



---

# Universidad de Valladolid

FACULTAD DE CIENCIAS SOCIALES, JURÍDICAS Y DE LA COMUNICACIÓN

Grado en Administración y Dirección de Empresas

TRABAJO FIN DE GRADO

## **DEPENDENCIA ESPACIAL EN TASAS DE DESEMPLEO: UN ANÁLISIS PARA ESPAÑA**

Presentado por SHEILA PALACIOS RUBIO

Tutelado por ALFONSO MORALDE BLAS

Segovia, 12 de Noviembre de 2017



## ÍNDICE

<b>INTRODUCCIÓN</b> .....	<b>4</b>
<b>CAPITULO 1</b>	
<b>ANTECEDENTES BIBLIOGRÁFICOS</b>	
<b>1.1. Métodos de medición de la dependencia espacial</b> .....	<b>8</b>
<b>1.2. Referencias Bibliográficas</b> .....	<b>8</b>
<b>CAPITULO 2</b>	
<b>BASE DE DATOS Y METODOLOGÍA</b>	
<b>2.1. Base de datos</b> .....	<b>12</b>
<b>2.1.1. Variables</b> .....	<b>12</b>
<b>2.1.1.1. Desempleo</b> .....	<b>12</b>
<b>2.1.1.2. Retardo espacial sin ponderar</b> .....	<b>14</b>
<b>2.1.1.3. Retardo espacial ponderado</b> .....	<b>15</b>
<b>2.1.1.4. Peso sectorial</b> .....	<b>15</b>
<b>2.2. Metodología</b> .....	<b>16</b>
<b>CAPITULO 3</b>	
<b>RESULTADOS</b>	
<b>Resultados</b> .....	<b>20</b>
<b>CAPITULO 4</b>	
<b>CONCLUSIONES</b>	
<b>Conclusiones</b> .....	<b>28</b>
<b>REFERENCIAS BIBLIOGRAFICAS</b> .....	<b>30</b>
<b>ANEXO I</b> .....	<b>31</b>

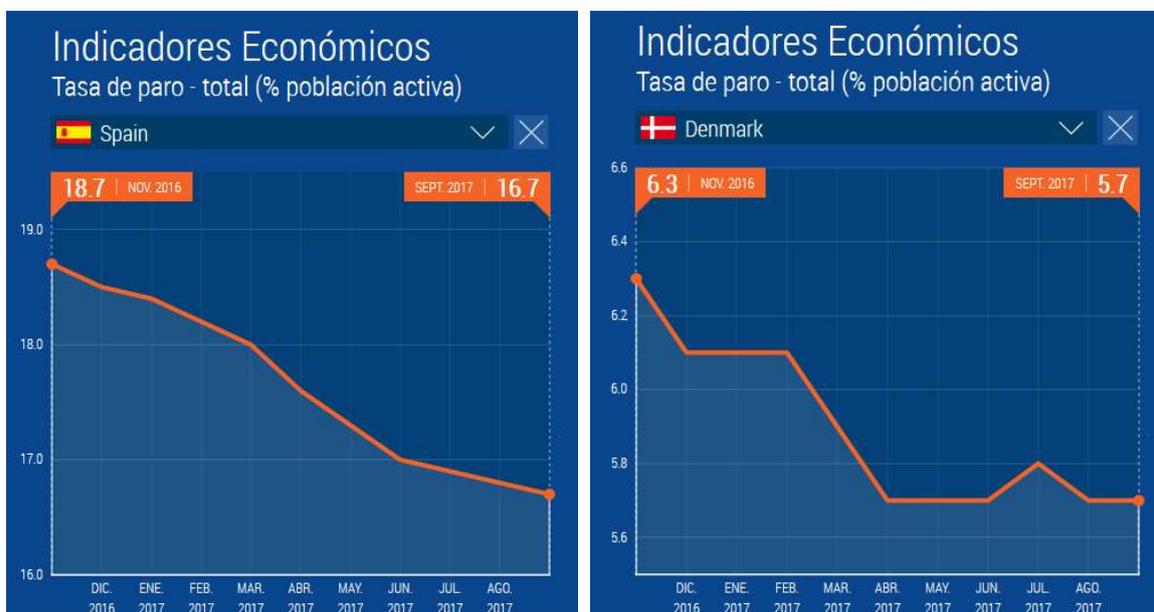
# INTRODUCCIÓN

El objetivo del presente trabajo fin de grado es analizar la influencia de los territorios vecinos sobre el valor alcanzado por la tasa de paro provincial. Este efecto está relacionado con el concepto de dependencia espacial que hace referencia a la influencia del territorio sobre cualquier variable económica.

La razón por la que he escogido este tema para mi trabajo fin de grado es porque gracias a las recomendaciones del tutor me adentró en la curiosidad de comprobar si realmente el entorno vecino influye o no en el valor de las tasas de paro. Al leer estudios he podido comprobar la importancia territorial a la hora de ver las tasas de desempleo; y gracias a estos estudios y a las conclusiones que de ellos se extraen, se puede argumentar la influencia del entorno sobre las variables económicas de un determinado territorio geográfico.

Por otro lado, la simple observación de los datos públicos permite apreciar de manera general tanto los niveles de las tasas de paro españolas, como las diferencias entre distintos territorios. Más concretamente, si se compara la tasa de paro de España de Septiembre de 2017 y la de Dinamarca para la misma fecha, se puede comprobar cómo los valores de España triplican los que presenta una de las economías más prósperas de la UE.

**Gráfico 1. Tasas de paro España y Dinamarca.**



Fuente: INE

Más adelante se verán datos más concretos del caso de España a nivel provincial, en las cuales se verán también las diferencias existentes entre los distintos territorios.

De forma concreta, en este estudio se van analizar los siguientes aspectos:

- Efecto del entorno sobre las tasas de paro: Se analizará el efecto de las tasas de paro de otras provincias vecinas sobre el valor que toma la provincia de referencia.
- Si existe influencia, comprobar si la influencia es simétrica para todas las provincias o ver si depende del peso de cada provincia. Es decir, se analizará si el tamaño de las provincias es relevante o no.

El presente trabajo va a estar organizado en 5 capítulos. En el primero de ellos hago referencia al concepto de dependencia espacial y a los estudios que he leído para adquirir más conocimientos y poder abordar de mejor forma la realización de este trabajo. En el segundo capítulo está constituido por dos subcapítulos, en el primero de ellos se explican los datos utilizados para desarrollar el trabajo; y en el segundo se definen las variables empleadas y la forma en que los datos han sido tratados para llegar a los resultados. El tercer capítulo muestra los resultados obtenidos, gracias al tratamiento de los datos. Finalmente, el penúltimo capítulo expone las conclusiones extraídas gracias a la interpretación de los resultados.



