograma IPDR13.07



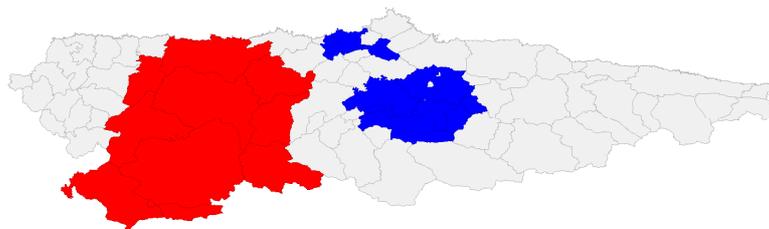
Mapa 14. Cartograma I1PIL13.07



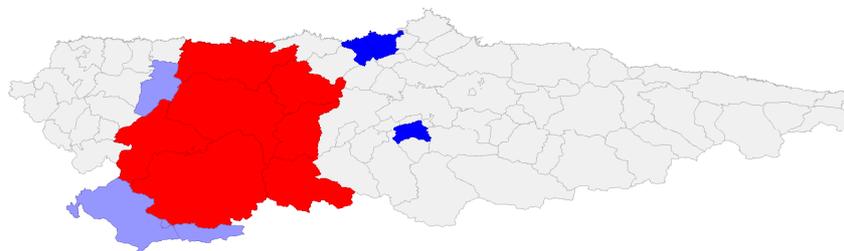
Mapa 15. Cartograma IPAC13.07



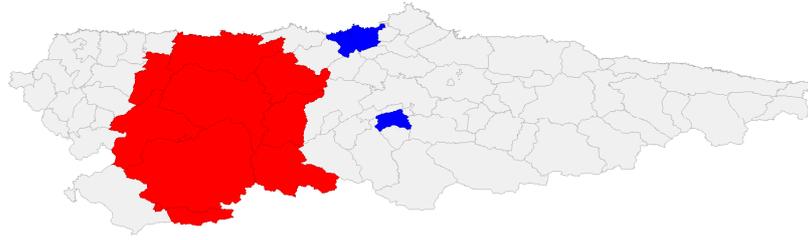
Mapa 16. Mapa LISA I de Moran IPDR13.07



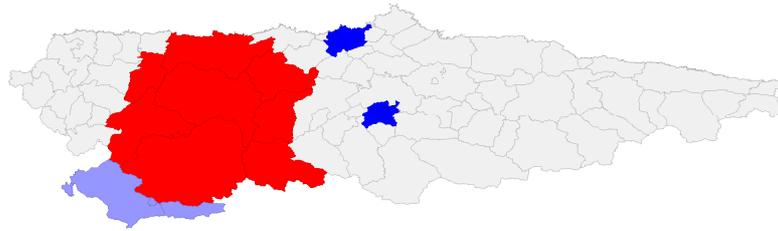
Mapa 17. Mapa LISA  $G_i$  IPDR13.07



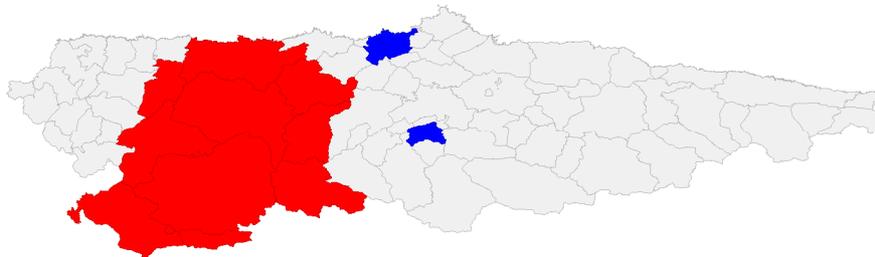
Mapa 18. Mapa LISA I de Moran I1PIL13.07



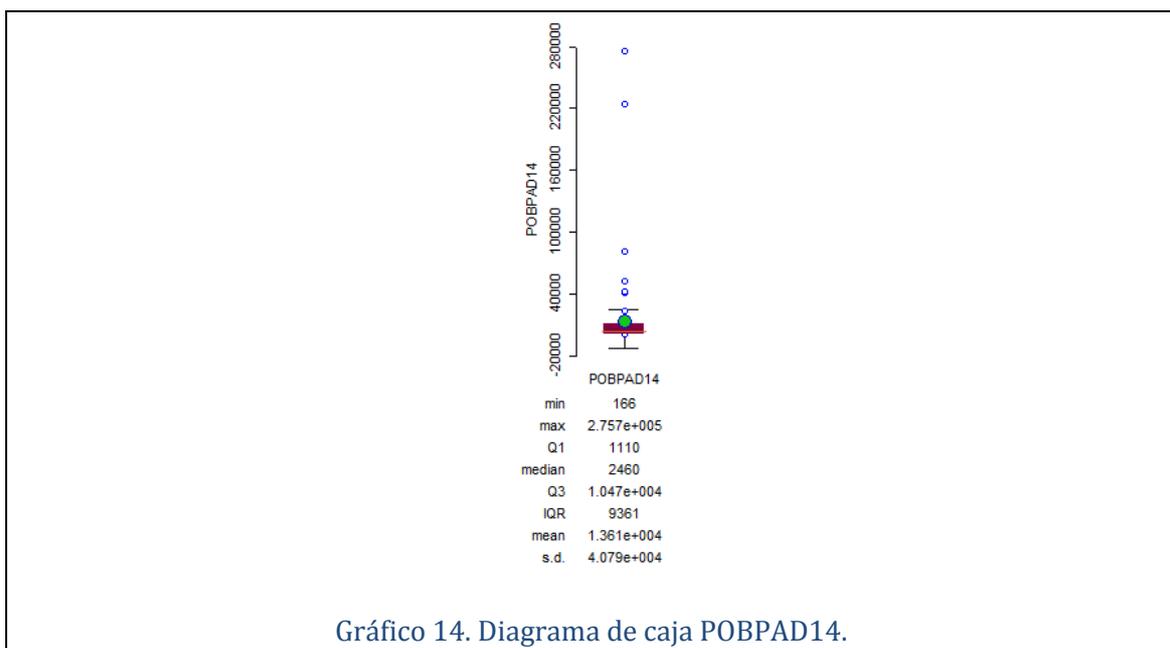
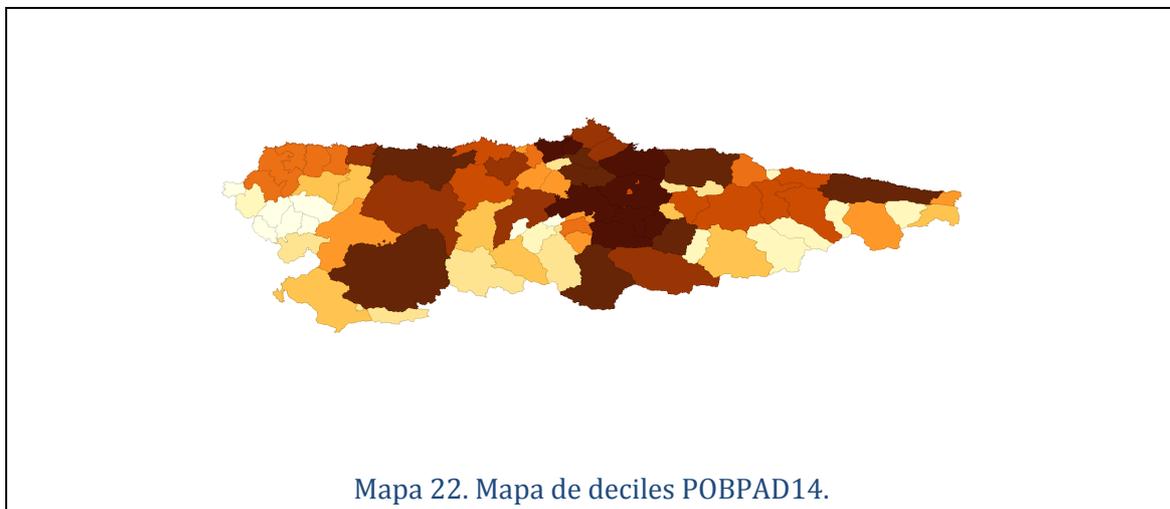
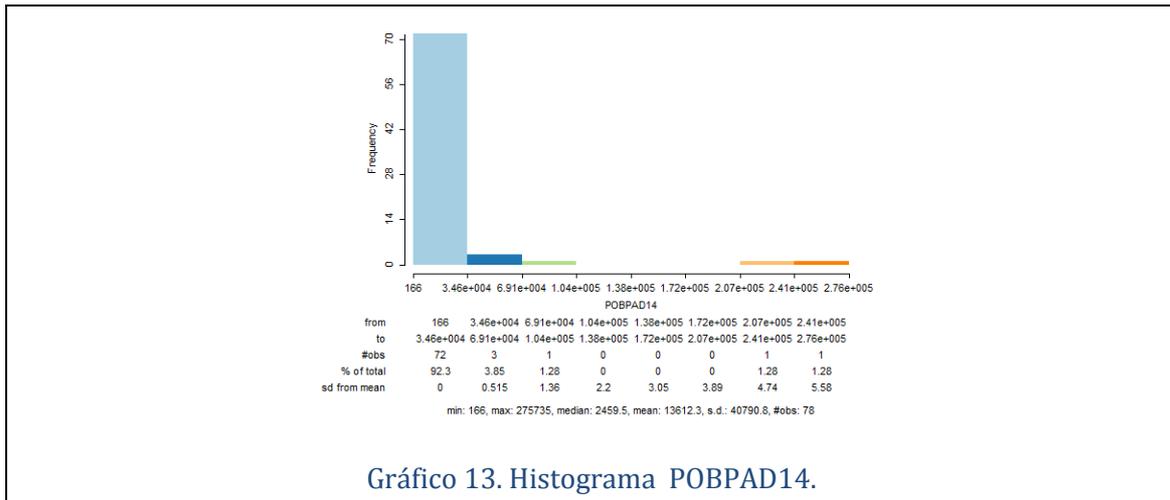
Mapa 19. Mapa LISA G<sub>1</sub>IPIL13.07

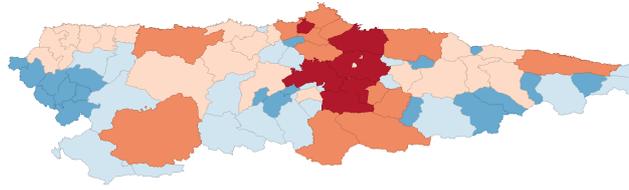


Mapa 20. Mapa LISA I de Moran IPAC13.07



Mapa 21. Mapa LISA G<sub>i</sub> IPAC13.07





Mapa 23. Mapa de caja POBPAD14.

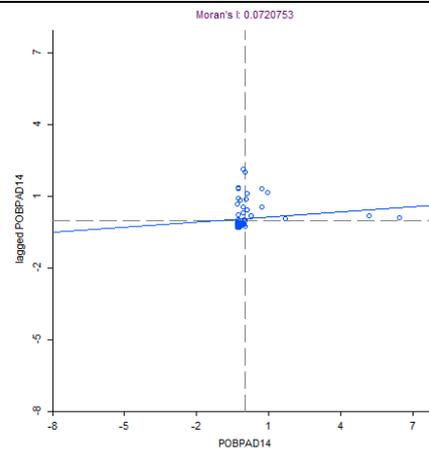


Gráfico 15. Diagrama de dispersión de Moran POBPAD14.

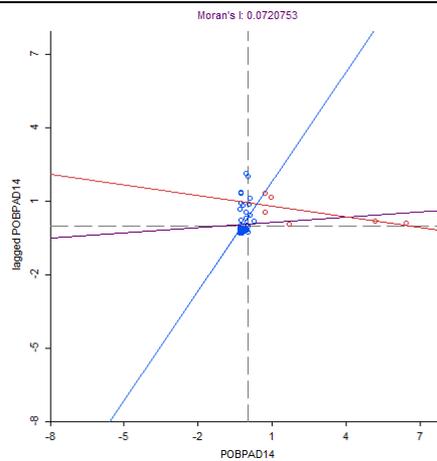
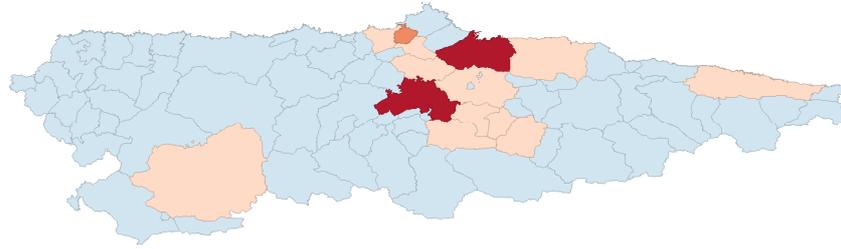
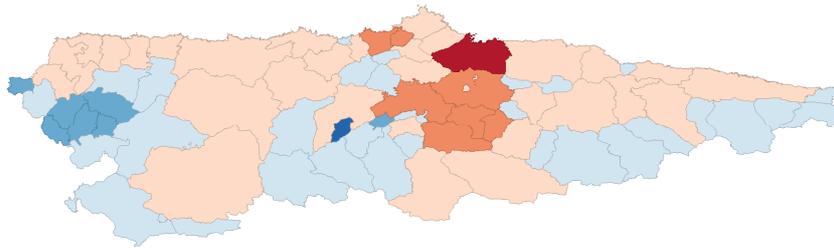


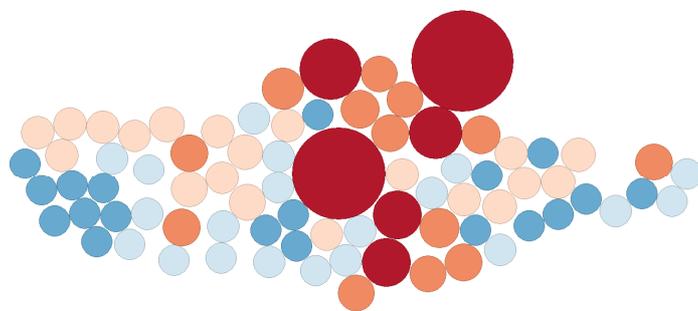
Gráfico 16. Diagrama de dispersión de Moran POBPAD14 con las rectas de regresión que eliminan valores atípicos.



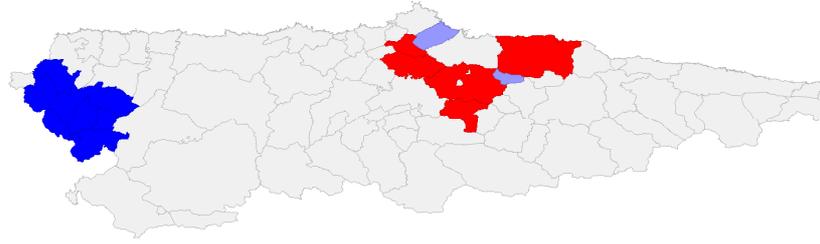
Mapa 24. Mapa de desviación típica POBPAD14.



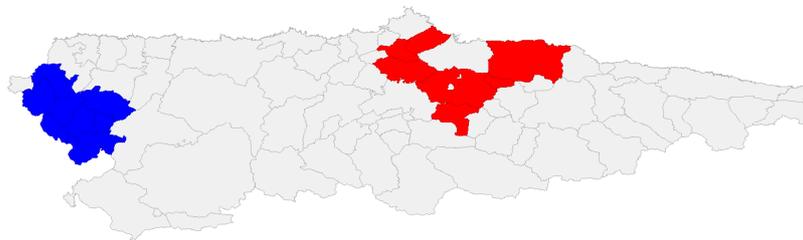
Mapa 24. Mapa de percentiles POBPAD14.



Mapa 25. Cartograma POBPAD14.



Mapa 26. Mapa LISA I de Moran POBPAD14.

Mapa 27. Mapa LISA  $G_i$  POBPAD14.Mapa 28. Mapa LISA  $G_i^*$  POBPAD14.

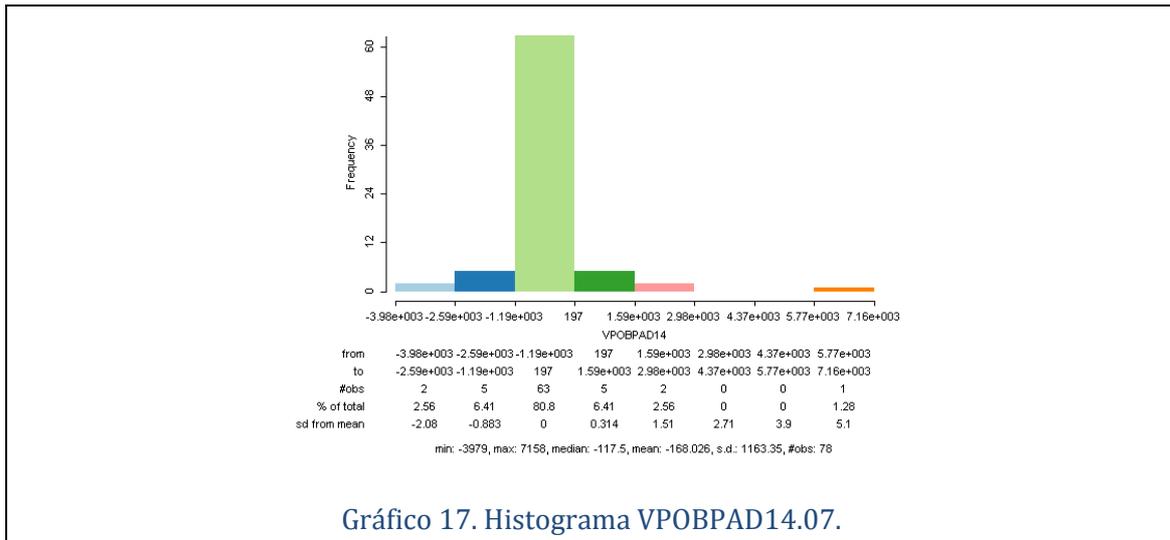
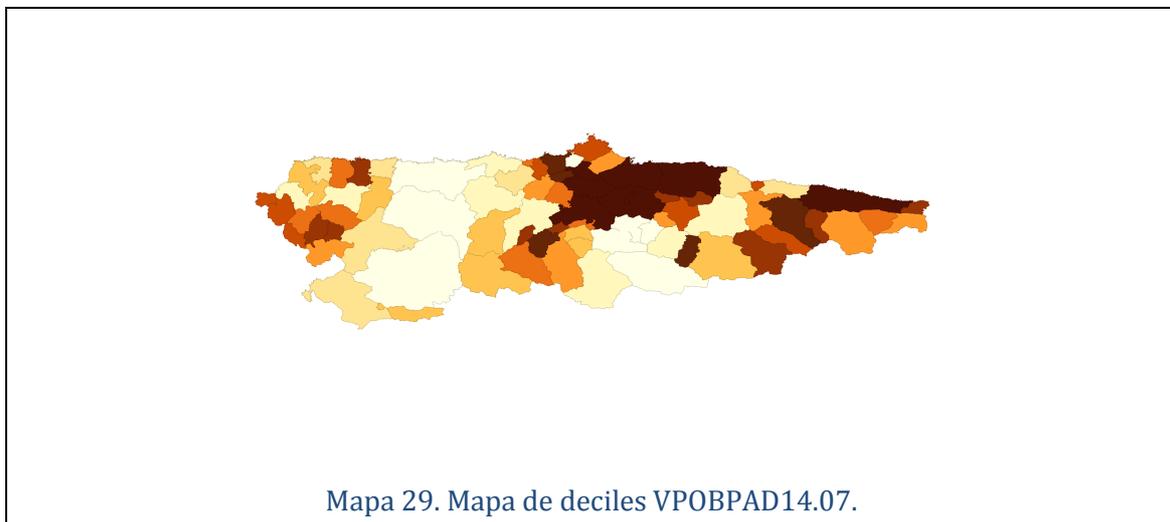


Gráfico 17. Histograma VPOBPAD14.07.



Mapa 29. Mapa de deciles VPOBPAD14.07.

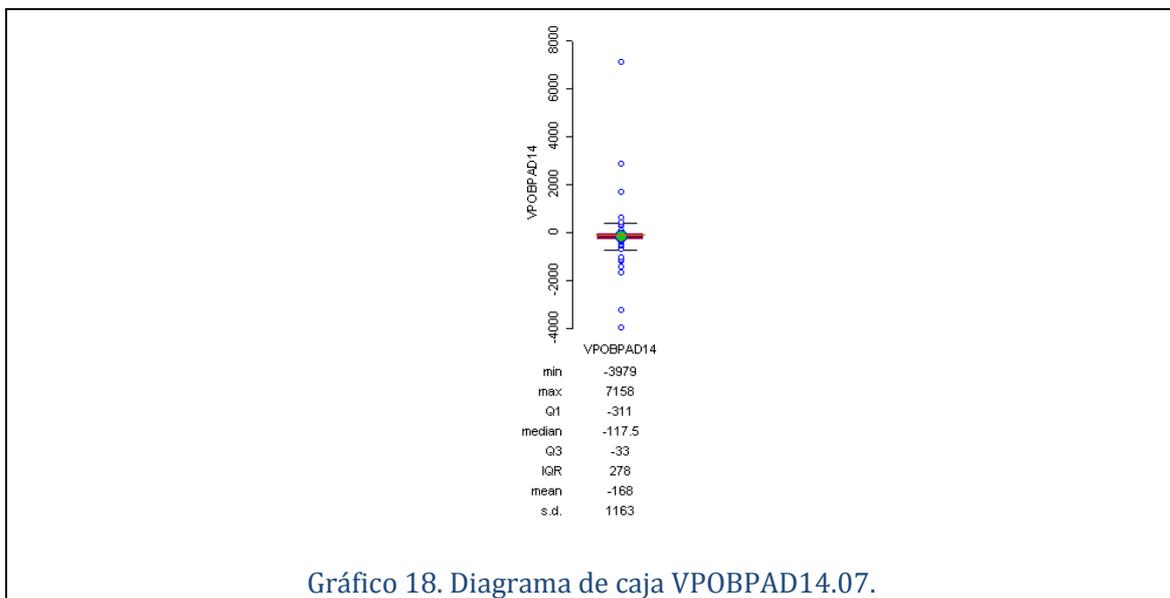
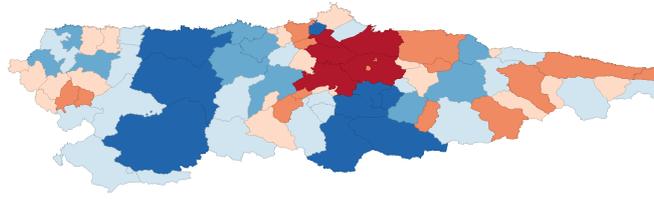


Gráfico 18. Diagrama de caja VPOBPAD14.07.



Mapa 30. Mapa de caja VPOBPAD14.07.

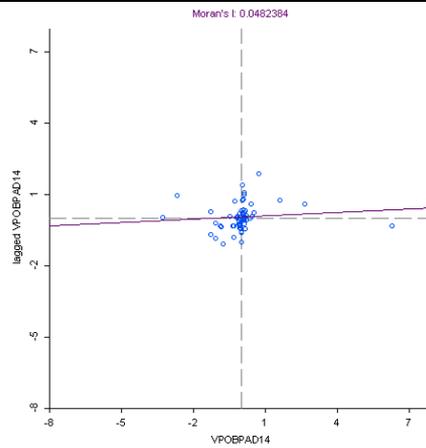


Gráfico 19. Diagrama de dispersión de Moran VPOBPAD14.07.

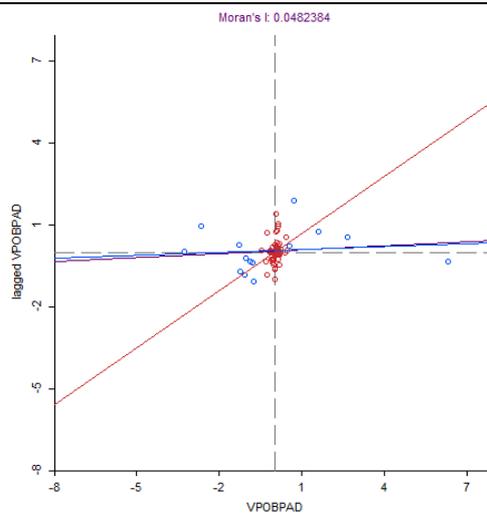
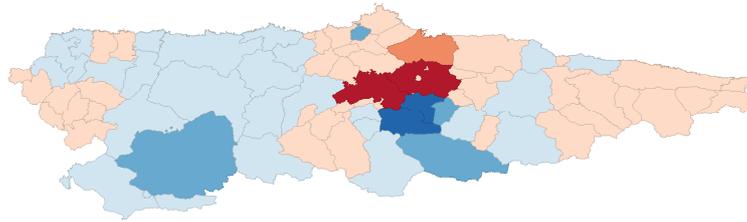
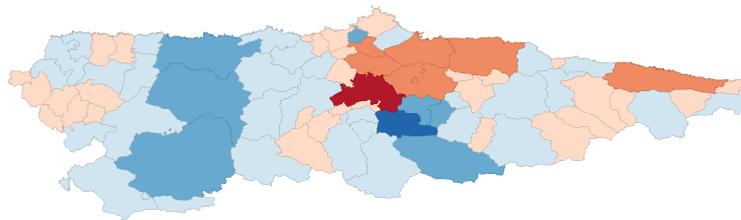


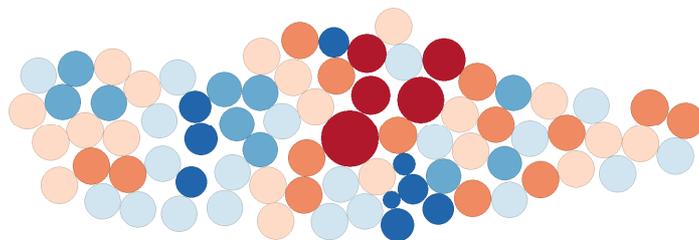
Gráfico 20. Diagrama de dispersión de Moran VPOBPAD14.07 con las rectas de regresión que eliminan valores atípicos.



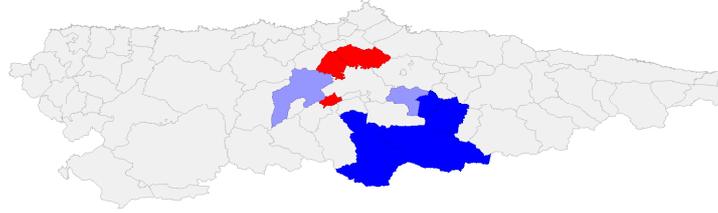
Mapa 31. Mapa de desviación típica VPOBPAD14.07.



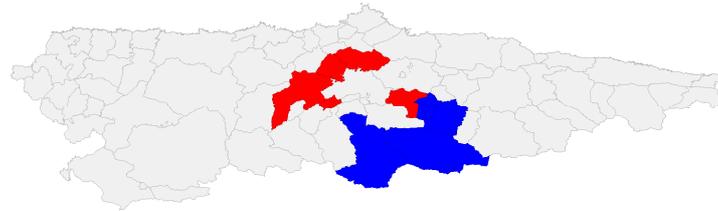
Mapa 32. Mapa de percentiles VPOBPAD14.07.



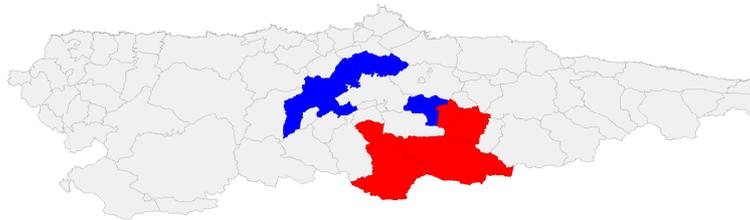
Mapa 33. Cartograma VPOBPAD14.07.



Mapa 34. Mapa LISA I de Moran VPOBPAD14.07.



Mapa 35. Mapa LISA  $G_i$  VPOBPAD14.07.



Mapa 36. Mapa LISA  $G_i^*$  VPOBPAD14.07.

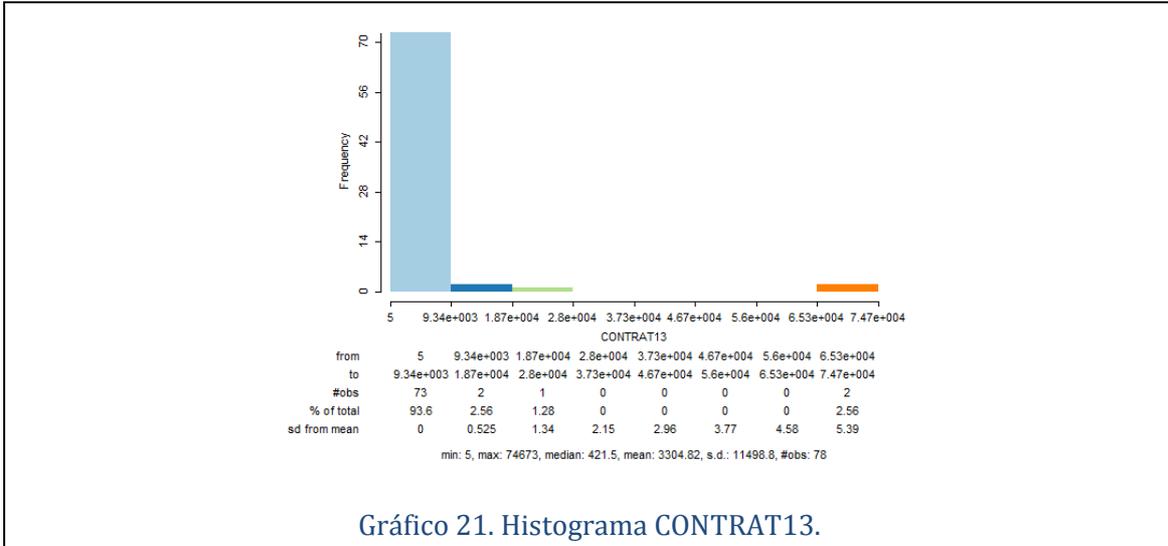


Gráfico 21. Histograma CONTRAT13.

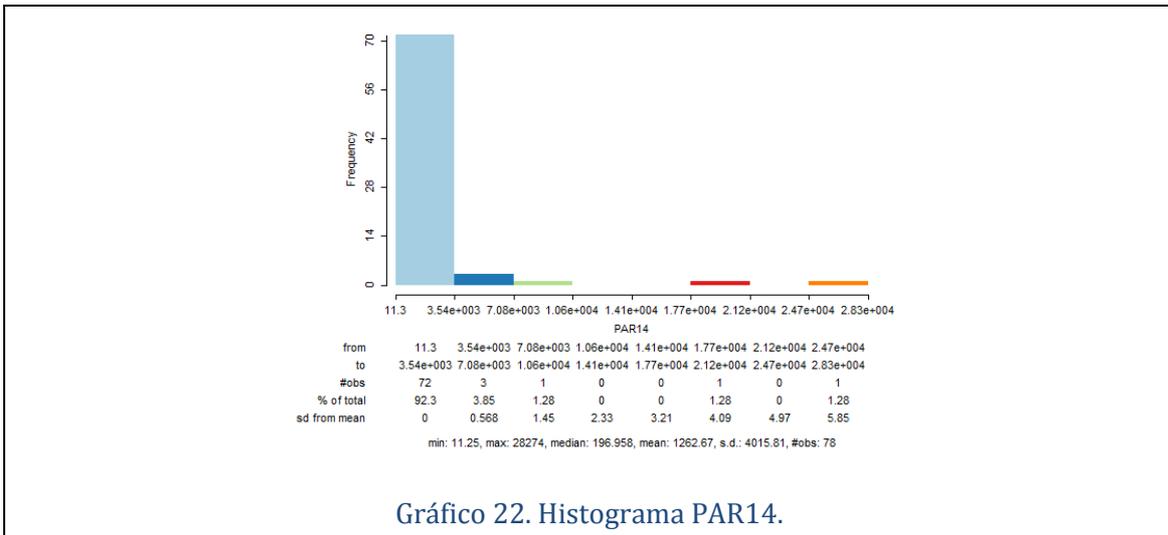


Gráfico 22. Histograma PAR14.

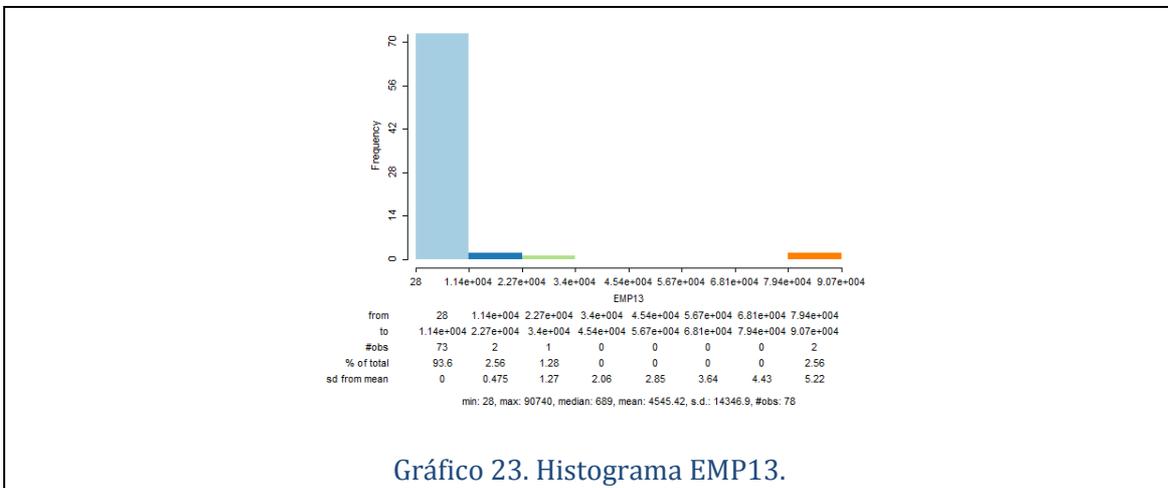


Gráfico 23. Histograma EMP13.

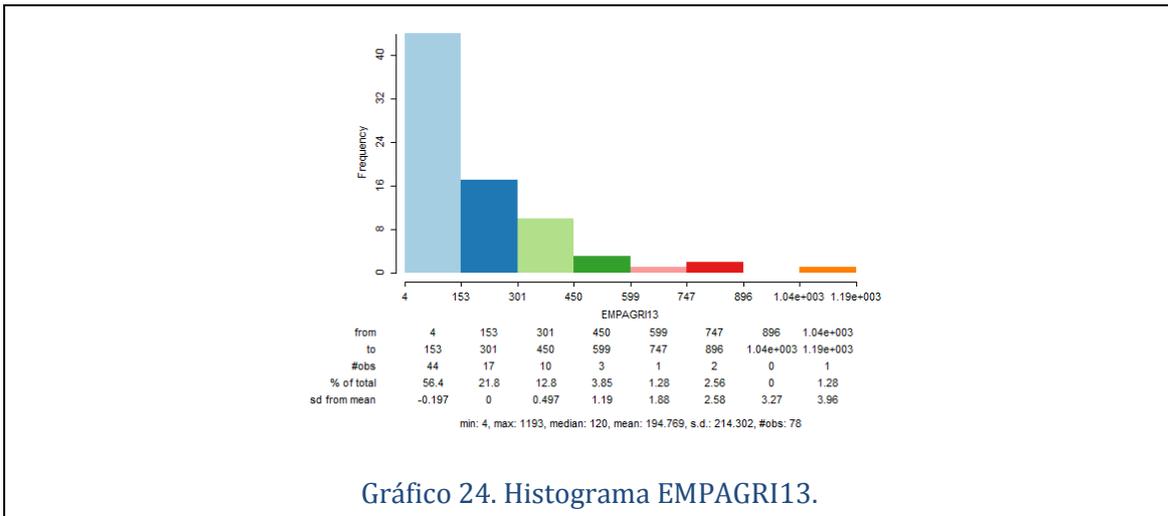


Gráfico 24. Histograma EMPAGRI13.

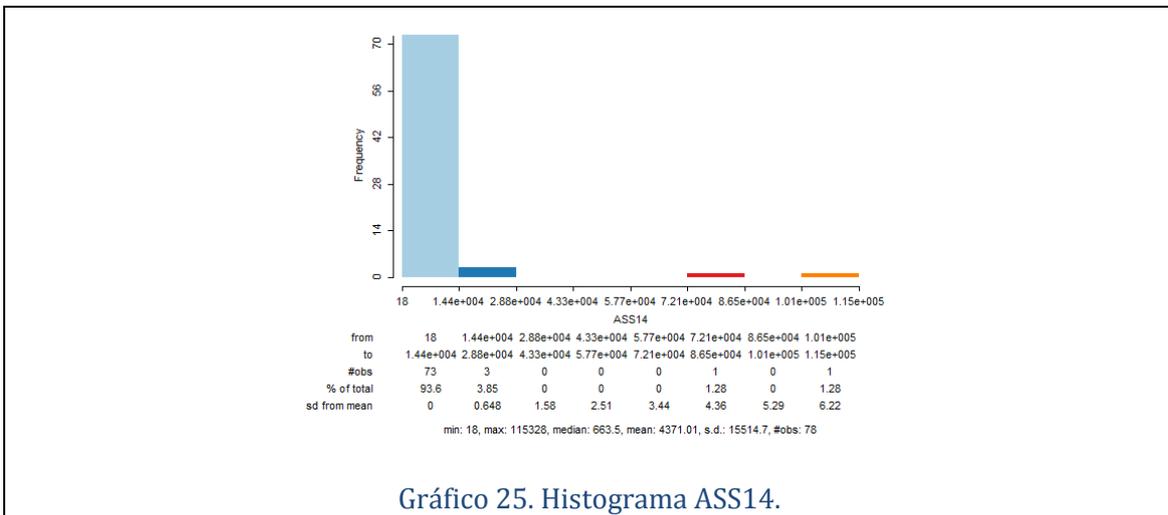
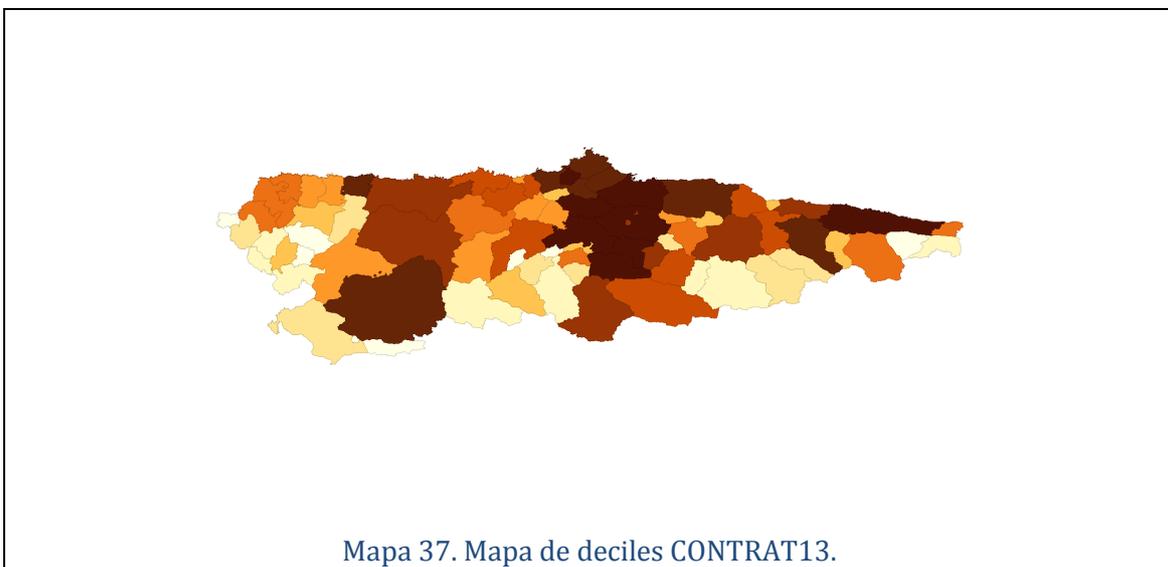
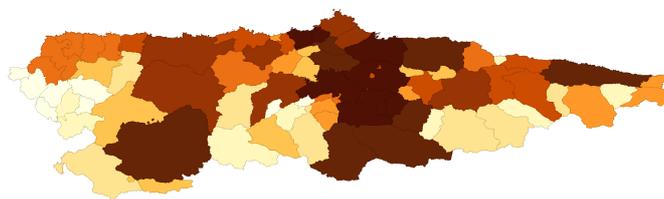


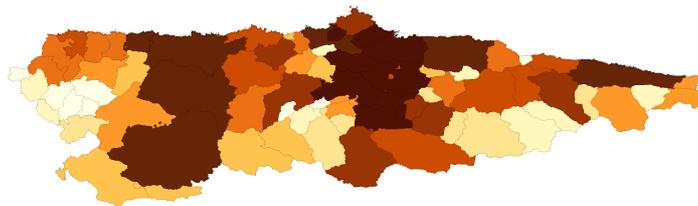
Gráfico 25. Histograma ASS14.



