
Universidade Vigo


# MÁSTER EN TÉCNICAS ESTADÍSTICAS

# Estimación Trimestral en Áreas Pequeñas de los Ocupados de la Encuesta de Población Activa Según su Actividad Económica

**Autor:** David Celdrán Bouzas.

**Directores:** María Esther López Vizcaíno.

María José Lombardía Cortiña.

Wenceslao González Manteiga.

**JULIO 2012**

# Agradecimientos

A María Esther López Vizcaíno por todo su tiempo invertido en dirigirme en este proyecto. Sin su apoyo y ayuda no podría haberlo acabado.

A todos los miembros del Instituto Galego de Estadística por toda la ayuda ofrecida en mi tiempo de estancia en prácticas y desarrollo de este trabajo.

A mis tutores Wenceslao González Manteiga y María José Lombardía Cortiña, en especial a ésta última por brindarme sus consejos y estar presente en todo momento.

A todos los profesores del Máster en Técnicas Estadísticas por sus clases magistrales que tanto han ayudado a mi formación.

A mis compañeros del Máster, en especial a María y Andrés, que me han ayudado en todo momento.

A mi familia, sobre todo a mi madre, y a mis amigos, por estar conmigo cada día.

A Ana, por todo.

# Índice General

# Capítulo 1

# Introducción

El muestreo estadístico, en contraposición con los censos, permite obtener información sobre materias muy dispares con un coste más bajo. El muestreo no se utiliza tan sólo para la obtención de estimaciones de la población, sino para estimar parámetros en una variedad de subpoblaciones (dominios). Saei y Chambers (2003) definieron los dominios (o áreas pequeñas) como una subdivisión de la población para la variable de interés. Jiang y Lahiri (2006) los definieron como una población para la cual no se pueden obtener estadísticos fiables debido a ciertas limitaciones de los datos disponibles. En general, los dominios se suelen definir como áreas geográficas, grupos socioeconómicos u otras subpoblaciones (Rao (2003)). Ejemplos de áreas pequeñas o dominios podrían ser:

- Áreas geográficas: Municipios, comarcas, distritos, islas.
- Grupos Socio-Demográficos: Grupos específicos por sexo, edad, raza.
- Otras subpoblaciones: Actividades económicas particulares, conjunto de grupos de empresas…

Hoy en día existe una creciente demanda de información cada vez más desagregada. Para responder a esta demanda cabría como soluciones aumentar el tamaño de la muestra, con el coste que conlleva, y utilizar estimadores basados en el diseño empleados en la estadística oficial, o tratar de utilizar técnicas de estimación más complejas como son las técnicas de estimación en áreas pequeñas.

La demanda de estadísticas a nivel de áreas pequeñas o dominios se ha consolidado estos últimos años entre los objetivos prioritarios de las instituciones estadísticas oficiales, aunque su uso se originó hace varios siglos. En efecto, Brackstone (1987) menciona la existencia de dicha estadística en el siglo XI en Inglaterra y en el

siglo XVII en Canadá. Todas estas primeras estadísticas para áreas pequeñas estaban basadas en el censo o en registros administrativos.

Pero bien es cierto que en los últimos tiempos el interés por la inferencia en áreas pequeñas ha aumentado enormemente, debido, entre otras cosas, a la creciente demanda tanto del sector público como del privado. Algunos de los ejemplos en el uso de estas técnicas son los siguientes:

- Fay y Herriot (1979) estimaron la renta per cápita en áreas con menos de 1000 habitantes en Estados Unidos, estimaciones que fueron utilizadas por el Treasury Department para determinar el reparto de fondos a los gobiernos locales dentro de los diferentes estados. Su método se empleó de nuevo en el año 2000 para realizar un modelo para producir estimaciones por condado de la pobreza infantil en Estados Unidos.
- Battese, Harter y Fuller (1988) estimaron la superficie de cultivo de cereal y soja en doce condados de Iowa North-Central, para la administración de programas federales involucrando pagos a granjeros con una producción baja.
- Malec, Davis y Cao (1999) estudiaron el predominio del sobrepeso en los adultos estadounidenses por estados.
- En el ámbito europeo un antecedente del uso de esta metodología se encuentra en la Oficina Nacional de Estadística del Reino Unido (Office for National Statistics, ONS) donde las estimaciones basadas en modelos de datos de desempleo a partir de su encuesta de fuerza de trabajo (Labor Force Survey, LFS) han sido aceptadas recientemente como estadística nacional.

La aceptación de las estimaciones en áreas pequeñas como estadística oficial, considerada como aquella que cumple todos los requisitos del Código de Buenas Prácticas de las estadísticas oficiales, es creciente. Esto conlleva nuevos retos tanto para la investigación en estos métodos como para la presentación y explicación adecuada de los resultados obtenidos a los usuarios.

Algunos ejemplos del uso de la estimación en áreas pequeñas por parte de los institutos de estadística españoles son:

- Estimación en áreas pequeñas con datos de la Encuesta de Población Activa en Canarias. (ISTAC).
- Estimación en áreas pequeñas de la Encuesta Industrial en el País Vasco. (EUSTAT).
- Estimación en áreas pequeñas de en la Encuesta de Población Activa en relación con la actividad en el País Vasco. (EUSTAT).
- Estimación en áreas pequeñas del importe neto de la cifra de negocio de los establecimientos industriales con menos de 20 trabajadores en Galicia. (IGE).

- Estimación en áreas pequeñas del ingreso medio mensual por comarca en los hogares gallegos. (IGE).

Uno de los grandes problemas existentes a la hora de realizar la estimación para áreas pequeñas es que los tamaños muestrales para los dominios de interés son típicamente pequeños o incluso cero, debido probablemente al diseño de la muestra. Esto provoca que al realizar la estimación para esos dominios por los métodos tradicionales basados en el diseño muestral de la encuesta (estimación directa) se obtengan estimadores con una varianza muy grande.

Es habitual emplear información auxiliar para definir los estimadores para áreas pequeñas, empleando por ejemplo valores de la variable de interés en otras áreas similares, o valores pasados de la variable en la misma área objeto de estudio (estimación indirecta). Esos valores se incorporan al proceso de estimación a través de un modelo implícito o explícito que proporciona un link para toda la información auxiliar.

Dentro de los estimadores indirectos se encuentran los basados en modelos explícitos que pese a que tienen una larga historia solo han recibido atención en las últimas décadas como aproximación a la estimación de las características de las áreas pequeñas. Asumen que la variabilidad entre los dominios de la variable respuesta puede ser enteramente explicada en términos de la correspondiente variabilidad de la información auxiliar (modelos de efectos fijos) o requieren la suposición de que la variabilidad específica del dominio permanece "inexplicada" aún después de haber considerado la información auxiliar, momento en que aparecen los modelos mixtos, incorporando de esta manera efectos aleatorios específicos de cada dominio, que explican las variaciones adicionales entre áreas en los datos, no explicadas por la parte de los efectos fijos del modelo considerado.

Los modelos lineales mixtos tienen una amplia gama de aplicaciones. En particular la habilidad de predecir una combinación lineal de efectos fijos y aleatorios es una de las propiedades más atractivas de este tipo de modelos. En una serie de escritos, Henderson (1948, 1949, 1959, 1963, 1973, 1975) desarrolló los estimadores BLUP (best linear unbiased prediction) para modelos mixtos, refiriéndose por "best" al mínimo error cuadrático medio entre todos los predictores lineales insesgados, "linear" significa que el predictor es una combinación lineal de los valores de la variable respuesta y "unbiased" indica que el valor esperado del error de predicción es cero. Sin embargo este método asume que las varianzas de los efectos aleatorios son conocidas, cuando en la práctica no lo son y se deben estimar a partir de los datos. Harville (1991), Robinson (1991) y Harville (1992) describen un predictor obtenido a partir de los BLUP donde las componentes de varianza desconocidas son reemplazadas por estimadores asociados: EBLUP (empirical best linear unbiased predictor), que son los estimadores que se van a utilizar en este estudio.

Dentro de los estimadores basados en el modelo, los modelos de áreas pequeñas se podrían clasificar en dos grandes clases o tipos, dependiendo de la disponibilidad de la información auxiliar: modelos a nivel de área y modelos a nivel de unidad:

- Modelo a nivel de área: Se emplean cuando se dispone de información auxiliar sólo a nivel de área. Relaciona el estimador directo del área con una covariable específica del área. Se asume que el estimador directo del área obedece a un modelo de población.
- Modelo a nivel de unidad o modelo de errores anidado: Se emplean cuando se dispone de información auxiliar sobre las unidades individuales de la población. Relaciona los valores de la variable en estudio con el valor de la covariable medida en cada unidad.

En este trabajo solo se dispone de información a nivel de área, por lo que se emplearán este tipo de modelos.

Se introduce a continuación alguna de la notación que se empleará a lo largo del trabajo. Se utilizará el índice $d$ para los dominios, $d = 1, \dots, D$, siendo $D$ el número total de dominios. Se utiliza $Y$ para hacer referencia a la variable respuesta y $x_i$ para cada una de las variables auxiliares $i = 1, \dots, p$, con $p$ el número de variables auxiliares consideradas en el estudio.

En los modelos a nivel de área las variables auxiliares, se asume que están disponibles tanto para las áreas muestreadas como para las no muestreadas. Estos modelos asumen que la media de la población del área pequeña $\bar{Y}_d$ está relacionada con $x_d$ a través de un modelo lineal con efectos aleatorios de área $v_d$:

$$\bar{Y}_d = x'_d\beta + v_d, \quad d = 1, \dots, D$$

donde $\beta$ es el p-vector de los parámetros de regresión y los $v_d$ son los efectos aleatorios que están incorrelados con media cero y varianza $\sigma_v^2$ desconocida. La normalidad en los $v_d$ es algo que se asume a menudo.

Se considera que el estimador directo $\hat{\bar{Y}}_d$ de $\bar{Y}_d$ está disponible siempre y cuando el tamaño muestral del área $n_d \geq 1$, y es habitual suponer que $\hat{\bar{Y}}_d = \bar{Y}_d + e_d$ donde los errores muestrales $e_d$ son independientes $N(0, \psi_d)$ con varianza $\psi_d$ conocida.

De esta forma obtenemos el modelo lineal mixto a nivel de área propuesto por Fay y Herriot en 1979:

$$\hat{\bar{Y}}_d = x'_d\beta + v_d + e_d \qquad (1)$$

Tras ajustar el modelo por máxima verosimilitud, el estimador EBLUP del total del dominio $d$, basado en el modelo es:

$$\hat{Y}_d^{FH} = N_d\hat{\bar{Y}}_d^{FH}$$

donde el estimador Fay-Herriot es una combinación convexa del estimador directo y del estimador sintético, es decir:

$$\hat{\bar{Y}}_d^{FH} = \hat{\gamma}_d \hat{\bar{Y}}_d + (1 - \hat{\gamma}_d) x_d \hat{\beta}\,, \qquad donde\ \hat{\gamma}_d = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_d^2}$$

Sin embargo, en el campo de la estimación en áreas pequeñas, los datos están a menudo disponibles simultáneamente para muchas áreas pequeñas pero también posiblemente para unos pocos períodos de tiempo, como por ejemplo en la Encuesta de Población Activa (EPA). La disponibilidad de esta información a lo largo del tiempo ha servido para introducir esta información en el modelo y para mejorar los estimadores de áreas pequeñas.