## **CAPITULO 1**

### **ANTECEDENTES BIBLIOGRÁFICOS**

Este primer capítulo está estructurado en dos subcapítulos. En el primero de ellos se aborda la definición de la dependencia espacial para conocer el efecto que vamos a estudiar.

Po su parte, en el segundo subcapítulo se va hacer referencia a los estudios leídos previamente a la elaboración del presente trabajo y que tienen relación con los contenidos del presente trabajo.

## **1.1 MÉTODOS DE MEDICIÓN DE LA DEPENDENCIA ESPACIAL**

Para abordar el tema de la dependencia espacial primero tenemos que hacer referencia a los datos con los que se han trabajado.

Para el desarrollo de este trabajo se van a utilizar datos espaciales. Los datos espaciales son aquellos que tienen una referencia geográfica y que se pueden ver representados geográficamente. En este caso he utilizado datos de paro de provincias, de los cuales se pueden realizar representaciones gráficas como veremos en algún apartado sucesivo. Una vez que hemos explicado con que datos vamos a trabajar, ya podemos abordar el tema de la dependencia espacial.

Consideramos que existe dependencia espacial cuando la variable estudiada en el proceso está relacionada o se ve afectada por los valores que toman otras variables. Esta dependencia espacial es positiva si la evolución de la variable va en el mismo sentido que la de las provincias vecinas, o negativa si va en sentido contrario. En nuestro caso existirá dependencia espacial si la tasa de paro de una provincia se ve afectada o condicionada por las tasas de paro de las provincias vecinas. Si una provincia se rodea de otras con tasas de paro similares la dependencia será positiva y si lo hace de otras tasas de paro muy diferentes será negativa.

Para establecer el criterio de vecindad hemos considerado como vecinas las provincias que pertenecen a la misma Comunidad Autónoma a la que pertenece la provincia estudiada.

Existen casos en los cuales no tenemos provincias vecinas como son: Madrid, Asturias, Cantabria, Navarra, La Rioja, Murcia y las Islas Baleares; ya que son uniprovincias.

Al leer el estudio de Moreno Serrano y Vayá Valcarce (2002) compruebo que existen diversos índices para medir la dependencia espacial entre los que podemos destacar los proporcionados por Geary, Getis y Ord y por Moran. El índice al que más referencia hacían los estudios que he leído era el índice de Moran.

En nuestro caso no voy a utilizar estos índices ya que abordan de manera más profunda la dependencia espacial. El trabajo sigue una perspectiva diferente al escogido por Moreno Serrano y Vayá Valcarce (2002). El método utilizado en el presente trabajo se basa en regresar la tasa de paro de una provincia con respecto al valor medio de sus provincias vecinas.

## **1.2 REFERENCIA BIBLIOGRÁFICAS**

Antes de ponerme a la realización de las estimaciones he recopilado información como he señalado antes de distintos estudios para conocer más el tema desde una visión más general y poder entender de forma completa que estaba realizando y ver si se cumplían los objetivos de este trabajo.

## ANTECEDENTES BIBLIOGRÁFICOS

El primero de ellos, y el que más se acerca al tema tratado fue publicado por Moreno Serrano y Vayá Valcarce (2002). Definen el surgimiento de la dependencia o autocorrelación espacial por la relación de un valor de una variable con su valor en otro u otros lugares del espacio. Señalan que la econometría espacial surgió como una disciplina derivada de la econometría general, y que gracias a esta subdivisión se pueden trabajar con datos que presentan problemas de heterogeneidad y/o dependencia espacial.

Tratan de comprobar si existe dependencia entre regiones europeas para la productividad laboral, y extraen dos conclusiones:

- La primera de ellas es que sí existe dependencia espacial para regiones vecinas.
- Como segunda conclusión dicen que existen importantes problemas de dependencia espacial en la ecuación clásica de convergencia que tienen que ser considerados de manera explícita ya que afectan a los resultados. Y animan a los investigadores a realizar una revisión sobre las técnicas aplicadas en el trabajo.

En el estudio realizado por Alonso, Del Río y Toharia (2009) concluyen que existen diferencias en las tasas de desempleo en función del tamaño de los distintos municipios. Pertenecientes al territorio español. Alonso et al. (2009) distinguen también entre municipios pequeños, en los cuales los niveles de dispersión de las tasas de paro son mayores; municipios medianos en los que las tasas de paro son más elevadas y municipios grandes en los que las tasas y la dispersión toman menores niveles en comparación con los demás. También destacan que las ciudades grandes existe menos desempleo que en las ciudades medianas, en las que señalan que más del 25% de la población está en situación de desempleo.

En el año 2013, Buendía Azorín y Sánchez de la Vega en su estudio sobre La distribución del desempleo en las provincias españolas: Un análisis con datos de panel mediante el filtrado de espacial, del cual desarrollan las siguientes conclusiones:

- Como principal conclusión determinan que existe una persistente desigualdad en las tasas de desempleo de las provincias.
- Las tasas de paro relativas son caracterizadas por una creciente y fuerte dependencia espacial en la distribución provincial.
- Las provincias vecinas, que tienen características específicas similares con la provincia que se estudia en cada caso influyen en la tasa de paro de dicha provincia.

Y en el último estudio en el que me he fijado realizado por Cuéllar, Martín y Moral (2016) han conseguido dividir la tasa de desempleo efectiva en dos componentes uno cíclico y otro natural.

Cuéllar et al. (2016) detectan una autocorrelación espacial global significativa y positiva en las tasas de desempleo provinciales y llegan a la conclusión de que en la península española se pueden distinguir dos grandes zonas: la zona 1 compuesta por el nordeste de la península en la que existe un bajo desempleo y la zona 2 en la que existe un mayor desempleo y está compuesta por las provincias situadas en el extremo sur de la península.



## **CAPITULO 2**

### **BASE DE DATOS Y METODOLOGÍA**

Este capítulo también se divide en dos apartados: uno donde se analiza la base de datos y otro dedicado a la metodología. En el primero de ellos da a conocer los datos con los que se han trabajado y explica de donde han sido proporcionados. En el segundo de ellos trata de explicar cómo se han trabajado con los datos para llegar a los resultados.

## 2.1 BASE DE DATOS

Lo primero para poder empezar a realizar las estimaciones es tener una base de datos con los que poder trabajar. Los datos empleados en este trabajo provienen del Instituto Nacional de Estadística (INE), más concretamente hemos empleado los proporcionados por la Encuesta de Población Activa (EPA). De aquí he extraído los datos para hallar la tasa de parados para cada provincia española de los cuatro trimestres para el año 2006 y el año 20013; también he extraído los datos de ocupados en miles de personas.