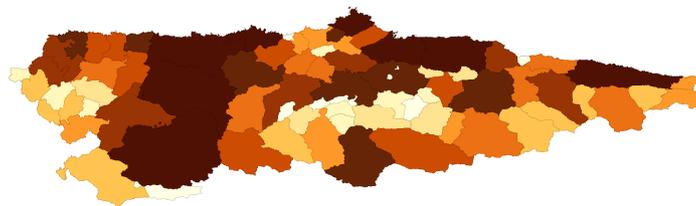
Mapa 37. Mapa de deciles CONTRAT13.



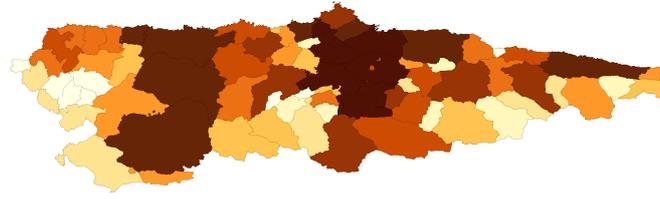
Mapa 38. Mapa de deciles PAR14.



Mapa 39. Mapa de deciles EMP13.



Mapa 40. Mapa de deciles EMPAGRI13.



Mapa 41. Mapa de deciles ASS14.

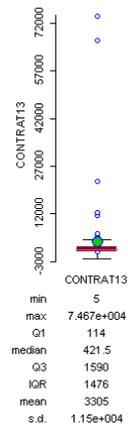


Gráfico 26. Diagrama de caja CONTRAT13.

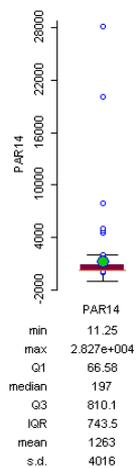


Gráfico 27. Diagrama de caja PAR14.

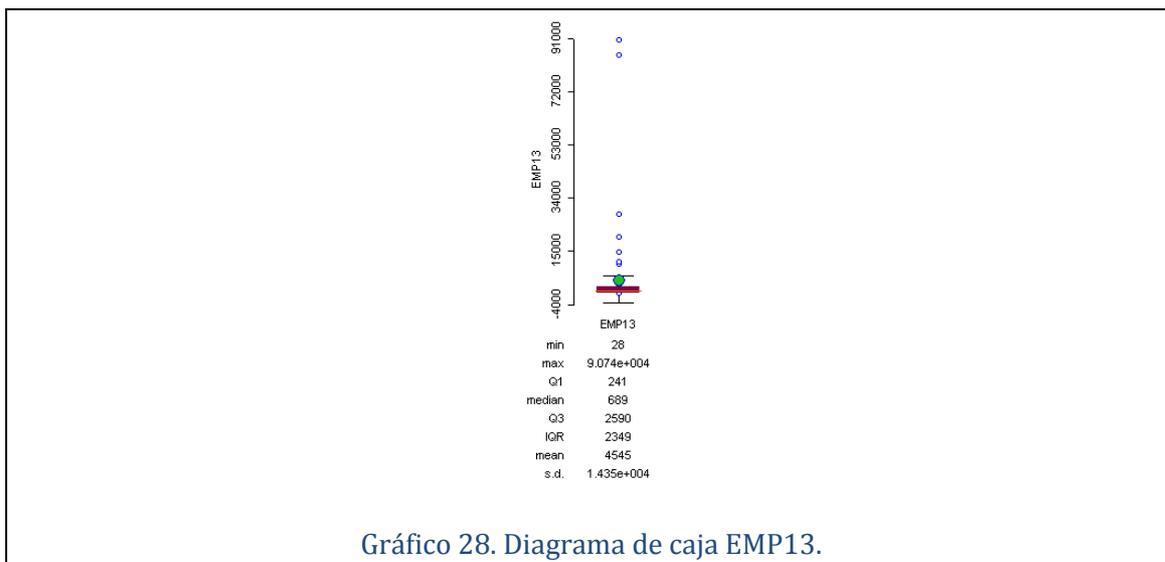


Gráfico 28. Diagrama de caja EMP13.

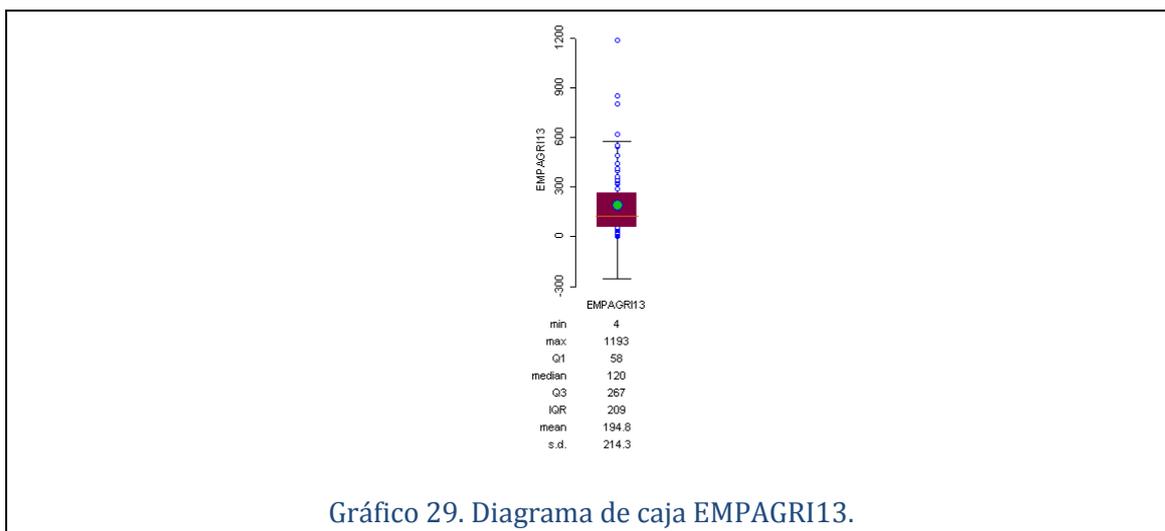


Gráfico 29. Diagrama de caja EMPAGRI13.

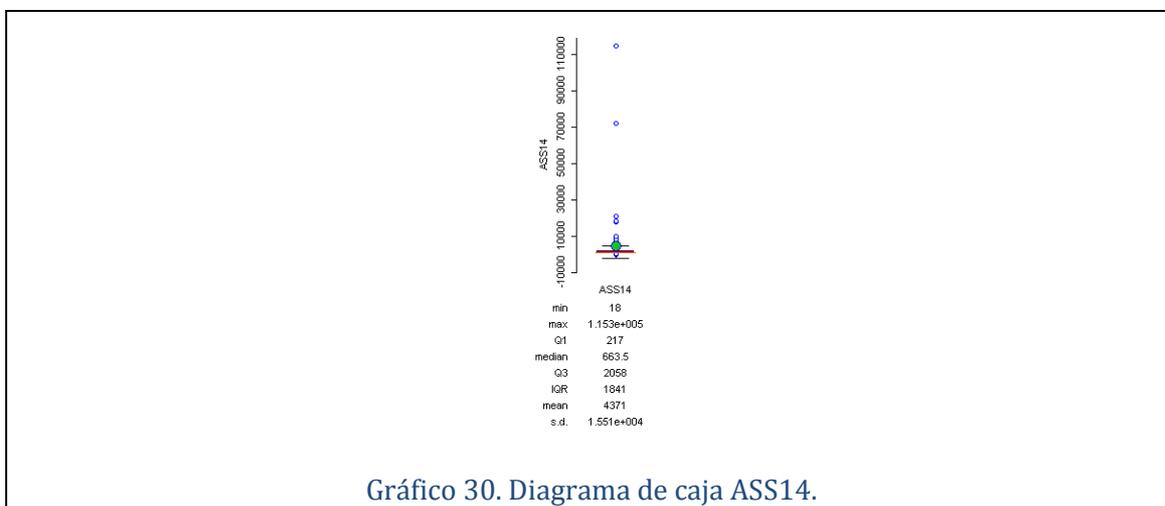
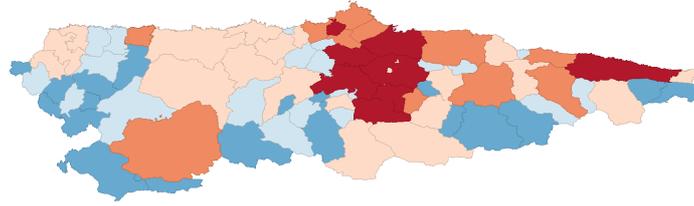
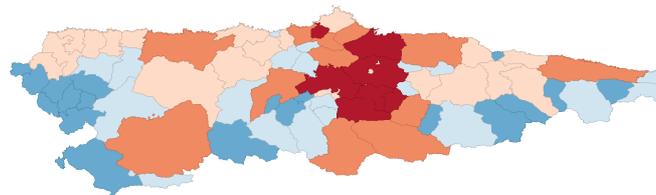


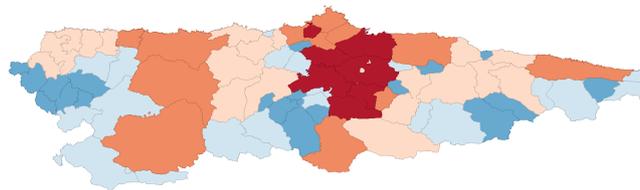
Gráfico 30. Diagrama de caja ASS14.



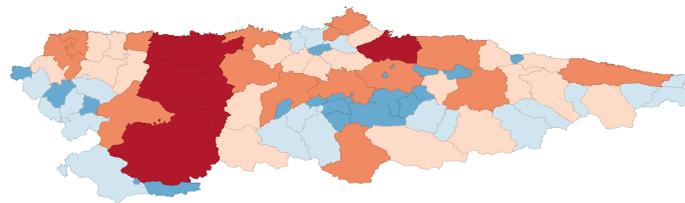
Mapa 42. Mapa de caja CONTRAT13.



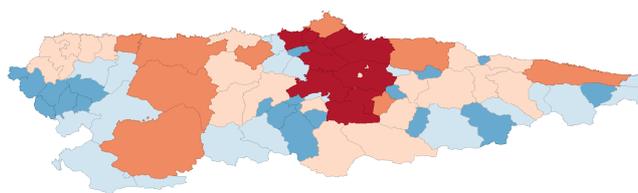
Mapa 43. Mapa de caja PAR14.



Mapa 44. Mapa de caja EMP13.



Mapa 45. Mapa de caja EMPAGRI13.



Mapa 46. Mapa de caja ASS14.

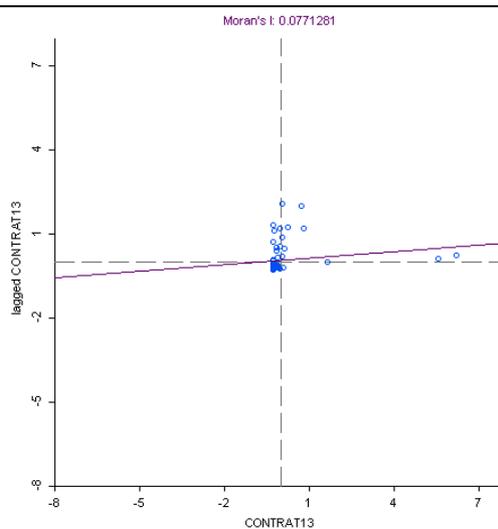
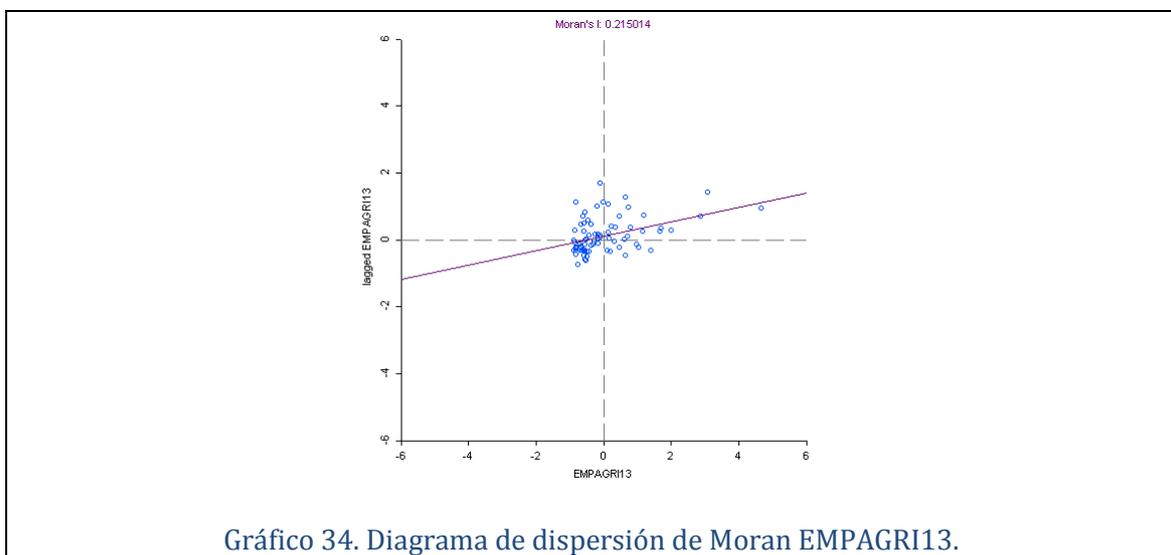
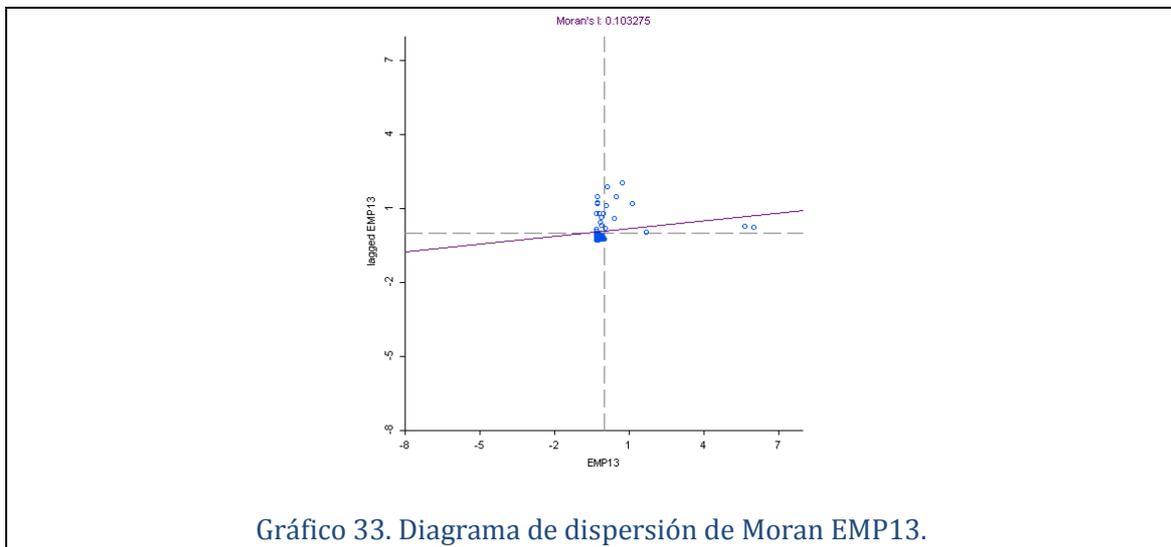
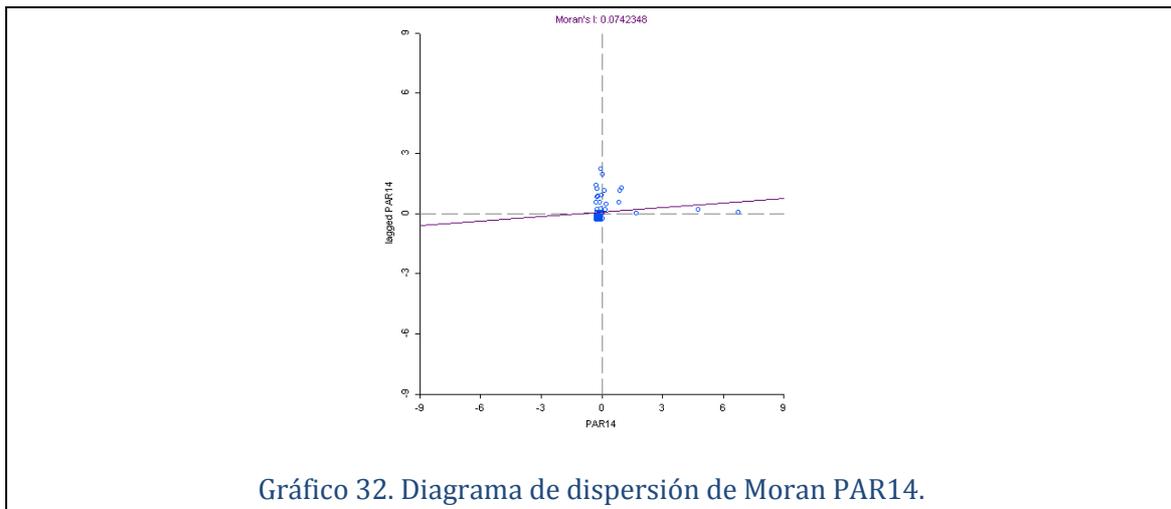
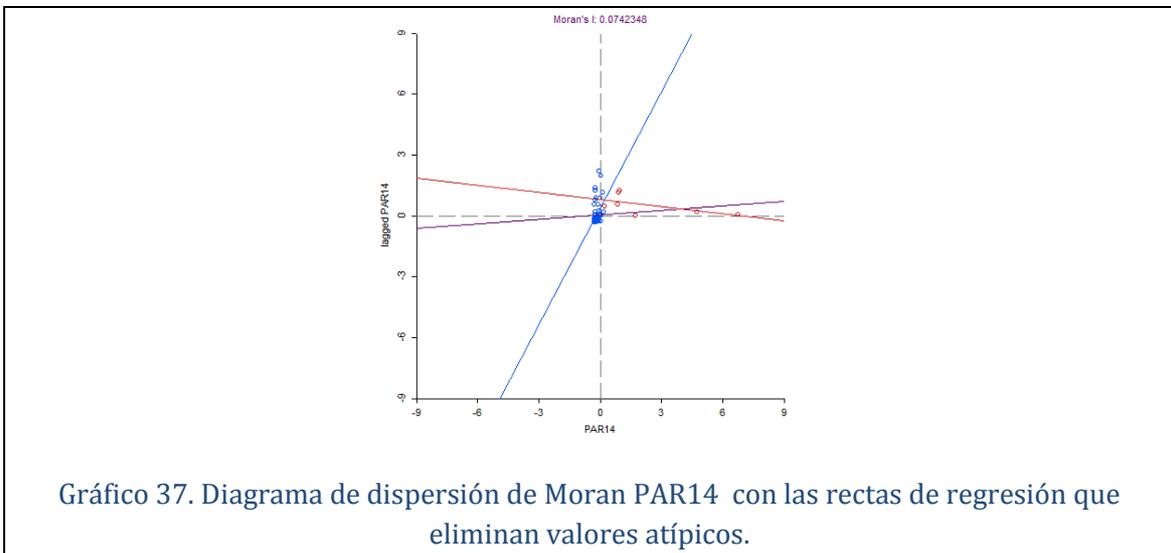
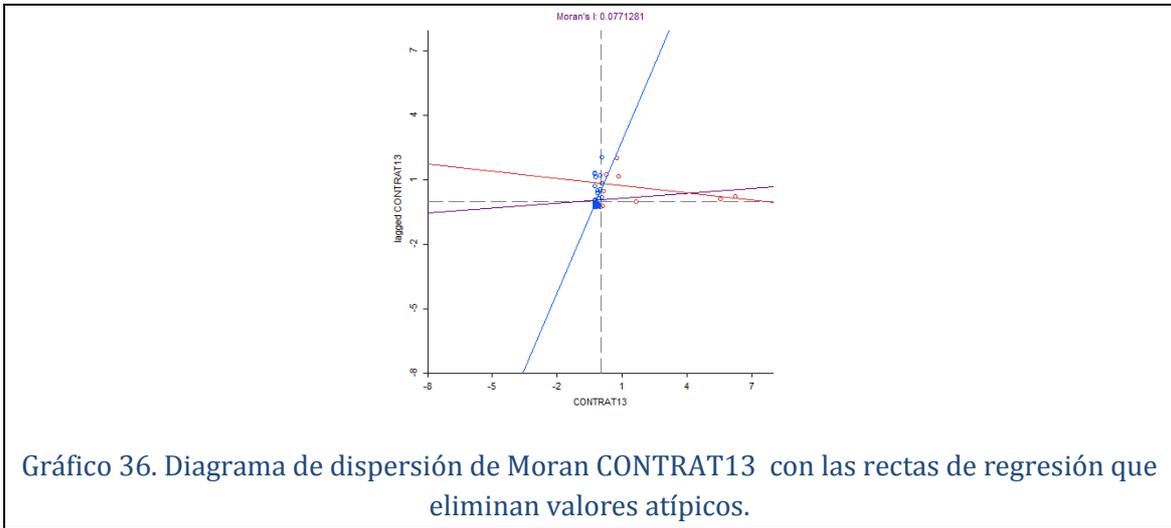
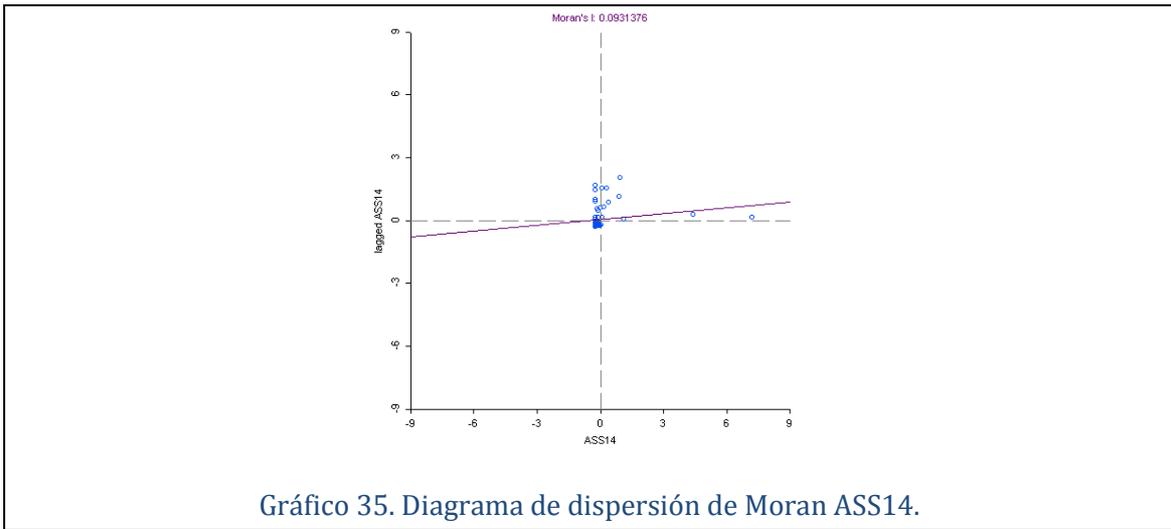


Gráfico 31. Diagrama de dispersión de Moran CONTRAT13.





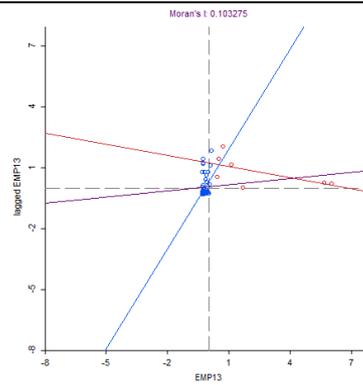


Gráfico 38. Diagrama de dispersión de Moran EMP13 con las rectas de regresión que eliminan valores atípicos.

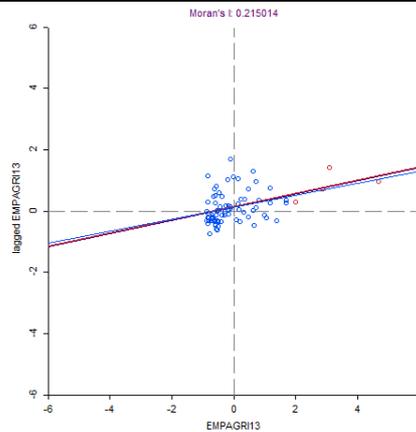


Gráfico 39. Diagrama de dispersión de Moran EMPAGRI13 con las rectas de regresión que eliminan valores atípicos.

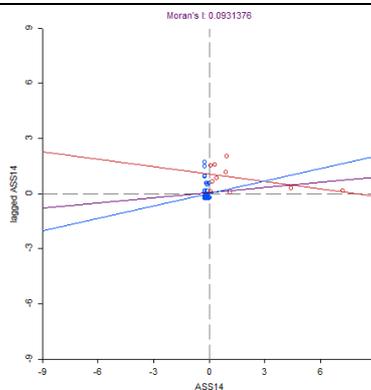
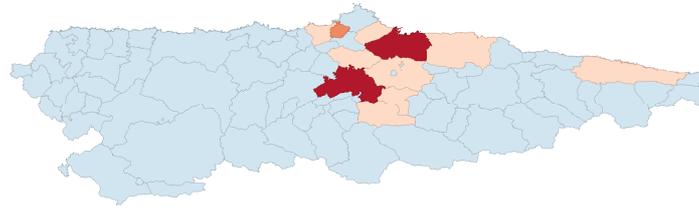
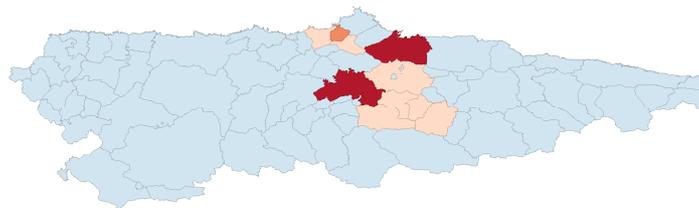


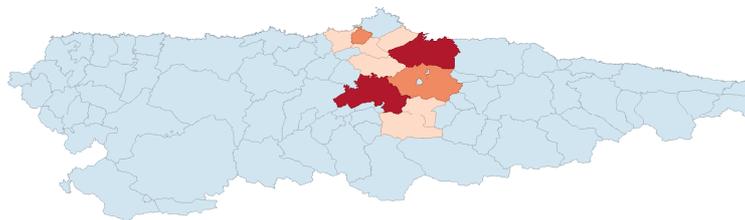
Gráfico 40. Diagrama de dispersión de Moran ASS14 con las rectas de regresión que eliminan valores atípicos.



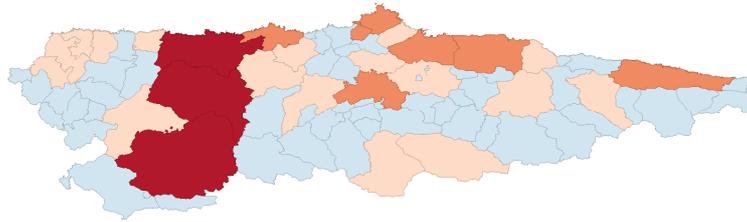
Mapa 47. Mapa desviación típica CONTRAT13.



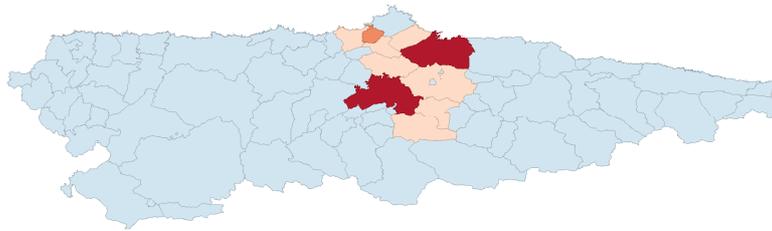
Mapa 48. Mapa desviación típica PAR14.



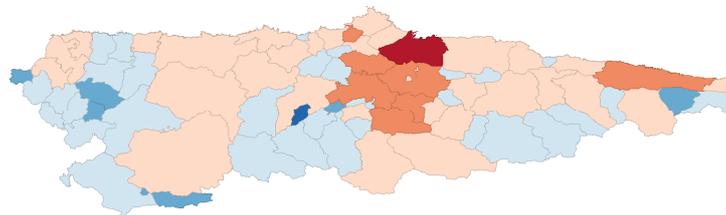
Mapa 49. Mapa desviación típica EMP13.



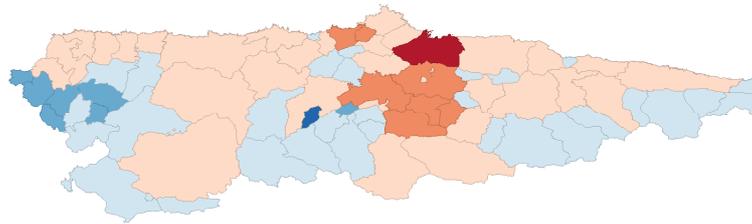
Mapa 50. Mapa desviación típica EMPAGRI13.



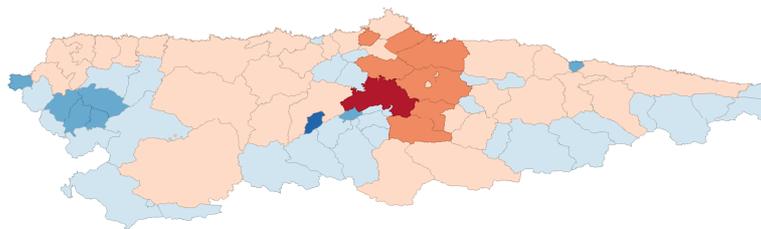
Mapa 51. Mapa desviación típica ASS14.



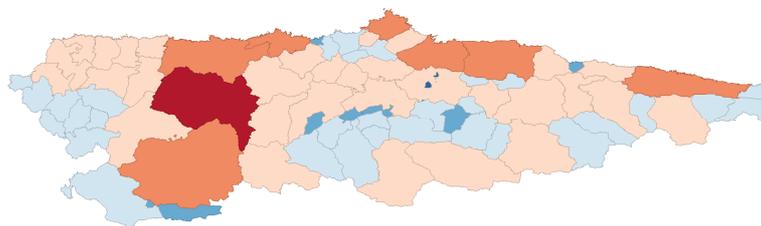
Mapa 52. Mapa de percentiles CONTRAT13.



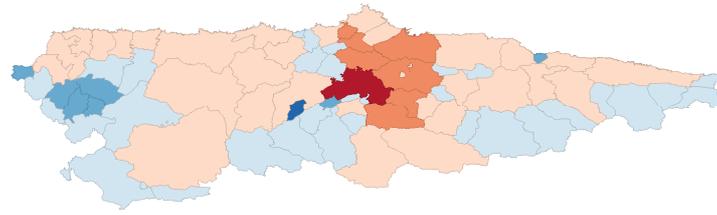
Mapa 53. Mapa de percentiles PAR14.



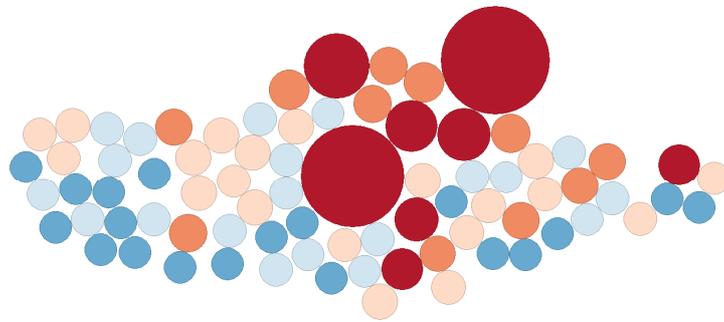
Mapa 54. Mapa de percentiles EMP13.



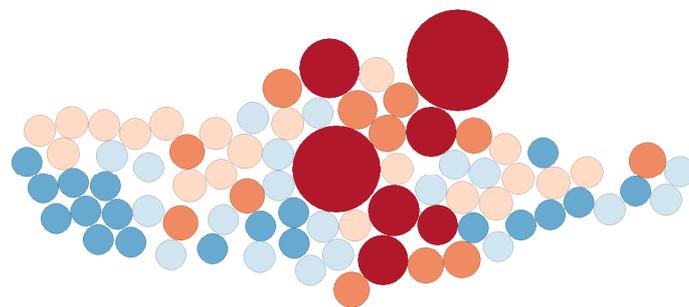
Mapa 55. Mapa de percentiles EMPAGRI13.



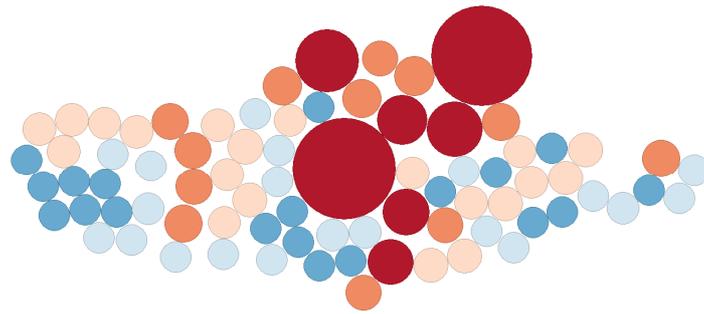
Mapa 56. Mapa de percentiles ASS14.



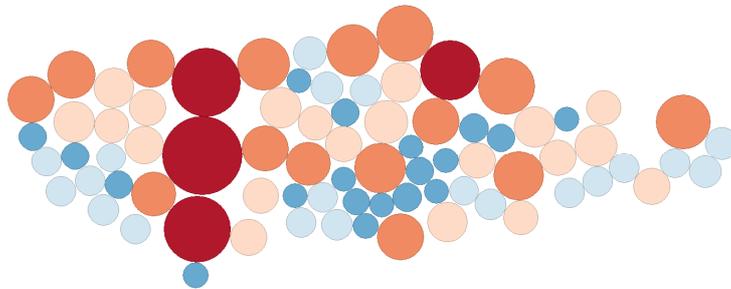
Mapa 57. Cartograma CONTRAT13.



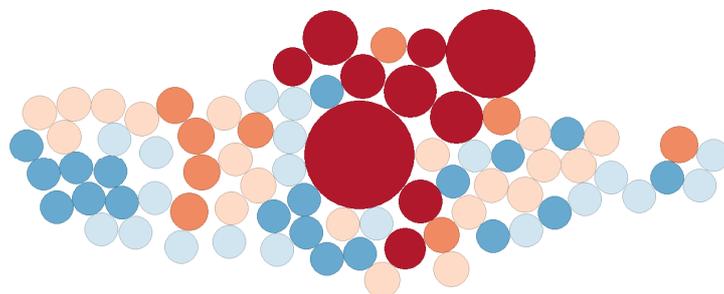
Mapa 58. Cartograma PAR14.



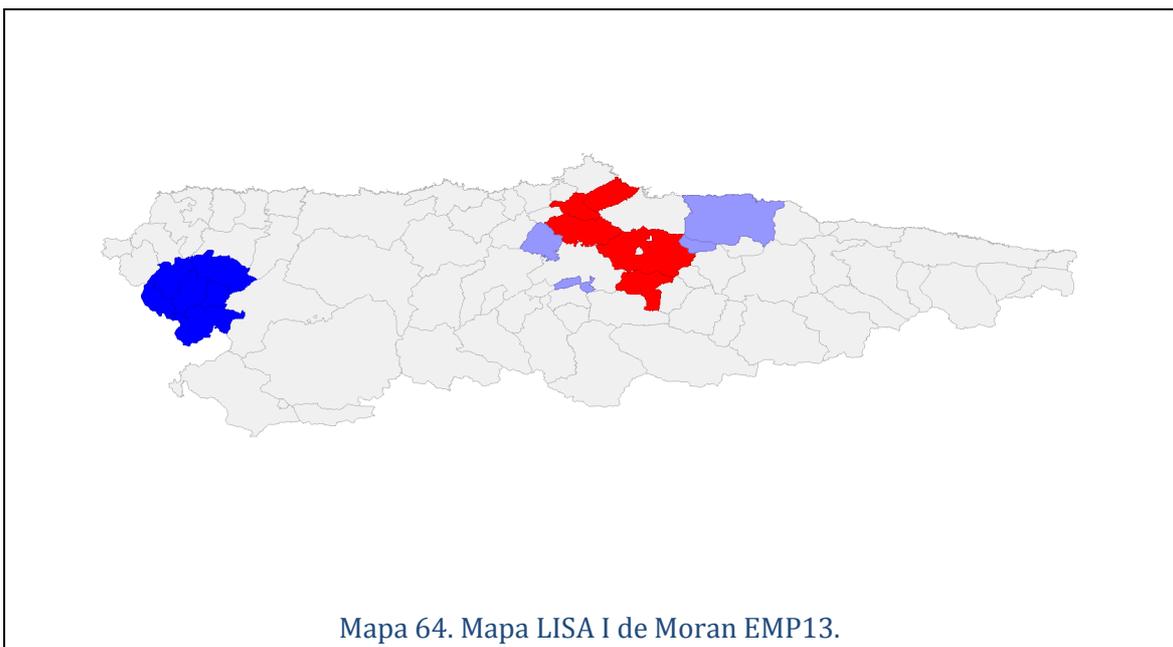
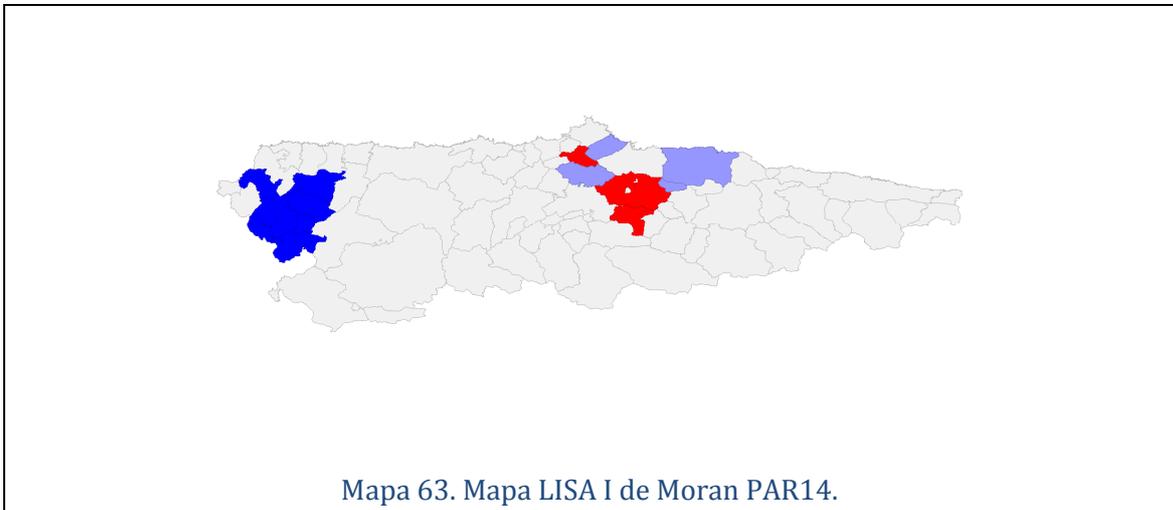
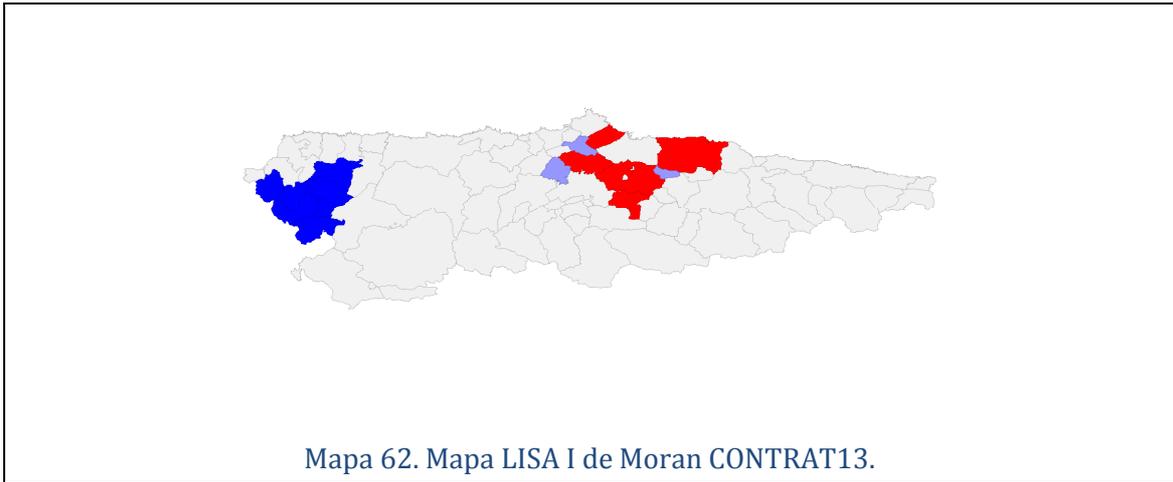
Mapa 59. Cartograma EMP13.