Rao y Yu (1994) proponen una forma simple de compartir información entre dominios y a lo largo del tiempo mediante la introducción en el modelo de efectos aleatorios que tengan en cuenta la variabilidad entre dominios y a lo largo del tiempo. Proponen de esta manera la extensión del modelo de Fay-Herriot (1). Sean $d = 1, \dots, D$ cada uno de los dominios objeto de estudio y $t = 1, \dots, m_d$ los instantes temporales a los que se hace referencia, el modelo considerado es:

$$\hat{\bar{Y}}_{dt} = x'_{dt}\beta + v_d + u_{dt} + e_{dt}, \quad d = 1, \dots, D, \qquad t = 1, \dots, m_d \qquad (2)$$

donde los efectos aleatorios $v_1, \dots, v_D$ están idénticamente distribuidos a una normal $N(0, \sigma_v^2)$, los efectos aleatorios entre los distintos instantes temporales del dominio $(u_{d1}, \dots, u_{dT})$ siguen un proceso autorregresivo AR(1), siendo los errores $(e_{11}, \dots, e_{DT})$ normalmente distribuidos con media 0 y varianza conocida $\Sigma$, y $\{v_d\}, \{u_{dt}\}, \{e_{dt}\}$ son independientes.

Trabajos que incluyen efectos a lo largo del tiempo son los de Ghosh, Nangia y Kim (1996) que proponen un modelo un poco más complicado a nivel de área con efectos de tiempo correlados para estimar el ingreso medio de familias de cuatro miembros para los cincuenta estados americanos y el distrito de Columbia; You y Rao (2000) y Datta, Lahiri y Maiti (2002) emplean el modelo de Rao y Yu pero reemplazando la distribución AR(1) por la de un paseo aleatorio; Datta, Lahiri, Maiti y Lu (1999) consideran un modelo similar pero añadiendo términos extra a los del modelo auxiliar para reflejar cambios estacionales en su aplicación, para así estimar tasas de desempleo mensuales en nueve estados americanos y el distrito de Columbia; You, Rao y Gambino (2001) aplican el modelo de Rao y Yu para estimar tasas de desempleo mensual para áreas censales metropolitanas en Canadá; Pfeffermann y Buck (2000) describen una clase general de modelos state-space.

El Instituto Galego de Estadística (IGE), consciente de la necesidad actual de realizar estadísticas de calidad, y ante la falta de coherencia que las fuentes disponibles son capaces de ofrecer cuando se trata de datos con una desagregación amplia, tiene como objetivo elaborar estadísticas en las que las áreas pequeñas sean la base para generar resultados óptimos para su posterior publicación como estadística oficial.

En particular, cuando se trata de estimar los ocupados que ofrece la Encuesta de Población Activa (EPA) por rama de actividad de la Clasificación Nacional de Actividades Económicas (CNAE 09), a menudo los resultados carecen de sentido. Tal

y como se observa en la Figura 1.1, el hecho de tener picos consecutivos tan amplios en la estimación de los ocupados en estas dos actividades, que hacen referencia a la hostelería, no parece del todo real, porque aunque se esperase una cierta estacionalidad, se observa que ésta no es la que produce estos picos y sí lo es la inexactitud que ofrecen los datos. Se esperaría una tendencia un poco más suave. Esto puede ser debido a que, en este caso, la actividad económica no es un criterio de estratificación de la EPA y por tanto su falta de representatividad provoca estos resultados.

Además, la encuesta genera unos errores muestrales muy grandes en aquellas actividades que presentan un número bajo o limitado de observaciones. Con motivo de reducir los errores muestrales en estadísticos de áreas pequeñas obtenidos a partir de la muestra, se puede combinar información muestral, registros administrativos/censos e incluso encuestas previas, y eso es lo que se va a tratar de hacer en este trabajo.

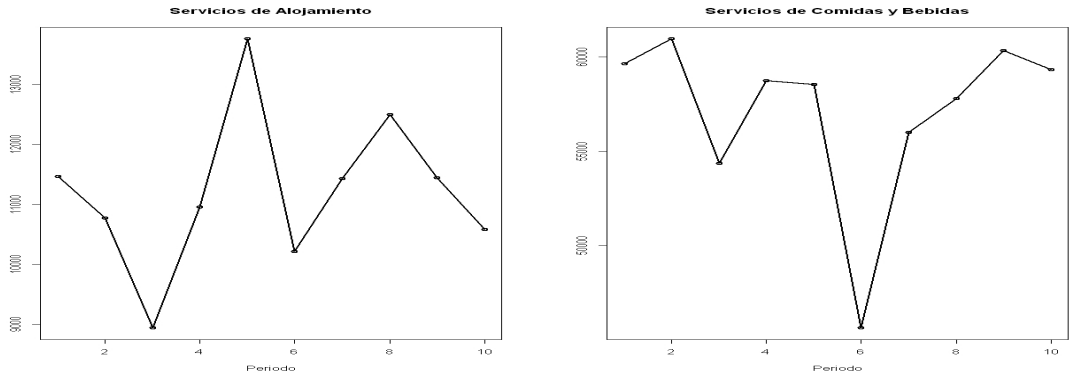


**Figura 1.1 Evolución de dos actividades económicas recogidas en la CNAE desde el tercer trimestre del año 2009 hasta el cuarto trimestre del año 2011.**

El presente estudio, que recoge una memoria de las prácticas realizadas en el IGE, tiene como objetivo la estimación trimestral de los ocupados, obtenidos a partir de la EPA en la comunidad autónoma gallega, según su actividad económica desde el tercer trimestre del año 2009 hasta el cuarto trimestre del año 2011, tomando como referencia la CNAE 09 a dos dígitos, la cual se muestra en el Anexo I.

Para ello se adaptan dos procedimientos basados en modelos al caso de la estimación en áreas pequeñas de los ocupados según actividad económica. Estos se obtienen a partir del modelo considerado por Rao y Yu (1994). Se considerarán dos enfoques diferentes, aunque ambos considerando estimadores EBLUP: el primer enfoque está basado en un modelo lineal mixto con efectos de tiempo independientes y el segundo enfoque utiliza un modelo lineal mixto con efectos de tiempo correlados. Los dos emplean modelos a nivel área que usan los estimadores directos como variable respuesta, de esta manera se incorpora a los modelos la información del diseño muestral.

Teniendo en cuenta el modelo, se obtendrán estimadores de los errores cuadráticos medios (MSE), que es una medida de la eficiencia del estimador. Los estimadores de áreas pequeñas suelen ser sesgados, por eso la acuracidad de la

estimación es un tema fundamental en la estimación en áreas pequeñas. En este trabajo se utiliza la aproximación analítica de Prasad y Rao (1990).

Hay que tener en cuenta en este punto que para la validación de las estimaciones se va a tomar como referencia la ONS (2004), que considera como estadísticas aceptables aquellas que presentan un coeficiente de variación inferior al 20%, por lo que la estimación del MSE se presenta como factor fundamental a la hora de analizar los resultados obtenidos en este estudio.

Para la aplicación práctica de esta metodología se utilizará como referencia el proyecto europeo SAMPLE de Salvati, Giusti, Marchetti, Pratesi, Tzavidis, Molina, Morales, Esteban, Santamaría, Marhuenda, Pérez, Plagiarella y Chambers (2010), tanto para el estudio de los modelos como para la obtención del código de programación empleado en este trabajo.

El trabajo está organizado en cinco capítulos. Al primer capítulo de Introducción le sigue un capítulo en el cual se explicará cuál es el origen de los datos considerados en este estudio, a partir de los cuales se realizará un análisis exploratorio en el Capítulo 3 para ver si son aptos para la estimación en áreas pequeñas que se desea realizar. Una vez seleccionadas aquellas variables de interés se realizarán las estimaciones, Capítulo 4, a través de los dos enfoques comentados anteriormente y para finalizar se hará una breve conclusión en el Capítulo 5 y se propondrá una línea de investigación para completar lo expuesto en este trabajo.

# Capítulo 2

# Descripción de los Datos

Los datos que se han utilizado en este trabajo provienen del IGE. Recogen por una parte estimaciones del número de ocupados, obtenidas a partir de diversas encuestas, por otra parte el número de afiliados registrados en la Seguridad Social (SS) y por último el número de contratos registrados por rama de actividad económica.

La variable objeto de estudio va a ser el número de ocupados en la EPA y como variables auxiliares se emplearán el número de afiliados a la SS a final de mes, el número de ocupados estimado en función de diversas encuestas propias de cada sector de actividad y el número de contratos registrados.

Se dispone de datos trimestrales para todas las variables de interés desde el tercer trimestre del año 2009 hasta el cuarto trimestre del año 2010.

## 2.1 La Encuesta de Población Activa (EPA)

La fuente estadística principal en este trabajo, a partir de la cual se obtienen los ocupados por rama de actividad económica, es la EPA (INE, IGE (2005)).

La EPA es una investigación continua y de periodicidad trimestral dirigida a la población que reside en viviendas familiares, esto es, las utilizadas todo el año o la mayor parte del mismo como vivienda habitual o permanente. Está orientada a dar datos de las principales categorías poblacionales en relación con el mercado de trabajo (ocupados, parados, activos, inactivos) y a obtener clasificaciones de estas categorías según diversas características. También posibilita confeccionar series temporales homogéneas de resultados. Por último, al ser las definiciones y criterios utilizados

coherentes con los establecidos por los organismos internacionales que se ocupan de temas laborales, permite la comparación con datos de otras comunidades autónomas dentro del ámbito nacional o con otros países dentro del ámbito internacional.