### 2.1.1 VARIABLES

Para conocer más los datos con los que se han trabajado este capítulo se explican una por una las variables que se han utilizado para obtener algunos de los datos empleados a la hora de realizar las estimaciones.

#### 2.1.1.1. DESEMPLEO

Primero vamos a identificar la variable con la que se ha trabajado. El desempleo también conocido como paro se puede definir como el número de personas que están buscando un empleo activamente, esto quiere decir que forman parte del desempleo las personas que no tienen un trabajo y están buscando uno. No podemos considerar a todas las personas que no tengan un trabajo como desempleadas, ya que hay gente que no tiene trabajo pero tampoco lo está buscando.

La tasa de desempleo es un indicador porcentual de las personas que están desempleados que pertenecen a la población activa y se calcula de la siguiente forma:

$$Tasa\ de\ desempleo = \frac{Número\ de\ trabajadores\ desempleados}{Población\ activa}$$

Para poder conocer todas las variables que se incluyen en la fórmula hemos de destacar que la población activa es la suma de las personas que están empleadas y las personas que están desempleadas.

$$Población\ Activa = Ocupados + Desempleados$$

Los sistemas que se utilizan para medir el desempleo son dos:

- La Encuesta de Población Activa (EPA).
- El paro registrado.

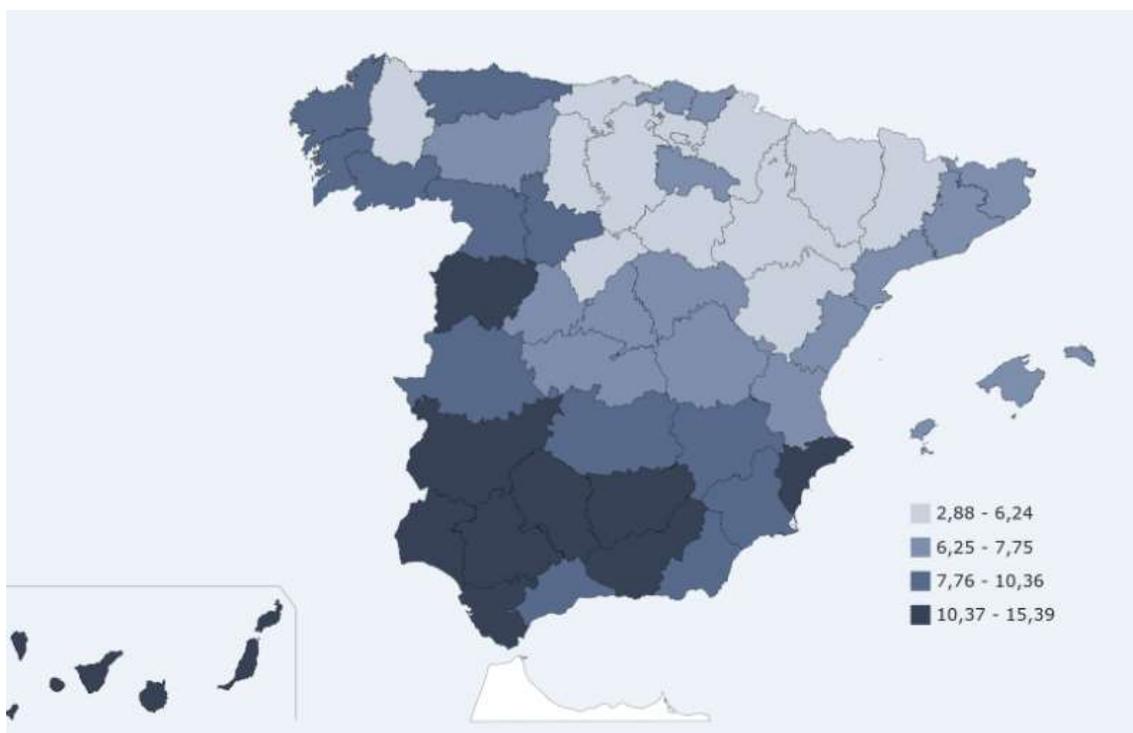
Los datos correspondientes a la tasa de desempleo que hemos utilizado para realizar las estimaciones en las que nos hemos basado han sido sacados de la base de datos del INE. Estos datos son proporcionados por la EPA, y según un estudio (Pérez Infante, 2008) se considera más fiable ya que entre los dos medidores anteriormente citados existen discrepancias.

Las diferencias que existen entre los dos medidores se deben a dos aspectos de acuerdo a Pérez Infante (2008), ¿Por qué el paro registrado supera al paro de la EPA?:

- Punto de vista metodológico:  
La EPA se elabora trimestralmente por el INE que realiza a ciertos hogares familiares a las personas en edad de trabajar y el Paro registrado se realiza mensualmente por el SEPE (Servicio Público de Empleo Estatal) en función de las personas registradas cada mes.
- Punto de vista perceptual:  
La EPA tiene que seguir la normativa fijada por la OIT (Organización Internacional del Trabajo) y también la fijada por la Comisión Europea, y el Paro registrado no tiene que seguir unas normativas tan estrictas.

Los datos de paro se corresponden con los años 2006 y 2013 para ver el efecto de la dependencia espacial en un año anterior a la crisis como es el 2006 y para un año posterior como es el 2013.