Mapa 60. Cartograma EMPAGRI13.

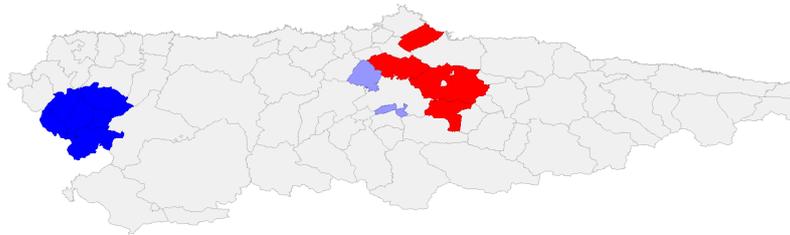


Mapa 61. Cartograma ASS14.

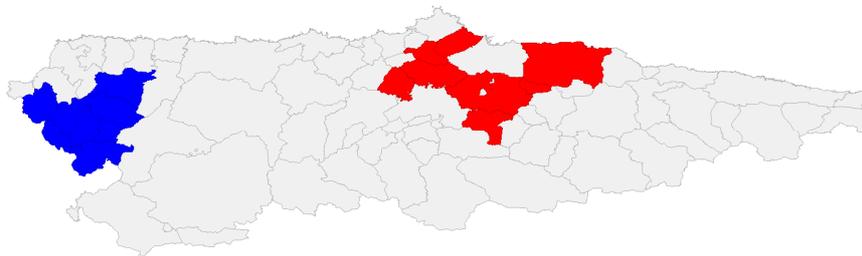




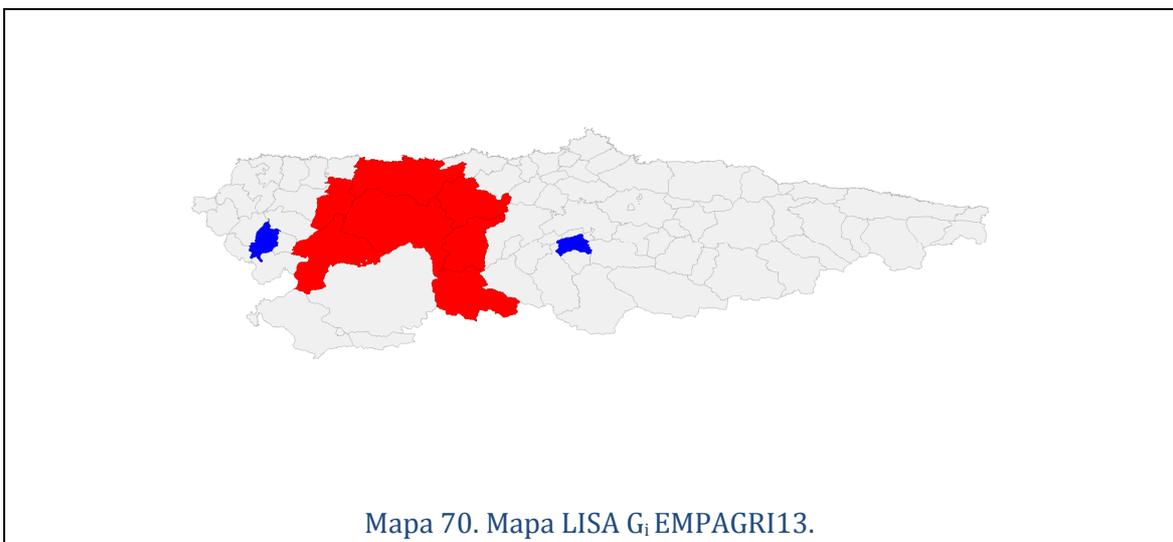
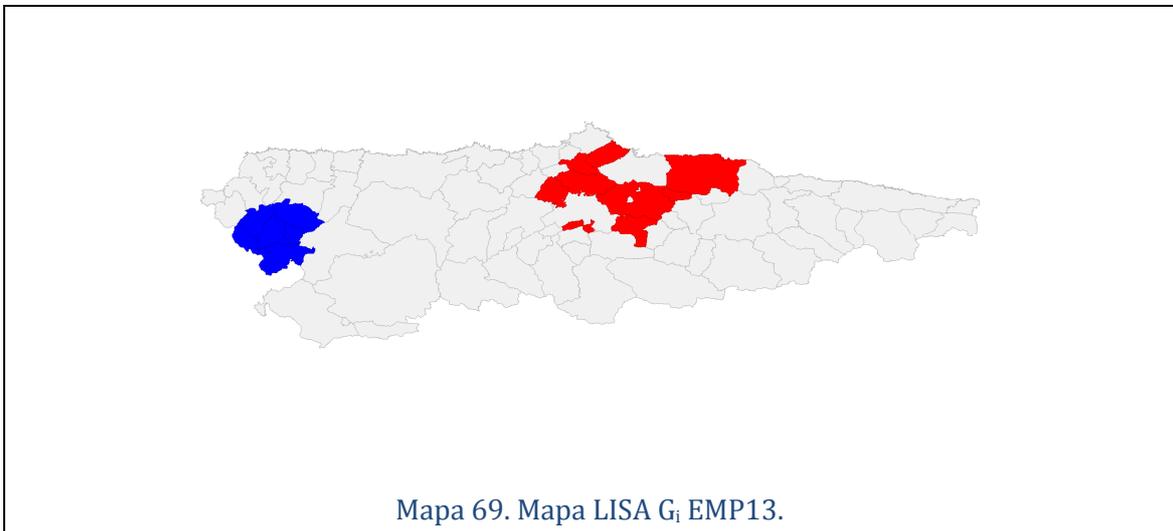
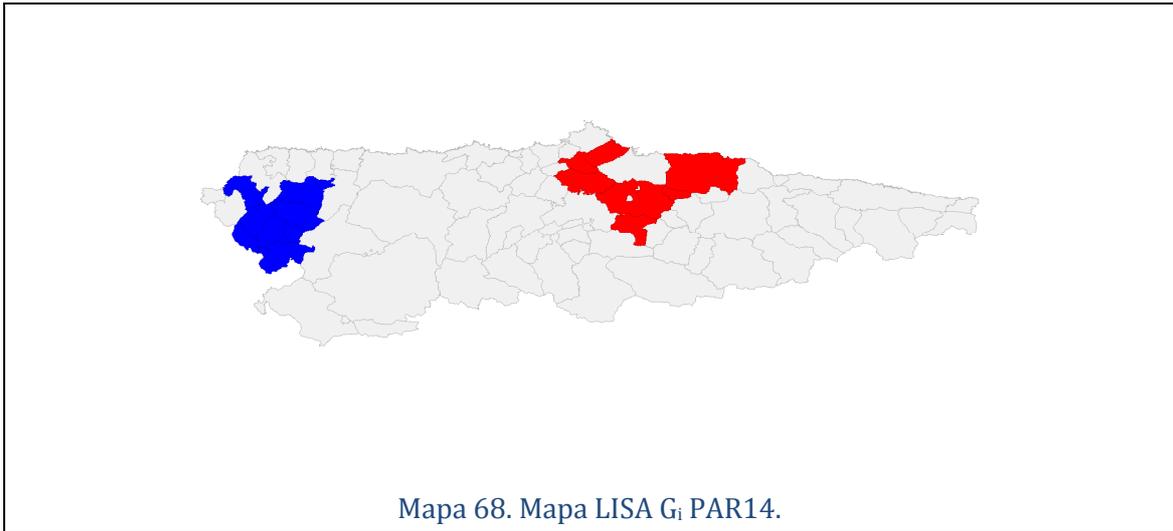
Mapa 65. Mapa LISA I de Moran EMPAGRI13.

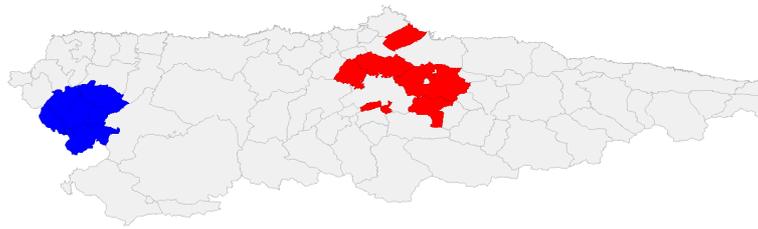


Mapa 66. Mapa LISA I de Moran ASS14.

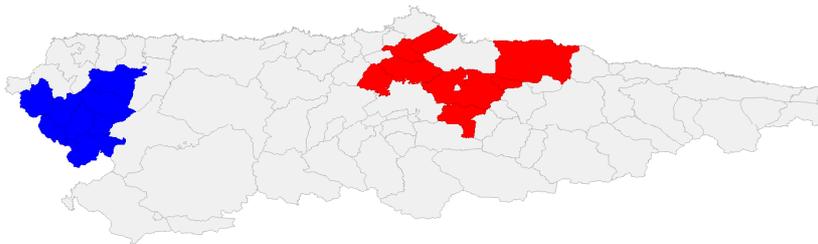


Mapa 67. Mapa LISA  $G_i$  CONTRAT13.





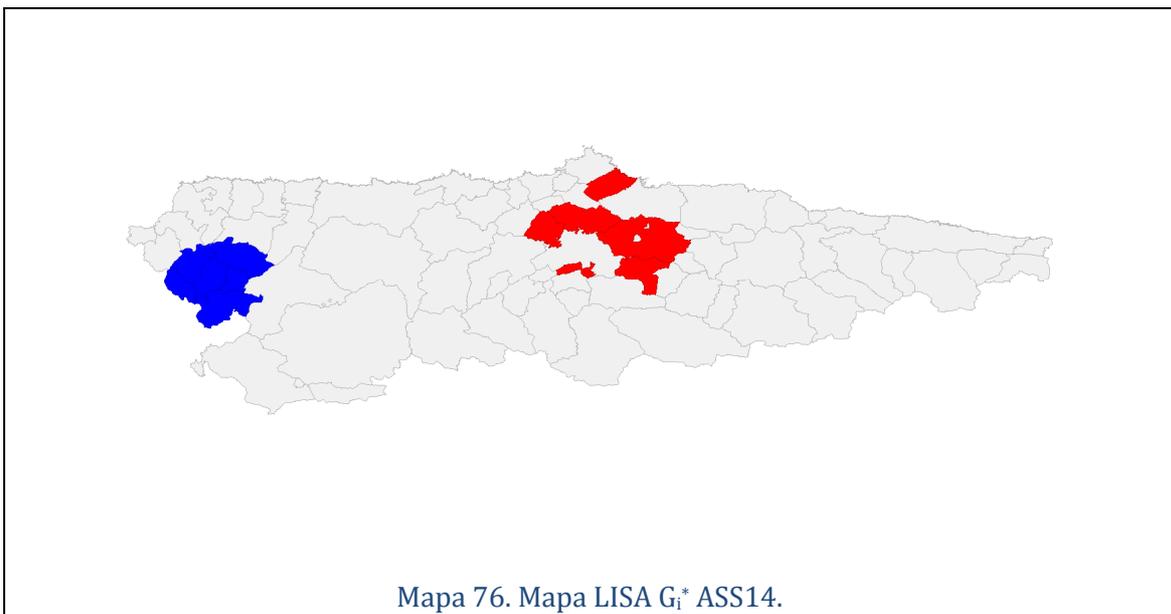
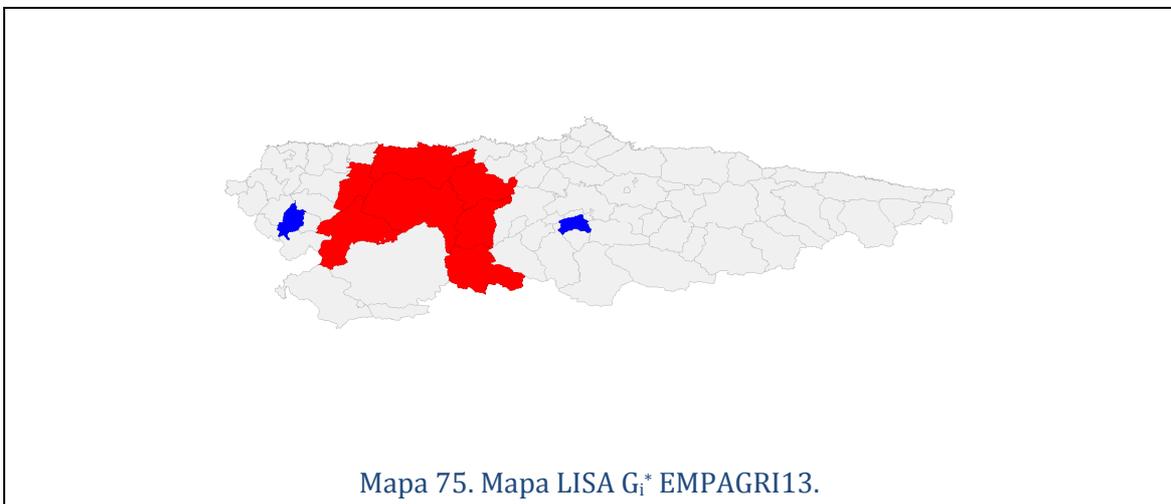
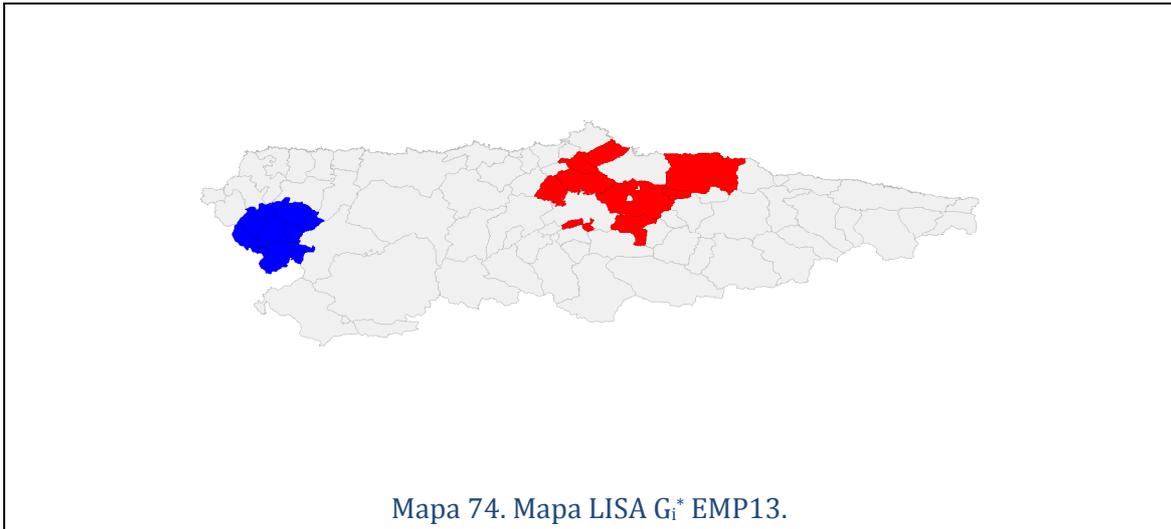
Mapa 71. Mapa LISA  $G_i$  ASS14



Mapa 72. Mapa LISA  $G_i^*$  CONTRAT13.



Mapa 73. Mapa LISA  $G_i^*$  PAR14.



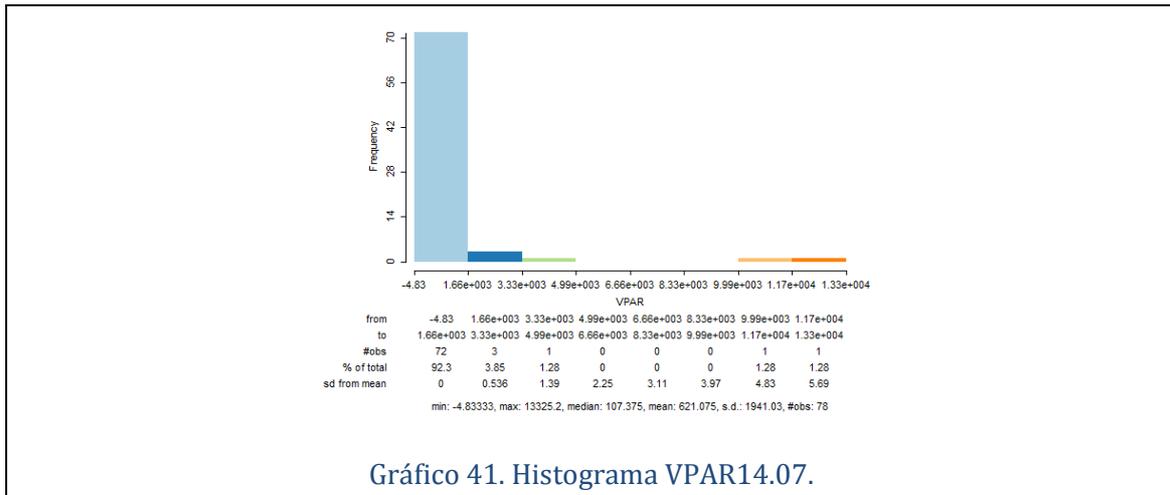


Gráfico 41. Histograma VPAR14.07.

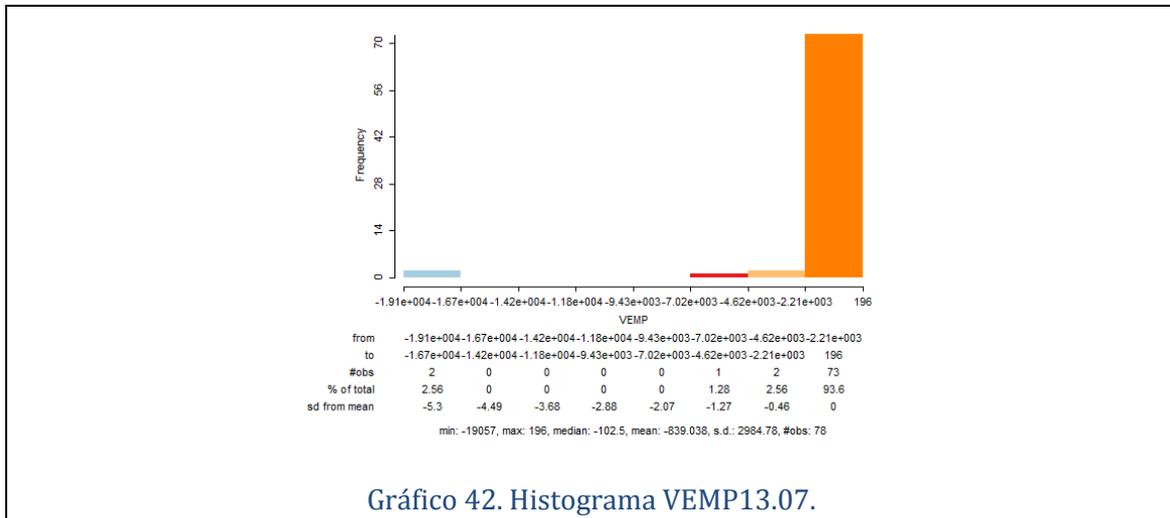


Gráfico 42. Histograma VEMP13.07.

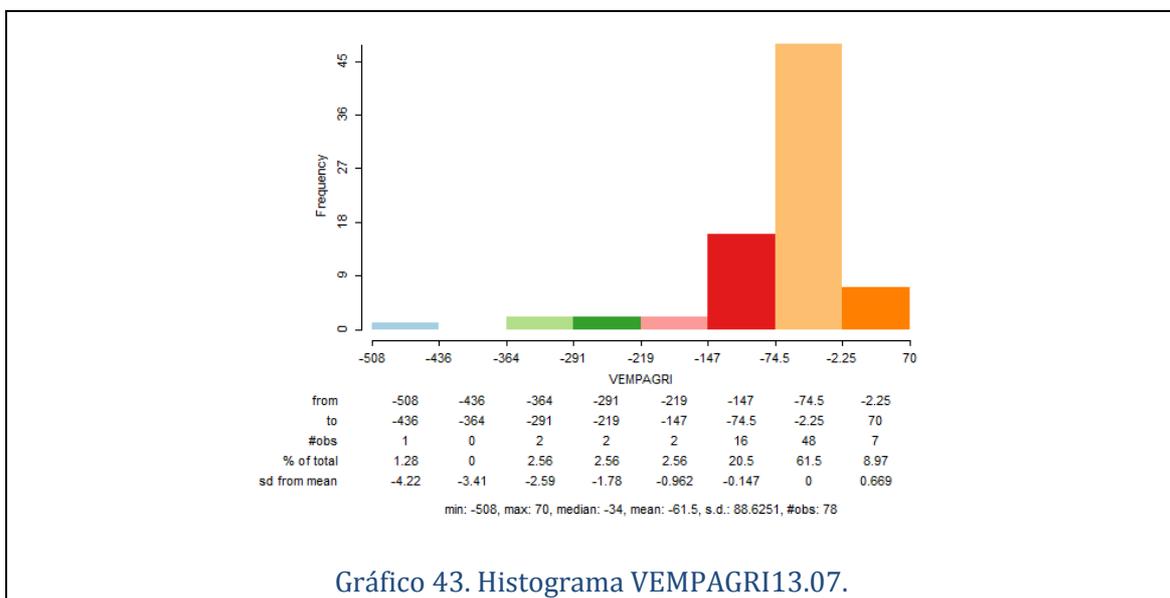
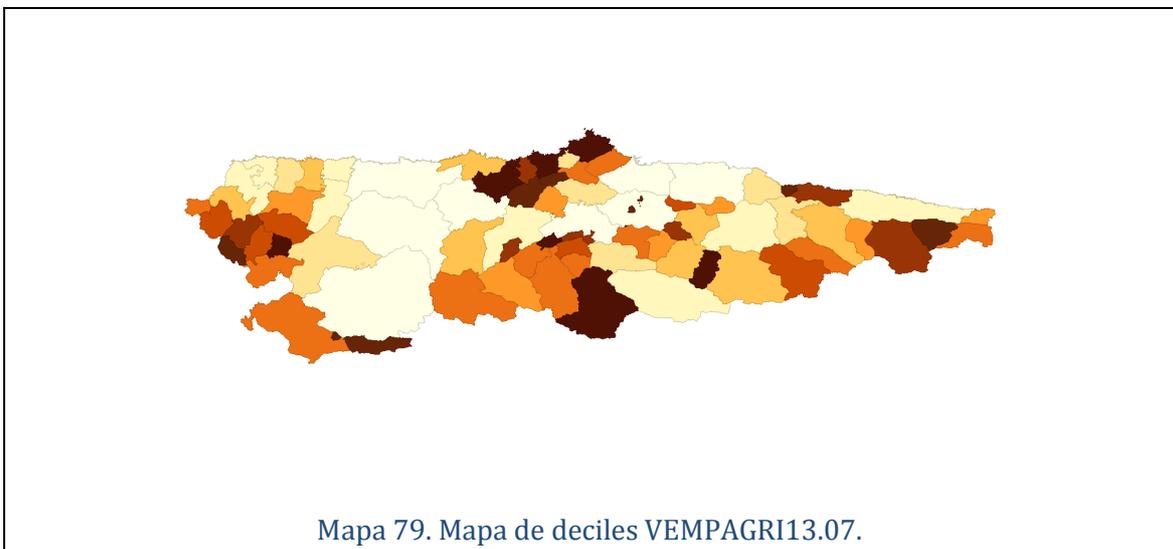
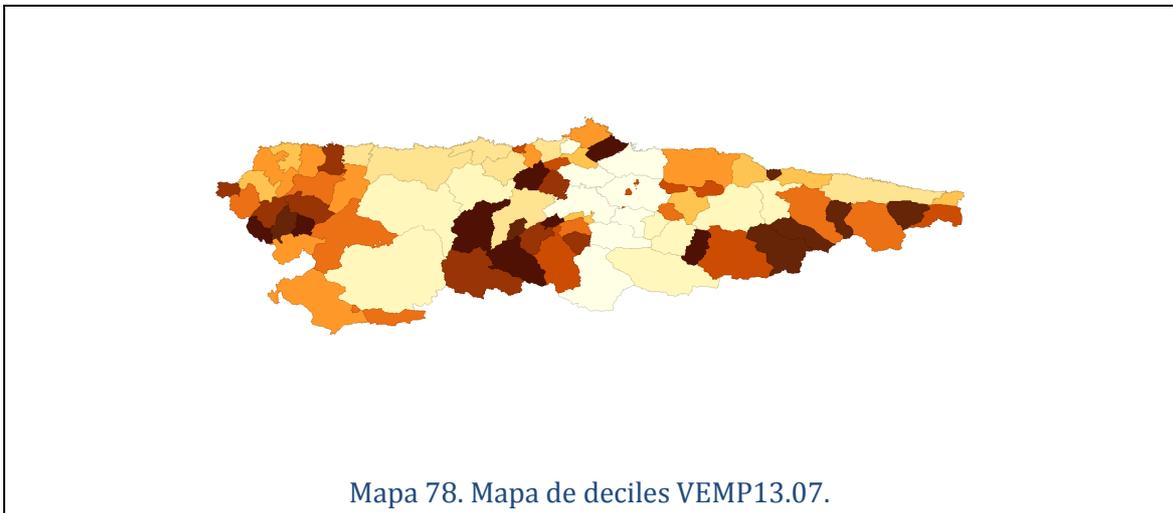
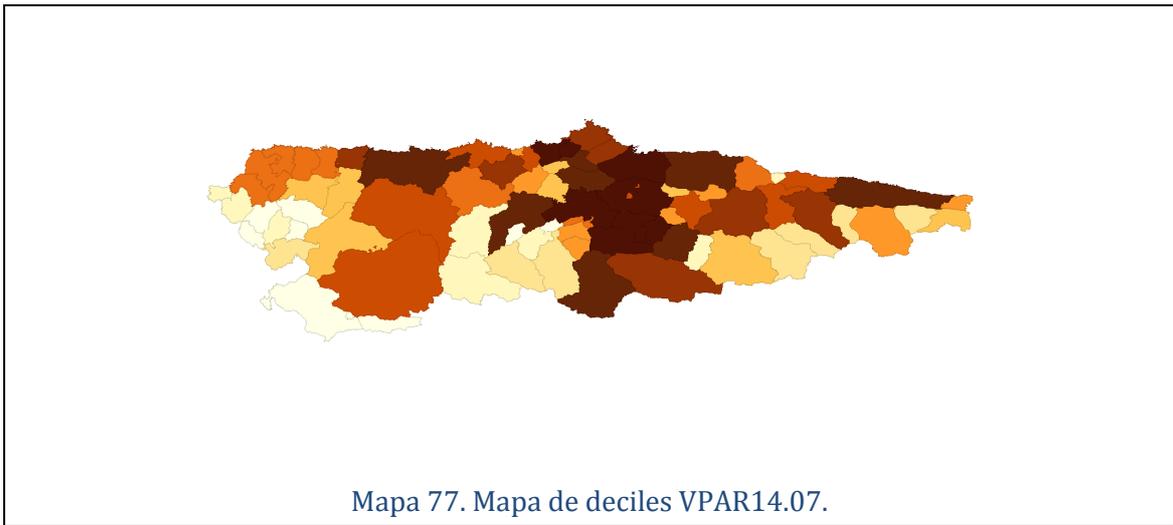


Gráfico 43. Histograma VEMPAGRI13.07.



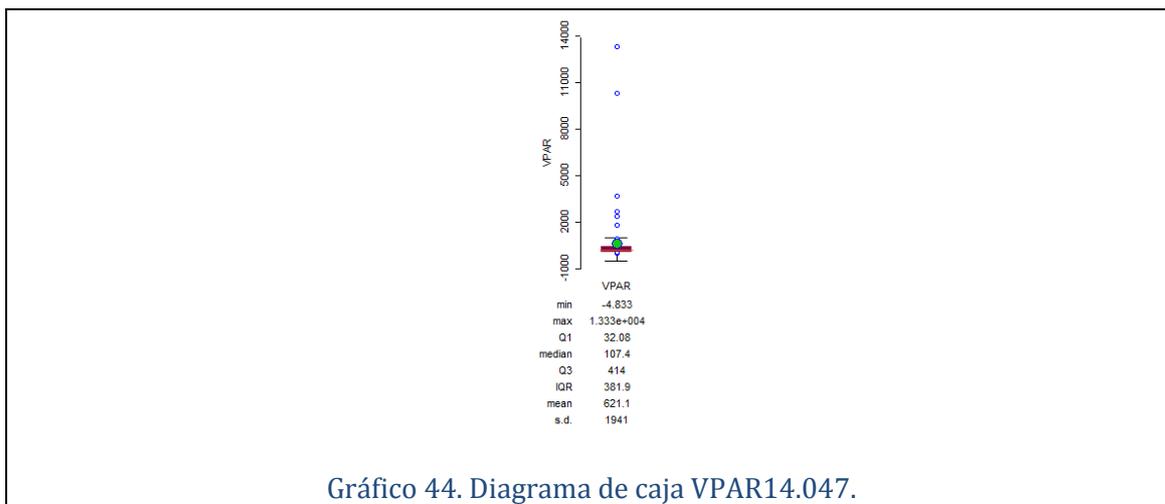


Gráfico 44. Diagrama de caja VPAR14.047.

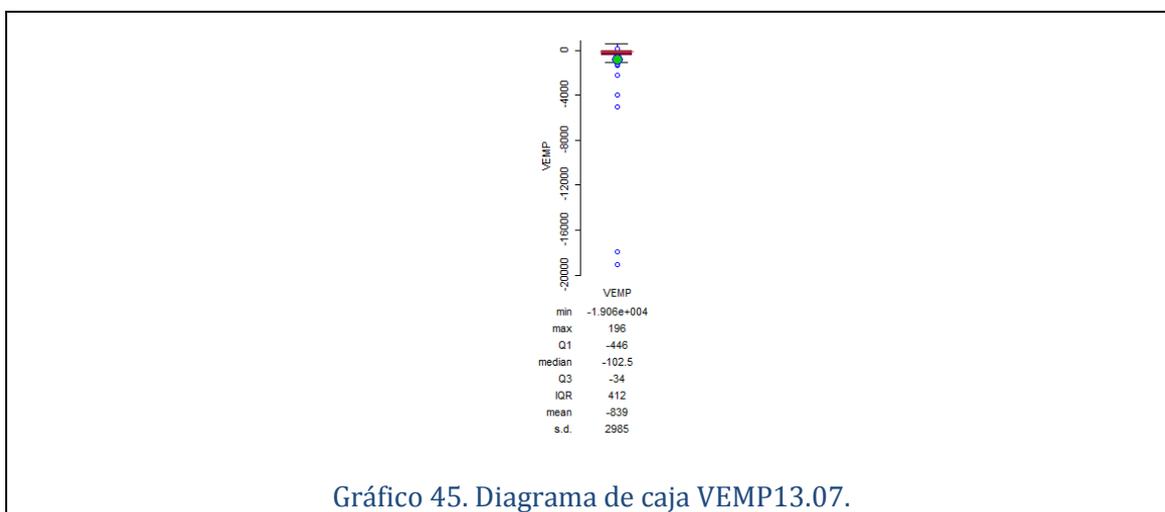


Gráfico 45. Diagrama de caja VEMP13.07.

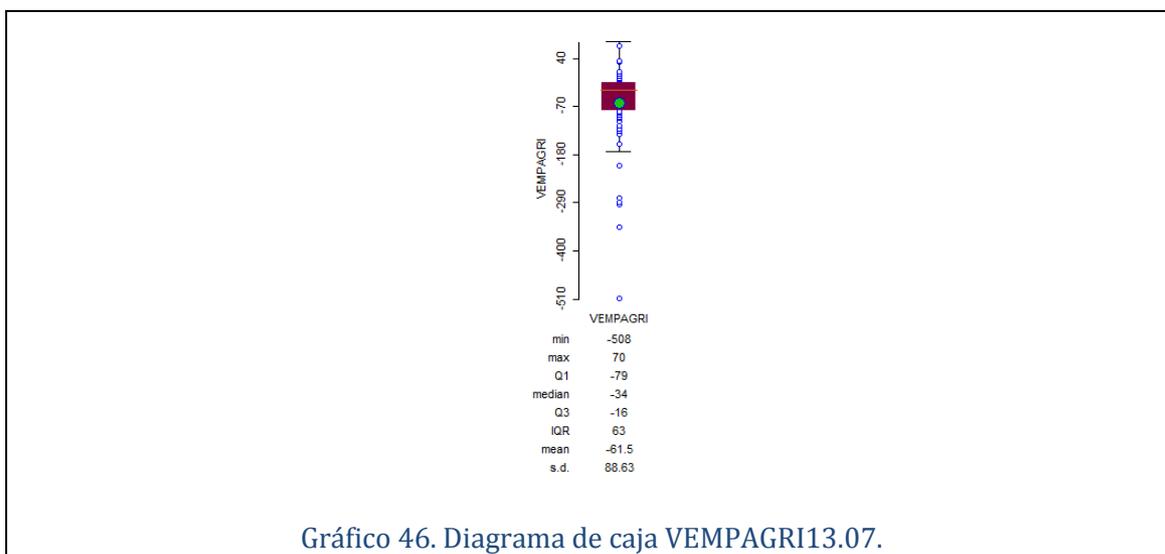
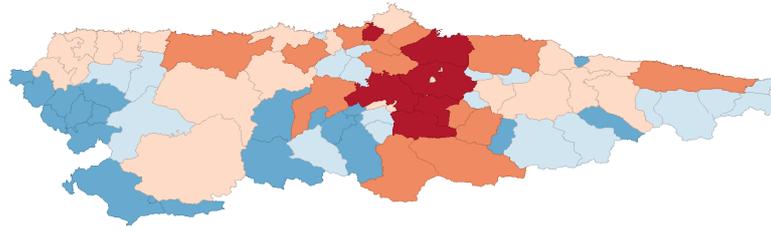
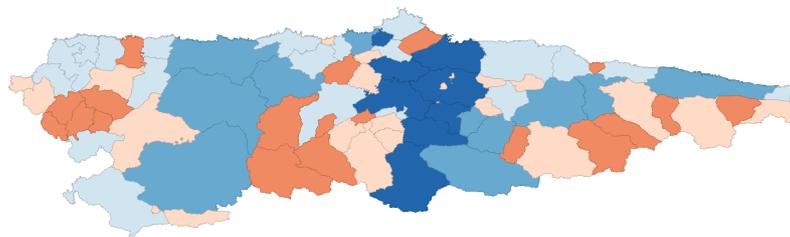


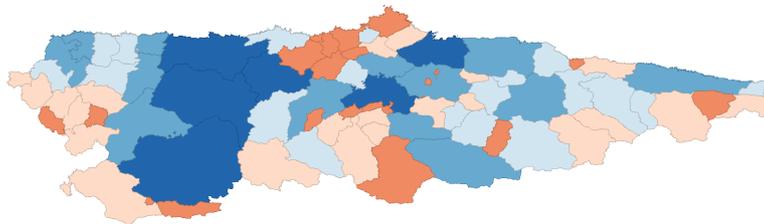
Gráfico 46. Diagrama de caja VEMPAGRI13.07.



Mapa 80. Mapa de caja VPAR14.07.



Mapa 81. Mapa de caja VEMP13.07.



Mapa 82. Mapa de caja VEMPAGRI13.07.

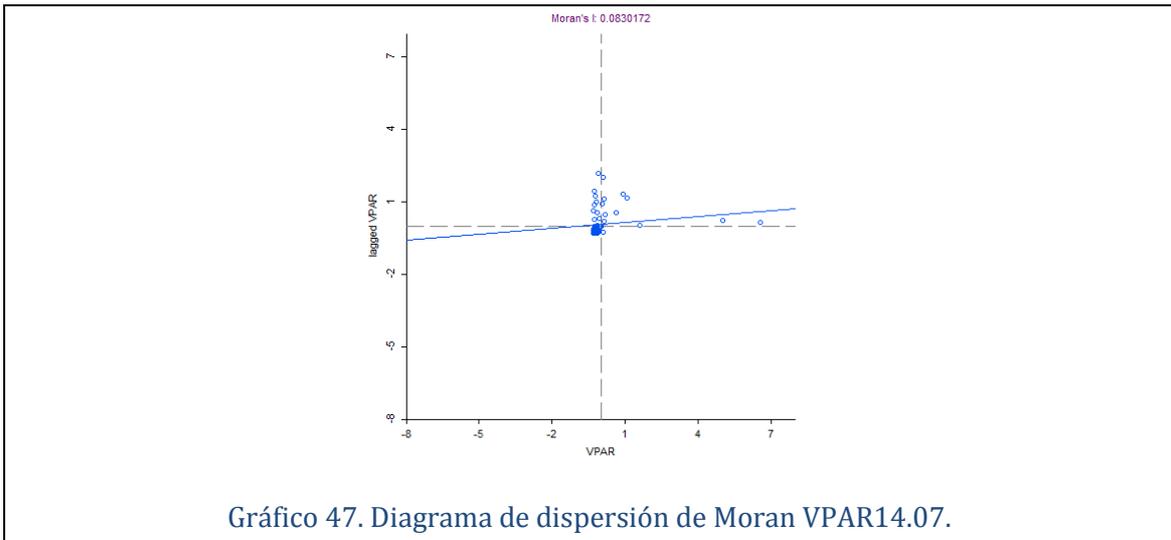


Gráfico 47. Diagrama de dispersión de Moran VPAR14.07.

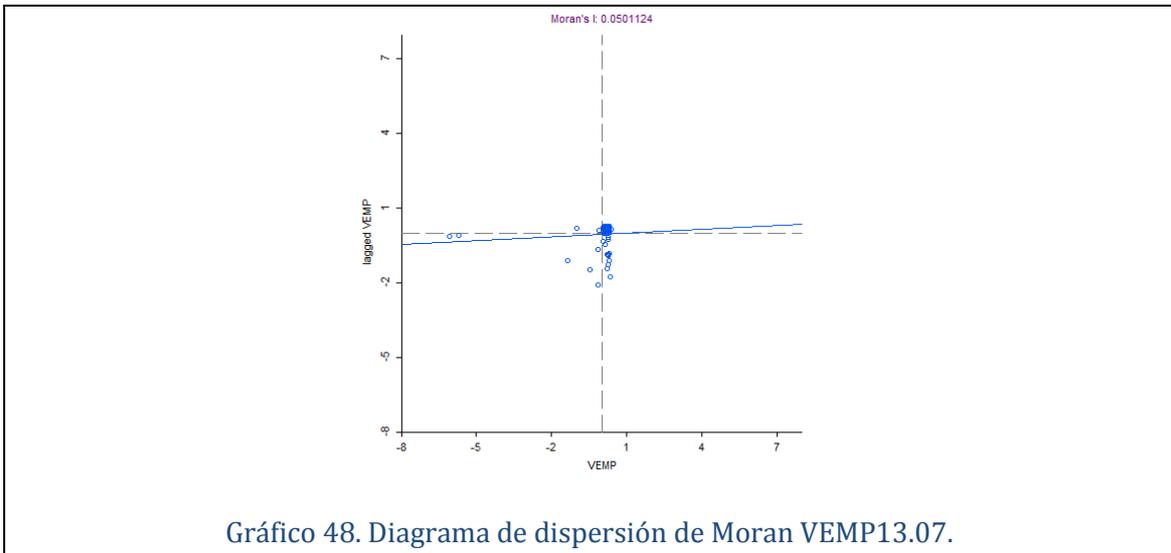


Gráfico 48. Diagrama de dispersión de Moran VEMP13.07.