Utiliza un muestreo aleatorio estratificado bietápico, con estratificación en las unidades de primera etapa que son las secciones censales (áreas geográficas perfectamente delimitadas). Los estratos se construyen siguiendo un criterio demográfico en el que se asigna a cada ayuntamiento su estrato correspondiente en función del tamaño del mismo. La clasificación por tamaño viene reflejada en la Tabla 2.1. Las unidades de segunda etapa son las viviendas, consideradas éstas como las utilizadas todo el año o la mayor parte de él como vivienda habitual, en las cuales no se realiza ningún submuestreo, de forma que se recoge información sobre todas las personas que residen en ellas.

| Estrato | Descripción |
|---|---|
| 1 | Ayuntamientos capital de provincia (A Coruña, Lugo, Ourense, Pontevedra). |
| 2 | Ayuntamientos autorrepresentados, importantes en relación con la capital (Santiago, Vigo). |
| 3 | Otros ayuntamientos autorrepresentados importantes en relación con la capital (Ferrol). |
| 4 | Ayuntamientos de 50.001 a 100.000 habitantes, excepto los anteriores. |
| 5 | Ayuntamientos de 20.001 a 50.000 habitantes. |
| 6 | Ayuntamientos de 10.001 a 20.000 habitantes. |
| 7 | Ayuntamientos de 5.001 a 10.000 habitantes. |
| 8 | Ayuntamientos de 2.001 a 5.000 habitantes. |
| 9 | Ayuntamientos con población menor o igual a 2.000 habitantes |

**Tabla 2.1. Descripción de los estratos empleados en la EPA.**

El número de secciones censales en el territorio gallego es de 468, una vez que han sido ampliadas en 234 unidades, gracias a la firma en julio de 2008 de un convenio de colaboración entre el INE y el IGE, lo que suponía un aumento del 100% de la muestra inicial. Con esta ampliación se llegan a evaluar cerca de 8.000 viviendas lo que supone finalmente una muestra total de unas 20.000 personas.

A la hora de analizar la EPA, hay que tener en cuenta una serie de definiciones que ayudan a clasificar a cada uno de los encuestados. De tal forma que se considera:

- **Población económicamente activa:** Conjunto de personas de 16 o más años que durante la semana de referencia suministran mano de obra para la producción de bienes y servicios económicos o que están disponibles y hacen gestiones para incorporarse a dicha producción.

- **Población ocupada:** La formada por todas aquellas personas de 16 o más años que durante la semana de referencia han tenido un trabajo por cuenta ajena o han ejercido una actividad por cuenta propia. Una persona se clasifica como ocupada cuando en la semana de referencia ha trabajado al menos una hora a cambio de un sueldo, salario u otra retribución conexa, o ha estado ausente del trabajo pero mantiene un fuerte vínculo con él.

- **Población parada:** Se considerarán paradas las personas de 16 o más años que se encuentren sin trabajo durante la semana de referencia, hayan tomado medidas concretas para buscar un trabajo y estén disponibles para trabajar. También se considerarán parados aquéllos que han estado sin trabajo en la semana de referencia pero no buscan empleo porque han encontrado uno al que se incorporarán en los tres meses posteriores a la semana de referencia.

- **Población económicamente inactiva:** Comprende a las personas de 16 o más años que no han sido clasificados ni en población ocupada ni en parada. Comprende a personas que se ocupan del hogar, estudiantes, jubilados o prejubilados, pensionistas por motivos distintos a jubilación o prejubilación, incapacitados para trabajar…

La EPA utiliza estimadores de razón a los que se le aplican técnicas de reponderación para que de ese modo se hagan coincidir los resultados que proporciona la encuesta con las proyecciones de población. Sea $S_p$ la submuestra perteneciente a la provincia p, $N_h$ la población de 16 años o más en el estrato h según las proyecciones demográficas y $n_h$ el número de individuos de la muestra en el estrato h, el estimador que utiliza la EPA para determinar el número de ocupados en la provincia $p$ es el siguiente:

$$\hat{Y}_p^{EPA} = \sum_{j \epsilon S_p} w_j * y_j$$

donde $w_j = \frac{N_h}{n_h}$ son los factores de elevación empleados en la EPA y representan la cantidad de individuos que supone cada observación recogida en la encuesta. La EPA no proporciona estimaciones oficiales para los dominios, en el caso aquí estudiado, las ramas de actividad económica recogidas en la CNAE 09 a dos dígitos. Una expresión equivalente para representar el estimador que calcula los ocupados por dominios es la siguiente:

$$\hat{Y}_d^{EPA} = \sum_{j \epsilon S_d} w_j * y_{dj} \quad d = 1, \dots, 79$$

Las varianzas de los estimadores para los dominios van se aproximan con la librería *"survey"* de R (T.Lumley 2004-2010), que utiliza fórmulas elementales de la varianza de una suma:

$$Var\left(\hat{Y}_d^{EPA}\right) = \sum_j w_j^2 * Var\left(\hat{Y}_{dj}\right) + 2 \sum_{i<j} w_i w_j \, Cov\left(\hat{Y}_{di}, \hat{Y}_{dj}\right) \quad d = 1, \dots, 79$$

$$CV\left(\hat{Y}_d^{EPA}\right) = \frac{\sqrt{Var\left(\hat{Y}_d^{EPA}\right)}}{\hat{Y}_d^{EPA}} \times 100 \quad d = 1, \dots, 79 \qquad (3)$$

Denotaremos por $Y_d$ , con $d$ el dominio que indica la actividad económica a la que hace referencia, el número de ocupados EPA. Dicha variable ha sido desagregada por la rama de actividad económica a la cual pertenece la persona ocupada, según la CNAE 09 a dos dígitos, habiendo eliminado aquellas ramas de actividad para las cuales no se disponía de datos suficientes para la estimación, bien fuese porque la actividad no tiene presencia en la comunidad autónoma o porque aun existiendo asalariados en la misma, la EPA no ha podido recoger datos de ocupados en alguno de los trimestres objeto de estudio.

La Tabla 2.2 muestra cuáles han sido las actividades eliminadas del estudio debido a lo comentado anteriormente.

| Código Actividad | Descripción |
|---|---|
| **05** | Extracción de antracita, hulla y lignito. |
| **06** | Extracción de crudo de petróleo y gas natural. |
| **07** | Extracción de minerales metálicos. |
| **12** | Industria del tabaco. |
| **37** | Recogida y tratamiento de aguas residuales. |
| **39** | Actividades de descontaminación y otros servicios de gestión de residuos. |
| **63** | Servicios de información. |
| **98** | Actividades de los hogares como productores de bienes y servicios para uso propio. |
| **99** | Actividades de organizaciones y organismos extraterritoriales |

**Tabla 2.2 Actividades económicas por código CNAE 09 eliminadas del estudio.**

## 2.2 Variables Auxiliares

Se presentan a continuación, el conjunto de variables auxiliares que se van a considerar a la hora de explicar la evolución del número de ocupados EPA. Algunas de ellas han sido obtenidas a través de la página oficial del IGE mientras que otras han sido elaboradas a través de diversos procedimientos comentados a continuación.

### 2.2.1 Afiliaciones a la Seguridad Social (SS)

Además de la EPA, que estima periódicamente el empleo, existe otra fuente de información estadística que proporciona información periódica sobre la ocupación: las afiliaciones a la SS.

Las afiliaciones a la SS es una estadística administrativa que no tiene como objetivo directo estimar el empleo sino registrar las afiliaciones y cotizantes a la SS. Es de periodicidad mensual refiriéndose los datos al último día de cada mes y a la media del mes. En este trabajo se utilizarán los datos a último día de mes.

El dato de afiliación recoge el número de trabajadores en alta, que realizan una actividad laboral propiamente dicha (no se incluyen parados ni estudiantes) por lo que quedan obligados a cotizar al Sistema Público para la protección de, al menos, las situaciones de jubilación, invalidez y muerte.

El dato de afiliación, como todos los procedentes de registros, es sensible a las variaciones legislativas que le afectan. El fichero puede contener además cotizantes ficticios que buscan obtener una pensión de jubilación en el futuro o el derecho a percibir el subsidio de paro.

Hay personas ocupadas según EPA que no están obligadas a cotizar a SS, como es el caso de algunas ayudas familiares (esta categoría de trabajadores han de estar afiliados a la SS sólo cuando trabajan de forma habitual, fundamental y directa). También hay personas que tienen la obligación de estar afiliadas a la SS y no lo están

Dada la periodicidad trimestral de la EPA, y la periodicidad mensual de esta variable, se han tenido que trimestralizar estos datos. Para ello se ha empleado la media aritmética de esta variable cada tres meses.

### *2.2.2* Ocupados por Actividad Económica Recogidos en Diversas Encuestas Propias de Cada Sector de Actividad

Esta variable se construye a través de diferentes encuestas propias de los sectores de actividad económica (ver anexo de actividades que forman parte de cada sector económico). Las encuestas que conforman esta variable no recogen información de todos los sectores económicos, sino que están referenciadas a algunos en particular.

La unión del conjunto de encuestas es lo que forma esta variable. Se han considerado:

- **Encuesta Industrial de Empresas (EIE)**
  El objetivo fundamental de esta encuesta es proporcionar una información precisa y fiable de las principales características estructurales y de actividad de los diversos sectores que constituyen la actividad industrial. Es una encuesta que se lleva a cabo con carácter anual, y los datos solicitados se refieren al año natural objeto de la encuesta. La población objeto de estudio es el conjunto de empresas con una o más personas ocupadas remuneradas, y cuya actividad principal figura incluida en las Secciones B a E de la CNAE 09 (ver anexo), entendiéndose por actividad principal de la empresa aquella que genere mayor valor añadido. Es decir, la encuesta cubre las industrias extractivas; las industrias manufactureras; el suministro de energía eléctrica, gas, vapor y aire acondicionado y el suministro de agua, actividades de saneamiento, gestión de residuos y descontaminación.
  Se obtienen numerosos datos como el importe neto de la cifra de negocios, las personas remuneradas y las no remuneradas, las ventas netas de productos o las horas trabajadas… La variable que interesa para este estudio es el número de personas ocupadas, las cuales vienen consideradas como el conjunto de personas fijas y eventuales, que en el año de referencia de los datos se encontraban ejerciendo una labor remunerada o no, para la empresa, y perteneciendo y siendo pagadas por ésta, es decir, no incluye a

las personas no asalariadas (autónomos). Al obtener los datos de forma anual, es necesario desagregar la información proporcionada. Para ello, se emplea el método de desagregación con indicadores de Chow y Lin (1971).

- **Encuesta Anual de Servicios (EAS)**

  Tiene como objetivo estudiar las características estructurales y económicas de las empresas pertenecientes al sector servicios, el más importante de la economía en términos de Producto Interior Bruto (PIB) y creación de empleo. Cubre aquellas actividades relacionadas con el comercio, transporte, hostelería, sector inmobiliario, servicios a empresas (asesorías, abogados) y la parte privada de cultura y deporte, pero deja fuera a una serie de sectores como el sector financiero, los seguros y los servicios de no mercado: sanidad, educación, cultura, deportes,…, es decir, todo lo referente a la Administración Pública.

  El periodo de referencia para la encuesta es el año. Los datos relativos al número de establecimientos y empleo se solicitan a 30 de septiembre. Recoge diversas características de las empresas, tales como: actividad principal, naturaleza jurídica, periodo de actividad, número de locales, variables sobre la estructura del empleo y datos contables, como compras y gastos, ingresos, operaciones de capital e impuestos. Esta encuesta obtiene información del personal ocupado (a 30 de septiembre) según su remuneración, el tipo de jornada laboral y el sexo, además puede ofrecer una evolución del personal por trimestre y otro tipo de clasificaciones.