**Mapa 2. Tasa de paro para el trimestre 4 del año 2006**

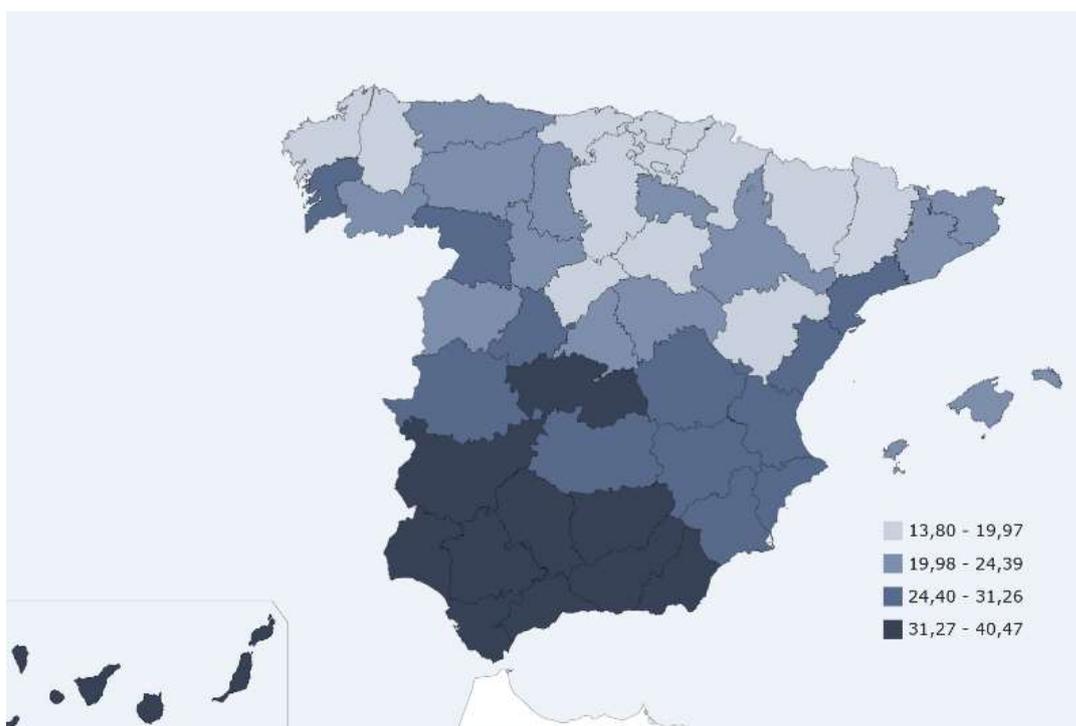


Fuente: INE

Desde un punto de vista meramente geográfico, el mapa uno nos permite comprobar cómo en representación geográfica de las tasas de paro para el año 2006 existen dos zonas muy diferenciadas dependiendo de la tasa de desempleo de las mismas: la zona norte en la que se concentran tasas más bajas y en la zona sur donde podemos ver que las tasas son más alta.

Si nos fijamos en el mapa 2 nos muestra la misma situación pero para el año 2013 en la que se siguen distinguiendo dos zonas muy claras. La diferencia que podemos encontrar entre ambos mapas son los valores que toma la variable. En el año 2006 las tasas de paro son más bajas debido a que España se encontraba en una época de crecimiento. Sin embargo en el año 2013 España se encontraba en una época de recesión.

**Mapa 2. Tasa de paro para el trimestre 4 del año 2013**



Fuente: INE

Hoy en día es un tema de actualidad las tasas de desempleo que hay en España tomando un valor del 18,75% en el primer trimestre de 2017. Es un tema de actualidad debido a que España entró en crisis en el 2008 debido a una burbuja especulativa de la construcción situando la tasa de paro para ese año en un 13,79% en el último trimestre del año. En nuestro estudio queremos ver el efecto para un año anterior a esta situación de crisis y un año posterior; por ello hemos escogido el año 2006 y el año 2013. En el año 2006 la tasa de paro se situaba en torno a un 8,26% y en el año 2013 se correspondía con un 25,73%. En esta progresión en el tiempo se observa como la tasa de paro se incrementó bastante y a día de hoy la tasa sigue siendo bastante alta.

## 2.1.1.2. RETARDO ESPACIAL SIN PONDERAR

Para hallar esta variable he trabajado con los datos obtenidos de la tasa de desempleo de todas las provincias que pertenecen a cada comunidad. En concreto, esta variable mide la media de la tasa de paro de todas las provincias pertenecientes a una misma comunidad autónoma y que son distintas de la de referencia. Está sin ponderar ya que no tenemos en cuenta el distinto tamaño que tiene cada provincia, sino que consideramos que todas las provincias tienen el mismo peso. El cálculo concreto de esta variable se hace de acuerdo a la siguiente expresión:

$$\text{Ret. esp. sin ponderar} = \frac{\sum \text{Tasa de paro de las provincias vecinas}}{\text{N}^{\circ} \text{ de las provincias de la CC. AA a la que pertenece}}$$

## 2.1.1.3. RETARDO ESPACIAL PONDERADO

Esta variable es muy similar a la anterior, pero ahora tenemos en cuenta que no todas las provincias son iguales. Con esta variable conseguimos dar a cada provincia el peso que le corresponde en función de su tamaño, ya que se tiene en cuenta la tasa de desempleo de cada provincia. Para calcular el peso correspondiente a cada una de ellas utilizamos una ponderación basada en el número de ocupados. En concreto, el peso de cada provincia se obtiene de acuerdo a la siguiente expresión:

$$\text{Ponderación ocupados} = \frac{\text{Ocupados (miles de personas) de la provincia estudiada}}{\sum \text{Ocupados (miles de personas) de la CC. AA a la que pertenece}}$$

A partir de esas ponderaciones, el retardo espacial ponderado se obtiene de la siguiente manera:

$$\text{Ret. esp. ponderado} = \frac{\sum (\text{Pond ocupados provincias vecinas} \times \text{Tasa paro provincias vecinas})}{\sum \text{Ponderación provincias vecinas}}$$

## 2.1.1.4. PESO SECTORIAL

Finalmente, también se construye una variable que refleje la distinta composición sectorial de cada territorio. Para ello extraemos del INE el número de ocupados que pertenecen a cada sector de actividad de cada provincia en miles de personas. Una vez con estos datos hallamos el tanto por ciento que tiene cada sector en cada provincia, con esto estamos considerando ahora el peso de cada sector económico de cada provincia y con ello podremos ver si el pertenecer a un sector concreto influye en la tasa de paro.

Para hallar el tanto por ciento de cada sector utilizamos la siguiente fórmula que voy a poner para el caso de la agricultura pero sería para cada uno de los sectores igual:

$$\% \text{ Agricultura} = \frac{N^{\circ} \text{ de ocupados Agricultura (miles de personas) de la provincia}}{\sum N^{\circ} \text{ ocupados de cada sector de la provincia}}$$

## 2.2 METODOLOGÍA

Con este trabajo se intenta medir la influencia del entorno en la tasa de desempleo. Para ello un primer paso es escoger un criterio de vecindad. Como se ha comentado previamente hemos considerado vecinas a las provincias que pertenecen a la misma Comunidad Autónoma. Por lo tanto las comunidades uniprovinciales no se les asigna vecindad.

Con los datos extraídos del INE he creado un libro de Excel, en el cuál he calculado la tasa de paro anual para cada provincia para después calcular las variables que voy a introducir en las estimaciones de los modelos para proceder al cumplimiento de los objetivos del presente trabajo. Una vez creada la base de datos que contiene las variables necesarias para el estudio se han importado los datos al programa econométrico Eviews.

Con este programa se estiman diferentes modelos econométricos, en este caso concreto se han creado 4 modelos distintos para los años 2006 y 2013 de los cuales se extraen 8 estimaciones distintas. Cada modelo econométrico se realiza mediante la estimación de mínimos cuadrados. Dichos modelos se componen por una variable independiente y una o varias variables dependientes.