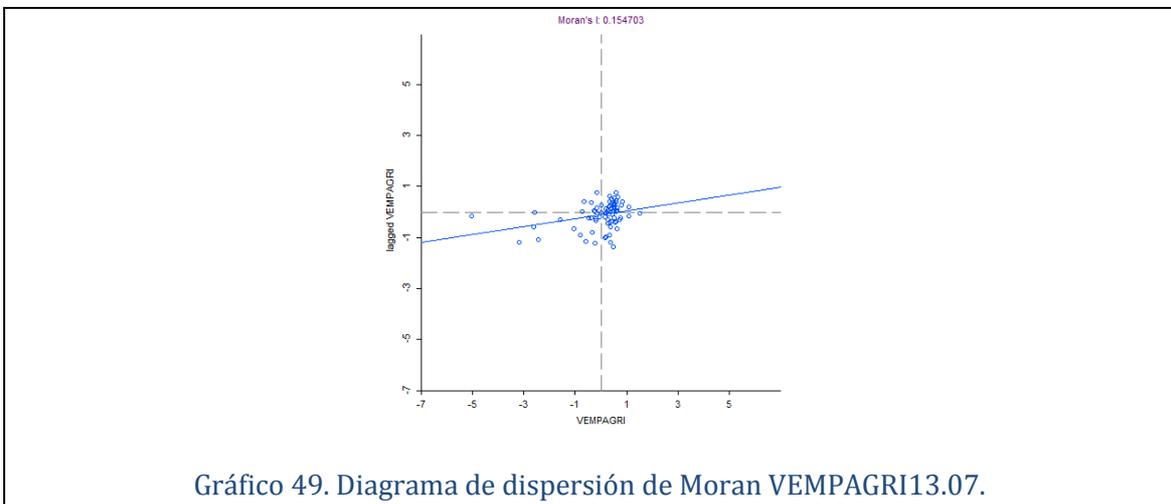


Gráfico 49. Diagrama de dispersión de Moran VEMPAGRI13.07.

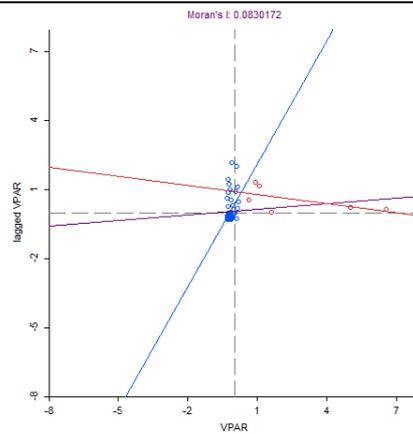


Gráfico 50. Diagrama de dispersión de Moran VPAR14.07 con las rectas de regresión que eliminan valores atípicos.

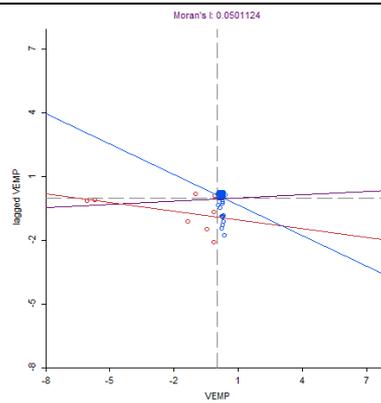


Gráfico 51. Diagrama de dispersión de Moran VEMP13.07 con las rectas de regresión que eliminan valores atípicos.

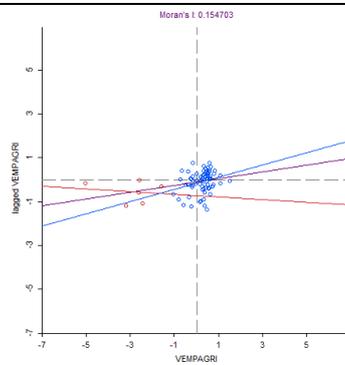
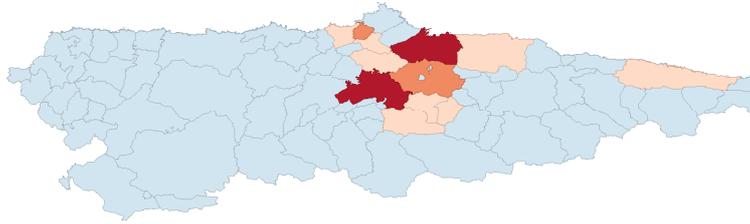
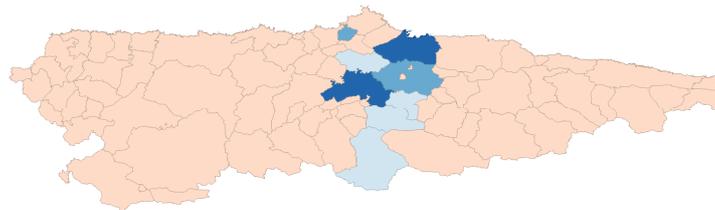


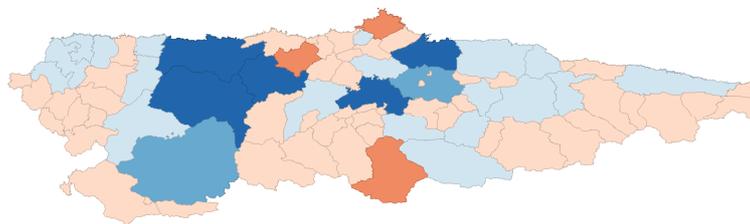
Gráfico 52. Diagrama de dispersión de Moran VEMPAGRI13.07 con las rectas de regresión que eliminan valores atípicos.



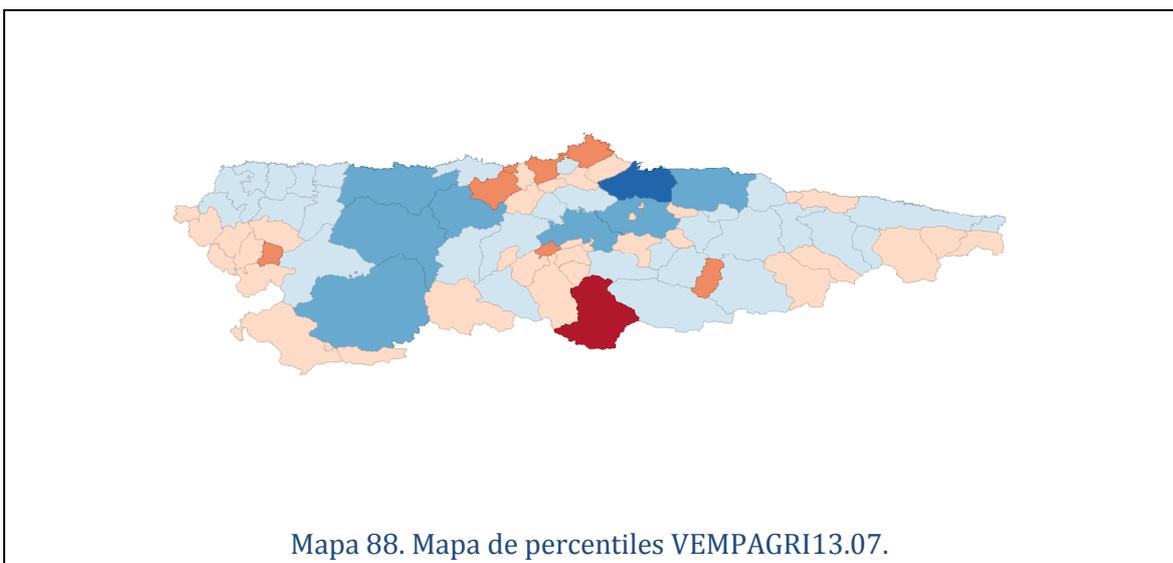
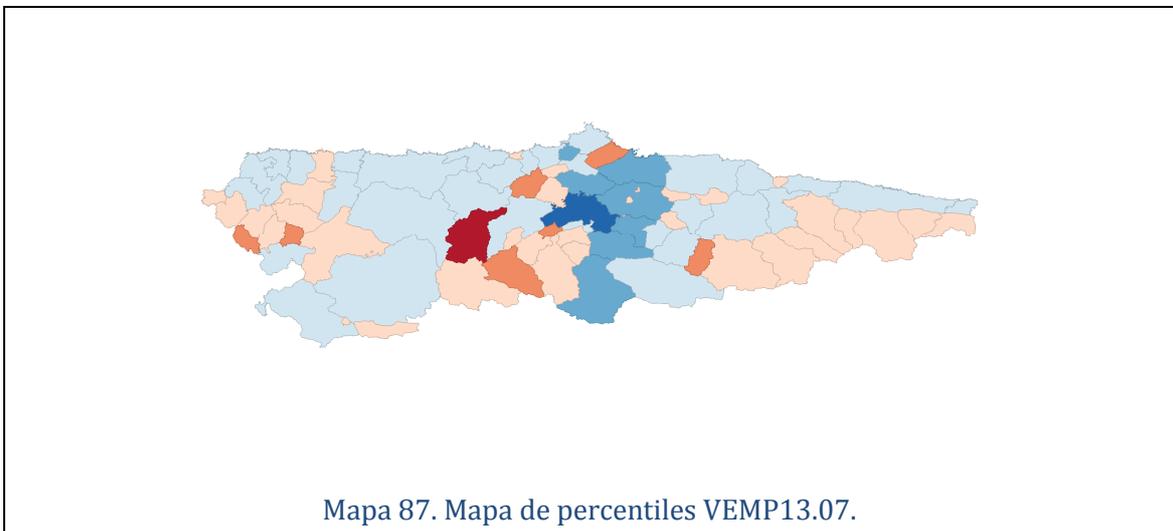
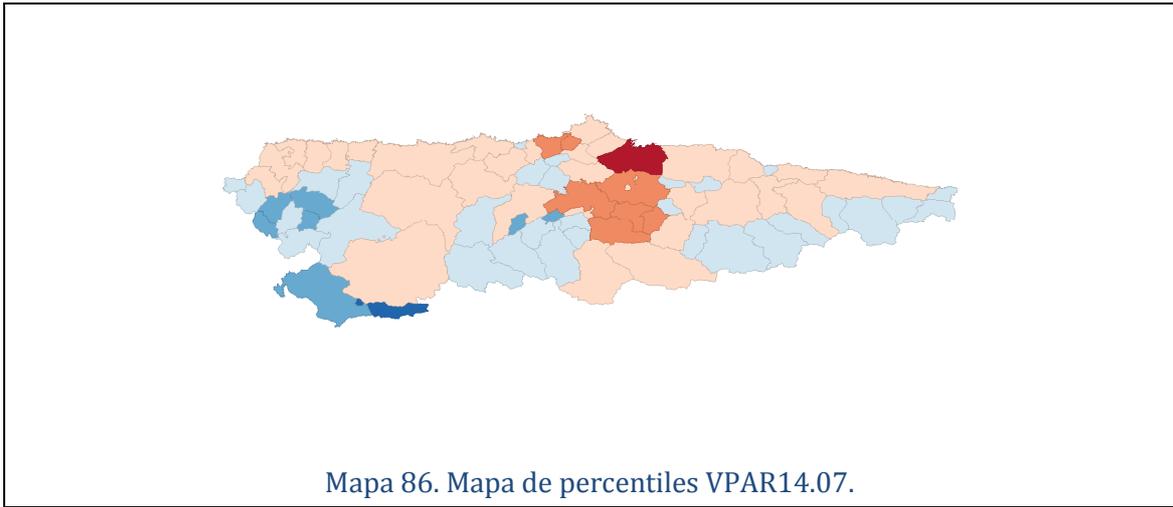
Mapa 83. Mapa de desviación típica VPAR14.07.

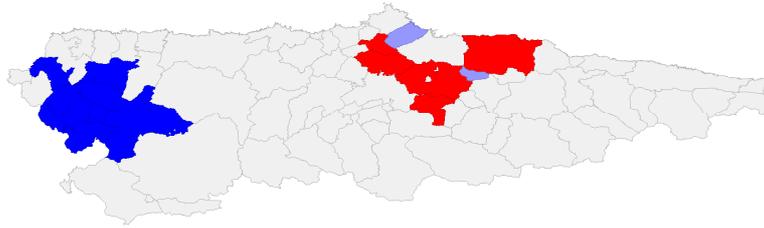


Mapa 84. Mapa de desviación típica VEMP13.07.

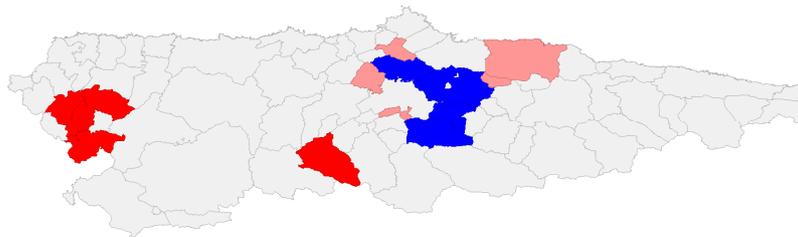


Mapa 85. Mapa de desviación típica VEMPAGRI13.07.

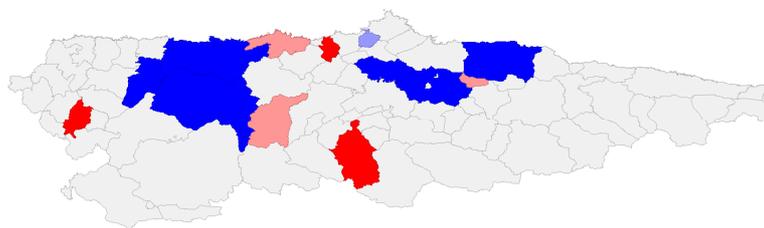




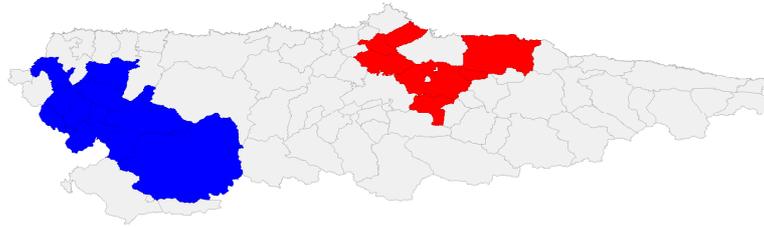
Mapa 89. Mapa LISA I de Moran VPAR14.07.



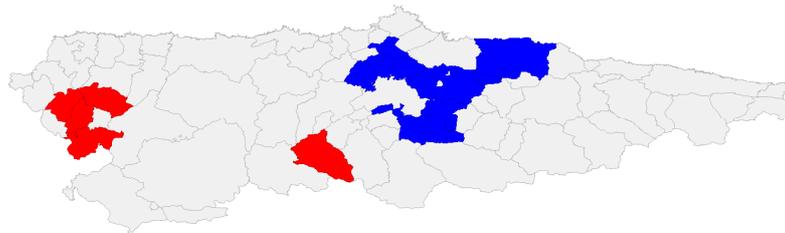
Mapa 90. Mapa LISA I de Moran VEMP13.07.



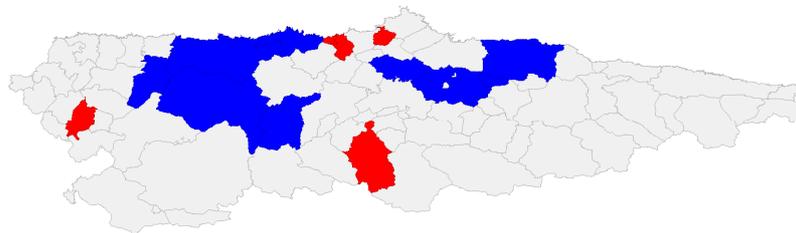
Mapa 91. Mapa LISA I de Moran VEMAGRI13.07.



Mapa 92. Mapa LISA  $G_i$  VPAR14.07.



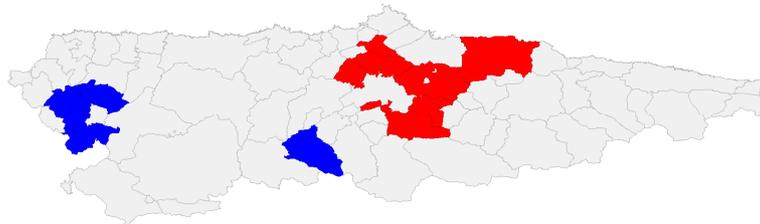
Mapa 93. Mapa LISA  $G_i$  VEMP13.07.



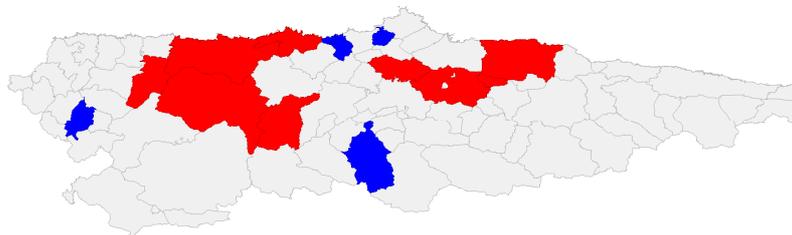
Mapa 94. Mapa LISA  $G_i$  VEMPAGRI13.07.



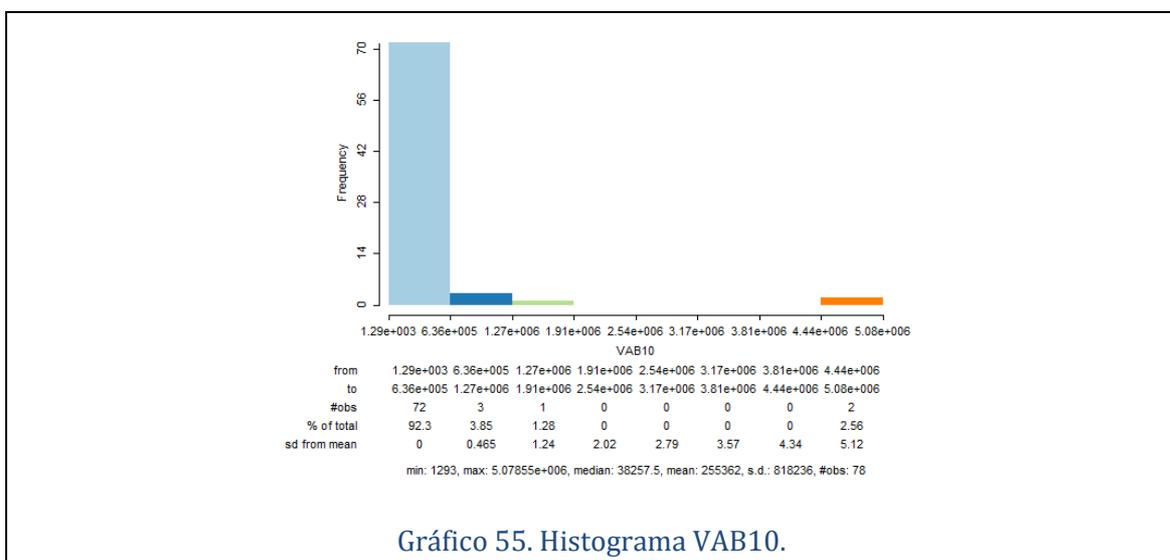
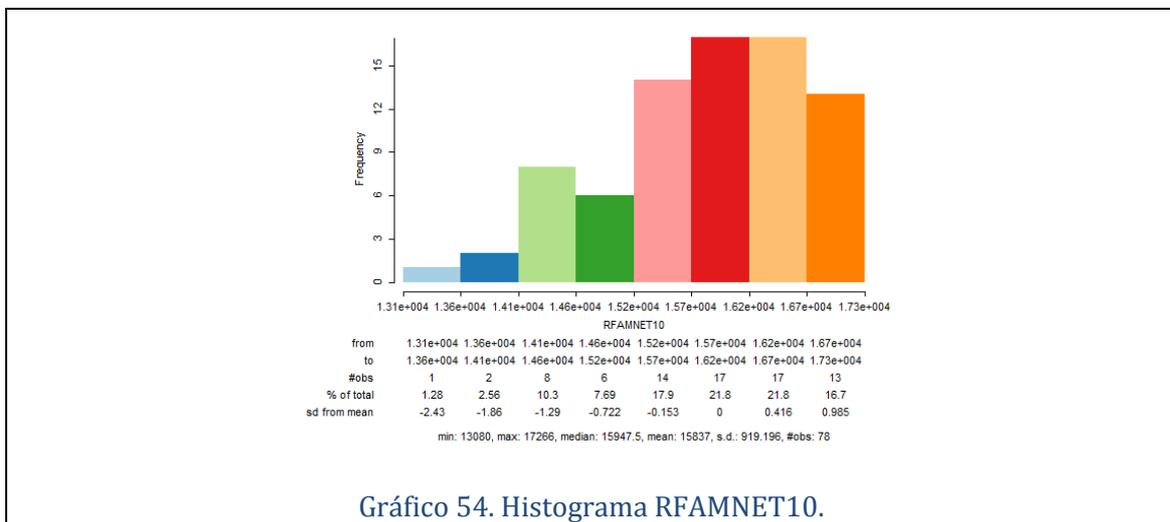
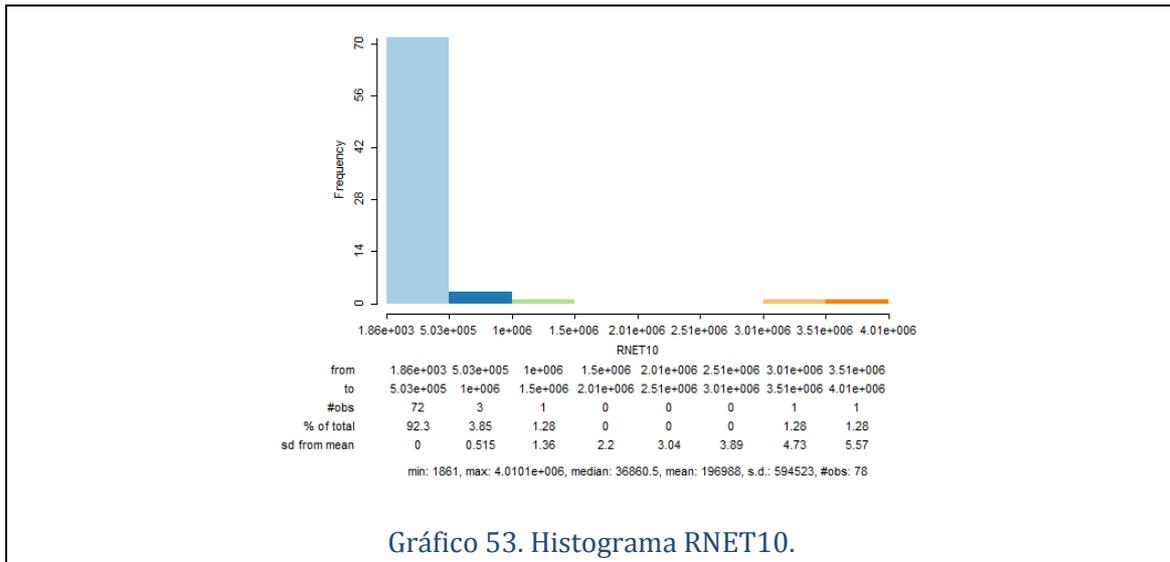
Mapa 95. Mapa LISA  $G_i^*$  VPAR14.07.



Mapa 96. Mapa LISA  $G_i^*$  VEMP13.07.



Mapa 97. Mapa LISA  $G_i^*$  VEMPAGRI13.07.



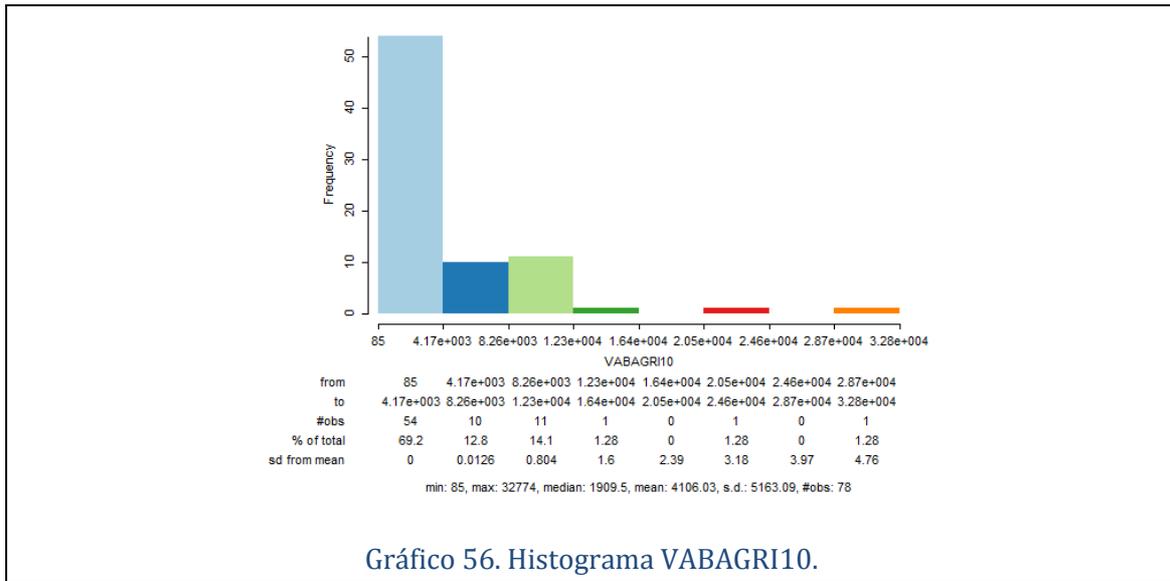
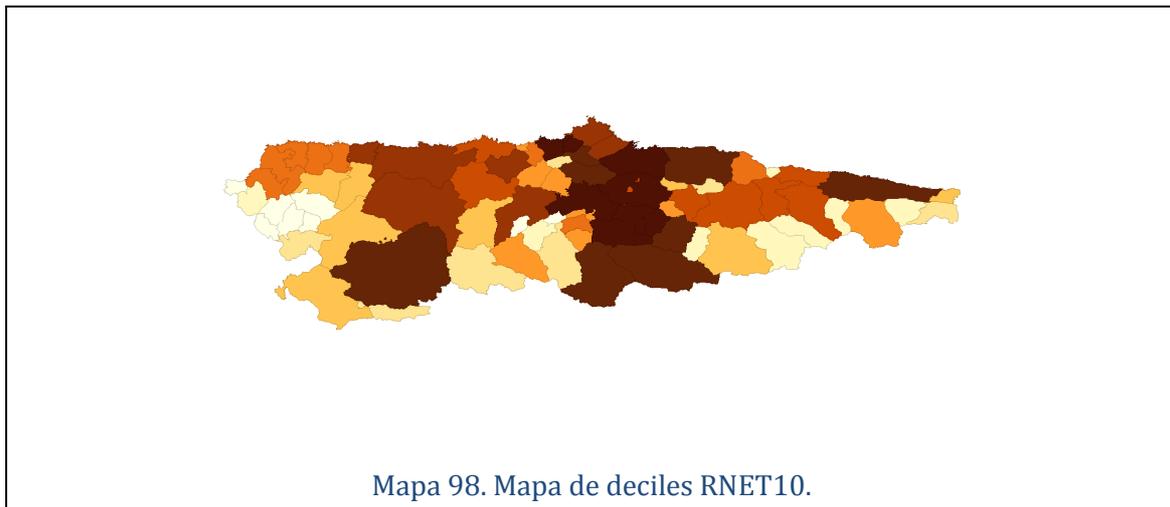
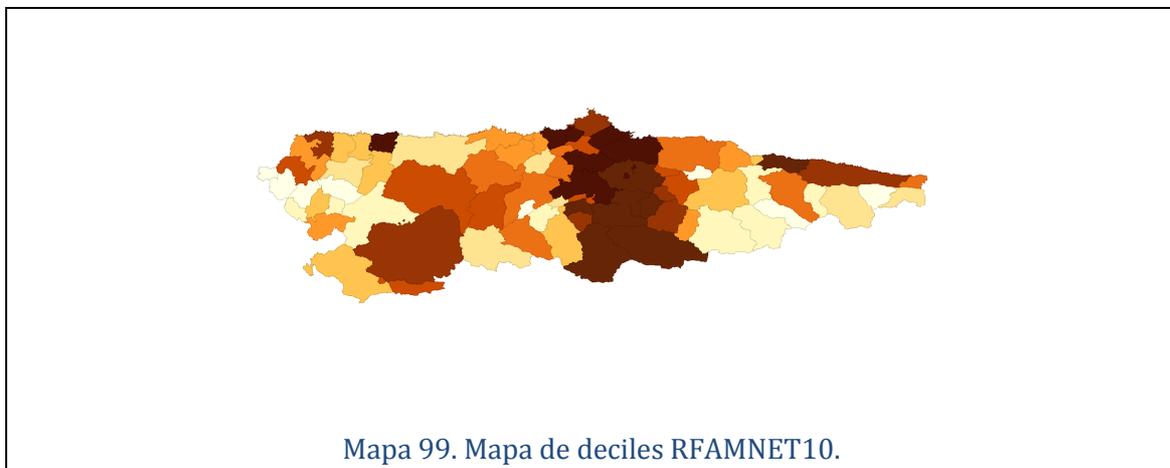


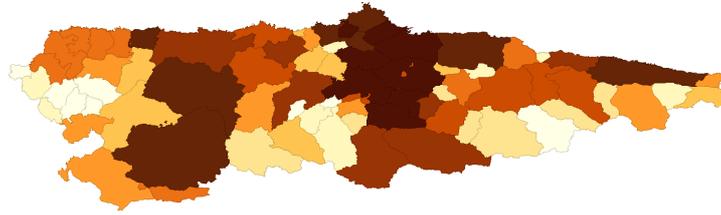
Gráfico 56. Histograma VABAGRI10.



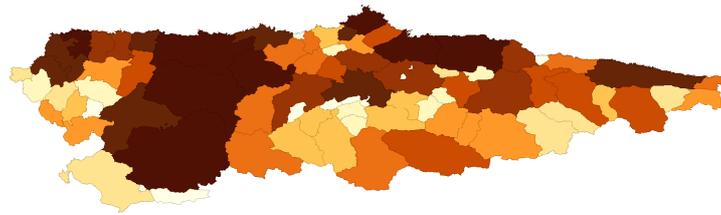
Mapa 98. Mapa de deciles RNET10.



Mapa 99. Mapa de deciles RFAMNET10.



Mapa 100. Mapa de deciles VAB10.



Mapa 101. Mapa de deciles VABAGRI10.

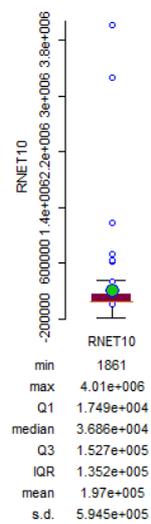
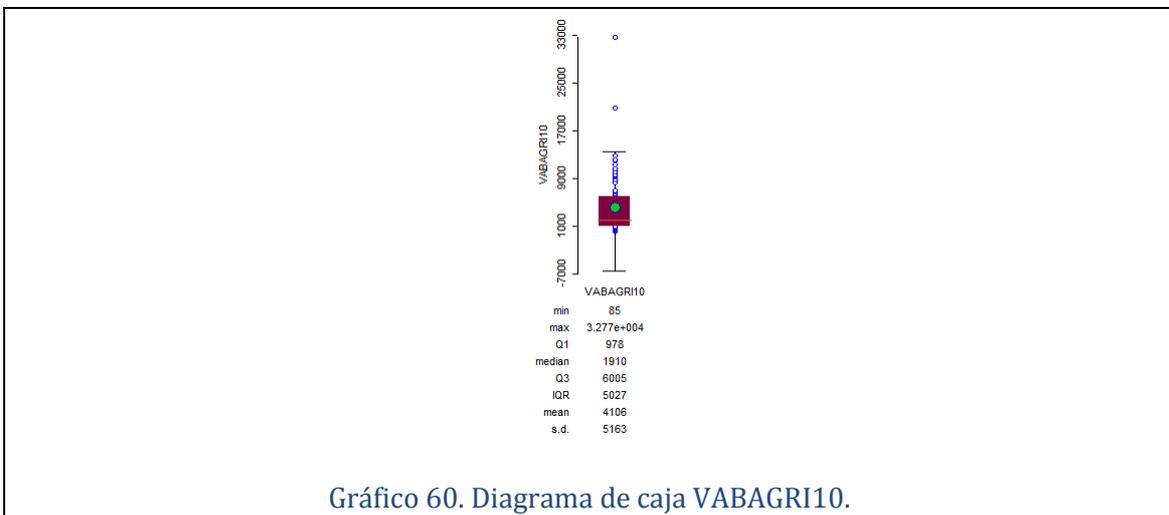
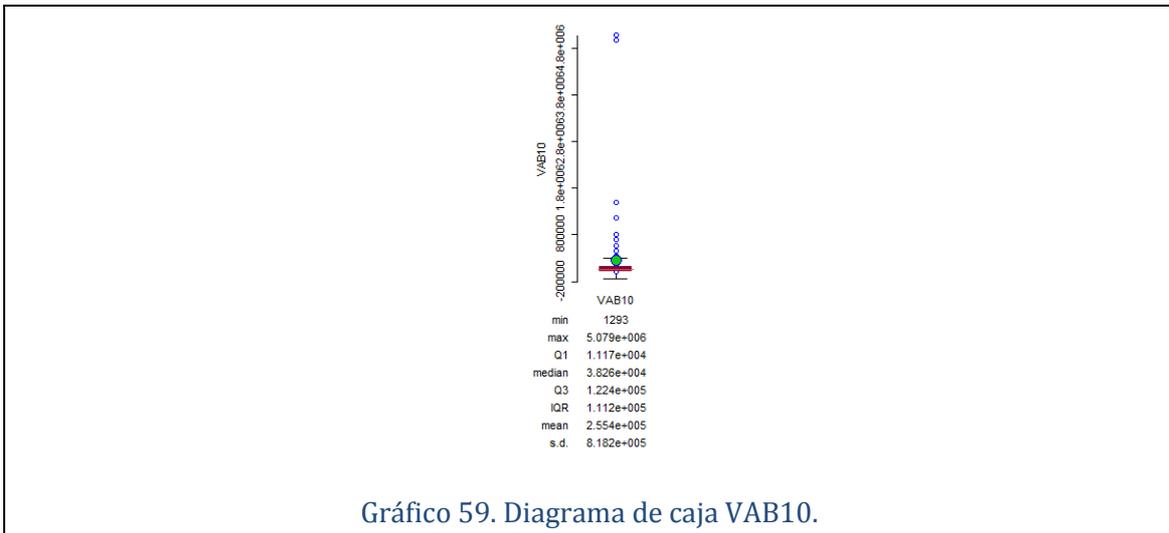
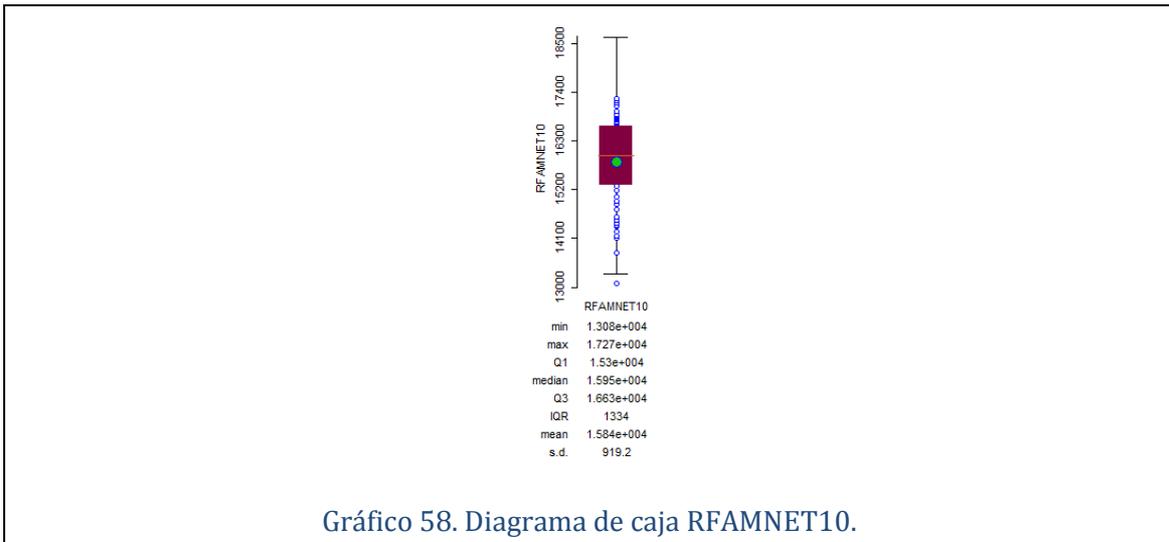
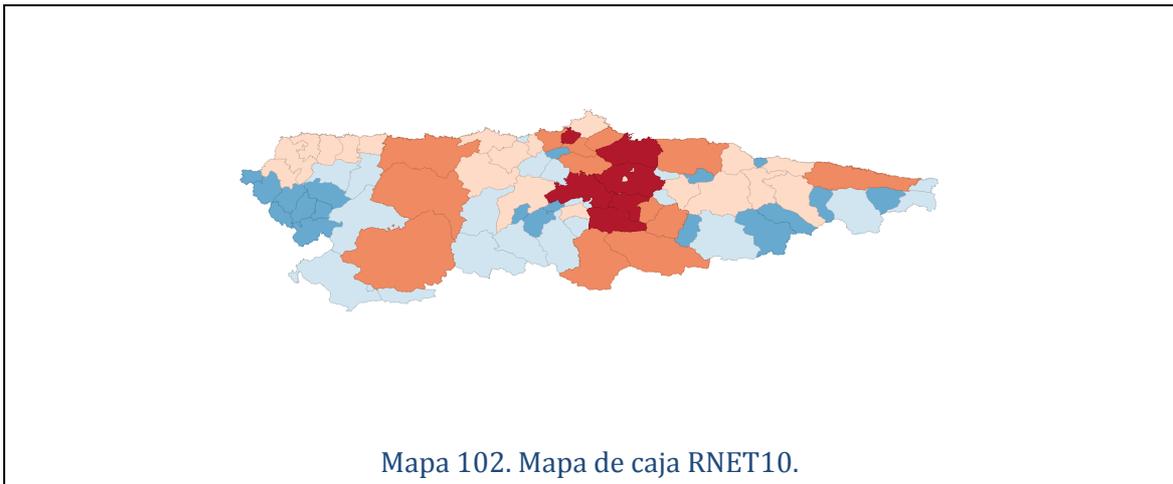
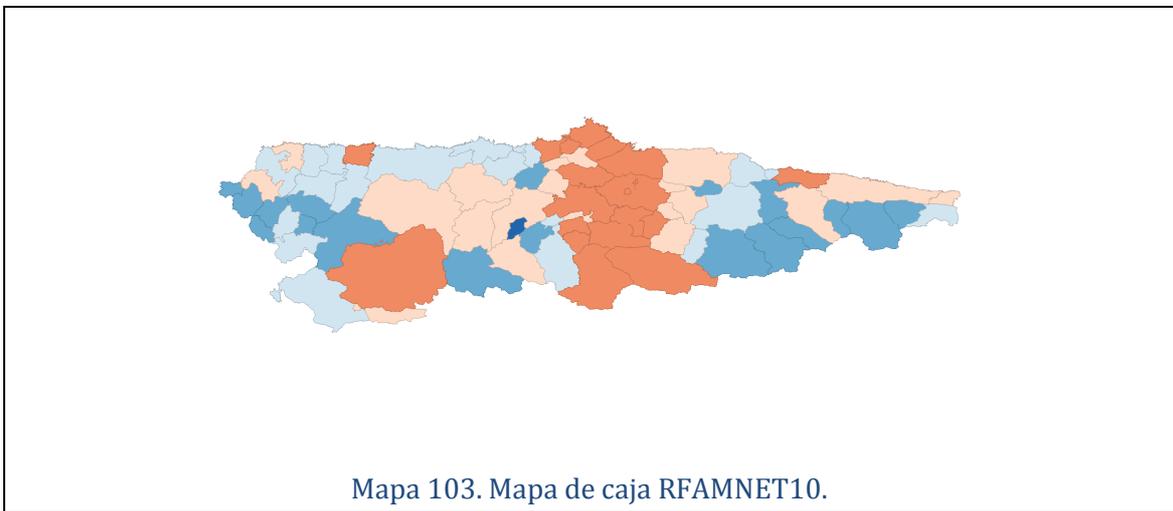


Gráfico 57. Diagrama de caja RNET10.