  De igual forma que con la encuesta anterior es necesario desagregar la información para disponer de datos trimestrales.

- **Encuesta Trimestral de Coste Laboral (ETCL):**

  El objetivo fundamental de esta encuesta es conocer la evolución del coste laboral medio por trabajador y por hora efectiva de trabajo. Lo que pretende la ETCL es proporcionar el coste laboral medio por trabajador y mes, el coste laboral medio por hora efectiva de trabajo, y el tiempo trabajado y no trabajado. También permite obtener información salarial, información de costes no salariales y un mejor conocimiento del tiempo trabajado y no trabajado así como su estructura.

  Cubre todas las secciones de la CNAE, excepto el sector primario (agricultura, pesca, ganadería… ), las actividades de los hogares como productores de servicios para uso propio y las actividades de organizaciones y organismos extraterritoriales.

  La encuesta trata el estudio de los asalariados por los que haya existido la obligación de cotizar al menos un día durante el mes de referencia, con independencia de su modalidad contractual y su jornada de trabajo. De igual forma que la EIE, quedarían fuera del estudio los no asalariados o autónomos.

- **Registro Central de Personal (RCP)**
  Es una publicación incluida en el Plan Estadístico Nacional elaborada por el Ministerio de Hacienda y Administraciones Públicas. Esta publicación de carácter semestral presenta un panorama genérico de la distribución de los ocupados o personal al servicio de la Administración Pública Estatal, las Administraciones de las comunidades autónomas, la Administración local (ayuntamientos, diputaciones…) y las universidades.

- **Memoria de la Delegación del Gobierno para Galicia**
  Da a conocer de forma pormenorizada, las cifras y proyectos de los diferentes organismos que integran la Administración General del Estado en Galicia. La memoria aporta datos pormenorizados sobre personal de la Administración General del Estado, inversiones en infraestructuras y otras áreas, estadísticas de seguridad ciudadana y vial, entre otros muchos.
  Tiene periodicidad anual y contiene una información más exhaustiva de los datos de empleo público a nivel autonómico.

- **Estadística de Personal Universitario**
  Elaborado por la Subdirección General de Análisis, Estudios y la Prospectiva Universitaria de la Secretaría General de Universidades del Ministerio de Educación, reúne de forma anual datos que ofrecen una visión global de los Recursos Humanos dedicados a funciones de gestión técnica, económica y administrativa, así como el apoyo, asesoramiento y asistencia en el desarrollo de las funciones de la universidad.

Los datos considerados para el sector industrial provienen de la Encuesta Industrial de Empresas (EIE) y de la Encuesta Trimestral de Coste Laboral (ETCL); los datos considerados para el sector de la construcción se recogen de la (ETCL); los datos para el sector servicios se recogen gracias a la Encuesta Anual de Servicios (EAS), la ETCL, el Registro Central de Personal, la Memoria de la Delegación del Gobierno para Galicia y la Estadística de Personal Universitario; y el sector primario, ante la carencia de encuestas que pudiesen emplearse, se ha optado por la introducción del número de afiliados a la SS, con el riesgo que esto conlleva de cara a problemas estadísticos.

### *2.2.3* Contratos Registrados

Esta última variable recoge el número de nuevos contratos que se registran mensualmente por cada una de las ramas de actividad económica objeto de estudio. Los datos los proporciona la Consellería de Traballo e Benestar al IGE. El Estatuto de los trabajadores, en su Artículo 16, establece que los empresarios tienen la obligación de comunicar a las oficinas de empleo de la Xunta de Galicia el contenido de los contratos de trabajo celebrados y las prórrogas de estos, en los diez días siguientes a su concertación. Los datos que se emplean corresponden a los contratos iniciales

registrados, se excluyen por tanto, las prórrogas y los contratos convertidos en indefinidos. Como los datos son mensuales, esta variable se ha agregado trimestralmente.

# Capítulo 3

# Análisis Exploratorio

En un procedimiento previo a la elaboración de los modelos que traten de explicar el comportamiento de los ocupados EPA, es conveniente presentar un análisis exploratorio de las variables.

En este capítulo se presentan una serie de gráficos y medidas propias para cada variable estudiada. Se estudiará la presencia de multicolinealidad entre las variables, así como la normalidad en la variable respuesta, hipótesis básica para poder emplear métodos de estimación paramétricos, que son los que en el Capítulo 4 se utilizarán.

Denotamos como:

| | |
|---|---|
| $Y$ | *estimador directo del número de ocupados en la EPA* |
| $X_1$ | *número de afiliados a la SS a final de mes* |
| $X_2$ | *número de ocupados según encuestas propias para cada sector* |
| $X_3$ | *número de nuevos contratos registrados* |

Se empieza el estudio representando unos diagramas de dispersión para observar cómo se comporta cada par de valores $(x_i, y_i)$. Los diagramas se representarán tanto para las variables en su totalidad como para las variables agrupadas por el sector de actividad: sector primario, industria, construcción y servicios (ver en anexo I las actividades que forman parte de cada sector económico).

La Figura 3.1 estudia la relación lineal entre la variable respuesta y cada una de las covariables consideradas, tanto en la forma original de las mismas como una vez realizada una transformación logarítmica. En ella se observa una clara relación lineal de las variables explicativas respecto a la variable respuesta tanto para los diagramas

que muestran las variables tomadas en su forma original como en aquellos que tratan las variables en su forma logarítmica. A medida que aumenta el valor de una de ellas también le sucede lo mismo a la variable dependiente.

Si se considera el caso en el que la variable respuesta toma su valor logarítmico pero no lo hace así la variable explicativa a la que se enfrenta, se pierde la forma lineal que se tiene en los otros casos.

Cabe señalar, en los diagramas de dispersión de los ocupados EPA respecto de los nuevos contratos registrados (Y respecto de $X_3$) una acumulación de puntos en el margen inferior derecho de cada uno de los tres gráficos correspondientes que provocan una modificación de la pendiente de la recta de regresión a través de un efecto palanca. Esto se produce al tirar ese conjunto de datos de la recta de regresión hacia abajo de forma notable al encontrarse alejados del centro de gravedad. Este efecto se debe a que dichos datos representan la rama de actividad 78 de la CNAE 09 que son las actividades relacionadas con el empleo, y esto incluye a las empresas de trabajo temporal, por lo que es comprensible que se dé el caso de tener muchos contratos en un periodo de tiempo pequeño sin que esto suponga que el número de ocupados aumente en la misma proporción.

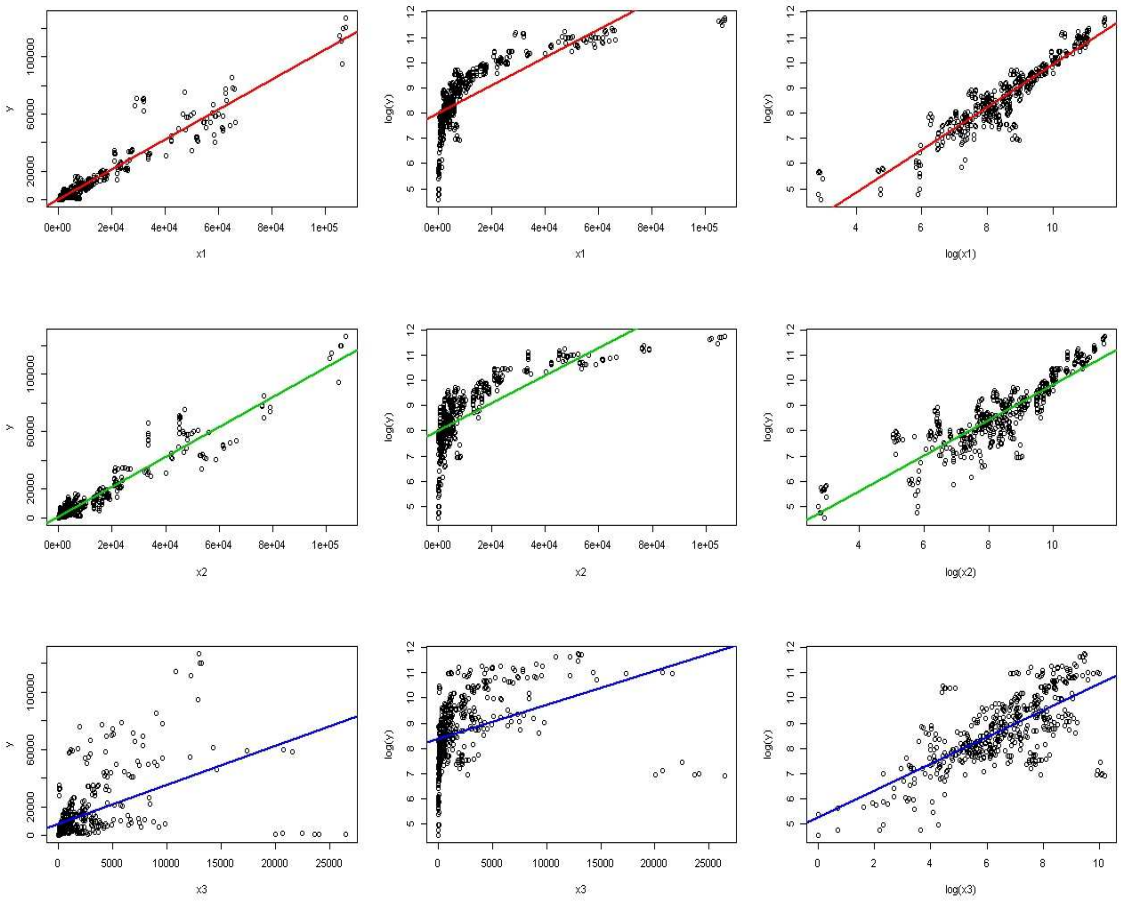


**Figura 3.1. Diagramas de dispersión de la variable respuesta respecto a cada variable explicativa, de la respuesta en logaritmos respecto a la explicativa y de la respuesta en logaritmos respecto a la explicativa en logaritmos.**

Además de lo considerado anteriormente, también se muestra en la Figura 3.1 un comportamiento similar de las variables $X_1$ y $X_2$ al enfrentarlas a la variable respuesta que puede llevar en última instancia a problemas de multicolinealidad entre las covariables.

Las Figuras 3.2, 3.3, 3.4 y 3.5 representan de nuevo los diagramas de dispersión pero en este caso diferenciados por el sector principal de la actividad económica al que hacen referencia: agricultura o sector primario, industria, construcción y servicios respectivamente.

Sorprende ver, en la Figura 3.2, la relación lineal descendente que existe para el sector primario entre la variable respuesta y la variable referente a los nuevos contratos registrados.