Un modelo econométrico sirve para ver si una o varias variables explican otra variable que se denominaría variable dependiente. En este caso la variable dependiente es la tasa de paro y las variables explicativas dependerán del modelo en el que nos fijemos, pero serán: el retardo espacial sin ponderar, el retardo espacial ponderado, ocupados sector servicios, ocupados industria, y ocupados construcción.

Las variables que hacen referencia al sector económico las hemos utilizado para apoyarnos en la composición sectorial de cada provincia y ver el mayor o menor peso de un sector influye también en la tasa de paro.

Al realizar los modelos nos basamos en el método de los mínimos cuadrados y comparamos nuestros resultados asociados a cada variable utilizando el estadístico T-Student que es el siguiente:

$$t = \frac{\widehat{\beta}_x}{S_{\widehat{\beta}_x}} \sim t_{N-k-1}$$

$H_0: \beta_1 = 0$  Variable No Significativa

$H_1: \beta_1 \neq 0$  Variable Significativa

Para ello vamos a comparar con un p-valor de 0,05 esto quiere decir que si el de la variable estudiada es mayor que 0,05 se acepta la hipótesis nula  $H_0$ , entonces estaríamos aceptando que la variable no es significativa y esto quiere decir que si

quitamos la variable del modelo el modelo podría seguir siendo válido ya que no afecta en nuestro modelo.

Si el p-valor es menor que 0,05 estaríamos rechazando la hipótesis nula  $H_0$ , y aceptaríamos  $H_1$ , lo que quiere decir es que la variable si es significativa y que tiene importancia en nuestro modelo.

Para la significación conjunta (esto quiere decir para ver si es un buen modelo o no) utilizamos el estadístico F de Snedecor que se define de la siguiente manera:

$$F = \frac{e' r e_r - e' e}{S^2 k} \sim F_{N-k-1}^k$$

$$H_0: \begin{pmatrix} \beta_{x1} \\ \beta_{x2} \\ \beta_{x3} \\ \beta_{xn} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \text{ Modelo No Significativo en conjunto}$$

$$H_1: \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_{xn} \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \text{ Modelo Significativo en conjunto}$$

Para hacer las comparaciones utilizamos el mismo valor crítico (p-valor de 0,05). Aceptaríamos la hipótesis nula  $H_0$ , cuando el p-valor sea mayor de 0,05, esto quiere decir que el modelo no es bueno y no seguiríamos estudiando dicho modelo ya que las conclusiones que se obtengan de él no van a ser válidas. Por lo contrario si el p-valor es menor de 0,05 estaríamos rechazando  $H_0$ , y aceptaríamos  $H_1$  que quiere decir que el modelo es significativo en conjunto y que el modelo es bueno, por lo tanto las conclusiones que se saquen de dicho modelo estarían bien.

Con el R-squared medimos el porcentaje de la varianza de la tasa de paro que es explicada por el modelo.



## **CAPITULO 3**

### **RESULTADOS**

Una vez que tenemos las variables definidas procedemos a trabajar con ellas en el programa econométrico Eviews.

En el ANEXO I adjunto los modelos que he obtenido de las estimaciones realizadas por si es de interés para alguien y quisiera comprobar resultados.

A continuación extraigo para cada estimación realizada los datos más importantes a tener en cuenta:

### MODELO 1

En la tabla 1 se representan los resultados del modelo 1. En ella se incluyen dos estimaciones una correspondiente al año 2006 y otra al año 2013. En este modelo incluimos únicamente la variable dependiente RETARDO SIN PONDERAR, considerando que todas las provincias tienen el mismo peso, sin darles importancia al número de ocupados que tiene cada una.

El modelo concreto a estimar es el recogido en la siguiente ecuación:

$$TASA DE PARO_t = \beta_0 + \beta_1 RET SIN PONDERAR_t + \varepsilon$$

Tabla1. Estimación de la tasa de paro sin tener en cuenta el peso de las provincias.

AÑO	VARIABLE	COEFICIENTE	P-VALOR	R <sup>2</sup>	P- VALOR F-STATISTIC
2006	RET SIN PONDERAR	0,053429	0,1416	0,044451	0,141643
2013	RET SIN PONDERAR	0,083282	0,0075	0,139623	0,007519

Fuente: Elaboración propia

En este modelo correspondiente al año 2006 podemos ver como el modelo no es significativo en conjunto porque tiene un p-valor = 0,141643, esto quiere decir que no es un buen modelo y no se seguirá estudiando, ya que rechazaríamos la hipótesis nula estaríamos aceptando que nuestro modelo no es significativo, aunque la variable residual si sea significativa individualmente con un valor de 0,0000.

En el modelo para el año 2013 vemos como el modelo si es significativo en conjunto, con un p-valor = 0,007519 < 0,05; por lo tanto es un buen modelo, las conclusiones que se obtengan son válidas.

## RESULTADOS

Podemos ver como aunque al introducir la variable aunque no consideremos el peso de cada provincia es significativa con un p-valor de 0,0075, lo que quiere decir es que por cada unidad que aumente el retardo espacial ponderado aumentaría en 0,0832829 unidades la tasa de paro de 2013.

La variabilidad de la tasa de paro está explicada en un 13,96% por las variables independientes.

## MODELO 2

El siguiente modelo sólo se diferencia del anterior en que ahora vamos a tener en cuenta del peso de cada provincia, ahora tenemos en cuenta el tamaño de cada provincia dando mayor peso a las provincias más grandes.

El modelo genérico a estimar es el siguiente:

$$\text{TASA DE PARO}_t = \beta_0 + \beta_1 \text{RET PONDERADO}_t + \varepsilon$$

Tabla2. Estimación de la tasa de paro teniendo en cuenta el peso de las provincias.

AÑO	VARIABLE	COEFICIENTE	P-VALOR	R <sup>2</sup>	P- VALOR F-STATISTIC
2006	RET PONDERADO	0,335724	0,0005	0,224074	0,000517
2013	RET PONDERADO	0,400078	0,0000	0,413991	0,000000

Fuente: Elaboración propia

Los datos correspondientes a las estimaciones de 2006 y 2013 aparecen recogidos en la tabla 2. En este nuevo modelo para el año 2006 podemos ver que ya es un modelo bueno, ya que su p-valor = 0,000517 < 0,05 y rechazamos H0, aceptando H1 que es un modelo significativo en conjunto. También podemos observar que la variable RETARDO ESPACIAL PONDERADO sí que es significativa individualmente por lo que podemos decir que afecta nuestra variable dependiente; si aumentamos en una unidad el retardo espacial ponderado aumentará la tasa de paro en 0,335724 unidades. Si nos fijamos en Adjusted R-Squared podemos decir que la variabilidad de la tasa de paro está explicada en un 22,41% por las variables independientes.