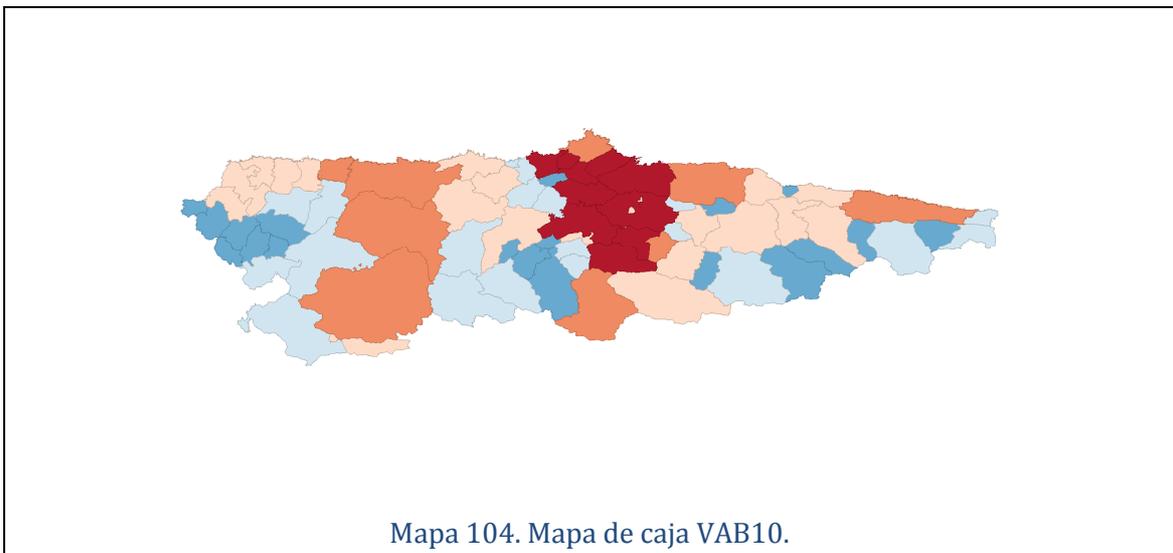




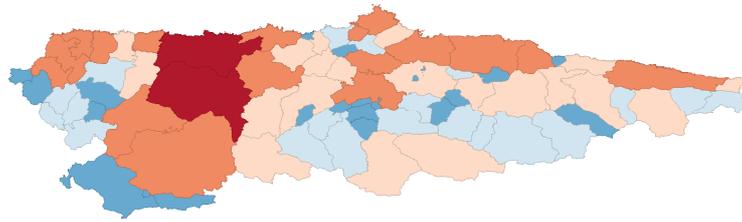
Mapa 102. Mapa de caja RNET10.



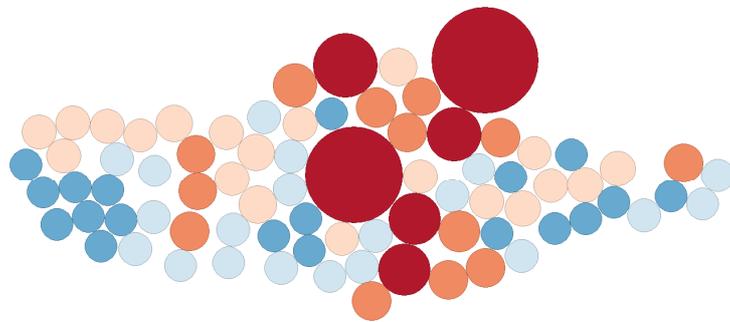
Mapa 103. Mapa de caja RFAMNET10.



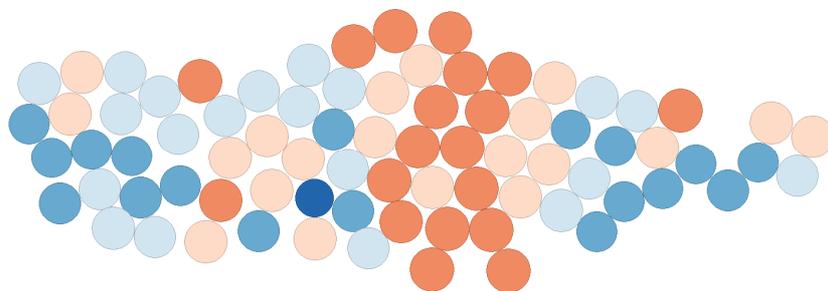
Mapa 104. Mapa de caja VAB10.



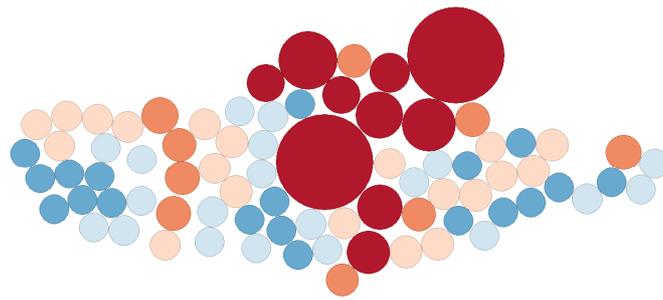
Mapa 105. Mapa de caja VABAGRI10.



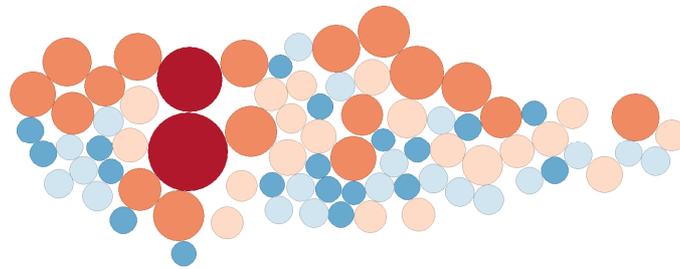
Mapa 106. Cartograma RNET10.



Mapa 107. Cartograma RFAMNET10.



Mapa 108. Cartograma VAB10.



Mapa 109. Cartograma VABAGRI10.

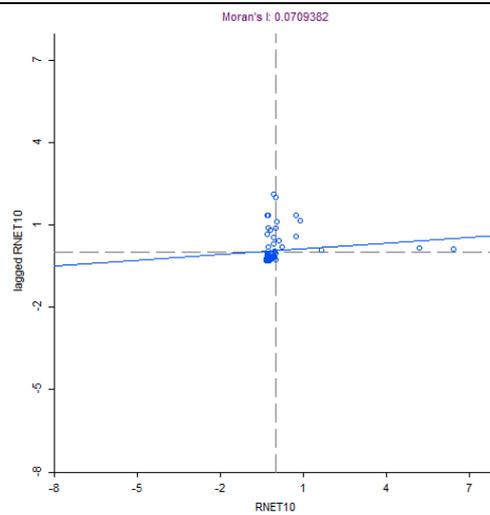
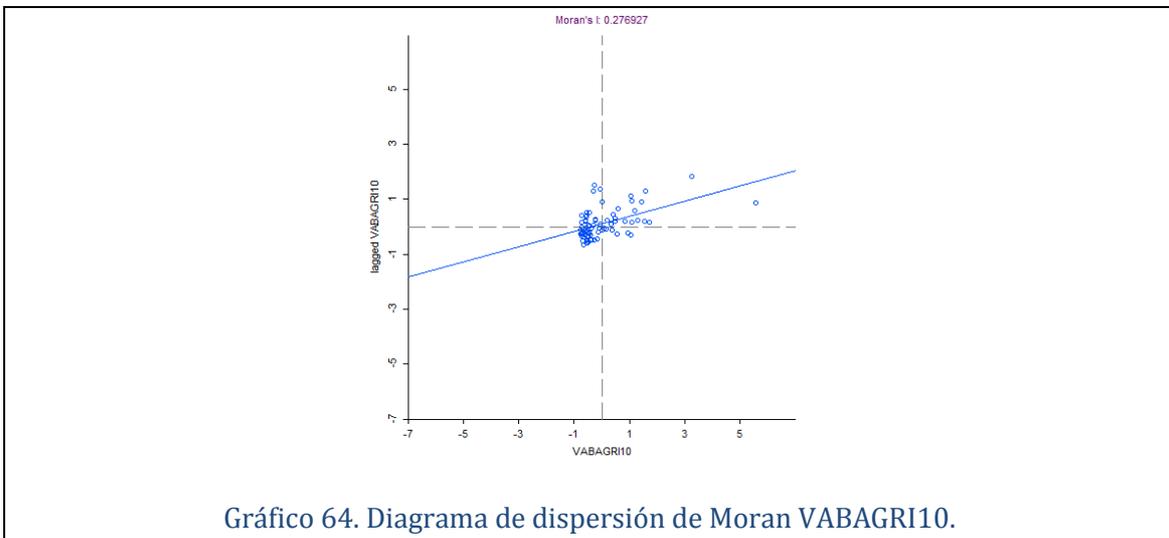
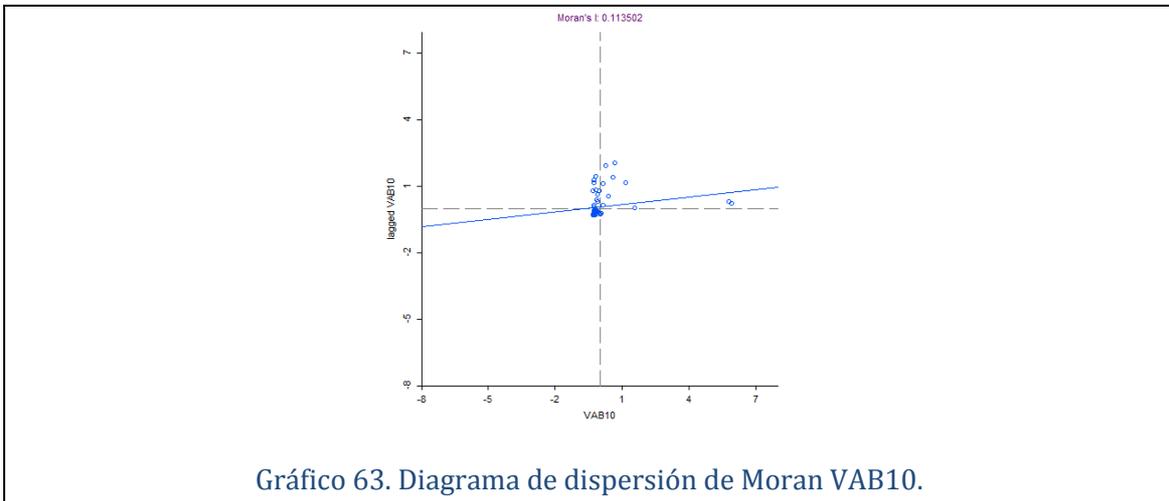
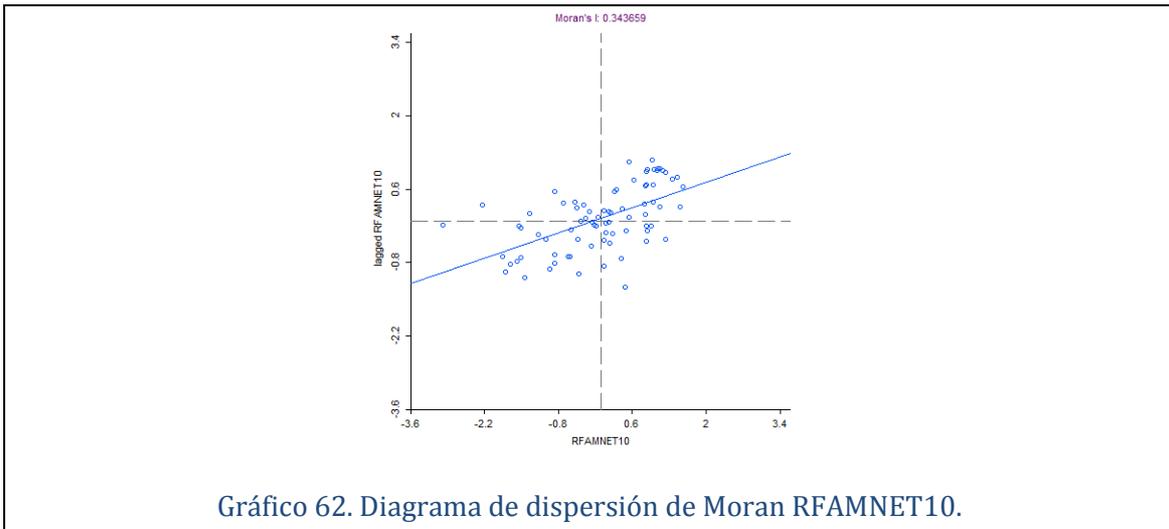


Gráfico 61. Diagrama de dispersión de Moran RNET10.



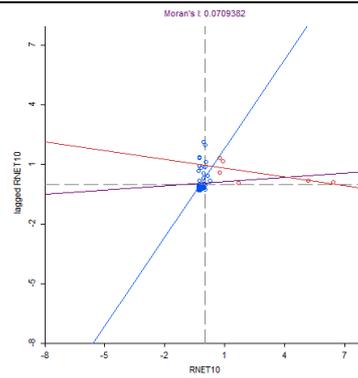


Gráfico 65. Diagrama de dispersión de Moran RNET10 con las rectas de regresión que eliminan valores atípicos.

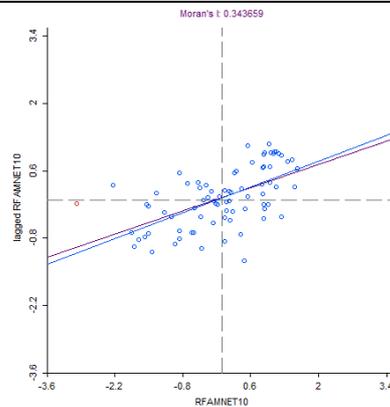


Gráfico 66. Diagrama de dispersión de Moran RFAMNET10 con las rectas de regresión que eliminan valores atípicos.

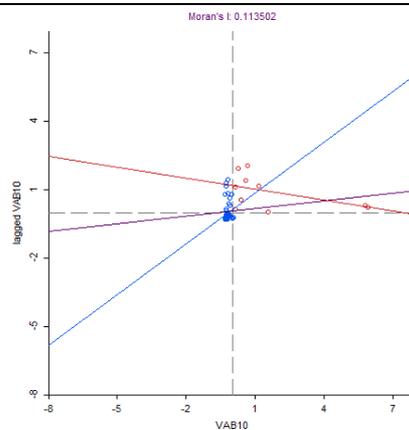


Gráfico 67. Diagrama de dispersión de Moran VAB10 con las rectas de regresión que eliminan valores atípicos.

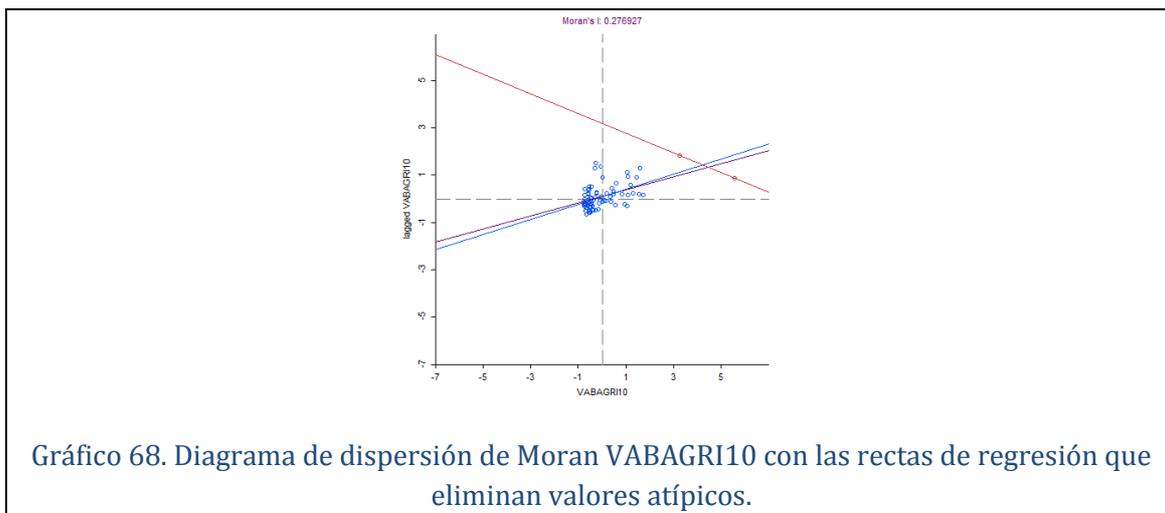
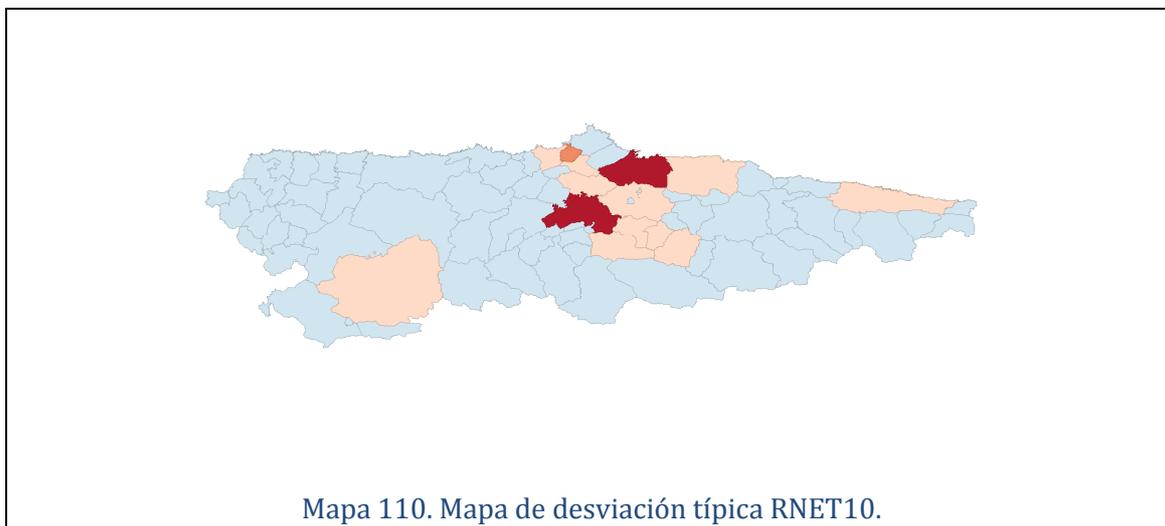
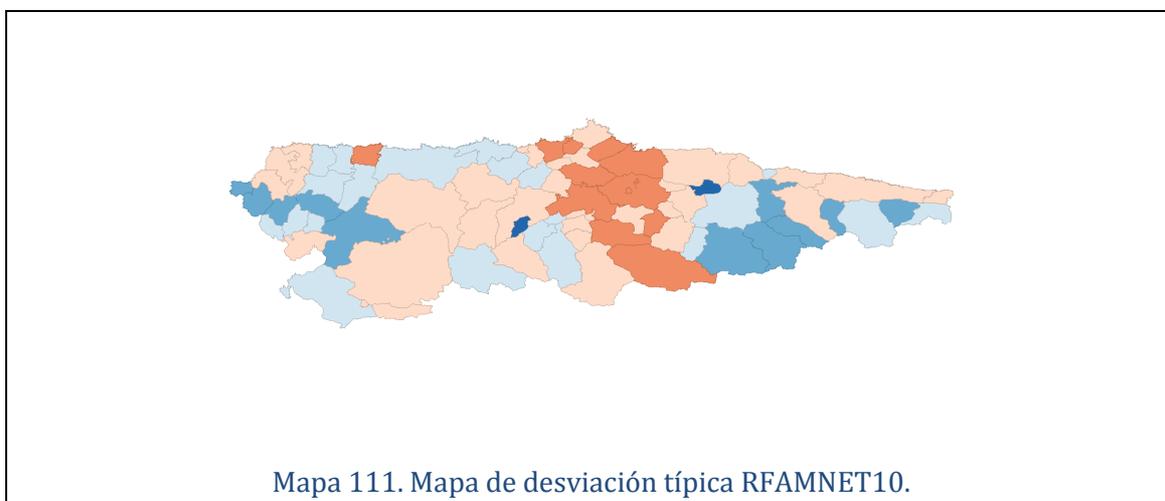


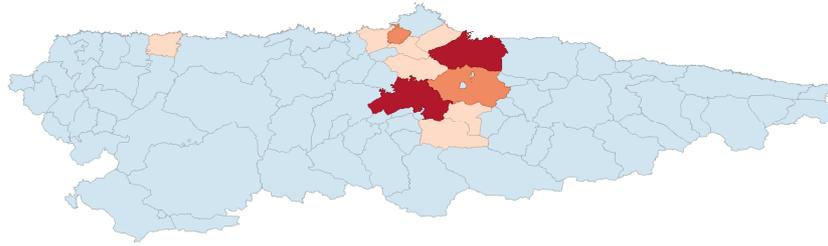
Gráfico 68. Diagrama de dispersión de Moran VABAGRI10 con las rectas de regresión que eliminan valores atípicos.



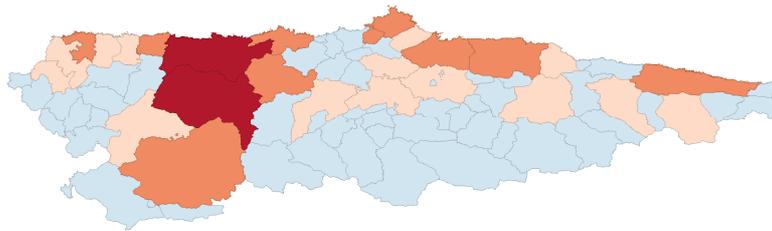
Mapa 110. Mapa de desviación típica RNET10.



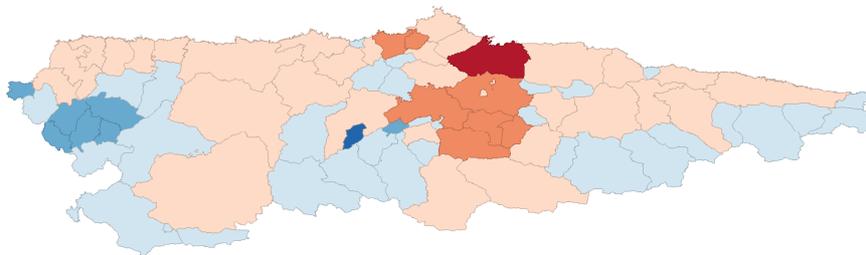
Mapa 111. Mapa de desviación típica RFAMNET10.



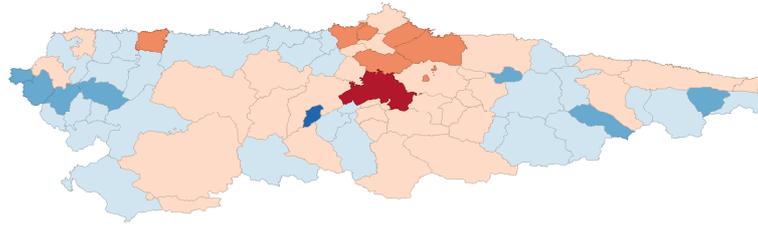
Mapa 112. Mapa de desviación típica VAB10.



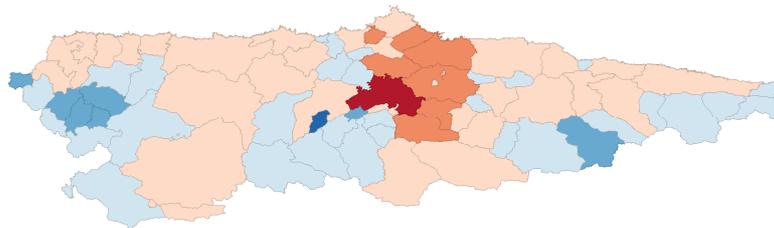
Mapa 113. Mapa de desviación típica VABAGRI10.



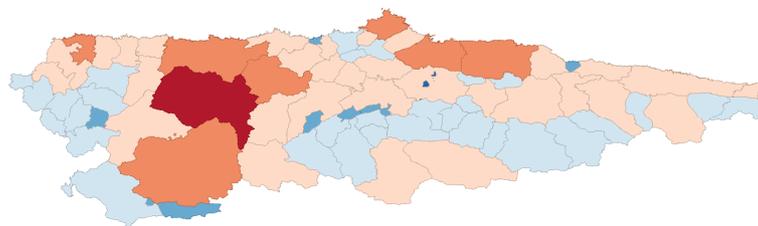
Mapa 114. Mapa de percentiles RNET10.



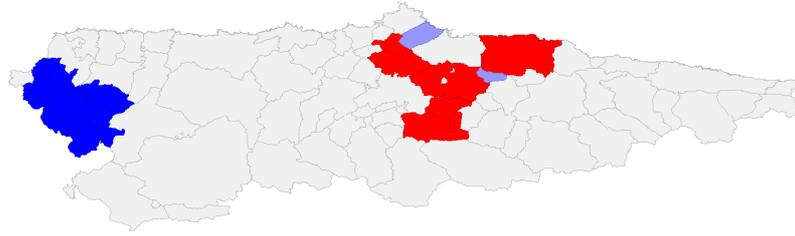
Mapa 115. Mapa de percentiles RFAMNET10.



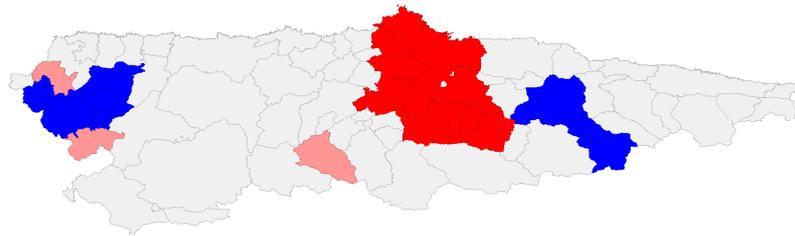
Mapa 116. Mapa de percentiles VAB10.



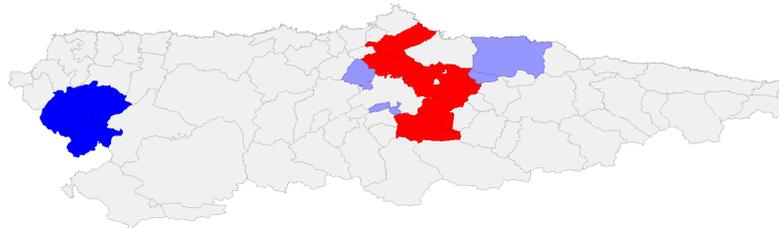
Mapa 117. Mapa de percentiles VABAGRI10.



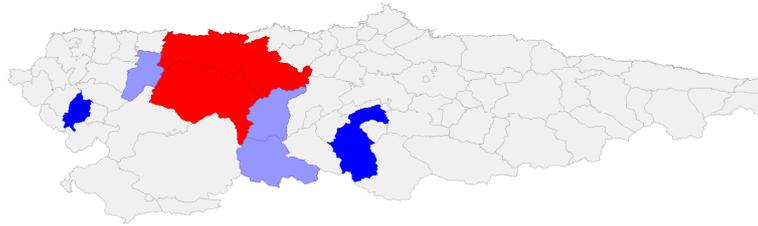
Mapa 118. Mapa LISA I de Moran RNET10.



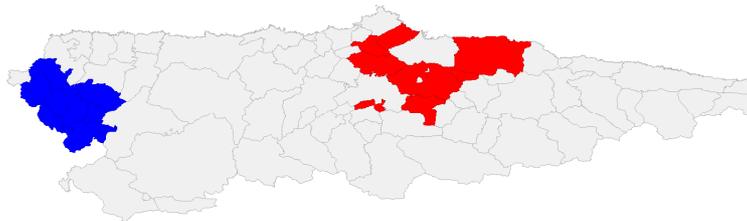
Mapa 119. Mapa LISA I de Moran RFAMNET10.



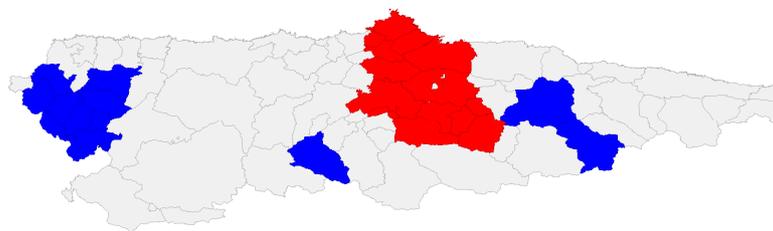
Mapa 120. Mapa LISA I de Moran VAB10.



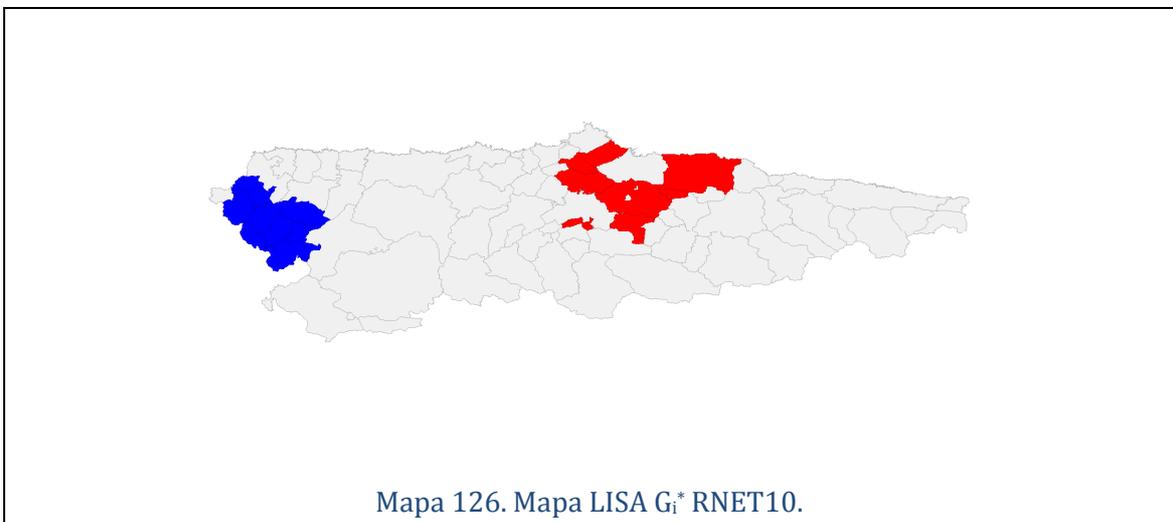
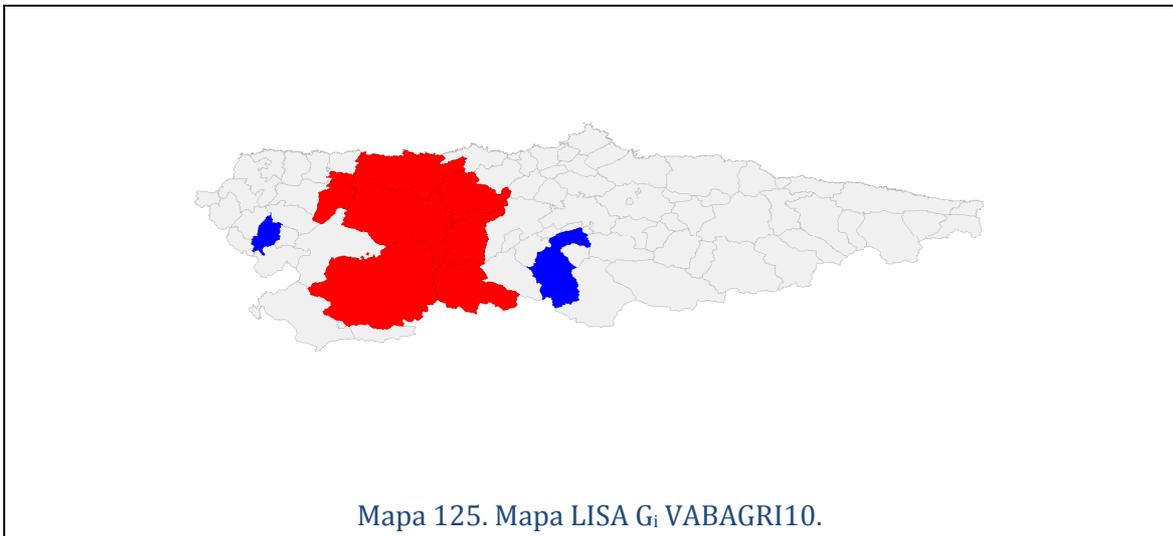
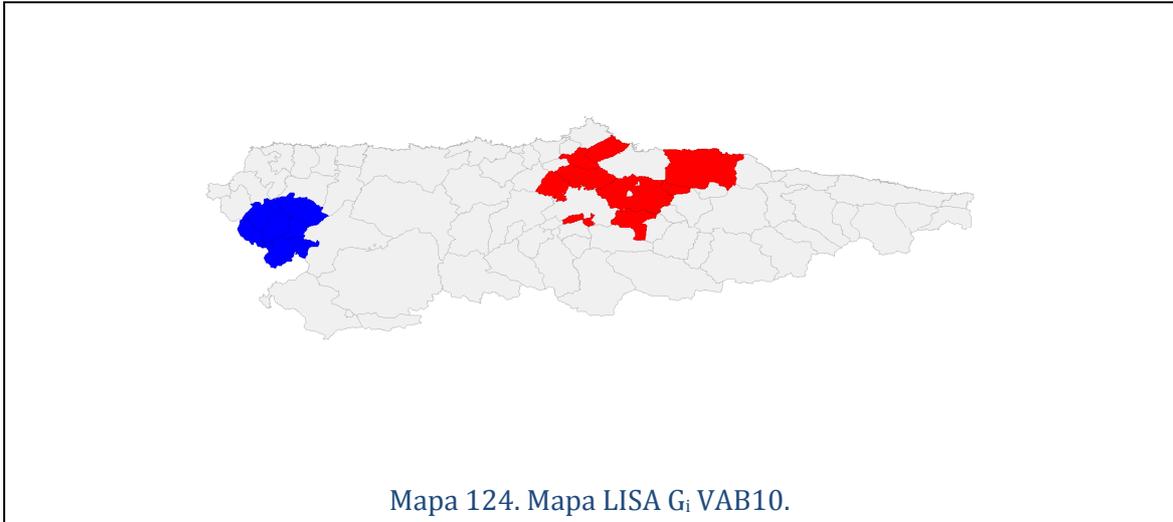
Mapa 121. Mapa LISA I de Moran VABAGRI10.

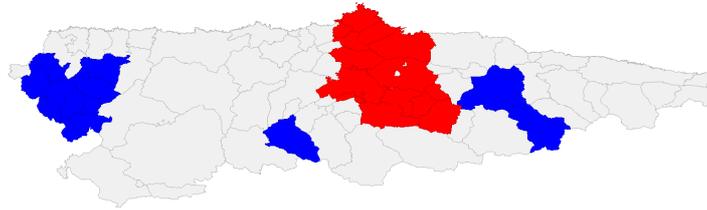


Mapa 122. Mapa LISA G<sub>1</sub> RNET10.

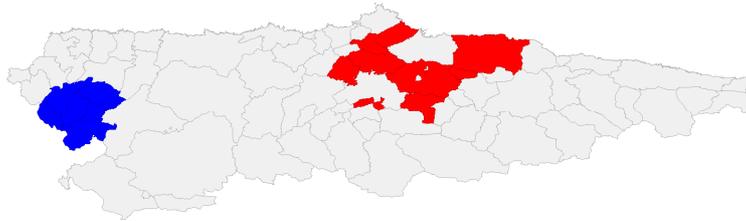


Mapa 123. Mapa LISA G<sub>1</sub> RFAMNET10.

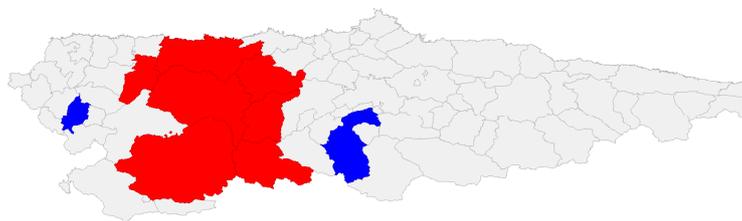




Mapa 127. Mapa LISA  $G_i^*$  RFAMNET10.



Mapa 128. Mapa LISA  $G_i^*$  VAB10.



Mapa 129. Mapa LISA  $G_i^*$  VABAGRI10.

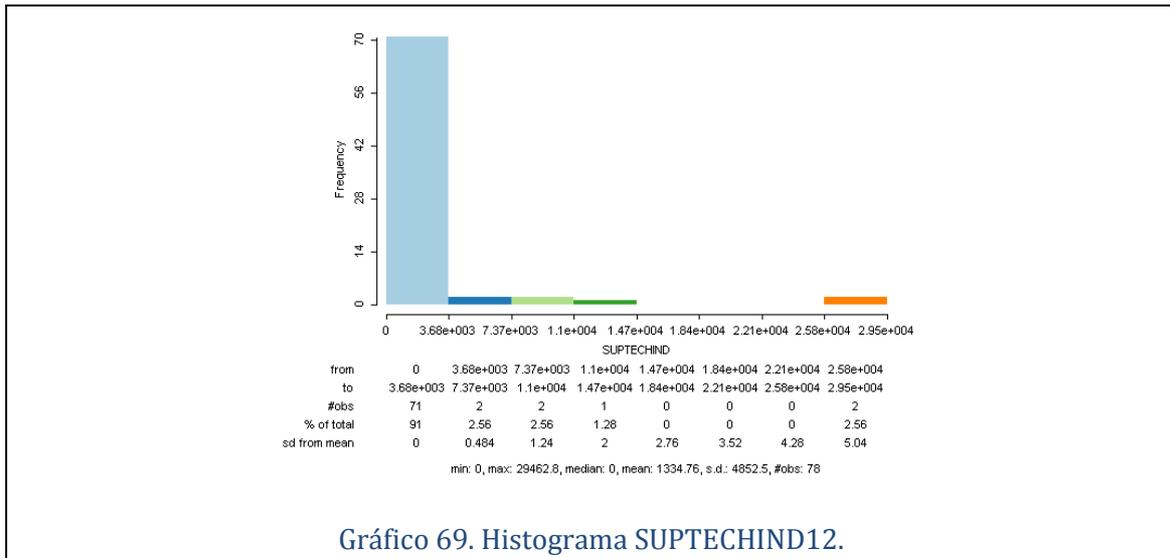


Gráfico 69. Histograma SUPTECHIND12.