**Sector Primario**

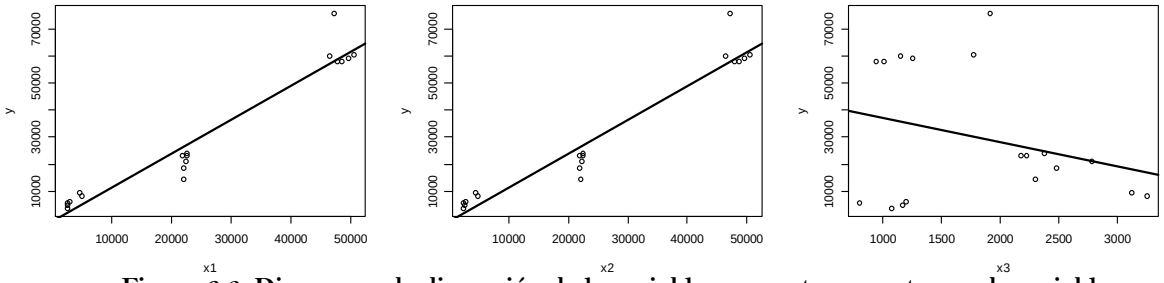


**Figura 3.2. Diagramas de dispersión de la variable respuesta respecto a cada variable explicativa para el sector de la agricultura.**

**Industria**

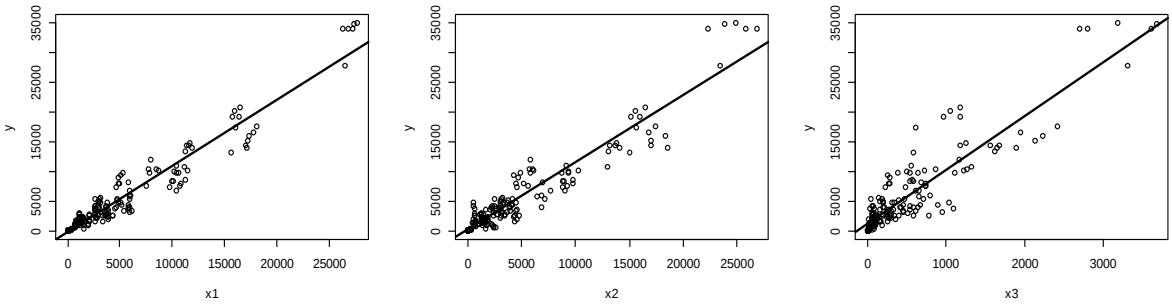


**Figura 3.3. Diagramas de dispersión de la variable respuesta respecto a cada variable explicativa para el sector de la industria.**

**Construcción**

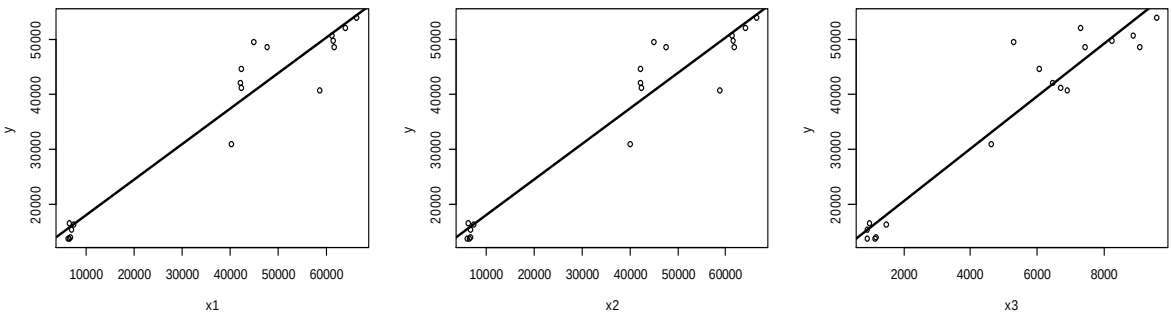


**Figura 3.4. Diagramas de dispersión de la variable respuesta respecto a cada variable explicativa para el sector de la construcción.**

**Servicios**

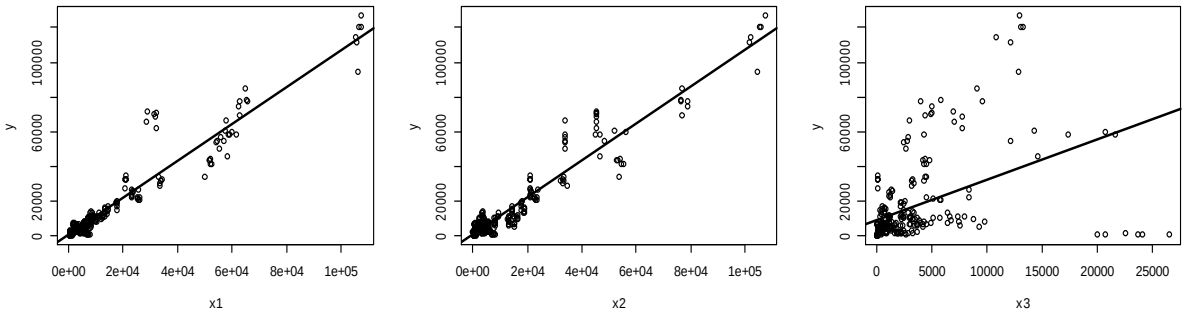


**Figura 3.5. Diagramas de dispersión de la variable respuesta respecto a cada variable explicativa para el sector de los servicios.**

Además, se pone de manifiesto en la Figura 3.5, en el gráfico de la derecha, el poder que tienen los datos correspondientes a los contratos de trabajo temporal. En el diagrama correspondiente se observa con claridad la capacidad de influencia de estos datos sobre la recta, cambiando con claridad su pendiente.

Es comprensible pensar, a tenor de los gráficos anteriores, que las variables presentan una determinada dependencia entre ellas. Cuando esto sucede se interpreta como que existe correlación entre las mismas, y es importante ver cuál es el grado de correlación presente. De tal forma se calculan los coeficientes de correlación entre las variables, que vienen representados en la Tabla 3.1.

| | Método Pearson | Método Spearman |
|---|---|---|
| $\mathbf{Y}$ , $\boldsymbol{X_1}$ | $\rho = 0.958$ | $\rho = 0.907$ |
| $\mathbf{Y}$ , $\boldsymbol{X_2}$ | $\rho = 0.959$ | $\rho = 0.859$ |
| $\mathbf{Y}$ , $\boldsymbol{X_3}$ | $\rho = 0.503$ | $\rho = 0.675$ |
| $\boldsymbol{X_1}$ , $\boldsymbol{X_2}$ | $\boldsymbol{\rho = 0.977}$ | $\rho = 0.948$ |
| $\boldsymbol{X_1}$ , $\boldsymbol{X_3}$ | $\rho = 0.554$ | $\rho = 0.759$ |
| $\boldsymbol{X_2}$ , $\boldsymbol{X_3}$ | $\rho = 0.543$ | $\rho = 0.691$ |
| $\log(\mathbf{Y})$ , $\boldsymbol{X_1}$ | $\rho = 0.761$ | $\rho = 0.907$ |
| $\log(\mathbf{Y})$ , $\boldsymbol{X_2}$ | $\rho = 0.754$ | $\rho = 0.859$ |
| $\log(\mathbf{Y})$ , $\boldsymbol{X_3}$ | $\rho = 0.376$ | $\rho = 0.675$ |

**Tabla 3.1. Coeficientes de correlación entre las variables.**

Con los resultados expuestos en la Tabla 3.1 se puede concluir que existe una correlación positiva entre cualquiera de las variables, siendo esta más fuerte entre la variable correspondiente a los afiliados $X_1$ y la que hace referencia a los ocupados según diversas encuestas sectoriales $X_2$. La correlación positiva hace que cuando una de las variables aumenta por encima de su media, la otra variable también lo haga.

Es preocupante la correlación tan fuerte que se produce entre las variables $X_1$ y $X_2$, que puede llevar a problemas de multicolinealidad y hacer que $|X'X| \approx 0$ generando problemas en las varianzas de los estimadores en donde se incluye la inversa de $|X'X|$, provocando varianzas muy elevadas que finalmente acaban por afectar a la precisión de las estimaciones y al análisis estructural.

Se puede detectar la multicolinealidad mediante el número de condición $\kappa(\mathrm{X})$ que es igual a la raíz cuadrada de la razón entre la raíz característica más grande $\lambda_{max}$ y la raíz característica más pequeña $\lambda_{min}$ de la matriz $|\mathrm{X}'\mathrm{X}|$.

$$\kappa(\mathrm{X}) = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}}$$

Considerando los datos de este estudio tenemos los siguientes valores:

$$\lambda_{max} = 701445658 \quad \lambda_{min} = 7916873$$

Belsley, Kuh y Welsch (1980) estiman que dependencias débiles se muestran con índices que están en torno los valores 5 y 10, mientras que un valor fuerte de dependencia aparece cuando el valor del número de condición es superior a 30.

$$\kappa(\mathrm{X}) = \sqrt{\frac{701445658}{7916873}} = 9.41$$

El valor obtenido es próximo a 10, por lo que hay indicios de una multicolinealidad moderada. Este hecho, sumado al de que el coeficiente de correlación entre las variables $X_1$ y $X_2$ sea próximo a 1 hace pensar que sacar una variable del estudio es una buena opción para tratar de solucionar este problema.

Se elimina por tanto del estudio la variable referente a las encuestas propias de cada sector de actividad $(X_2)$, puesto que en algunos casos no disponía de buenas estimaciones y además añadía una limitación al trabajo, dado que era la variable para la cual se disponía de información para un período de tiempo más corto.

Ahora pues, se continuará con el análisis pero ampliando el horizonte temporal, puesto que para las variables restantes se dispone de un mayor número de datos de los que existía para la variable eliminada. Se va a tratar de aprovechar este hecho, estimando a partir de ahora los ocupados EPA para los datos pertenecientes al periodo que va desde julio de 2009 hasta diciembre de 2011.

Se renombra la variable nuevos contratos registrados como $(X_2)$.

La normalidad es una de las hipótesis básicas para decidir si se pueden utilizar procedimientos de inferencia paramétricos o si por el contrario hay que emplear procedimientos no paramétricos. Debido a su importancia, resulta necesario un estudio de la misma. De tal forma, se van a representar los QQ-Plots correspondientes a la variable respuesta en su forma original y en su forma logarítmica, y para verificar su presencia se va a contrastar esta hipótesis, utilizando para ello dos métodos: el método de Shapiro-Wilk y el método de Kolmogorov-Smirnov-Lilliefors.