En el modelo para 2013 podemos ver también que es significativo en conjunto, es un buen modelo con un p-valor = 0,0000. Al fijarnos en la variable RETARDO ESPACIAL PONDERADO es significativa individualmente, si aumentamos en una unidad el retardo espacial ponderado aumentará en 0,400078 unidades la tasa de paro.

Al fijarnos en R-squared podemos decir que la variabilidad de la tasa de paro está explicada en un 41,13% por las variables independientes.

En conjunto podemos decir que existe dependencia espacial positiva entre las tasas de paro pertenecientes a provincias de una misma comunidad autónoma. Esta dependencia es muy superior a la del modelo 1. Esto implica que el peso de las provincias sí que es importante a la hora de medir la importancia del entorno. Finalmente también se comprueba que no hay grandes diferencias entre que estemos en una fase recesiva o expansiva a la hora de ver la importancia del espacio.

### MODELO 3

En la tabla 3 se muestran los resultados del modelo 3. En este modelo volvemos a introducir la variable retardo espacial sin ponderar, volvemos a darle la misma importancia a todas las provincias, no tenemos en cuenta su peso; también introducimos nuevas variables para tener en cuenta la composición sectorial de cada provincia y ver si este hecho afecta también a la tasa de paro.

El modelo concreto a estimar es:

$$\text{TASA PARO}_t = \beta_0 + \beta_1 \text{RET ESPC SIN POND}_t + \beta_2 \text{OCUP SERV}_t + \beta_3 \text{OCUP IND}_t + \beta_4 \text{OCUP CONS}_t + \varepsilon$$

Tabla3. Estimación de la tasa de paro sin tener en cuenta el peso de las provincias y teniendo en cuenta la composición sectorial

AÑO	VARIABLE	COEFICIENTE	P-VALOR	R <sup>2</sup>	P- VALOR F-STATISTIC
<b>2006</b>	RET ESPC SIN PONDERAR	0,011548	0,7381	0,348191	0,000581
	OCUP SERV	1,031168	0,9001		
	OCUP IND	-26,55104	0,0097		
	OCUP CONS	-7,175376	0,6639		
<b>2013</b>	RET ESPC SIN PONDERAR	0,038077	0,2123	0,378409	0,000215
	OCUP SERV	-2,574259	0,8879		
	OCUP IND	-60,77208	0,0074		
	OCUP CONS	-74,03433	0,3022		

Fuente: Elaboración propia

## RESULTADOS

Los resultados que se muestran en la tabla 3 que corresponde con 2006 se puede observar que es significativo en conjunto con un p-valor = 0,000581. Al fijarnos en las variables independientes vemos que sólo hay una significativa OCUPADOS INDUSTRIA con un p-valor = 0,0097. Las demás variables que hemos introducido no son significativas, si las quitamos del modelo no pasaría nada, incluso podría llegar a mejorar nuestro modelo.

Si aumentamos en una unidad el retardo espacial sin ponderar la tasa de paro aumentará en 0,011548 unidades; si aumentamos en una unidad los ocupados servicios la tasa de paro aumentará en 1,031158 unidades; si aumentamos en una unidad los ocupados industria la tasa de paro disminuirá en 26,55104 unidades; y si aumentamos en una unidad los ocupados construcción la tasa de paro disminuirá en 7,175376 unidades.

Si nos fijamos en R-Squared podemos decir que la variabilidad de la tasa de paro está explicada en un 34,81% por las variables independientes.

El modelo para 2013 podemos decir que con un p-valor = 0,000215 < 0,05, el modelo es significativo en su conjunto, con esto extraemos la conclusión de que el modelo es bueno, aunque si nos fijamos en las variables independientes podemos ver como sólo hay una variable significativa individualmente.

Si aumentamos en una unidad el retardo espacial sin ponderar la tasa de paro aumentará en 0,038077 unidades; si aumentamos en una unidad los ocupados servicios la tasa de paro disminuirá en 2,574259 unidades; si aumentamos en una unidad los ocupados industria la tasa de paro disminuirá en 60,77208 unidades; y si aumentamos en una unidad los ocupados construcción la tasa de paro disminuirá en 70,03433 unidades.

El R-squared nos dice que la variabilidad de la tasa de paro está explicada en un 37,84% por las variables independientes.

## MODELO 4

El siguiente modelo sólo se diferencia del anterior en que ahora vamos a tener en cuenta el peso de cada provincia, ahora tenemos en cuenta el tamaño de cada provincia dando mayor peso a las provincias más grandes.

En el último modelo volvemos a considerar el peso de cada provincia y también introducimos las variables para ver si la composición sectorial de cada provincia afecta a la tasa de paro o no.

El modelo genérico a estimar es el siguiente:

$$TASA PARO_t = \beta_0 + \beta_1 RET ESPC POND_t + \beta_2 OCUP SERV_t + \beta_3 OCUP IND_t + \beta_4 OCUP CONS_t + \varepsilon$$

Tabla4. Estimación de la tasa de paro teniendo en cuenta el peso de las provincias y teniendo en cuenta la composición sectorial

AÑO	VARIABLE	COEFICIENTE	P-VALOR	R <sup>2</sup>	P- VALOR F-STATISTIC
2006	RET ESPC PONDERADO	0,236668	0,0088	0,439903	0,000024
	OCUP SERV	3,762330	0,6094		
	OCUP IND	-20,71663	0,0244		
	OCUP CONS	-7,365062	0,6300		
2013	RET ESPC PONDERADO	0,306046	0,0001	0,552388	0,000000
	OCUP SERV	3,033378	0,8394		
	OCUP IND	-39,56521	0,0351		
	OCUP CONS	-71,62081	0,2168		

Fuente: Elaboración propia

El modelo par el año 2006 es significativo en conjunto con un p-valor = 0,000024. Al mirar las variables podemos ver que este modelo es mejor que el anterior, ya que hay dos variables significativas individualmente el Retardo espacial ponderado y los ocupados industria.

Si aumentamos en una unidad el tardo espacial ponderado aumentará en 0,236668 unidades las tasa de paro; si aumentamos en una unidad los ocupados servicios la tasa de paro aumentará en 3,762330 unidades; si aumentamos en una unidad los ocupados industria la tasa de paro disminuirá en 20,71663 unidades; y si aumentamos en una unidad los ocupados construcción la tasa de paro disminuirá en 7,365062 unidades.

Al ver R-squared nos dice que la variabilidad de la tasa de paro está explicada en un 43,99% por las variables independientes.

El modelo para 20013 obtiene un p-valor = 0,000000 < 0,05 el modelo es significativo en su conjunto, el modelo es bueno. Al fijarnos en las variables independientes vemos como el retardo espacial ponderado y los ocupados industria son las únicas variables significativas.