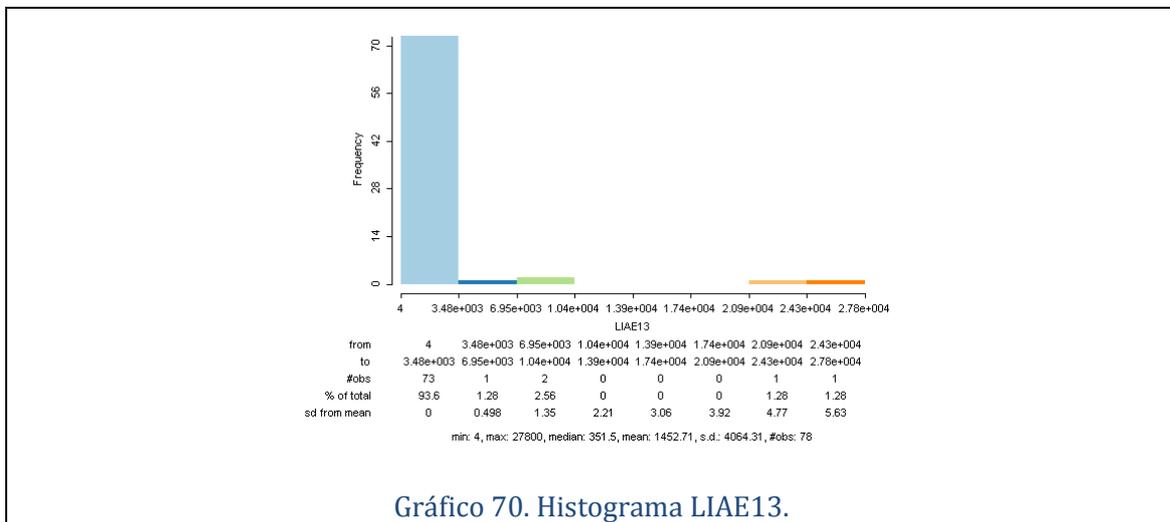


Gráfico 70. Histograma LIAE13.

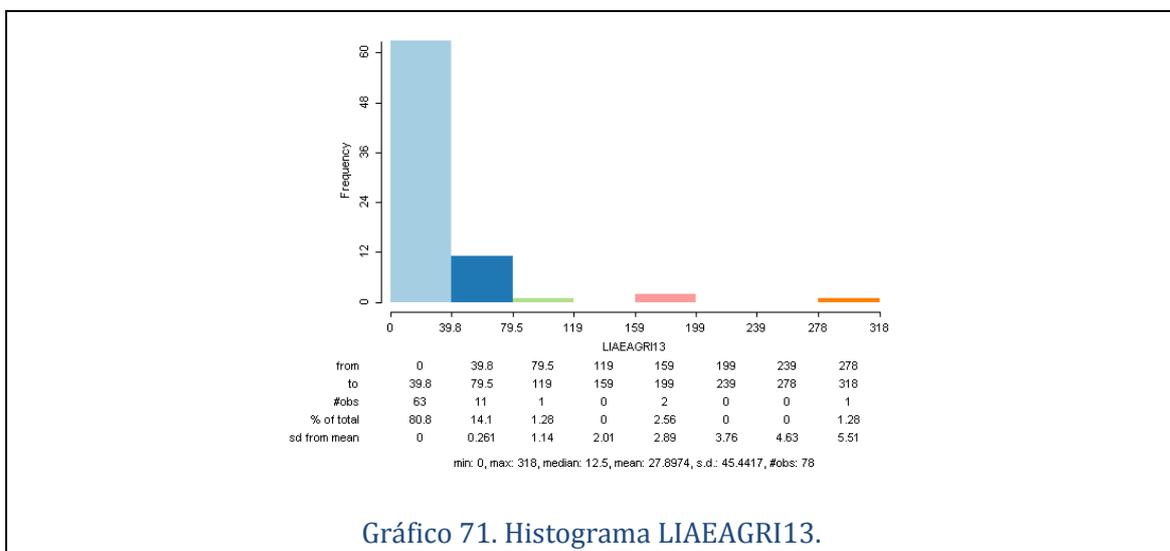
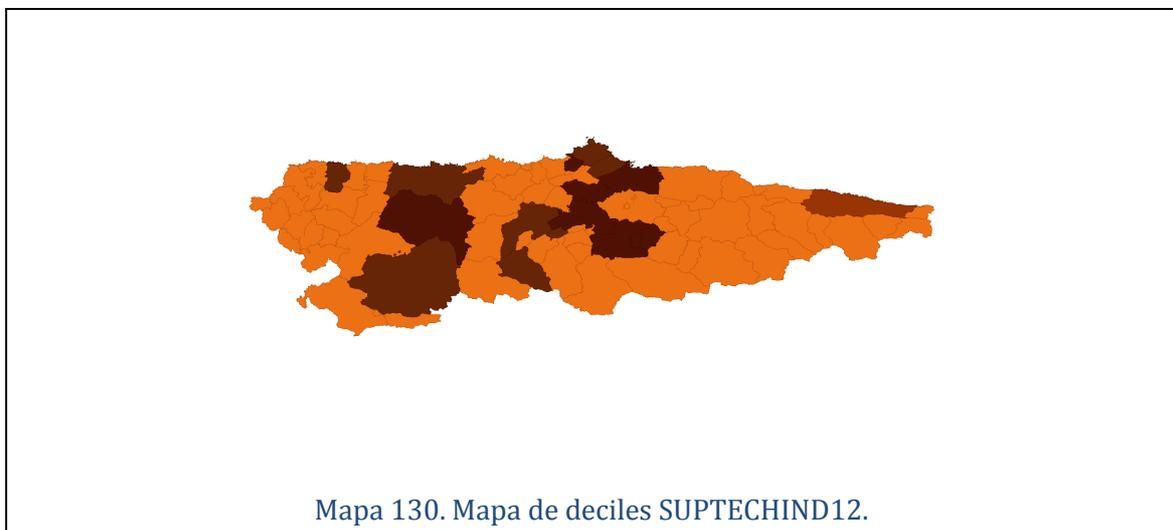
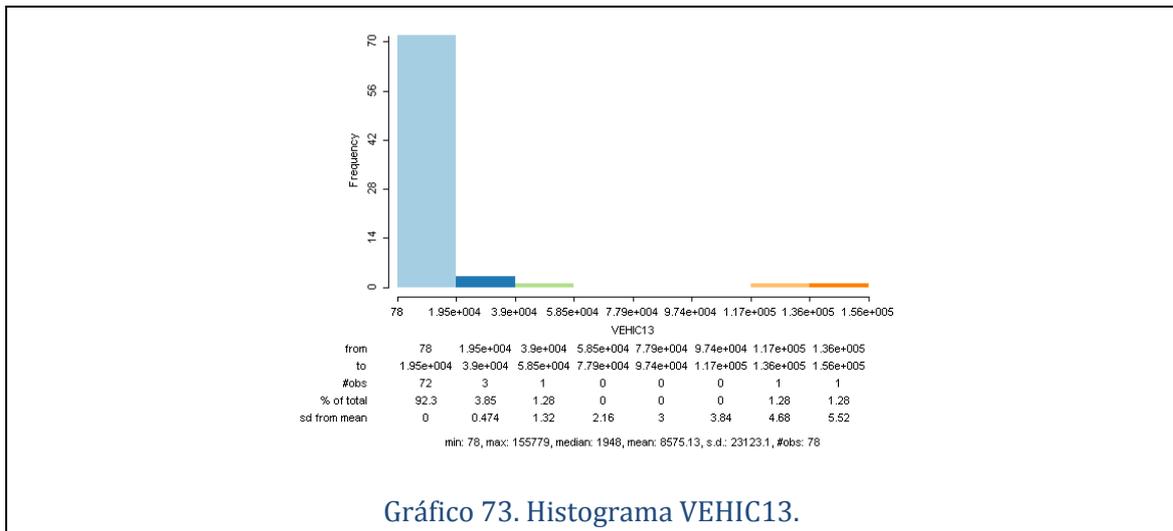
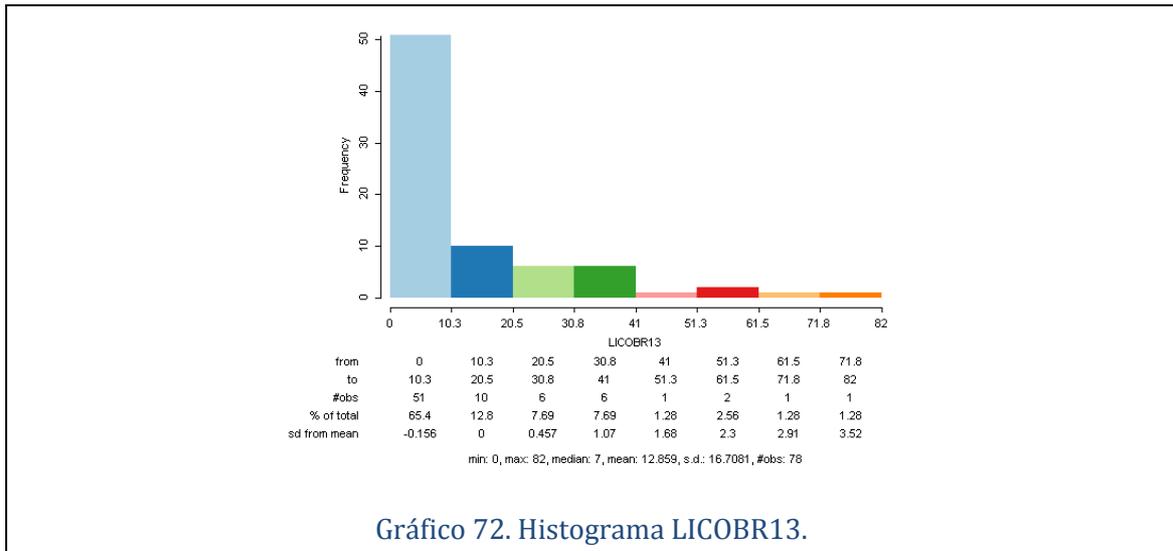
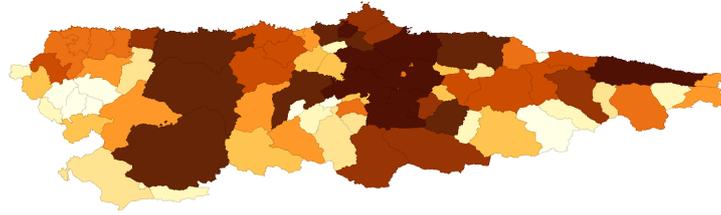
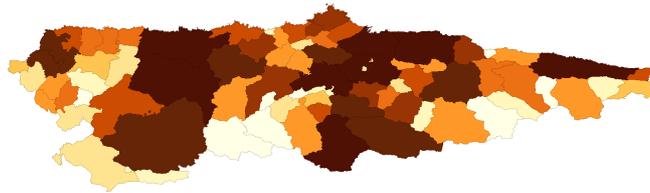


Gráfico 71. Histograma LIAEAGRI13.

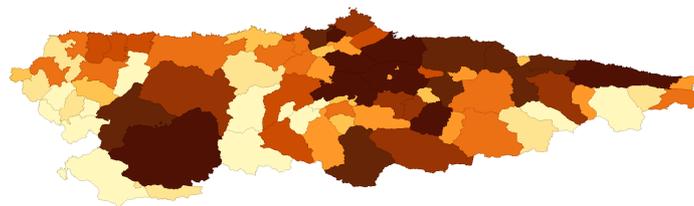




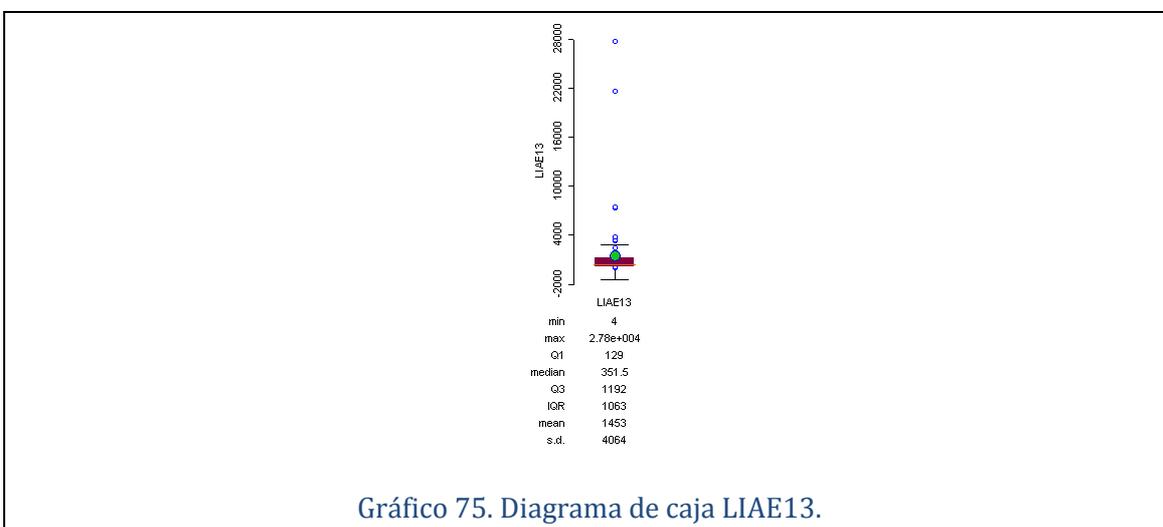
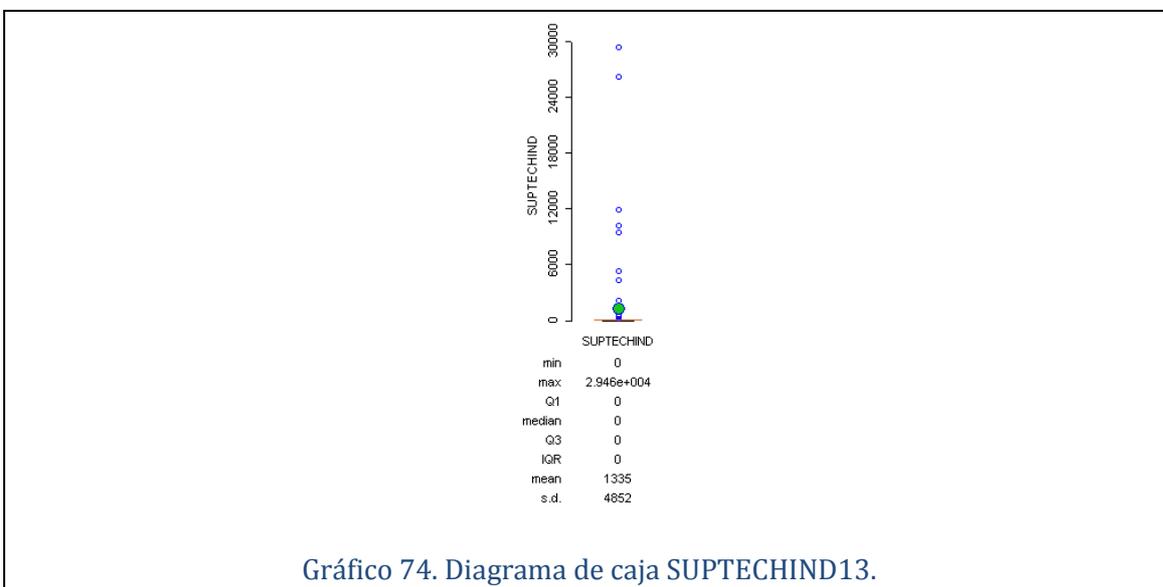
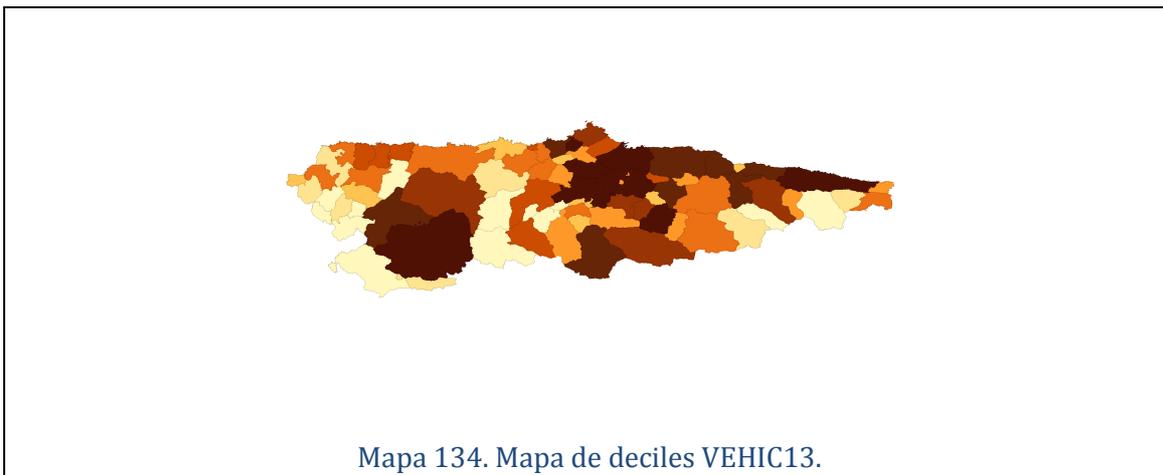
Mapa 131. Mapa de deciles LIAE13



Mapa 132. Mapa de deciles LIAEAGRI13.



Mapa 133. Mapa de deciles LICOB13.



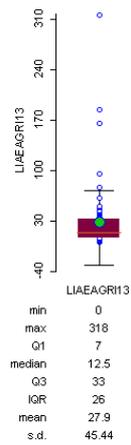


Gráfico 76. Diagrama de caja LIAEAGR13.

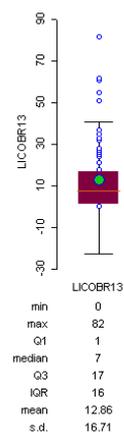


Gráfico 77. Diagrama de caja LICOBR13.

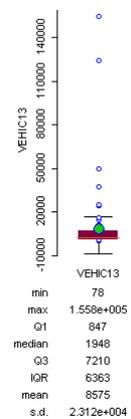
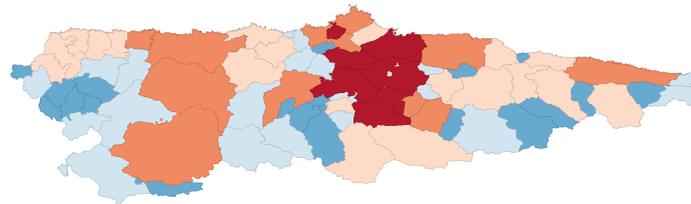


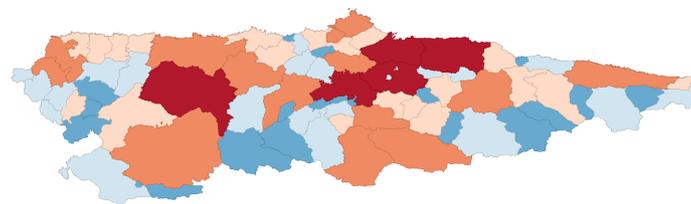
Gráfico 78. Diagrama de caja VEHIC13.



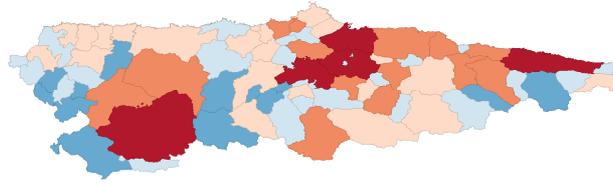
Mapa 135. Mapa de caja SUPTECHIND13.



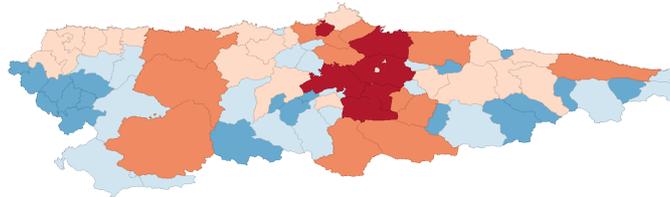
Mapa 136. Mapa de caja LIAE13.



Mapa 137. Mapa de caja LIAEAGRI13.



Mapa 138. Mapa de caja LICOB13.



Mapa 139. Mapa de caja VEHIC13.

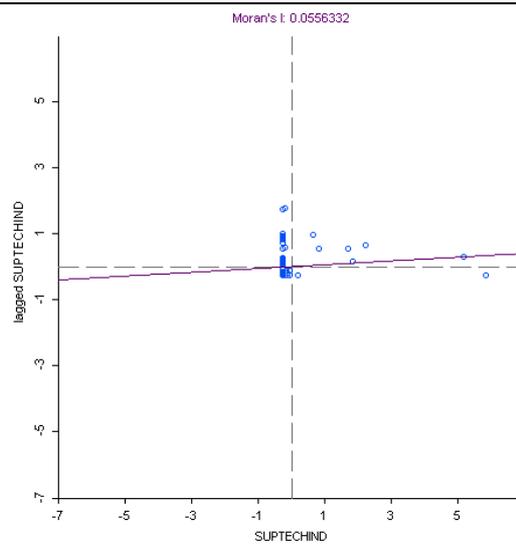


Gráfico 79. Diagrama de dispersión de Moran SUPTECHIND12.

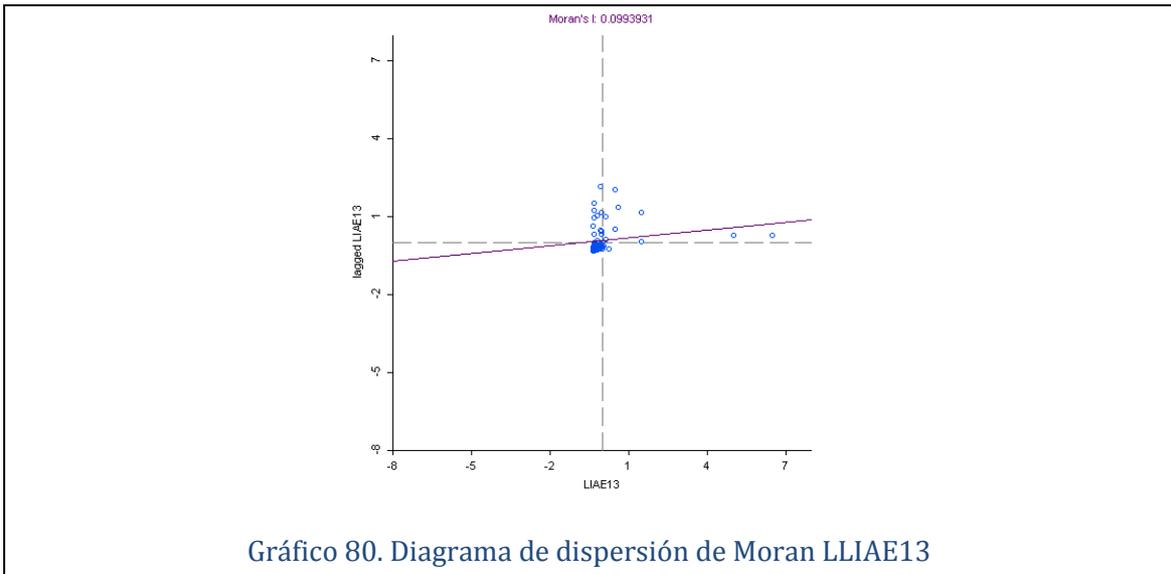


Gráfico 80. Diagrama de dispersión de Moran LIAE13

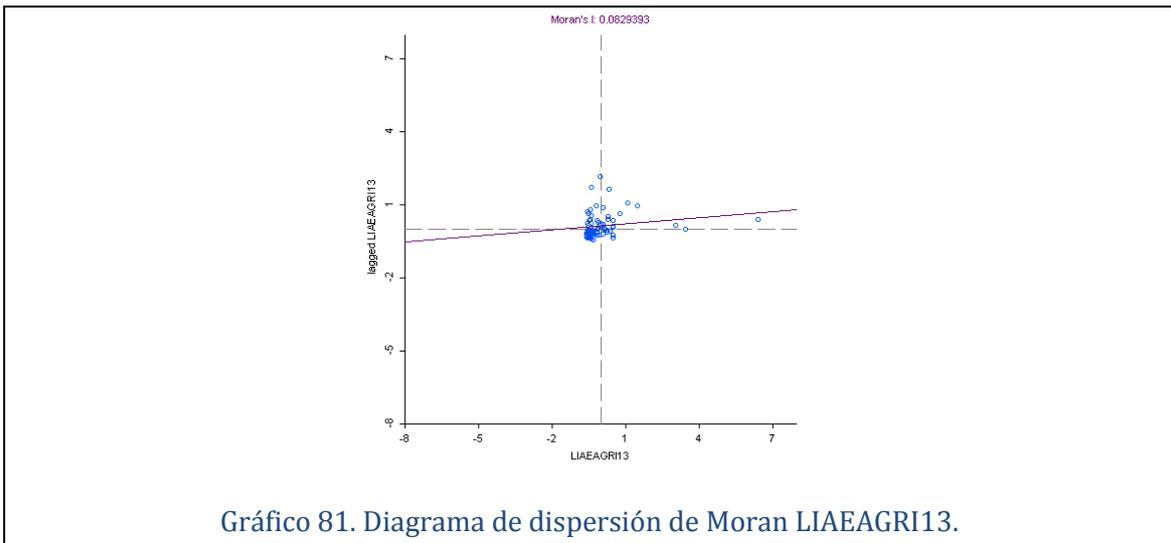


Gráfico 81. Diagrama de dispersión de Moran LIAEAGRI13.

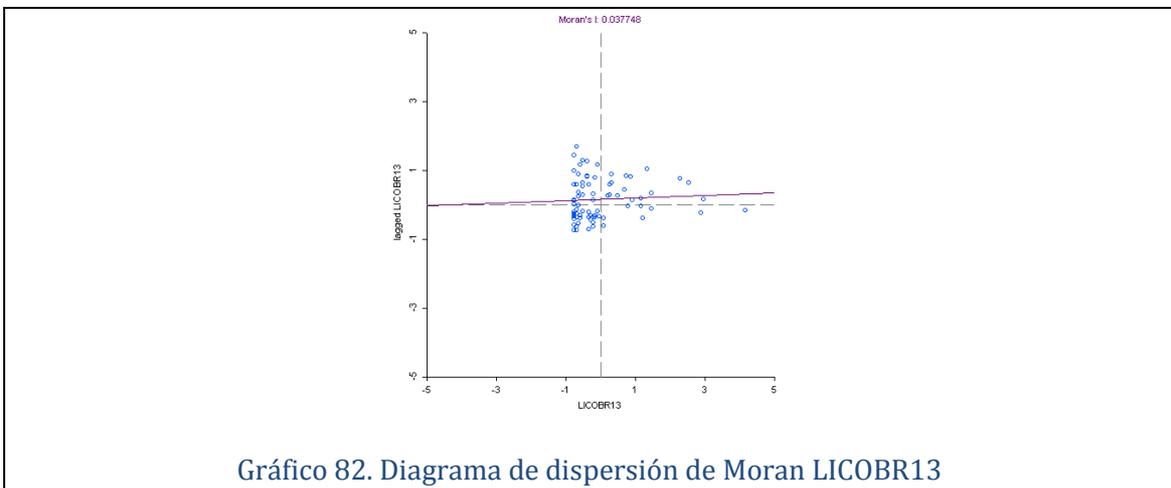


Gráfico 82. Diagrama de dispersión de Moran LICOB13

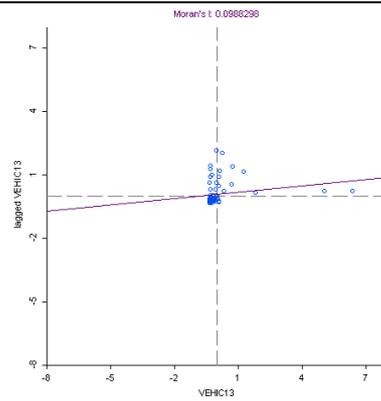


Gráfico 83. Diagrama de dispersión de Moran VEHIC13.

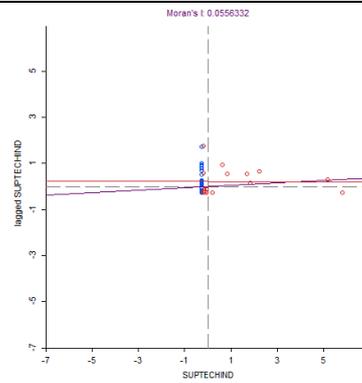


Gráfico 84. Diagrama de dispersión de Moran SUPTECHIND12 con las rectas de regresión que eliminan valores atípicos.

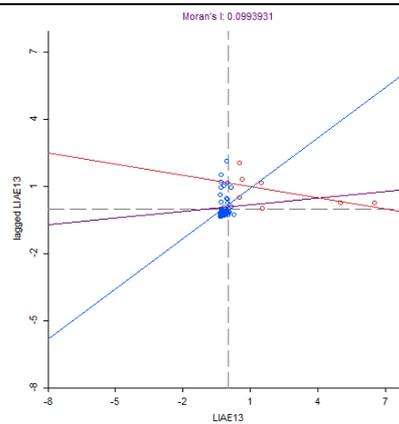


Gráfico 85. Diagrama de dispersión de Moran LIAE13 con las rectas de regresión que eliminan valores atípicos.

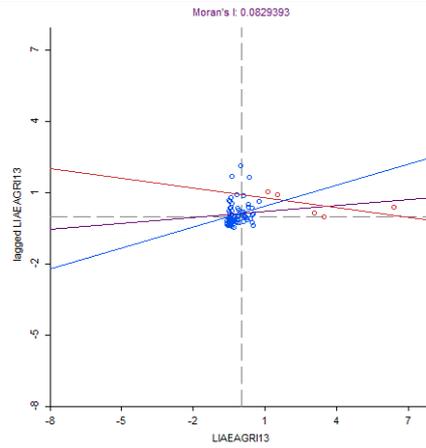


Gráfico 86. Diagrama de dispersión de Moran LIAEAGRI13 con las rectas de regresión que eliminan valores atípicos.

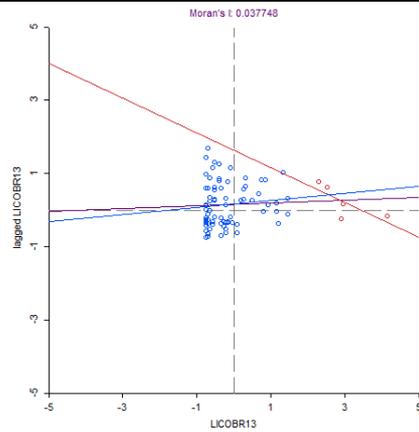


Gráfico 87. Diagrama de dispersión de Moran LICOBRI13 con las rectas de regresión que eliminan valores atípicos.

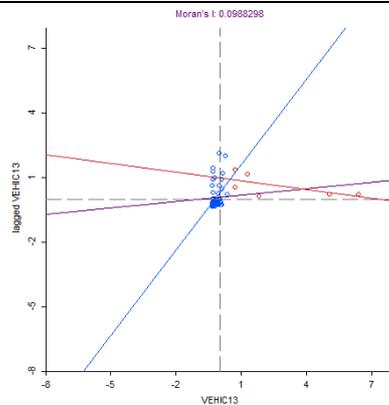
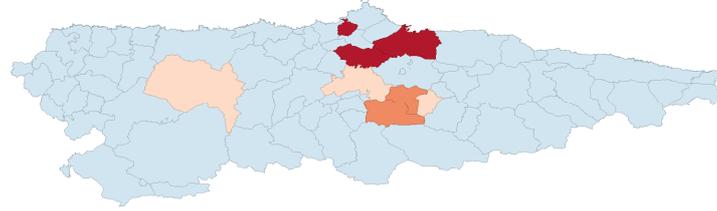
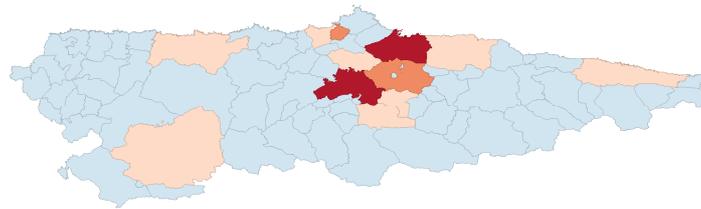


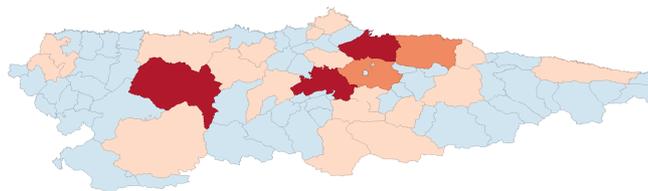
Gráfico 88. Diagrama de dispersión de Moran VEHIC13 con las rectas de regresión que eliminan valores atípicos.



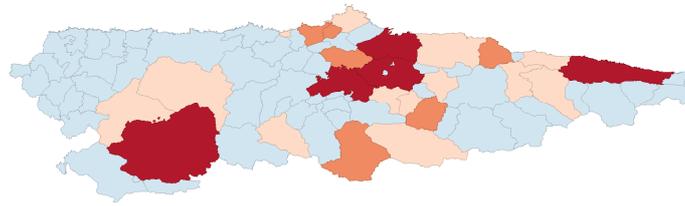
Mapa 140. Mapa de desviación típica SUPTECHIND12.



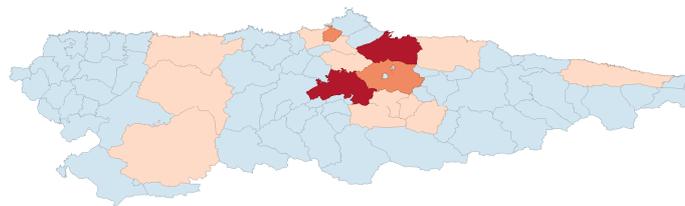
Mapa 141. Mapa de desviación típica LIAE13.



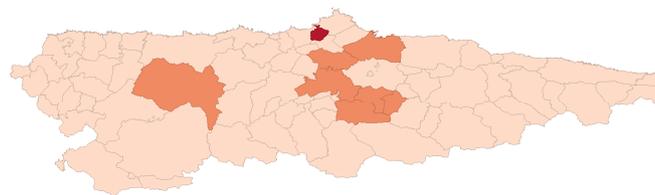
Mapa 142. Mapa de desviación típica LIAEAGRI13.



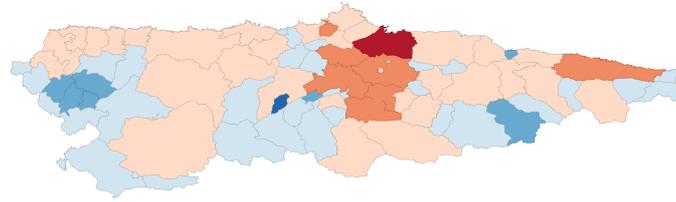
Mapa 143. Mapa de desviación típica LICOB13.



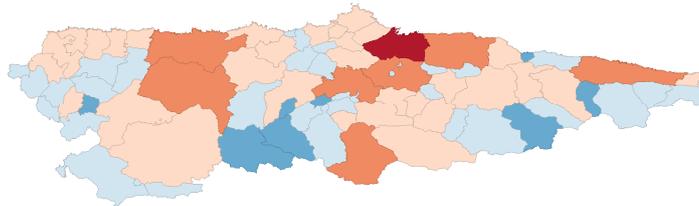
Mapa 144. Mapa de desviación típica VEHIC13.



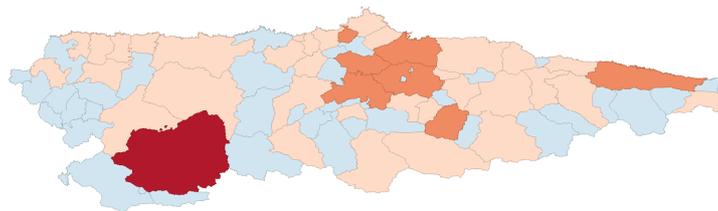
Mapa 145. Mapa de percentiles SUPTECHIND12.



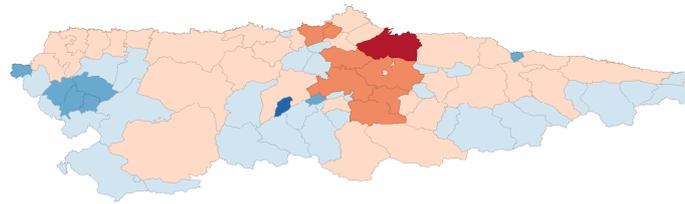
Mapa 146. Mapa de percentiles LIAE13.



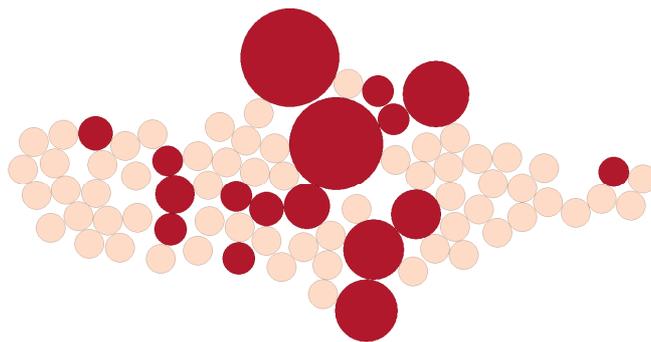
Mapa 147. Mapa de percentiles LIAEAGRI13.



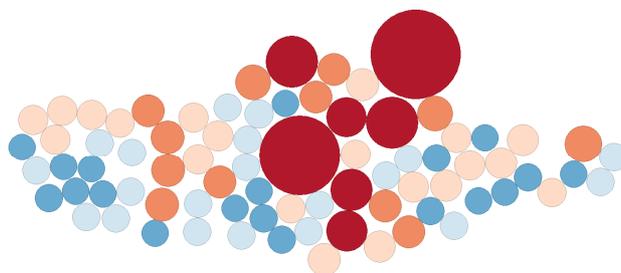
Mapa 148. Mapa de percentiles LICOB13.