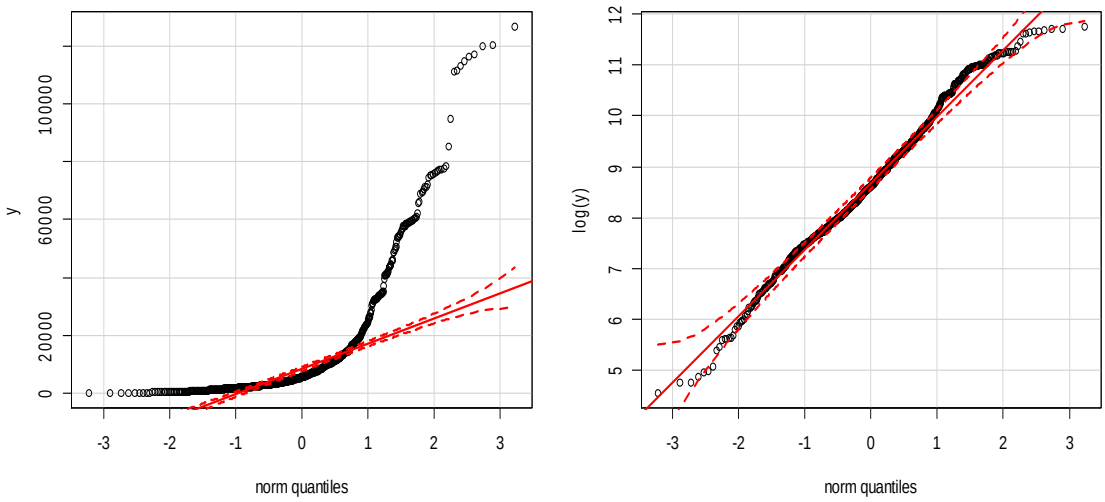


**Figura 3.6. Estudio de la normalidad de los ocupados EPA y de log(ocupadosEPA).**

A tenor de lo que muestra la Figura 3.6, los datos de los ocupados EPA no presentan un comportamiento normal en su forma original, pero sí parece que lo muestran en la forma logarítmica. Otro método visual que se puede emplear para observar la presencia o ausencia de normalidad es el histograma. La Figura 3.7 muestra el histograma del logaritmo de la variable respuesta. Se observa en sus densidades un comportamiento parecido al de una normal, con mayor densidad en el centro y con poca densidad en las colas.

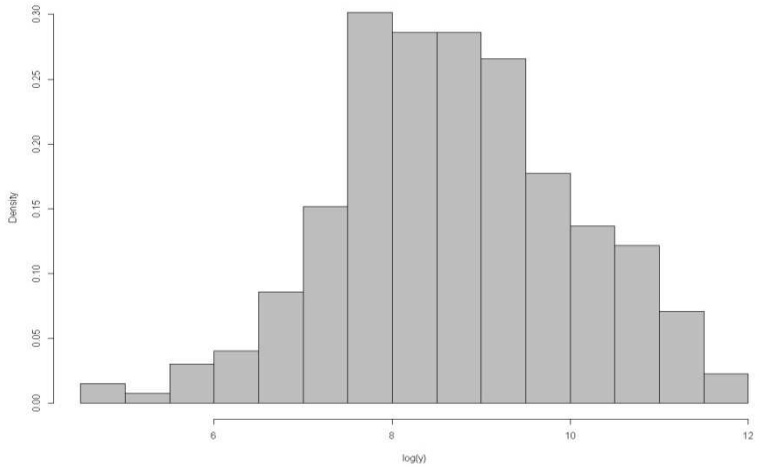


**Figura 3.7. Histograma del logaritmo de los ocupados EPA.**

A continuación se presenta la tabla con los estadísticos obtenidos en los contrastes de normalidad y el p-valor asociado a cada uno. Tan sólo se le ha aplicado el test de normalidad a la variable en su forma logarítmica puesto que el método gráfico del QQ-Plot ya nos hacía ver que en su forma original no cumplía la hipótesis de comportamiento gaussiano.

| | Shapiro Wilk | Kolmogrov-Smirnov-Lilliefors |
|---|---|---|
| **log (Y)** | W = 0.99, p-valor = 0.00 | D = 0.03, p-valor = 0.02 |

**Tabla 3.2. Resultados de los contrastes de normalidad para log(Y).**

La Tabla 3.2 muestra que se podría aceptar la hipótesis de normalidad por el método de Kolmogorov-Smirnov-Lilliefors para la variable log(Y) con un $\alpha = 0.025$. Este es un buen resultado puesto que estamos tratando con muestras muy grandes, y estos contrastes de normalidad suelen dar peores resultados cuando las muestras son de este tipo.

Para observar el error de muestreo que arrastra la EPA y así justificar el uso de técnicas de estimación en áreas pequeñas, es interesante realizar un diagrama de dispersión del tamaño de la muestra de la EPA y el coeficiente de variación del estimador directo en los diferentes dominios, obtenidos con la expresión (3) tal y como se muestra en la Figura 3.8. Se pone de manifiesto que cuanto menor es el tamaño de muestra, mayor es el coeficiente de variación asociado al mismo, sobre todo comparándolo con el coeficiente de variación deseable o aceptable para tales datos del 20% (ONS, 2004) que viene representado por la línea horizontal discontinua. Por tanto se hace necesario el uso de técnicas de estimación en áreas pequeñas.

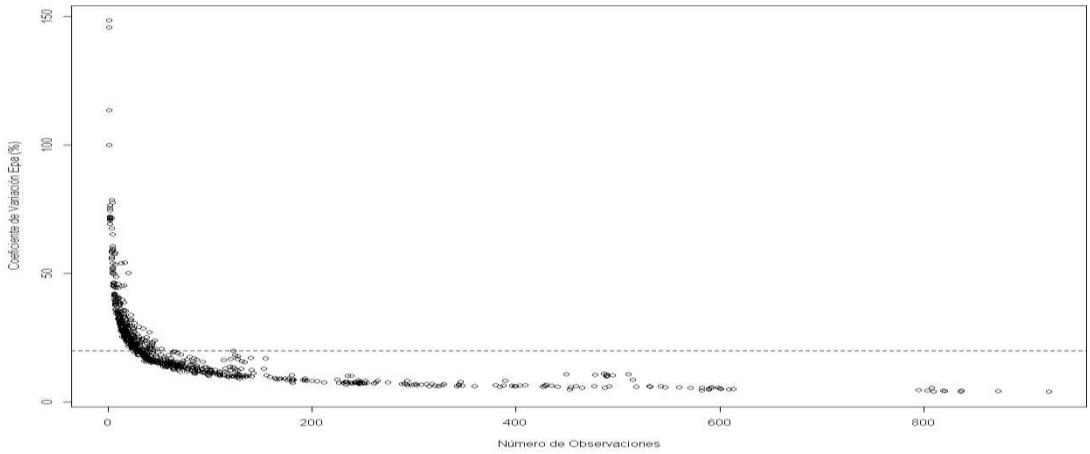


**Figura 3.8. Diagrama de dispersión del coeficiente de variación de los ocupados EPA respecto al tamaño de muestra.**

Para mostrar de un modo más preciso la incidencia que las áreas pequeñas tienen en el incremento de los coeficientes de variación, se van a representar unos diagramas de caja por sectores de actividad: sector primario, construcción, industria y servicios que pongan de manifiesto que cuando hay pocas observaciones el coeficiente de variación es mayor.

En la Figura 3.9 se observa que las actividades que presentan menos observaciones son las pertenecientes a los sectores de la industria y los servicios. Estos sectores, pese a que presentan observaciones con muchos datos para determinados dominios, tienen un valor para la mediana bajo, por lo que tienen asociados muchas áreas pequeñas. Se observa que son precisamente estos sectores los que tienen asociados unos coeficientes de variación más grandes, llegando a niveles de casi un 150% para algunos dominios.

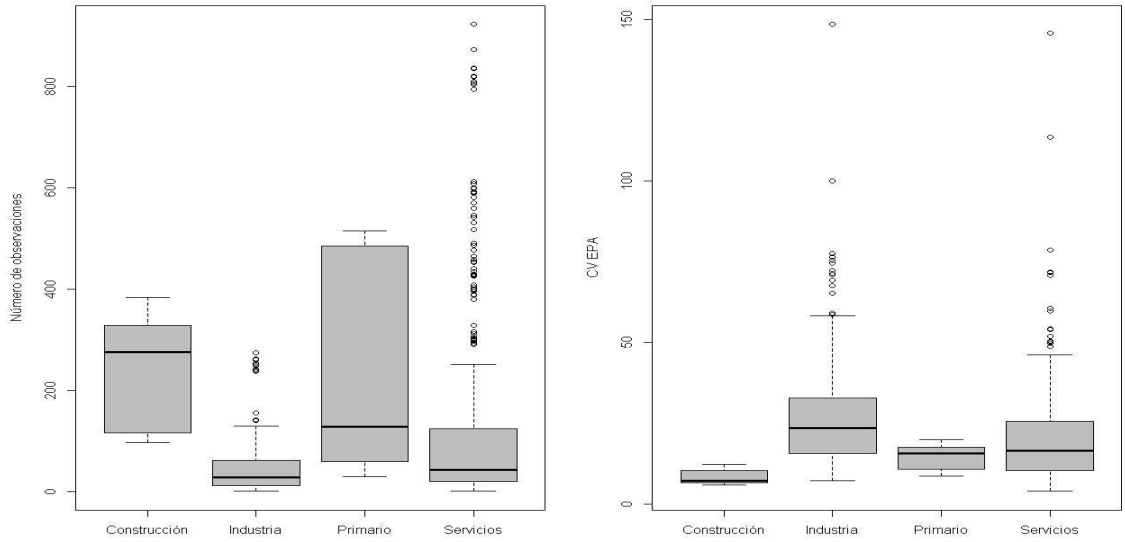


**Figura 3.9. Diagrama de caja del tamaño de muestra (izquierda) y del coeficiente de variación (derecha) en cada sector de actividad.**

Para observar mejor esta serie de medidas se presentan unas tablas indicativas de las mismas.

| Sector | Min | Max | Media | Mediana |
|---|---|---|---|---|
| **Construcción** | 97 | 384 | 242.3 | 276 |
| **Industria** | 1 | 274 | 44.63 | 28 |
| **Sector Primario** | 30 | 515 | 223.5 | 129 |
| **Servicios** | 1 | 923 | 120.7 | 44 |
| **Total** | 1 | 923 | 102.3 | 40 |

**Tabla 3.3. Distribución del tamaño muestral por sectores**

| Sector | Min | Max | Media | Mediana |
|---|---|---|---|---|
| **Construcción** | 6.026 | 12.3 | 8.247 | 7.255 |
| **Industria** | 7.28 | 148.3 | 27.84 | 23.61 |
| **Sector Primario** | 8.703 | 19.89 | 14.64 | 15.74 |
| **Servicios** | 4.026 | 145.7 | 19.78 | 16.6 |
| **Total** | 4.026 | 148.3 | 22 | 18.1 |

**Tabla 3.4. Distribución del coeficiente de variación por sectores.**

Si bien es cierto que los coeficientes de variación de las actividades económicas pertenecientes al sector primario y a la construcción no llegan a superar el nivel aceptable del 20%, las actividades económicas que están integradas en los otros dos sectores de actividad presentan coeficientes de variación mucho más elevados, tal como se ve reflejado en las medidas de la Tabla 3.4. En el peor de los casos se llega a un coeficiente de 148.3%, que está asociado a una actividad económica que tiene únicamente una observación (Actividades de apoyo a la industria extractiva (Actividad 09)). De esta forma, se pone de manifiesto la importancia de la estimación en áreas pequeñas de cara a solventar este problema.