Si aumentamos en una unidad el retardo espacial ponderado la tasa de paro aumentará en 0,0306046 unidades; si aumentamos en una unidad los ocupados

## RESULTADOS

servicios la tasa de paro aumentará en 3,033378 unidades; si aumentamos en una unidad los ocupados industria la tasa de paro disminuirá en 39,56521 unidades; y si aumentamos en una unidad los ocupados construcción la tasa de paro disminuirá en 71,62081 unidades.

El R-squared nos dice que la variabilidad de la tasa de paro está explicada en un 55,23% por las variables independientes.



## **CAPITULO 4**

### **CONCLUSIONES**

Una vez que hemos explicado los modelos, llega la hora de sacar las conclusiones.

Desde un punto de vista descriptivo los datos ponen de manifiesto que existen diferencias territoriales como hemos podido comprobar en los mapas que se incluyen en el subcapítulo 2.1 Base de datos, concretamente en el apartado 2.1.1.1 Desempleo. Esta diferencia territorial se representa en dos zonas concretas:

- La primera de ellas situada en el norte de España con tasas de paro más bajas.
- La segunda en la zona sur de la península en la que los valores que toma la variable son más elevados.

Otra conclusión podría ser la existencia de dependencia espacial, esto quiere decir que realmente la tasa de paro estudiada se ve influenciada por el valor que toma esta variable en el entorno vecino. Más concretamente la tasa de paro para una provincia determinada se ve afectada por la tasa de paro que tienen las provincias vecinas (en nuestro caso las provincias pertenecientes a la misma comunidad autónoma).

Hemos podido comprobar que en el modelo influye el peso de las provincias. Esto quiere decir que no podemos considerar que Segovia tenga el mismo peso que Valladolid. Tenemos que tener en cuenta el número de personas que tiene cada provincia.

Concretamente, podemos decir que las empresas se concentran en las ciudades grandes, siendo las ciudades más pequeñas las que tienen menos empresas. En las ciudades con mayor número de habitantes es más fácil de encontrar mano de obra para desarrollar la actividad diaria de las empresas. Geográficamente lo podemos ver representado en la presencia de grandes polígonos alrededor de las grandes ciudades como pueden ser Madrid, Valladolid, Valencia, etc.

Por otro lado esto puede ser consecuencia a que las actividades económicas se encuentran localizadas por zonas como se puede ver el caso en el que la industria se encuentra más localizada en las zonas norte de la península. Esto se trata de corregir con la composición sectorial.

La composición sectorial es importante porque mejora la especificación. Al introducir la composición sectorial se ve que el único que muestra un comportamiento diferencial es el sector servicios.

Si distinguimos entre épocas de crecimiento y épocas de recesión podemos ver como existe dependencia espacial en ambas, esto lo vemos gracias a que la variable dependiente se encuentra explicada por las variables independientes que introducimos en los modelos. En este caso cabe destacar que existe una mayor dependencia en épocas de recesión. Las estimaciones correspondientes al año 2013, época en la que España se encontraba en recesión, la variable estudiada se ve más relacionada con las variables independientes que hemos introducido.

## CONCLUSIONES

Este tipo de análisis puede tener importancia a la hora de planificar la aplicación de políticas fiscales o monetarias por parte de los gobiernos. A la hora de combatir los problemas del desempleo hay que tener en cuenta que el entorno vecino influye en el comportamiento de la tasa de paro. Por lo tanto, no se pueden aplicar políticas de forma individualizada para un territorio sin tener en cuenta lo que ocurre en los que consideramos vecinos.

## REFERENCIAS BIBLIOGRÁFICAS

Alonso-Villar, O., del Río, C., &Toharia, L. (2009). UN ANÁLISIS ESPACIAL DEL DESEMPLEO POR MUNICIPIOS\*. *Revista De Economía Aplicada*, 17(49), 47-80. Retrieved from <https://search.proquest.com/docview/214561532?accountid=14778>

Azorín, J. D. B. (2013). La distribución del desempleo en las provincias españolas: Un análisis con datos de panel mediante el filtrado espacial. *Investigaciones regionales: Journal of Regional Research*, (27), 143-154.

GUJARATI, D. N. y PORTER, D. (2010). "*Econometría*". Ed. McGraw-Hill. 5ª edición.

INE (Instituto Nacional de Estadística)

<http://www.ine.es/>

Mochón, F. (2006). Principios de economía (3ª ed.). McGraw-hill/interamericana de España, S. A. U, 2006.

Pérez Infante, J. I. (2008). ¿ Por qué el paro registrado supera al paro de la EPA?. *Revista del Ministerio de Trabajo y Asuntos Sociales*, (71), 53-65.

Román, Á. M., & de Blas, A. M. Descomposición del desempleo provincial en sus componentes estructural y coyuntural mediante un modelo de error compuesto y análisis de sus patrones espaciales.

Serrano, R. M., &Valcarce, E. V. (2002). Econometría espacial:: nuevas técnicas para el análisis regional. Una aplicación a las regiones europeas. *Investigaciones regionales*, (1), 83.

## ANEXO I

### ESTIMACIONES PARA 2006

Dependent Variable: TASA\_DE\_PARO\_2006

Method: LeastSquares

Date: 05/10/17 Time: 10:02

Sample (adjusted): 1 50

Included observations: 50 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
<b>RETARDO_ESPACIAL_SIN_PON</b>	0.053429	0.035755	1.494297	0.1416
C	8.184726	0.470236	17.40556	0.0000
R-squared	0.044451	Mean dependentvar		8.581900
Adjusted R-squared	0.024544	S.D. dependentvar		2.777251
S.E. of regression	2.742957	Akaikeinfocriterion		4.895128
Sum squared resid	361.1429	Schwarzcriterion		4.971609
Log likelihood	-120.3782	Hannan-Quinn criter.		4.924252
F-statistic	2.232922	Durbin-Watson stat		2.218108
Prob(F-statistic)	0.141643			

Dependent Variable: TASA\_DE\_PARO\_2006

Method: LeastSquares

Date: 05/10/17 Time: 10:04

Sample (adjusted): 1 50

Included observations: 50 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
<b>RETARDO_ESPACIAL_SIN_PON</b>	0.011548	0.034331	0.336381	0.7381
<b>__OCUPADOS_SERVICIOS_200</b>	1.031168	8.167235	0.126257	0.9001
<b>__OCUPADOS_INDUSTRIA_200</b>	-26.55104	9.823225	-2.702884	0.0097
<b>__OCUPADOS_CONSTRUCCION_</b>	-7.175376	16.40375	-0.437423	0.6639
C	13.33419	8.083470	1.649562	0.1060
R-squared	0.348191	Mean dependentvar		8.581900
Adjusted R-squared	0.290253	S.D. dependentvar		2.777251
S.E. of regression	2.339736	Akaikeinfocriterion		4.632593
Sum squared resid	246.3465	Schwarzcriterion		4.823795
Log likelihood	-110.8148	Hannan-Quinn criter.		4.705404
F-statistic	6.009669	Durbin-Watson stat		2.338019
Prob(F-statistic)	0.000581			