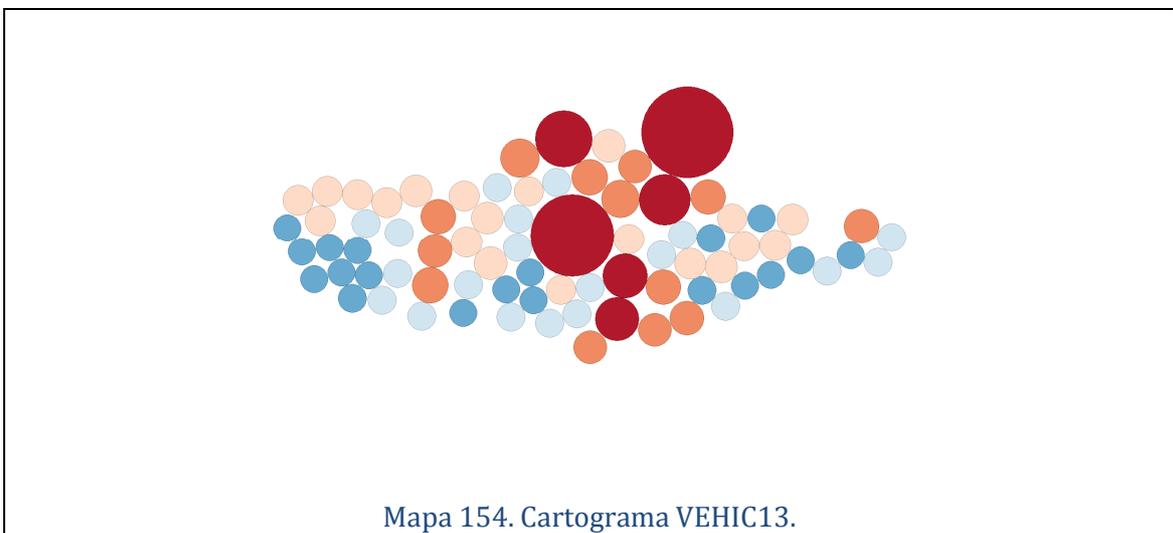
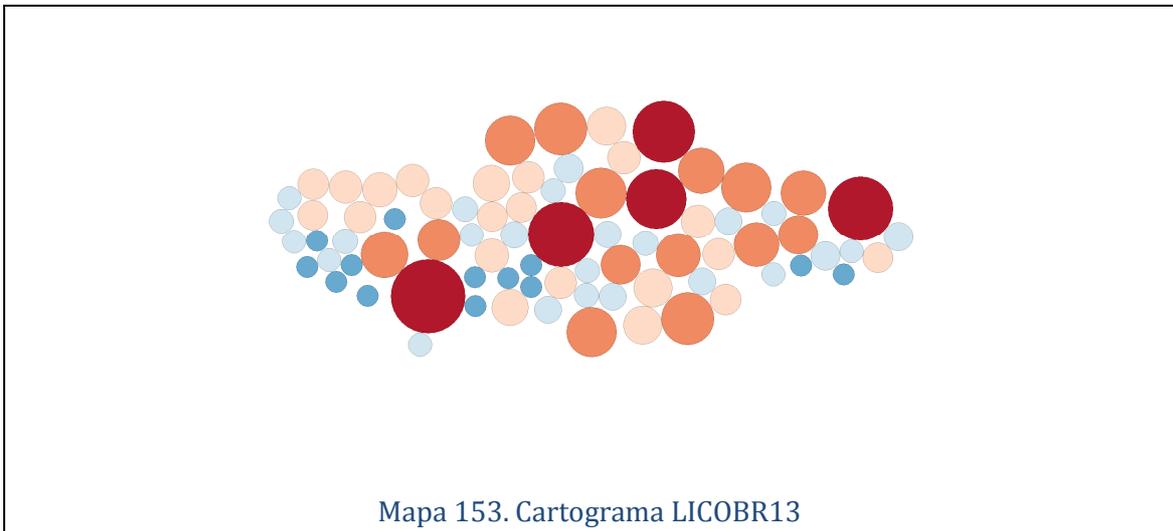
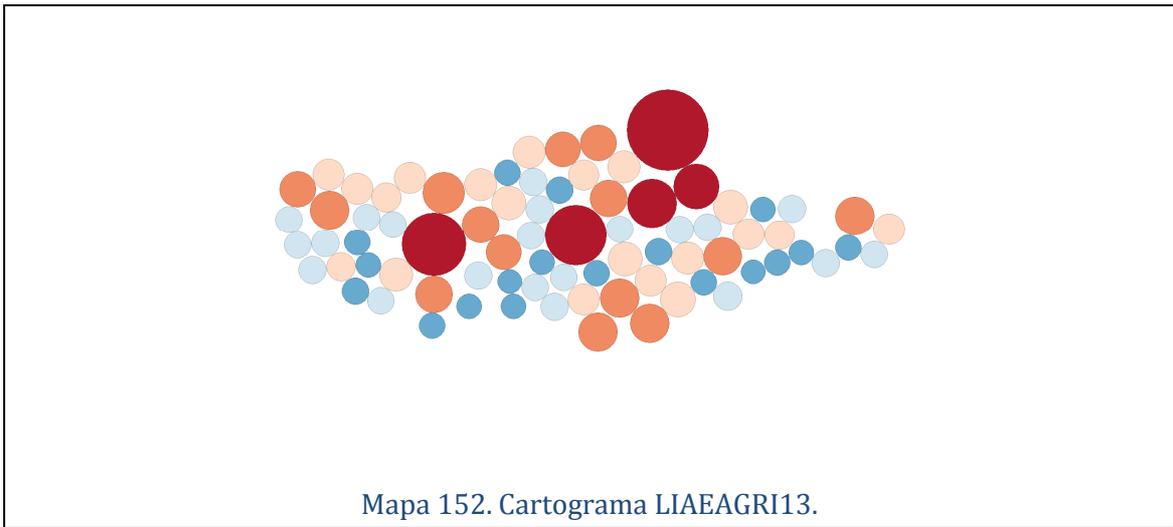
Mapa 149. Mapa de percentiles VEHIC13.

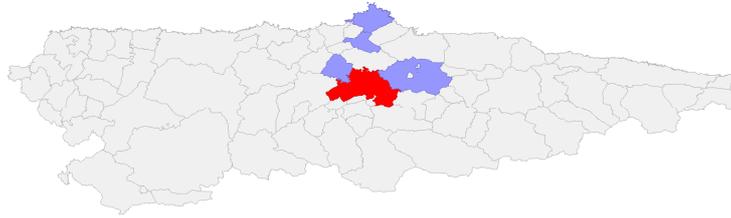


Mapa 150. Cartograma SUPTECHIND12.

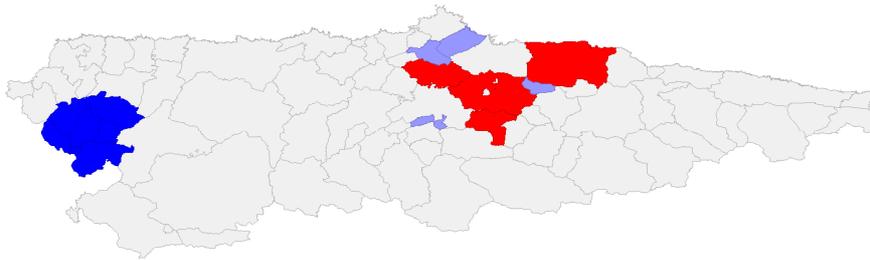


Mapa 151. Cartograma LIAE13.

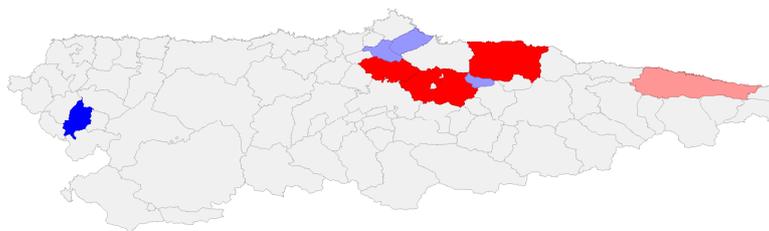




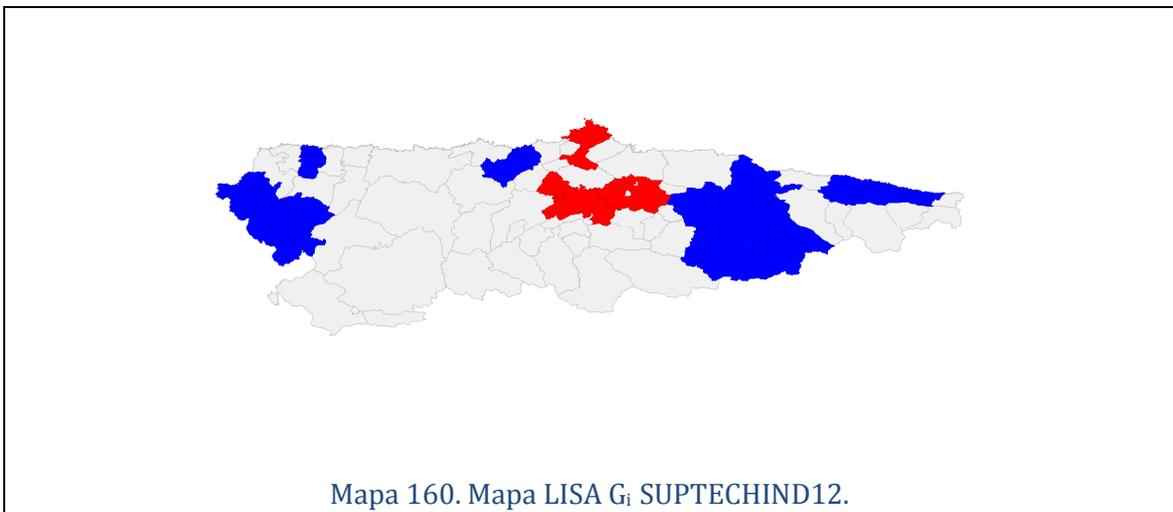
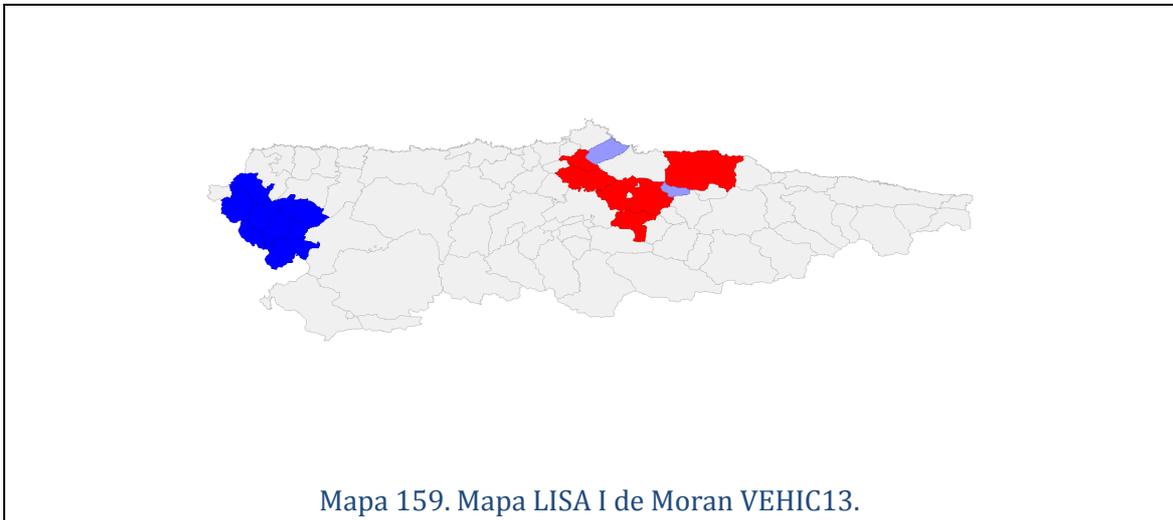
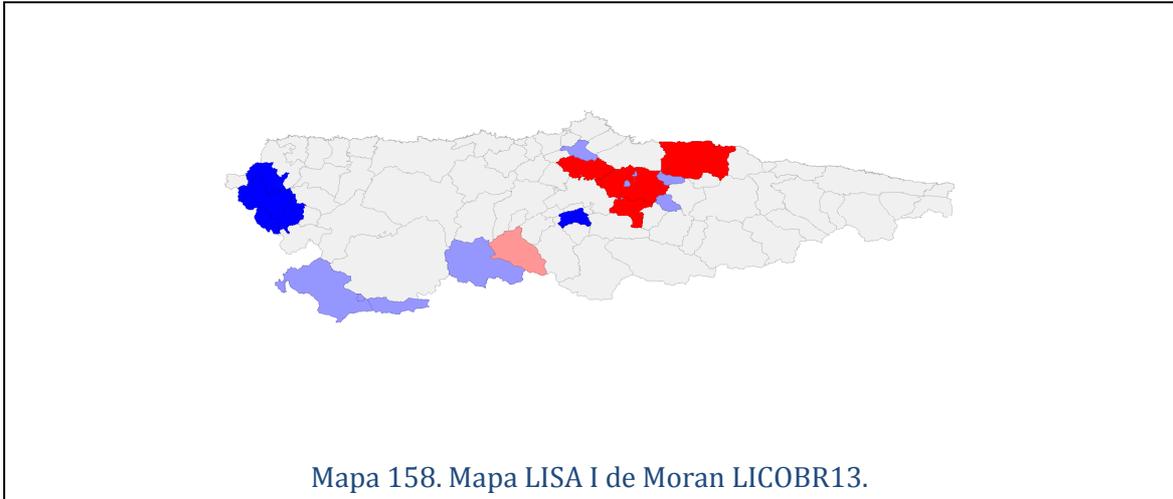
Mapa 155. Mapa LISA I de Moran SUPTECHIND12.



Mapa 156. Mapa LISA I de Moran LIAE13.

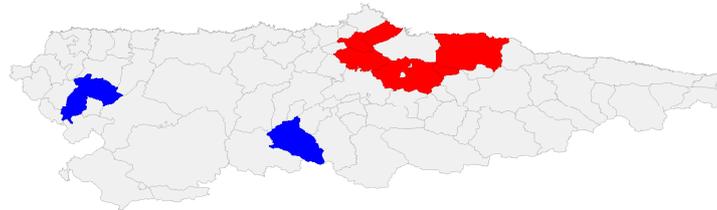


Mapa 157. Mapa LISA I de Moran LIAEAGRI13.

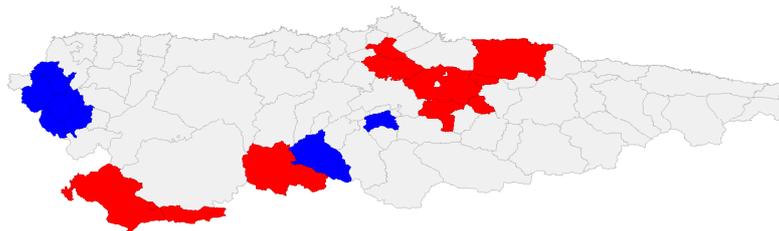




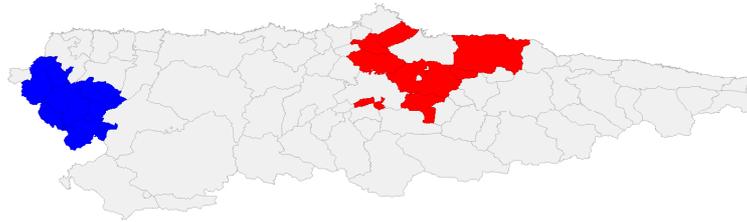
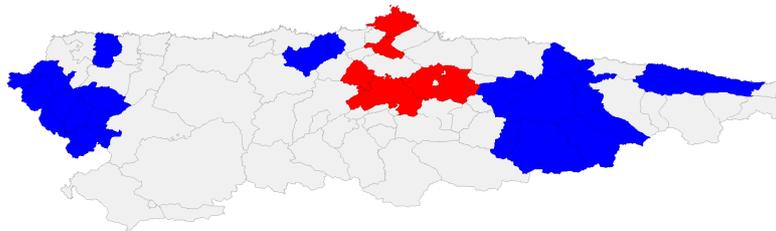
Mapa 161. Mapa LISA  $G_i$  LIAE13.

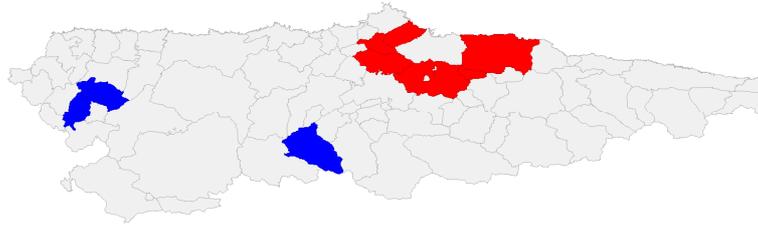


Mapa 162. Mapa LISA  $G_i$  LIAEAGRI13.

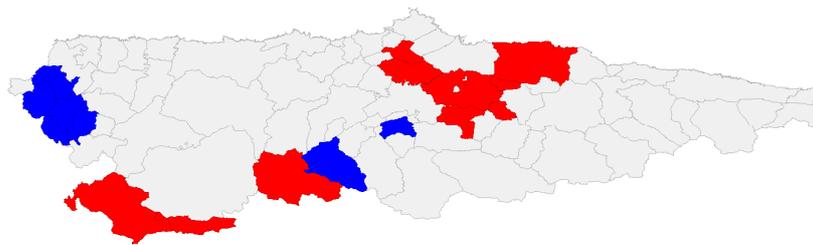


Mapa 163. Mapa LISA  $G_i$  LICOBR13.

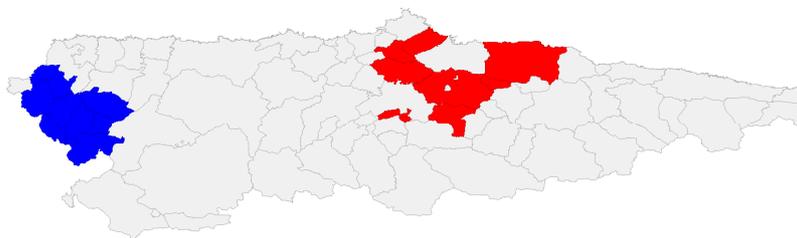
Mapa 164. Mapa LISA  $G_i$  VEHIC13.Mapa 165. Mapa LISA  $G_i^*$  SUPTECHIND12.Mapa 166. Mapa LISA  $G_i^*$  LIAE13.



Mapa 167. Mapa LISA  $G_i^*$  LIAEAGRI13.



Mapa 168. Mapa LISA  $G_i^*$  LICOBRI13.



Mapa 169. Mapa LISA  $G_i^*$  VEHIC13.

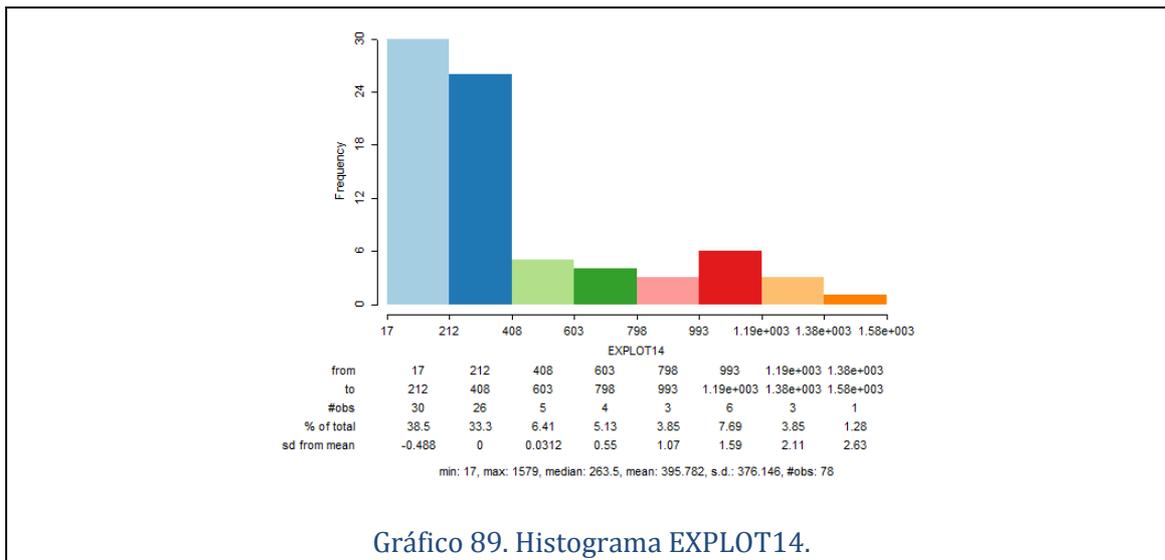


Gráfico 89. Histograma EXPLLOT14.

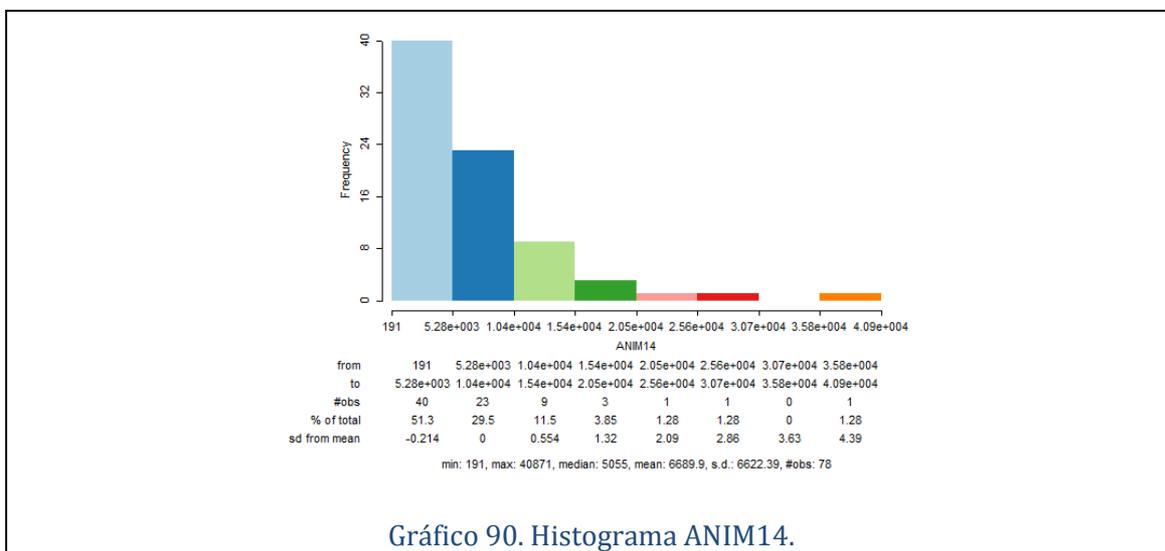


Gráfico 90. Histograma ANIM14.

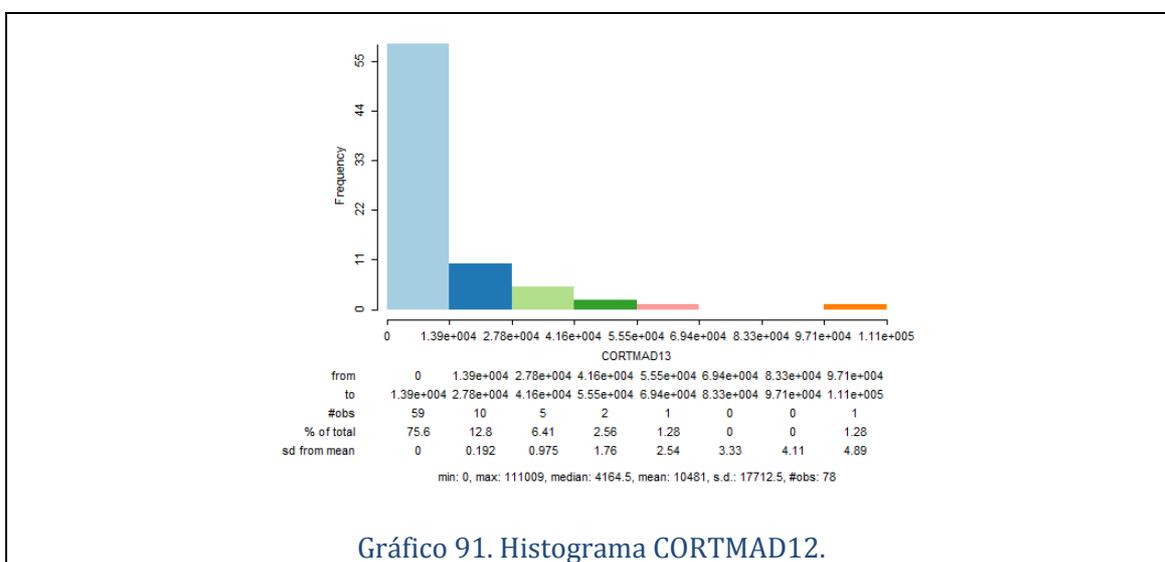
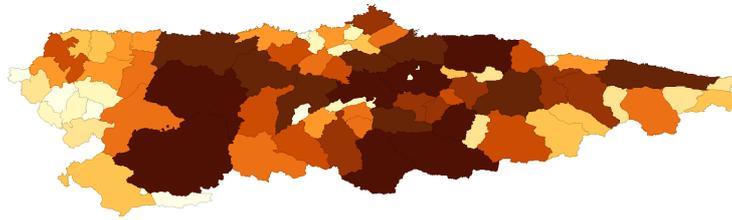
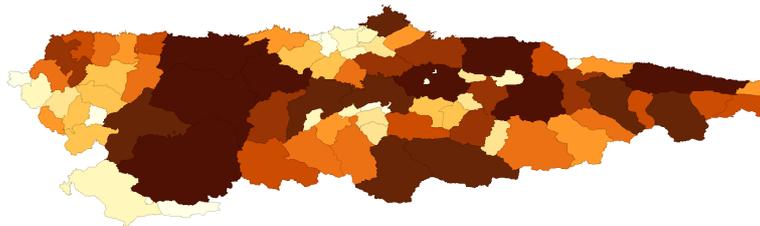


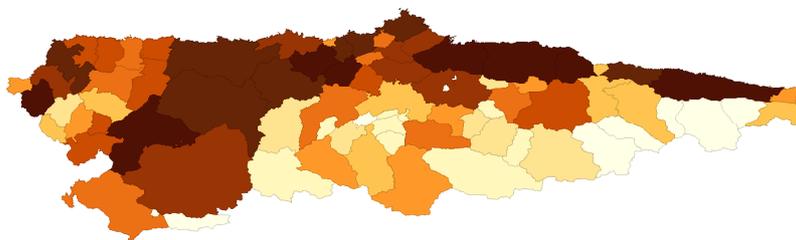
Gráfico 91. Histograma CORTMAD12.



Mapa 170. Mapa de deciles EXPL0T14.



Mapa171. Mapa de deciles ANIM14.



Mapa 172. Mapa de deciles CORTMAD12.

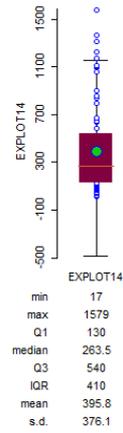


Gráfico 92. Diagrama de caja EXPLOT14.

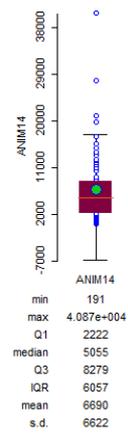


Gráfico 93. Diagrama de caja ANIM14.

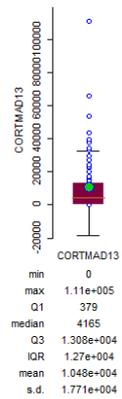
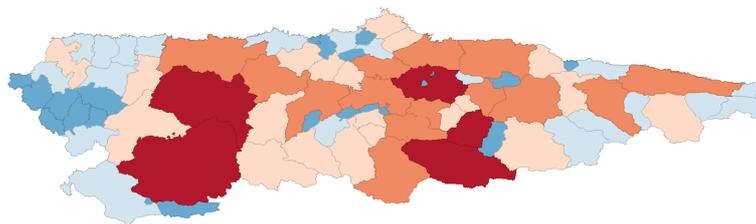
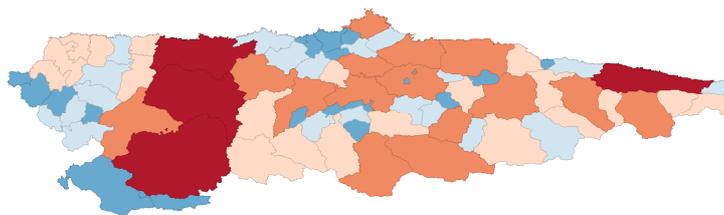


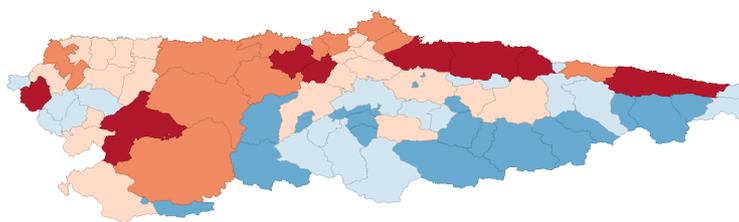
Gráfico 94. Diagrama de caja CORTMAD12.



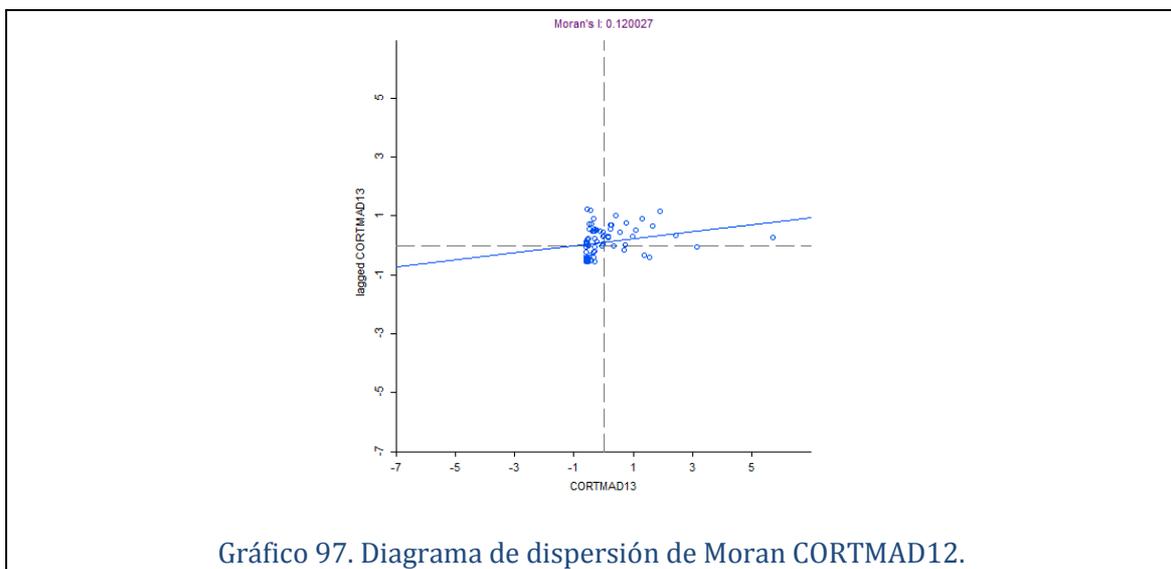
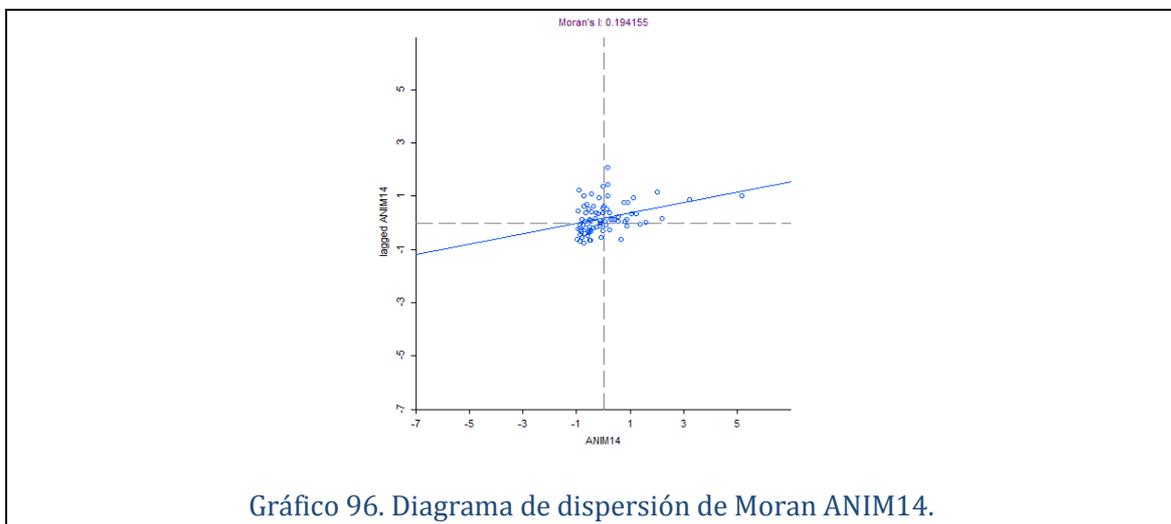
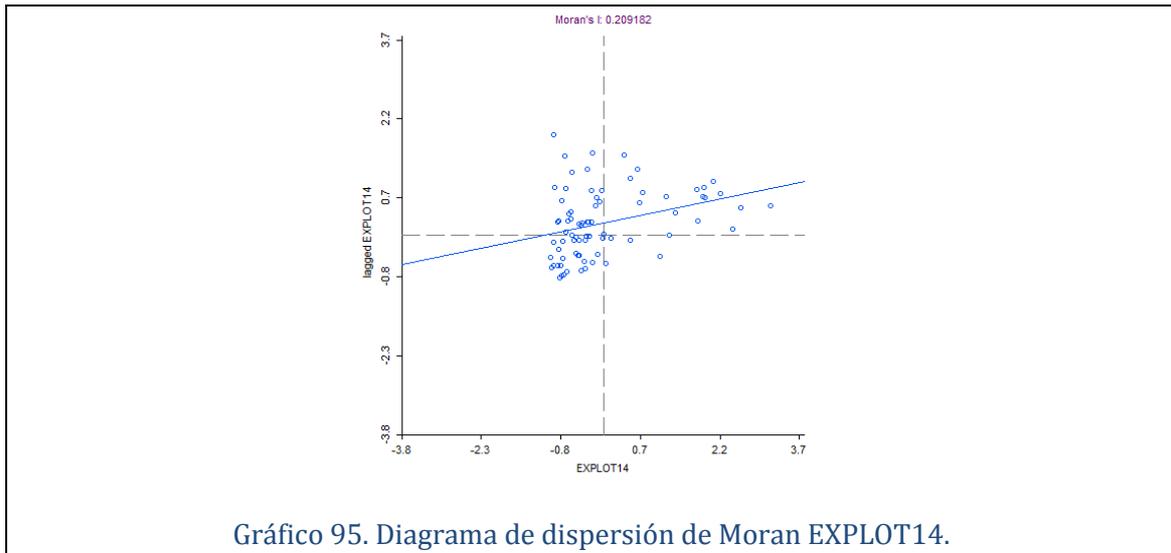
Mapa 173. Mapa de caja EXPLOT14.



Mapa 174. Mapa de caja ANIM14.



Mapa 175. Mapa de caja CORTMAD12.



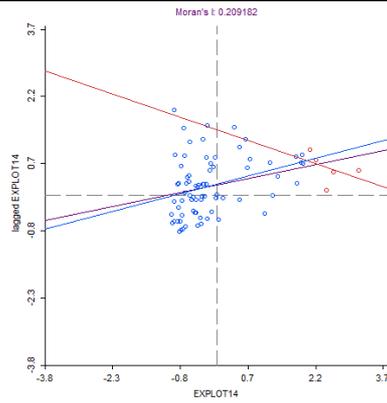


Gráfico 98. Diagrama de dispersión de Moran EXPL0T14 con las rectas de regresión que eliminan valores atípicos.

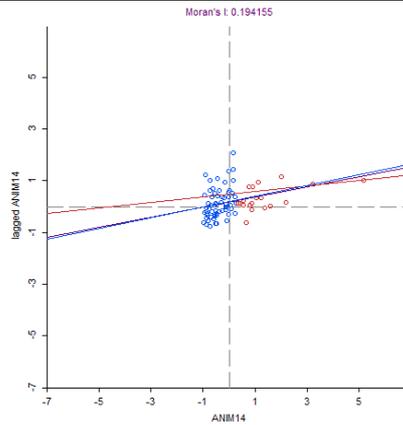


Gráfico 99. Diagrama de dispersión de Moran ANIM14 con las rectas de regresión que eliminan valores atípicos.

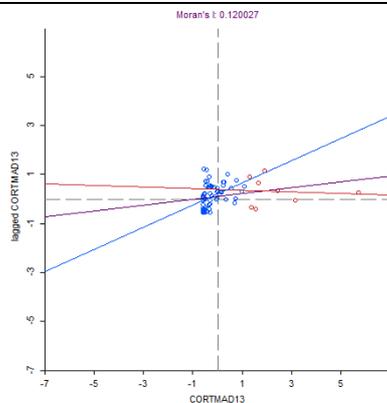
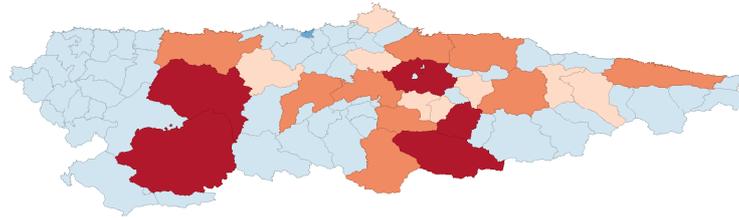
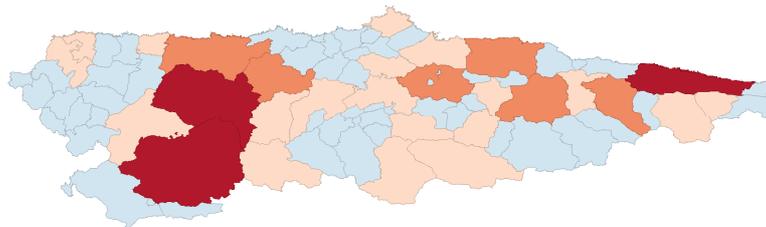


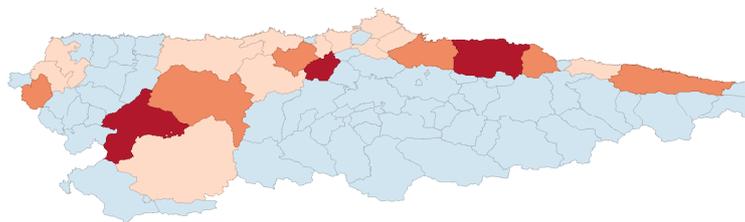
Gráfico 100. Diagrama de dispersión de Moran CORTMAD12 con las rectas de regresión que eliminan valores atípicos.



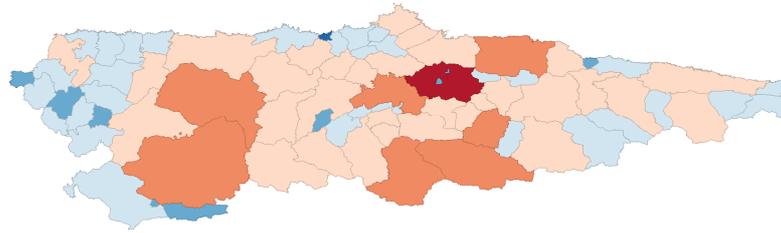
Mapa 176. Mapa de desviación típica EXPLOT14.



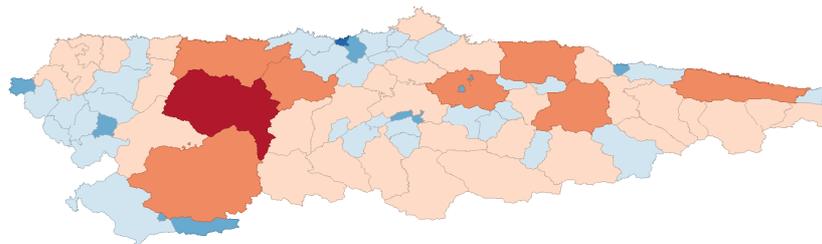
Mapa 177. Mapa de desviación típica ANIM14.



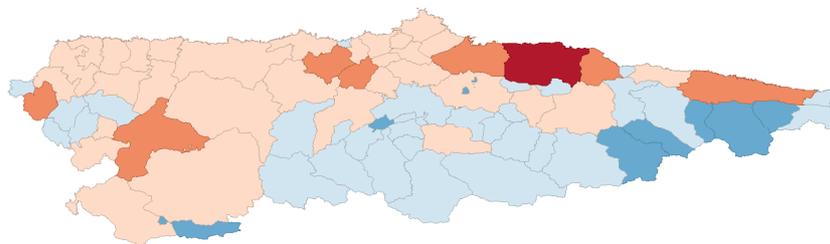
Mapa 178. Mapa de desviación típica CORTMAD12.