# Capítulo 4

# Metodología y Resultados

Este capítulo expone los dos modelos empleados a la hora de realizar la estimación con los datos reales objeto de estudio: el modelo a nivel de área con efectos de tiempo independientes y el modelo a nivel de área con efectos de tiempo correlados.

Se va a presentar cada uno de los modelos por separado junto con los resultados obtenidos para cada uno de ellos (ver resultados numéricos en Anexos II a V). Finalmente se realizará una comparación entre el estimador directo y los estimadores obtenidos con cada uno de los modelos.

## 4.1 Modelo 1. Modelo a Nivel de Área con Efectos de Tiempo Independientes.

Se considera el siguiente modelo (Modelo 1):

$$Y_{dt} = X_{dt}\beta + u_{dt} + e_{dt} \quad (4)$$

$$d = 1, \dots, D$$

$$t = 1, \dots, m_d$$

$Y_{dt}$ es el estimador directo de los ocupados EPA para el dominio d, $d = 1, \dots, D$ en el instante t, $t = 1, \dots, m_d$.

$X_{dt}$ es el vector que contiene los valores de las p variables auxiliares para el dominio d, $d = 1, \dots, D$ en el instante t, $t = 1, \dots, m_d$.

$u_{dt}$ son los efectos aleatorios tales que $u_{dt} \sim N(0, \sigma_u^2)$ $\sigma_u^2$ desconocidos

$e_{dt}$ son los errores $e_{dt} \sim N(0, \sigma_{dt}^2)$ $\sigma_{dt}^2$ son conocidos

$u_{dt}$ independientes de $e_{dt}$.

En notación matricial el modelo es $Y = X\beta + Zu + e$

donde $Y = \underset{1\le d\le D}{col}\ (Y_d)$ , $Y_d = \underset{1\le t\le m_d}{col}\ (Y_{dt})$ , $u = \underset{1\le d\le D}{col}\ (u_d)$ , $u_d = \underset{1\le t\le m_d}{col}\ (u_{dt})$ , $e = \underset{1\le d\le D}{col}\ (e_d)$ , $e_d = \underset{1\le t\le m_d}{col}\ (e_{dt})$ , $X = \underset{1\le d\le D}{col}\ (X_d)$ , $X_d = \underset{1\le t\le m_d}{col}\ (x_{dt})$ , $x_{dt} = \underset{1\le i\le p}{col}'(x_{dti})$ , $\beta = \underset{1\le i\le p}{col}\ (\beta_i)$ , $Z = I_M$ y $M = \sum_{d=1}^{D} m_d$ . En esta notación $u \sim N(0, V_u)$ y $e \sim N(0, V_e)$ son independientes con matrices de varianza – covarianza:

$V_u = \sigma_u^2 I_M$ , $I_M = \underset{1\le d\le D}{diag}\ (I_{md})$, $V_e = \underset{1\le d\le D}{diag}\ (\mathrm{V}_{ed})$ , $V_{ed} = \underset{1\le t\le m_d}{diag}\ (\sigma_{dt}^2)$ con $\sigma_{dt}^2$ conocidas.

## 4.1.1 Estimación del Modelo.

Según el modelo planteado en la sección anterior, los estimadores BLUE de $\beta$ y de $u$ son:

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}Y \quad \text{y} \quad \hat{u} = V_u Z'V^{-1}(Y - X\hat{\beta})$$

donde $Z = I_M$, $M = \sum_{d=1}^{D} m_d$ ,siendo $m_d$ el número de instantes temporales del dominio $d = 1, \dots, D$.

Se calcula la varianza de $Y$ como:

$$var(Y) = V = \sigma_u^2 \underset{1\le d\le D}{diag}\ (I_{md}) + V_e = \underset{1\le d\le D}{diag}\ \left(\sigma_u^2 I_{m_d} + V_{ed}\right) = \underset{1\le d\le D}{diag}\ (V_d)$$

Para calcular los $\hat{\beta}$ y $\hat{u}$ se han empleado las fórmulas:

$$\hat{\beta} = \left(\sum_{d=1}^{D} X_d' V_d^{-1} X_d\right)^{-1} \left(\sum_{d=1}^{D} X_d' V_d^{-1} Y_d\right) \quad , \quad \hat{u} = \sigma_u^2 \underset{1\le d\le D}{col}\ (V_d^{-1}(Y_d - X_d\hat{\beta}))$$

Para la estimación de la varianza de los efectos aleatorios $\sigma_u^2$, que es desconocida, se emplea el método de máxima verosimilitud restringida (REML). Como punto de partida del algoritmo REML se emplea el estimador de Henderson 3 de $\sigma_u^2$. Este estimador viene definido como:

$$\hat{\sigma}_{uH}^2 = \frac{Y'P_2Y - (M - p)}{tr\{P_2\}}$$

donde

$$P_2 = \underbrace{diag}_{1\le d\le D}(V_{ed}^{-1}) - \underbrace{col}_{1\le d\le D}(V_{ed}^{-1}X_d)Q_2\ \underbrace{col}_{1\le d\le D}{}'(X_dV_{ed}^{-1})$$

$$tr\{P_2\} = \sum_{d=1}^{D}\sum_{t=1}^{m_d}\sigma_{dt}^{-2} - \sum_{d=1}^{D} tr\{X'_dV_{ed}^{-2}X_dQ_2\}$$

$$Y'P_2Y = \sum_{d=1}^{D}\sum_{t=1}^{m_d}\sigma_{dt}^{-2}Y_{dt}^2 - \left(\sum_{d=1}^{D}Y'_dV_{ed}^{-1}X_d\right)Q_2\left(\sum_{d=1}^{D}Y'_dV_{ed}^{-1}X_d\right)'$$

$$Q_2 = \sum_{d=1}^{D}(X'_dV_{ed}^{-1}X_d)^{-1}$$

Los estimadores REML se estiman empleando el algoritmo de puntuación de Fisher con la siguiente fórmula.

$$\theta^{k+1} = \theta^k + F^{-1}(\theta^k)S(\theta^k)$$

donde $\theta = \sigma_u^2$. La puntuación REML y la cantidad de información de Fisher son:

$$S = S(\theta) = -\frac{1}{2}tr(P) + \frac{1}{2}Y'P^2Y \quad y \quad F = F(\theta) = \frac{1}{2}tr(P^2)$$

donde

$$P = \underbrace{diag}_{1\le d\le D}(V_d^{-1}) - \underbrace{col}_{1\le d\le D}(V_d^{-1}X_d)Q\ \underbrace{col'}_{1\le d\le D}(X'_dV_d^{-1})$$

$$tr(P) = \sum_{d=1}^{D} tr(V_d^{-1}) - \sum_{d=1}^{D} tr(X'_dV_d^{-2}X_dQ)$$

$$tr(P^2) = \sum_{d=1}^{D} tr(V_d^{-2}) - 2\sum_{d=1}^{D} tr(X'_dV_d^{-3}X_dQ) + tr\left\{\left(\sum_{d=1}^{D}X'_dV_d^{-2}X_d\right)Q\left(\sum_{d=1}^{D}X'_dV_d^{-2}X_d\right)Q\right\}$$

$$Y'P^2Y = \sum_{d=1}^{D}Y'_dV_d^{-2}Y_d - 2\left(\sum_{d=1}^{D}Y'_dV_d^{-1}X_d\right)Q\left(\sum_{d=1}^{D}X'_dV_d^{-2}Y_d\right) + \left(\sum_{d=1}^{D}Y'_dV_d^{-1}X_d\right)Q\left(\sum_{d=1}^{D}X'_dV_d^{-2}X_d\right)Q\left(\sum_{d=1}^{D}Y'_dV_d^{-1}X_d\right)'$$

$$Q = \left( \sum_{d=1}^{D} X'_d V_d^{-1} X_d \right)^{-1}$$

El estimador REML de $\beta$ es $\hat{\beta}_{REML} = \left(X'\hat{V}^{-1}X\right)^{-1} X'\hat{V}^{-1}Y$

Las distribuciones asintóticas de los estimadores REML de $\sigma_u^2$ y $\beta$ son:

$$\hat{\sigma}_u^2 \sim N\left(\theta, F^{-1}(\sigma_u^2)\right), \;\; \hat{\beta} \sim N_p(\beta, (X'V^{-1}X)^{-1})$$

A partir de las cuales obtenemos los intervalos de confianza asintóticos a un nivel $(1-\alpha)$ para $\sigma_u^2$ y $\beta_i$ , que son:

$$\hat{\sigma}_u^2 \mp z_{\alpha/2} v^{1/2}, \quad \hat{\beta}_i \mp z_{\alpha/2} q_{ii}^{1/2}, \qquad i = 1, \ldots, p$$

Donde $\hat{\sigma}_u^2 = \sigma_u^{2,(\kappa)}$, $v = F^{-1}\left(\sigma_u^{2,(\kappa)}\right)$, $\left(X'V^{-1}\left(\sigma_u^{2,(\kappa)}\right)X\right)^{-1} = \left(q_{ij}\right)_{i,j=1,\ldots,p}$, $\kappa$ es la iteración final del algoritmo de puntuación de Fisher y $z_\alpha$ es el cuantil $\alpha$ de la distribución normal estándar $N(0,1)$.

El p-valor para contrastar la hipótesis nula $H_0: \beta_i = 0$ es

$$p = 2P_{H_0}\left(\hat{\beta}_i > |\beta_0|\right) = 2P\left(N(0,1) > |\beta_0| / \sqrt{q_{ii}}\right)$$

En este trabajo el interés está en predecir $\mu_{dt} = X_{dt}\beta + u_{dt}$ con el predictor EBLUP $\hat{\mu}_{dt} = X_{dt}\hat{\beta} + \hat{u}_{dt}$ . Si no se tiene en cuenta el error $e_{dt}$, esto es equivalente a predecir $Y_{dt} = a'Y$ donde *a* es un vector con "1" en la celda $t + \sum_{l=1}^{d-1} m_l$ y "0" en las restantes celdas. Por tanto el total $Y_{dt}$ se estima con $\hat{Y}_{dt}^{EBLUP} = \hat{\mu}_{dt}$.