Dependent Variable: TASA\_DE\_PARO\_2006  
 Method: LeastSquares  
 Date: 05/10/17 Time: 10:03  
 Sample (adjusted): 1 50  
 Included observations: 50 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
<b>RETARDO_ESPACIAL_PONDERA</b>	0.335724	0.090173	3.723114	0.0005
<b>C</b>	6.066600	0.760666	7.975384	0.0000
R-squared	0.224074	Mean dependentvar		8.581900
Adjusted R-squared	0.207909	S.D. dependentvar		2.777251
S.E. of regression	2.471739	Akaikeinfocriterion		4.686899
Sum squared resid	293.2558	Schwarzcriterion		4.763380
Log likelihood	-115.1725	Hannan-Quinn criter.		4.716024
F-statistic	13.86158	Durbin-Watson stat		2.091357
Prob(F-statistic)	0.000517			

Dependent Variable: TASA\_DE\_PARO\_2006  
 Method: LeastSquares  
 Date: 05/10/17 Time: 10:04  
 Sample (adjusted): 1 50  
 Included observations: 50 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
<b>RETARDO_ESPACIAL_PONDERA</b>	0.236668	0.086418	2.738627	0.0088
<b>__OCUPADOS_SERVICIOS_200</b>	3.762330	7.311784	0.514557	0.6094
<b>__OCUPADOS_INDUSTRIA_200</b>	-20.71663	8.896279	-2.328685	0.0244
<b>__OCUPADOS_CONSTRUCCION_</b>	-7.365062	15.18576	-0.484998	0.6300
<b>C</b>	8.988638	7.363815	1.220650	0.2286
R-squared	0.439903	Mean dependentvar		8.581900
Adjusted R-squared	0.390117	S.D. dependentvar		2.777251
S.E. of regression	2.168895	Akaikeinfocriterion		4.480952
Sum squared resid	211.6848	Schwarzcriterion		4.672155
Log likelihood	-107.0238	Hannan-Quinn criter.		4.553763
F-statistic	8.835806	Durbin-Watson stat		2.115745
Prob(F-statistic)	0.000024			

## ESTIMACIONES PARA 2013

Dependent Variable: TASA\_DE\_PARO\_2013

Method: LeastSquares

Date: 05/10/17 Time: 10:08

Sample (adjusted): 1 50

Included observations: 50 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
<b>RETARDO_ESPACIAL_SIN_PON</b>	0.083282	0.029840	2.790973	0.0075
C	24.00455	1.121661	21.40090	0.0000
R-squared	0.139623	Mean dependentvar		25.83385
Adjusted R-squared	0.121699	S.D. dependentvar		6.867783
S.E. of regression	6.436329	Akaikeinfocriterion		6.600972
Sum squared resid	1988.464	Schwarzcriterion		6.677453
Log likelihood	-163.0243	Hannan-Quinncrier.		6.630096
F-statistic	7.789528	Durbin-Watson stat		2.358483
Prob(F-statistic)	0.007519			

Dependent Variable: TASA\_DE\_PARO\_2013

Method: LeastSquares

Date: 05/10/17 Time: 10:09

Sample (adjusted): 1 50

Included observations: 50 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
<b>RETARDO_ESPACIAL_SIN_PON</b>	0.038077	0.030096	1.265185	0.2123
<b>__OCUPADOS_SERVICIOS_201</b>	-2.574259	18.16280	-0.141732	0.8879
<b>__OCUPADOS_INDUSTRIA_201</b>	-60.77208	21.66369	-2.805250	0.0074
<b>__OCUPADOS_CONSTRUCCION_</b>	-74.03433	70.93804	-1.043648	0.3022
C	40.59023	18.22380	2.227319	0.0310
R-squared	0.378409	Mean dependentvar		25.83385
Adjusted R-squared	0.323157	S.D. dependentvar		6.867783
S.E. of regression	5.650159	Akaikeinfocriterion		6.395884
Sum squared resid	1436.593	Schwarzcriterion		6.587086
Log likelihood	-154.8971	Hannan-Quinncrier.		6.468695
F-statistic	6.848720	Durbin-Watson stat		2.151249
Prob(F-statistic)	0.000215			

Dependent Variable: TASA\_DE\_PARO\_2013  
Method: LeastSquares  
Date: 05/10/17 Time: 10:08  
Sample (adjusted): 1 50  
Included observations: 50 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
<b>RETARDO_ESPACIAL_PONDERA</b>	0.400078	0.068704	5.823225	0.0000
C	16.67075	1.743662	9.560770	0.0000
R-squared	0.413991	Mean dependentvar		25.83385
Adjusted R-squared	0.401782	S.D. dependentvar		6.867783
S.E. of regression	5.311856	Akaikeinfocriterion		6.216938
Sum squared resid	1354.359	Schwarzcriterion		6.293418
Log likelihood	-153.4234	Hannan-Quinn criter.		6.246062
F-statistic	33.90995	Durbin-Watson stat		2.185798
Prob(F-statistic)	0.000000			

Dependent Variable: TASA\_DE\_PARO\_2013  
Method: LeastSquares  
Date: 05/10/17 Time: 10:10  
Sample (adjusted): 1 50  
Included observations: 50 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
<b>RETARDO_ESPACIAL_PONDERA</b>	0.306046	0.068929	4.439996	0.0001
<b>__OCUPADOS_SERVICIOS_201</b>	3.033378	14.88510	0.203786	0.8394
<b>__OCUPADOS_INDUSTRIA_201</b>	-39.56521	18.21127	-2.172567	0.0351
<b>__OCUPADOS_CONSTRUCCION_</b>	-71.62081	57.18040	-1.252541	0.2168
C	27.14432	14.94126	1.816736	0.0759
R-squared	0.552388	Mean dependentvar		25.83385
Adjusted R-squared	0.512600	S.D. dependentvar		6.867783
S.E. of regression	4.794675	Akaikeinfocriterion		6.067529
Sum squared resid	1034.501	Schwarzcriterion		6.258731
Log likelihood	-146.6882	Hannan-Quinn criter.		6.140339
F-statistic	13.88337	Durbin-Watson stat		1.952403
Prob(F-statistic)	0.000000			