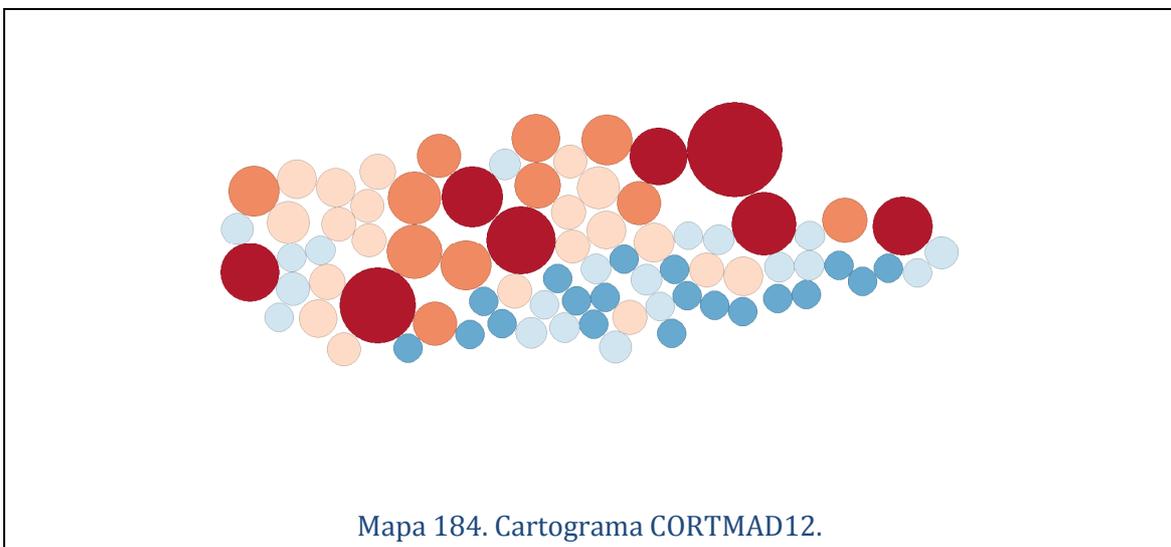
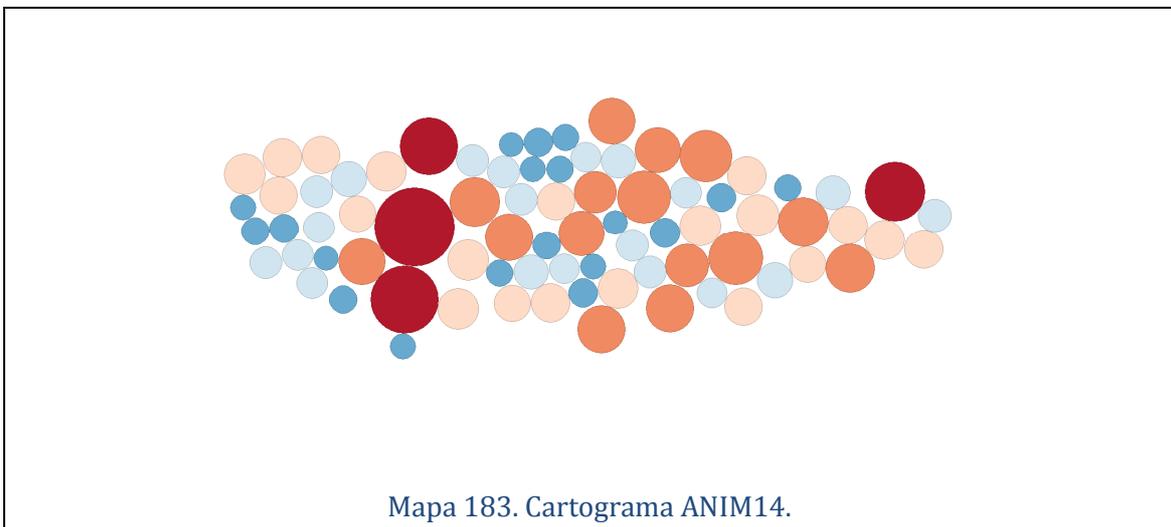
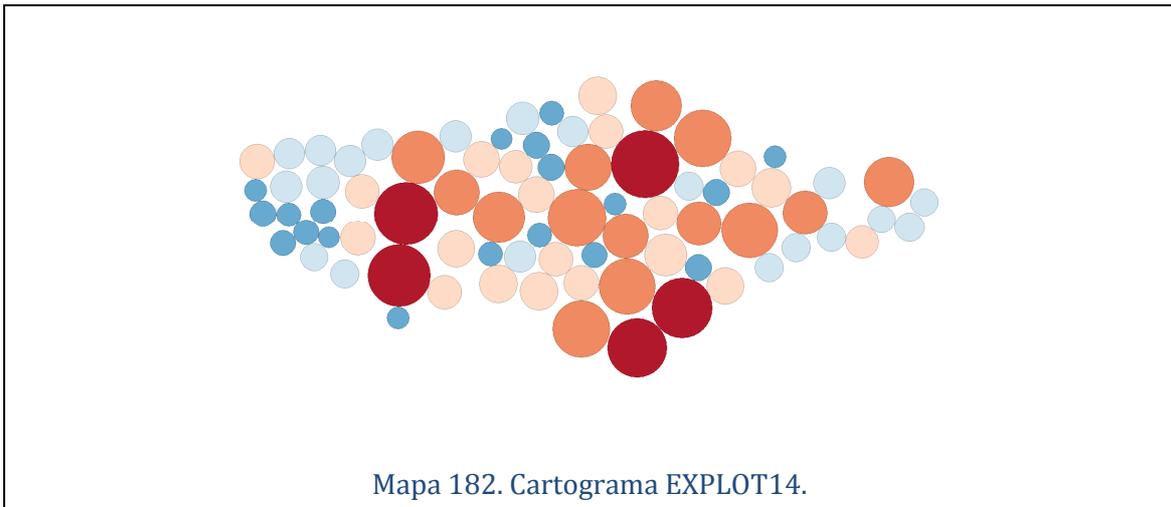
Mapa 179. Mapa de percentiles EXPL0T14.

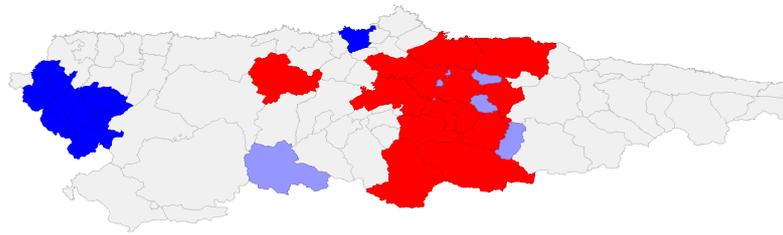


Mapa 180. Mapa de percentiles ANIM14.

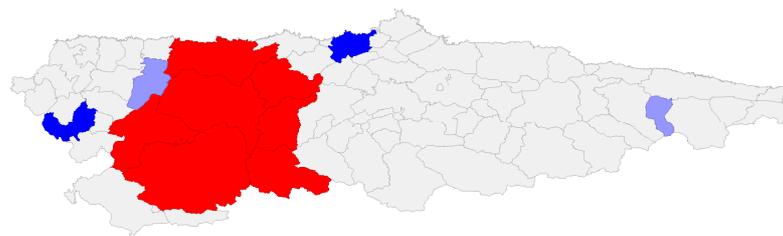


Mapa 181. Mapa de percentiles CORTMAD12.

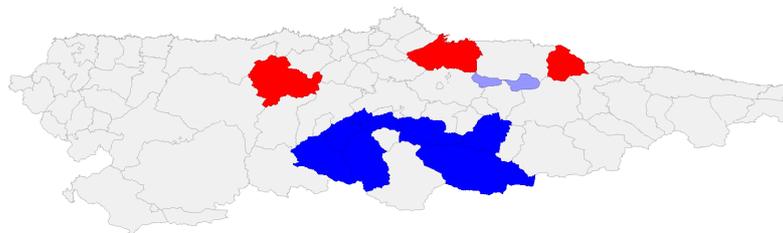




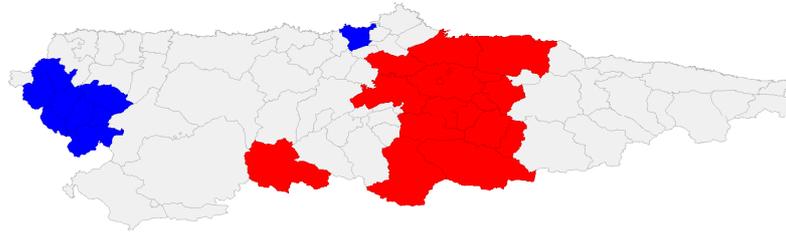
Mapa 185. Mapa LISA I de Moran EXPL0T14.



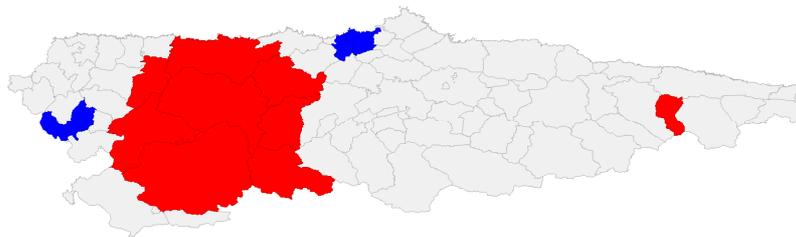
Mapa 186. Mapa LISA I de Moran ANIM14.



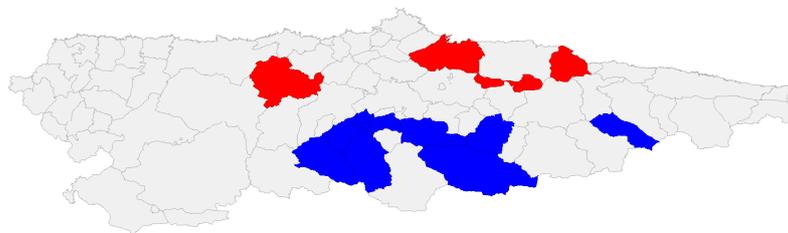
Mapa 187. Mapa LISA I de Moran CORTMAD12.



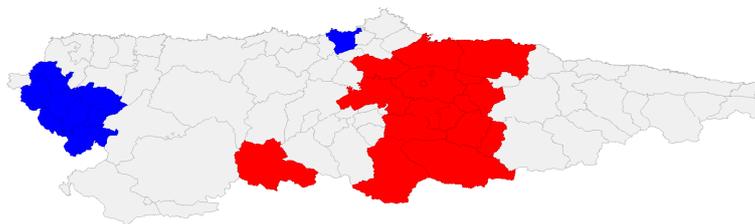
Mapa 188. Mapa LISA  $G_i$  EXPLOT14.



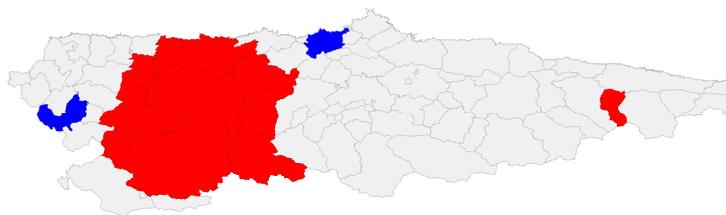
Mapa 189. Mapa LISA  $G_i$  ANIM14.



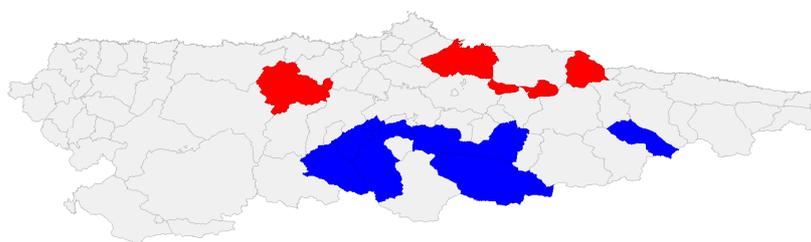
Mapa 190. Mapa LISA  $G_i$  CORTMAD12.



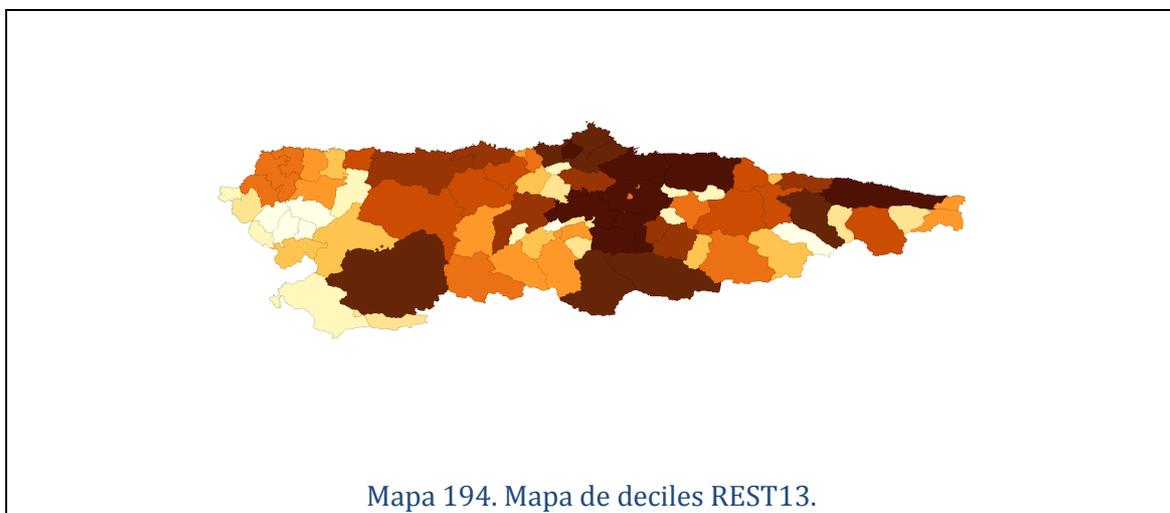
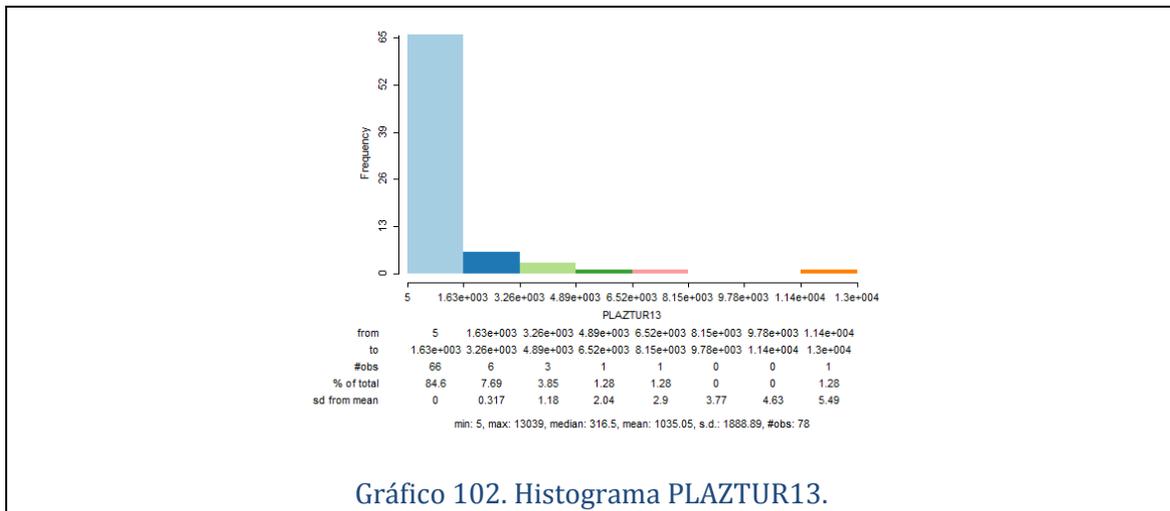
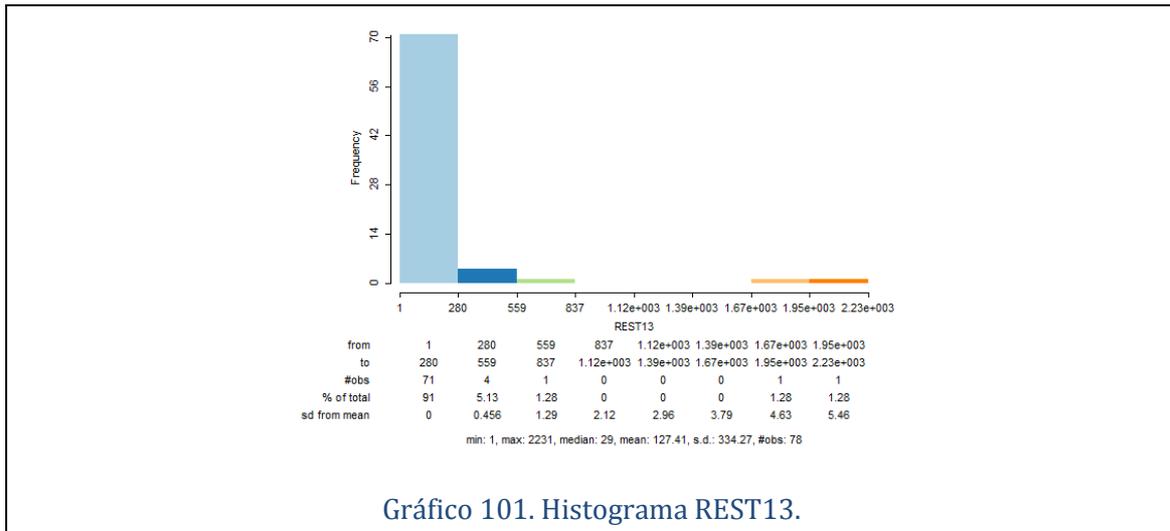
Mapa 191. Mapa LISA  $G_i^*$  EXPLOT14.

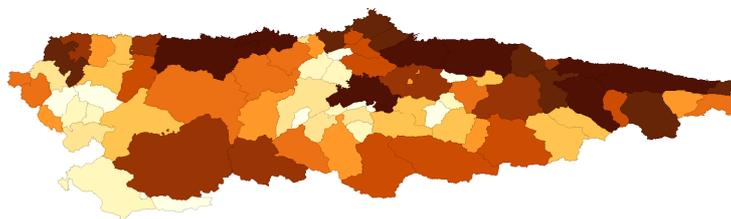


Mapa 192. Mapa LISA  $G_i^*$  ANIM14.



Mapa 193. Mapa LISA  $G_i^*$  CORTMAD12.





Mapa 195. Mapa de deciles PLAZTUR13.

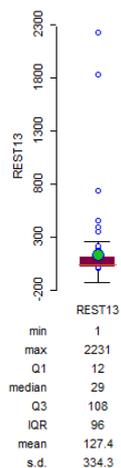


Gráfico 103. Diagrama de caja REST13.

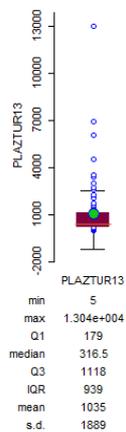
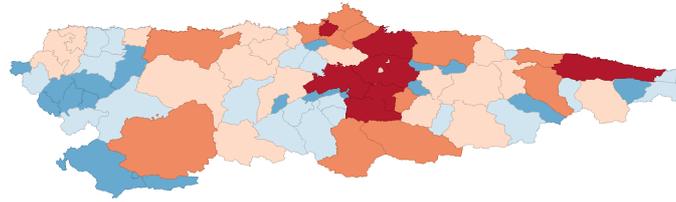
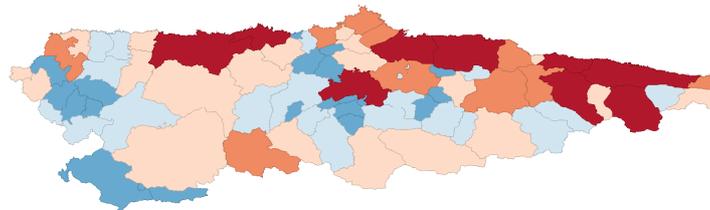


Gráfico 104. Diagrama de caja PLAZTUR13.



Mapa 196. Mapa de caja REST13.



Mapa 197. Mapa de caja PLAZTUR13.

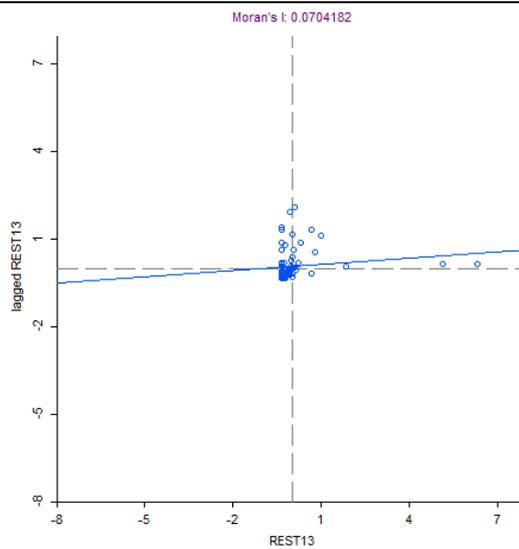
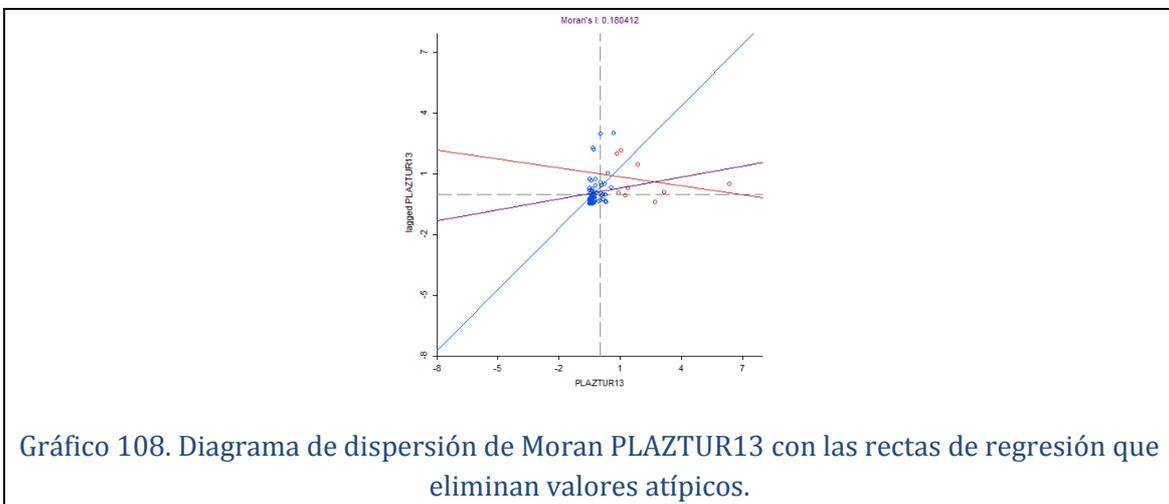
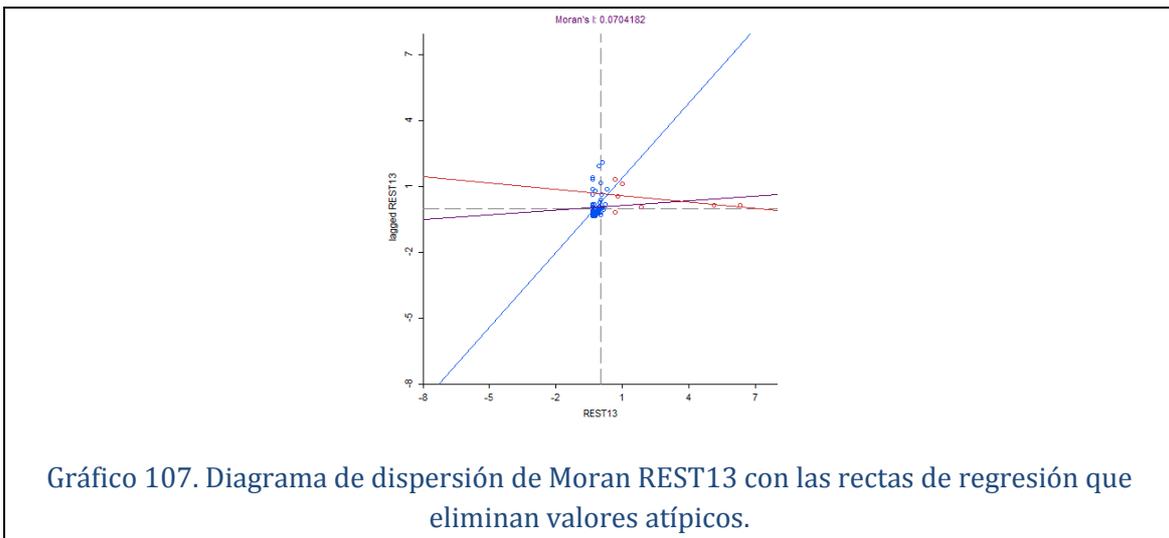
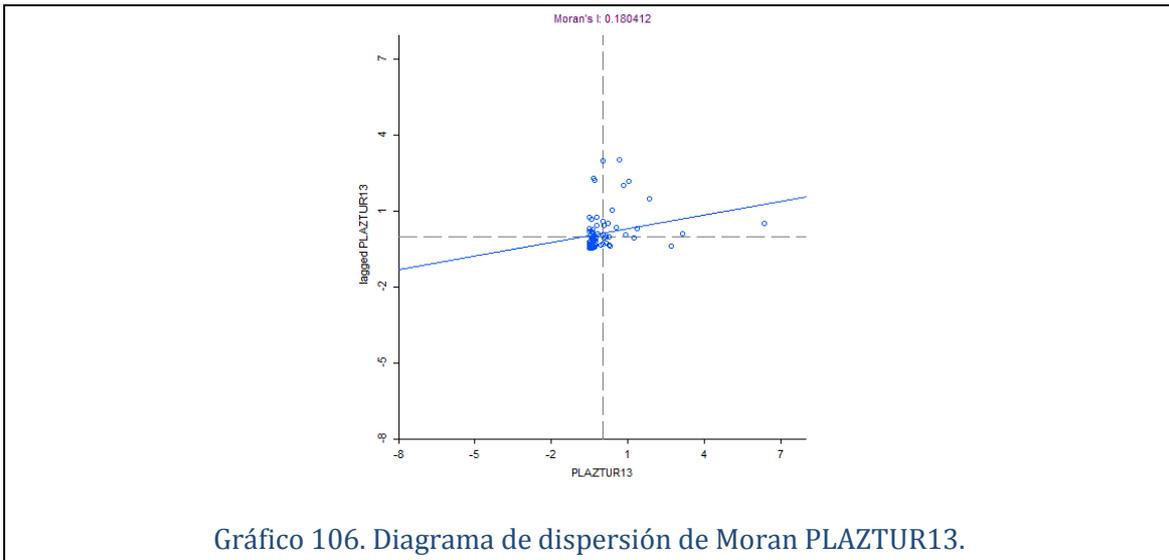
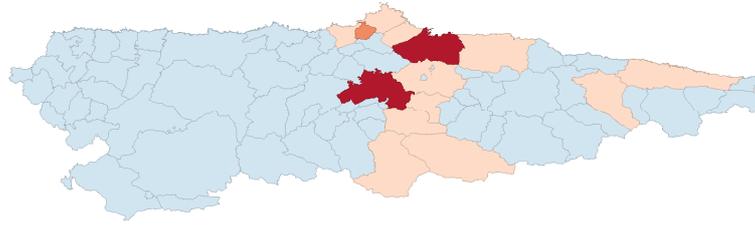
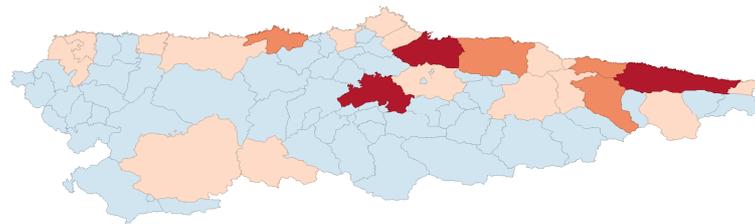


Gráfico 105. Diagrama de dispersión de Moran REST13.

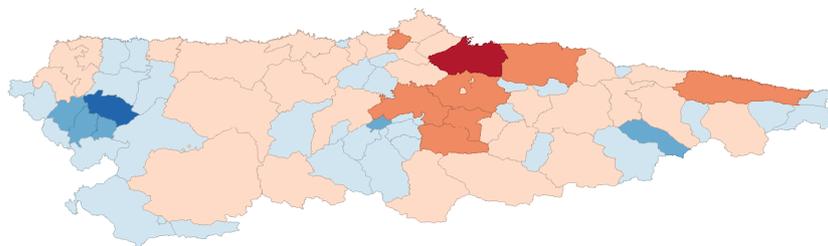




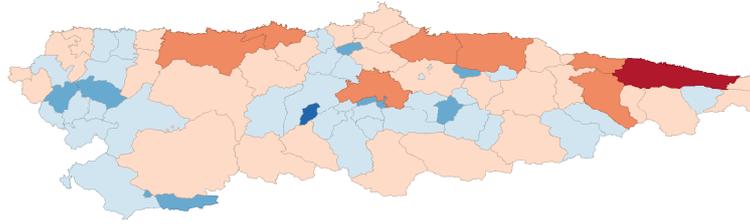
Mapa 198. Mapa de desviación típica REST13.



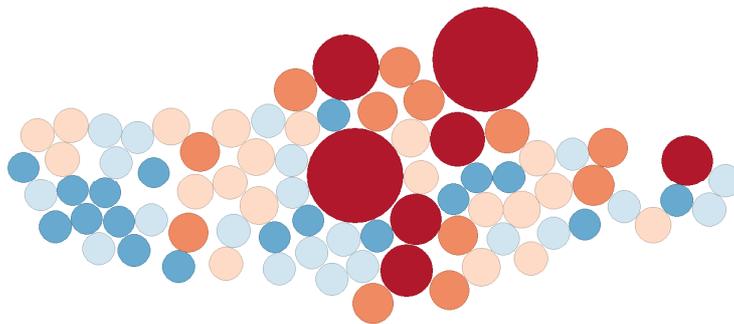
Mapa 199. Mapa de desviación típica PLAZTUR13.



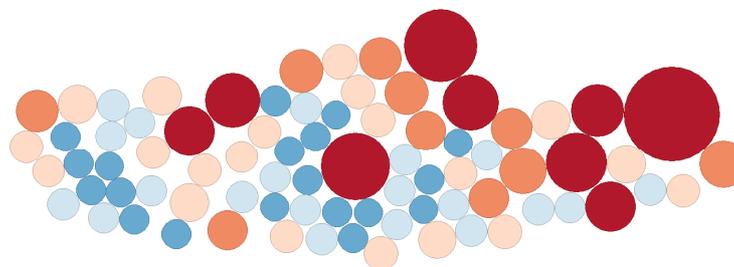
Mapa 200. Mapa de percentiles REST13.



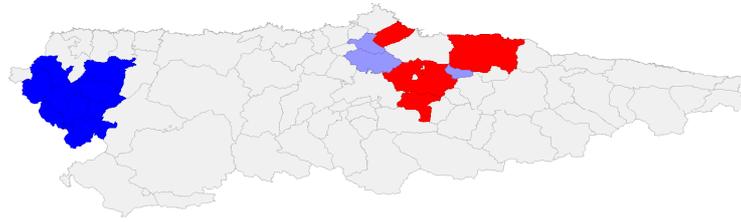
Mapa 201. Mapa de percentiles PLAZTUR13.



Mapa 202. Cartograma REST13.



Mapa 203. Cartograma PLAZTUR13.



Mapa 204. Mapa LISA I de Moran REST13.



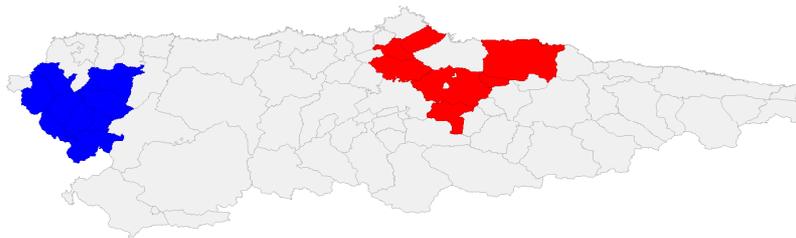
Mapa 205. Mapa LISA I de Moran PLAZTUR13.



Mapa 206. Mapa LISA G<sub>1</sub> REST13.



Mapa 207. Mapa LISA  $G_i$  PLAZTUR13.



Mapa 208. Mapa LISA  $G_i^*$  REST13.



Mapa 209. Mapa LISA  $G_i^*$  PLAZTUR13.

**ANEXO 2****CUADRO 4.3. TEST DE CHOW. MODELO 1. IPDR13.07**

<b>REGIMES DIAGNOSTICS - CHOW TEST</b>			
<b>VARIABLE</b>	<b>DF</b>	<b>VALUE</b>	<b>PROB</b>
CONSTANT	4	9.540	0.0489
IPDR13.07	4	14.664	0.0055
RNET10	4	67.452	0.0000
VEMP13.07	4	122.435	0.0000
lambda	4	46.122	0.0000
Global test	20	752.221	0.0000

**CUADRO 4.4. TEST DE CHOW. MODELO 1. I1PIL13.07**

<b>REGIMES DIAGNOSTICS - CHOW TEST</b>			
<b>VARIABLE</b>	<b>DF</b>	<b>VALUE</b>	<b>PROB</b>
CONSTANT	4	20.630	0.0004
I1PIL	4	25.841	0.0000
RNET10	4	112.432	0.0000
VEMP	4	223.313	0.0000
W_VPOBPAD	4	13.771	0.0081
lambda	4	30.254	0.0000
Global test	24	726.938	0.0000

CUADRO 4.5. TEST DE CHOW. MODELO 1. IPAC13.07

REGIMES DIAGNOSTICS - CHOW TEST			
VARIABLE	DF	VALUE	PROB
CONSTANT	4	12.680	0.0130
IPAC	4	24.074	0.0001
RNET10	4	162.785	0.0000
VEMP	4	227.782	0.0000
W_VPOBPAD	4	11.902	0.0181
lambda	4	38.622	0.0000
Global test	24	949.437	0.0000

CUADRO 4.6. TEST DE CHOW. MODELO 1. IPDR13.07. SIN OUTLIERS.

REGIMES DIAGNOSTICS - CHOW TEST			
VARIABLE	DF	VALUE	PROB
CONSTANT	4	3.655	0.4547
CONTRAT13	4	20.143	0.0005
IPDR	4	16.466	0.0025
VEMP	4	16.939	0.0020
lambda	4	2.106	0.7163
Global test	20	108.271	0.0000

CUADRO 4.7. TEST DE CHOW. MODELO 1. I1PIL13.07. SIN OUTLIERS.

REGIMES DIAGNOSTICS - CHOW TEST			
VARIABLE	DF	VALUE	PROB
CONSTANT	4	4.047	0.3997
I1PIL	4	7.395	0.1164
VAB10	4	12.823	0.0122
VEMP	4	18.361	0.0010
Global test	16	74.470	0.0000

CUADRO 4.8. TEST DE CHOW MODELO 1. IPAC13.07. SIN OUTLIERS.

REGIMES DIAGNOSTICS - CHOW TEST			
VARIABLE	DF	VALUE	PROB
CONSTANT	4	12.353	0.0149
CONTRAT13	4	26.807	0.0000
IPAC	4	14.034	0.0072
VEMP	4	4.379	0.3571
W_VPOBPAD	4	4.619	0.3287
lambda	4	8.950	0.0624
Global test	24	206.722	0.0000

CUADRO 4.9. TEST DE CHOW. MODELO 2. IPDR13.07

REGIMES DIAGNOSTICS - CHOW TEST			
VARIABLE	DF	VALUE	PROB
CONSTANT	4	8.118	0.0873
IPDR	4	5.287	0.2591
SUPTECHIND	4	6.496	0.1650
VAB10	4	106.596	0.0000
Global test	16	518.336	0.0000

CUADRO 4.10. TEST DE CHOW. MODELO 2. I1PIL13.07 e IPAC13.07

REGIMES DIAGNOSTICS - CHOW TEST			
VARIABLE	DF	VALUE	PROB
CONSTANT	4	12.613	0.0133
SUPTECHIND	4	14.933	0.0048
VAB10	4	197.582	0.0000
lambda	4	10.038	0.0398
Global test	16	345.766	0.0000

CUADRO 4.11. TEST DE CHOW. MODELO 2. IPDR13.07, I1PIL13.07 E IPAC13.07. SIN OUTLIERS.

REGIMES DIAGNOSTICS - CHOW TEST			
VARIABLE	DF	VALUE	PROB
CONSTANT	4	14.610	0.0056
VAB10	4	23.020	0.0001
Global test	8	46.694	0.0000

CUADRO 4.12. TEST DE CHOW. MODELO 3. IPDR13.07

REGIMES DIAGNOSTICS - CHOW TEST			
VARIABLE	DF	VALUE	PROB
CONSTANT	4	12.861	0.0120
ASS14	4	21.007	0.0003
IPDR	4	9.573	0.0483
VABAGRI10	4	6.555	0.1613
VEMP	4	29.506	0.0000
W_VEMPAGRI	4	11.493	0.0215
lambda	4	16.662	0.0022
Global test	28	500.018	0.0000

CUADRO 4.13. TEST DE CHOW. MODELO 3. I1PIL13.07

REGIMES DIAGNOSTICS - CHOW TEST			
VARIABLE	DF	VALUE	PROB
CONSTANT	4	39.628	0.0000
ASS14	4	5.789	0.2155
I1PIL	4	51.211	0.0000
VABAGRI10	4	9.754	0.0448
VEMP	4	17.470	0.0016
W_VEMPAGRI	4	19.092	0.0008
lambda	4	17.672	0.0014
Global test	28	482.575	0.0000

CUADRO 4.14. TEST DE CHOW. MODELO 3. IPAC13.07

REGIMES DIAGNOSTICS - CHOW TEST			
VARIABLE	DF	VALUE	PROB
CONSTANT	4	20.690	0.0004
ASS14	4	10.676	0.0305
IPAC	4	38.963	0.0000
VABAGRI10	4	11.441	0.0220
VEMP	4	17.475	0.0016
W_VEMPAGRI	4	11.758	0.0192
lambda	4	12.123	0.0165
Global test	28	565.992	0.0000

CUADRO 4.15. TEST DE CHOW. MODELO 3. IPDR13.07. SIN OUTLIERS.

REGIMES DIAGNOSTICS - CHOW TEST			
VARIABLE	DF	VALUE	PROB
CONSTANT	4	4.163	0.3845
IPDR	4	34.443	0.0000
VABAGRI10	4	77.213	0.0000
lambda	4	55.227	0.0000
Global test	16	165.475	0.0000

CUADRO 4.16. TEST DE CHOW. MODELO 3. I1PIL13.07. SIN OUTLIERS.

REGIMES DIAGNOSTICS - CHOW TEST			
VARIABLE	DF	VALUE	PROB
CONSTANT	4	10.811	0.0288
EXPLOT14	4	12.377	0.0148
I1PIL	4	8.512	0.0745
LIAEAGRI13	4	22.173	0.0002
VABAGRI10	4	25.452	0.0000
lambda	4	13.782	0.0080
Global test	24	162.261	0.0000

CUADRO 4.17. TEST DE CHOW. MODELO 4. IPDR13.07

REGIMES DIAGNOSTICS - CHOW TEST			
VARIABLE	DF	VALUE	PROB
CONSTANT	4	2.875	0.5790
IPDR	4	14.849	0.0050
VPAR	4	28.341	0.0000
Global test	12	58.341	0.0000

CUADRO 4.18. TEST DE CHOW. MODELO 4. I1PIL13.07

REGIMES DIAGNOSTICS - CHOW TEST			
VARIABLE	DF	VALUE	PROB
CONSTANT	4	2.531	0.6391
I1PIL	4	10.473	0.0332
VPAR	4	30.617	0.0000
lambda	4	10.588	0.0316
Global test	16	44.973	0.0001

CUADRO 4.19. TEST DE CHOW. MODELO 4. IPAC13.07

REGIMES DIAGNOSTICS - CHOW TEST			
VARIABLE	DF	VALUE	PROB
CONSTANT	4	1.620	0.8052
IPAC	4	11.273	0.0237
VPAR	4	42.871	0.0000
Global test	12	48.797	0.0000

CUADRO 4.20. TEST DE CHOW. MODELO 4. IPDR13.07. SIN OUTLIERS.

REGIMES DIAGNOSTICS - CHOW TEST			
VARIABLE	DF	VALUE	PROB
CONSTANT	4	16.998	0.0019
IPDR	4	20.035	0.0005
VABAGRI10	4	49.251	0.0000
lambda	4	43.058	0.0000
Global test	16	130.878	0.0000

CUADRO 4.21. TEST DE CHOW. MODELO 4. I1PIL13.07. SIN OUTLIERS.

REGIMES DIAGNOSTICS - CHOW TEST			
VARIABLE	DF	VALUE	PROB
CONSTANT	4	1.567	0.8146
I1PIL	4	10.137	0.0382
VPAR	4	21.256	0.0003
lambda	4	53.590	0.0000
Global test	16	80.505	0.0000

CUADRO 4.22. TEST DE CHOW MODELO 4. IPAC13.07. SIN OUTLIERS.

REGIMES DIAGNOSTICS - CHOW TEST			
VARIABLE	DF	VALUE	PROB
CONSTANT	4	0.627	0.9601
IPAC	4	17.594	0.0015
VPAR	4	19.222	0.0007
Global test	12	40.131	0.0001