### 4.1.2 Estimación del MSE.

Una parte muy importante en la estimación en áreas pequeñas es la estimación del error del estimador, en particular se trabaja con el MSE. Así, una vez estimados los parámetros del modelo se propone aplicar la metodología de Prasad y Rao (1990) para aproximar el MSE de $\hat{Y}_{dt}^{EBLUP}$ :

$$MSE\left(\hat{Y}_{dt}^{EBLUP}\right) \approx g_1(\sigma_u^2) + g_2(\sigma_u^2) + g_3(\sigma_u^2)$$

y el estimador de $MSE\left(\hat{Y}_{dt}^{EBLUP}\right)$ es

$$mse\left(\hat{Y}_{dt}^{EBLUP}\right) = g_1(\hat{\sigma}_u^2) + g_2(\hat{\sigma}_u^2) + 2g_3(\hat{\sigma}_u^2)$$

donde

$$g_1(\sigma_u^2) = \frac{\sigma_u^2 \sigma_{dt}^2}{\sigma_u^2 + \sigma_{dt}^2}$$

$$g_2(\sigma_u^2) = [a'_d X_d - \sigma_u^2 a'_d V_{ed}^{-1} X_d + \sigma_u^4 a'_d V_d^{-1} V_{ed}^{-1} X_d] Q [X'_d a_d - \sigma_u^2 X'_d V_{ed}^{-1} a_d + \sigma_u^4 X'_d V_{ed}^{-1} V_d^{-1} a_d]$$

$$\text{con } a_d = \underbrace{col}_{1 \le k \le m_d} (\delta_{tk})$$

$$g_3(\sigma_u^2) = qF^{-1}(\sigma_u^2), \qquad q = \frac{1}{\sigma_u^2 + \sigma_{dt}^2} - \frac{2\sigma_u^2}{(\sigma_u^2 + \sigma_{dt}^2)^2} + \frac{\sigma_u^4}{(\sigma_u^2 + \sigma_{dt}^2)^3}$$

y $F$ es la cantidad de información de Fisher calculada a través del REML a través de la ecuación del algoritmo de puntuaciones de Fisher.

### 4.1.3 Aplicación a Datos Reales.

En este apartado se aplica el modelo temporal con efectos de tiempo independientes a los datos descritos en el Capítulo 2. En esta aplicación contamos con $D = 79$ dominios y $m_d = 10$ para $d = 1, \dots, 79$.

Una vez estimado el modelo (4) se obtienen los resultados de la Tabla 4.1 para los efectos fijos y aleatorios. Cabe decir que se ha considerado finalmente la variable respuesta en su forma logarítmica, puesto que de esa manera cumplía con la propiedad de normalidad necesaria para emplear métodos paramétricos, los cuales se emplean en este modelo. Además, las variables explicativas también se han tomado en su forma logarítmica para considerarlas en las mismas unidades que la variable respuesta, puesto que es una transformación biyectiva y esto no afecta al comportamiento de las mismas.

Una vez realizados los contrastes de significatividad para los parámetros $\beta_i$: $\begin{matrix} H_0: \beta_i = 0 \\ H_1: \beta_i \neq 0 \end{matrix}$ se obtiene como resultado que los parámetros son significativos dentro del modelo.

| Parámetros | Valor | Error estándar | p-valor |
|---|---|---|---|
| $\boldsymbol{\beta_o}$ | 1.135 | 0.206 | 0.000 |
| $\boldsymbol{\beta_1}$ | 0.916 | 0.033 | 0.000 |
| $\boldsymbol{\beta_2}$ | 0.036 | 0.027 | 0.008 |
| $\boldsymbol{\sigma_u}$ | 0.129 | 0.017 | |

**Tabla 4.1. Estimación de los parámetros del Modelo 1 y p-valor asociado a cada uno de ellos.**

Se presentan a continuación los gráficos correspondientes en primer lugar al estimador directo frente al obtenido con el modelo, y en segundo lugar un zoom para aquellas observaciones que contienen un menor número de ocupados para ver cómo se comporta el modelo en los dominios pequeños.

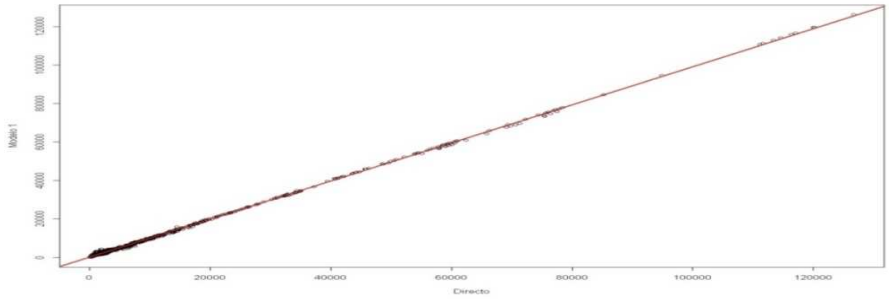


**Figura 4.1. Diagrama de dispersión del estimador directo frente al obtenido con el Modelo 1.**

La Figura 4.1 muestra que el ajuste aproxima los datos que tienen un gran número de observaciones muy bien, pero en el margen inferior izquierdo se observan una serie de valores que se mueven entre 0 y 10.000 ocupados que merecen un mayor análisis, por lo cual parece apropiado hacer un zoom sobre esa zona, el cual se puede ver en la Figura 4.2.

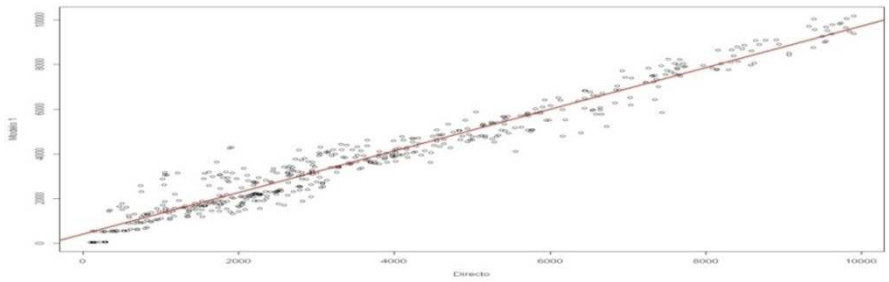


**Figura 4.2. Residuos del Modelo 1 representados por tamaño de muestra.**

La Figura 4.2 pone de manifiesto una mayor discrepancia entre el estimador directo y el obtenido mediante el modelo en los dominios con valores pequeños en el número de ocupados.

En la Figura 4.3 se observan los residuos del modelo ordenados por tamaño de muestra. Los residuos se han calculado mediante la siguiente expresión:

$$Res = \frac{Y_{dt} - \hat{Y}_{dt}^{EBLUP}}{\hat{Y}_{dt}^{EBLUP}} \qquad (5)$$

Se observa la presencia de residuos altos para tamaños de muestra pequeños y valores muy próximos a 0 en general.

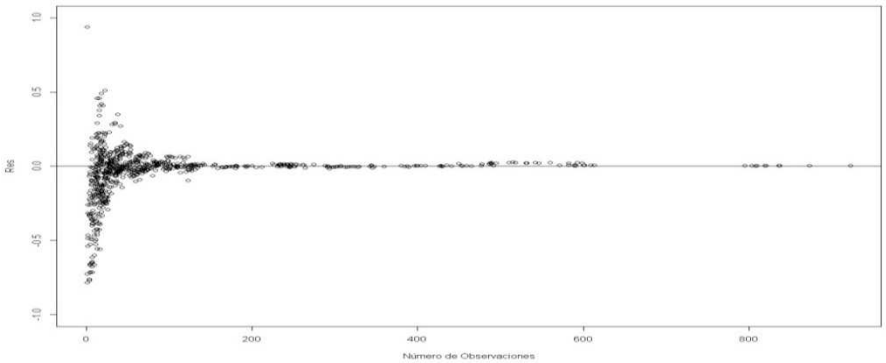


**Figura 4.3. Residuos del Modelo 1 en frente al tamaño muestral de los dominios.**

Se presentan ahora una serie de gráficos para los cuatro últimos trimestres que muestran el estimador directo y el estimador obtenido con el modelo una vez ordenados los datos según el tamaño de muestra.

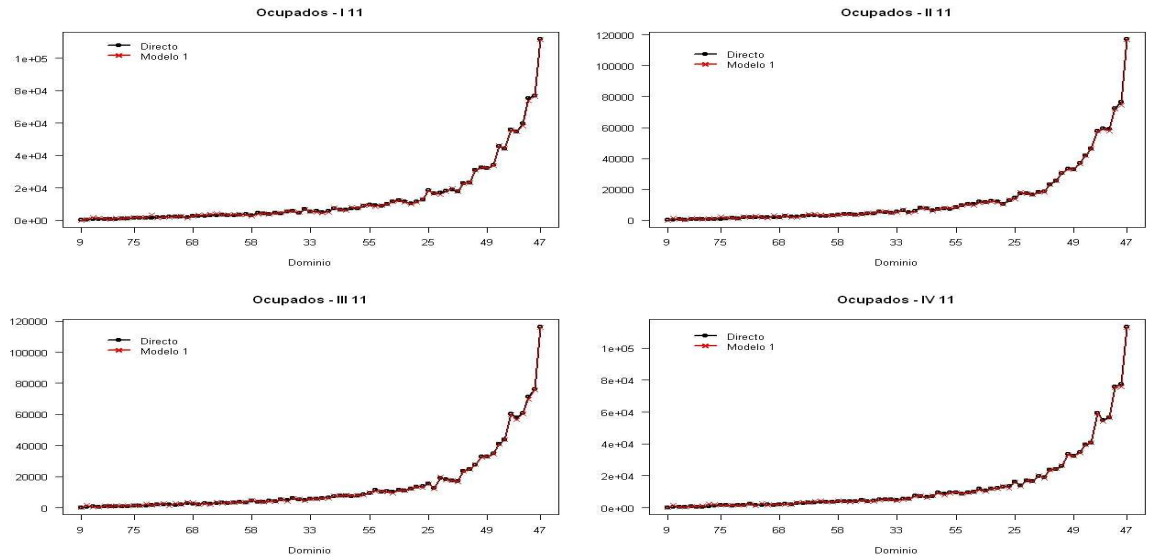


**Figura 4.4. Estimador directo y estimador obtenido por el Modelo 1 ordenado por tamaño de muestra para los cuatro trimestres del 2011.**

Del mismo modo que para la representación de la totalidad de los datos, se hace un zoom para observar cómo son las estimaciones del modelo para tamaños de muestra pequeños ya que cuando el tamaño es considerable los estimadores funcionan igual.

La Figura 4.5 muestra que las mayores variaciones entre el estimador directo y el estimador del modelo se producen en los dominios que tienen entre 1 y 4.000 observaciones, y parece que después ambos estimadores se aproximan bastante.

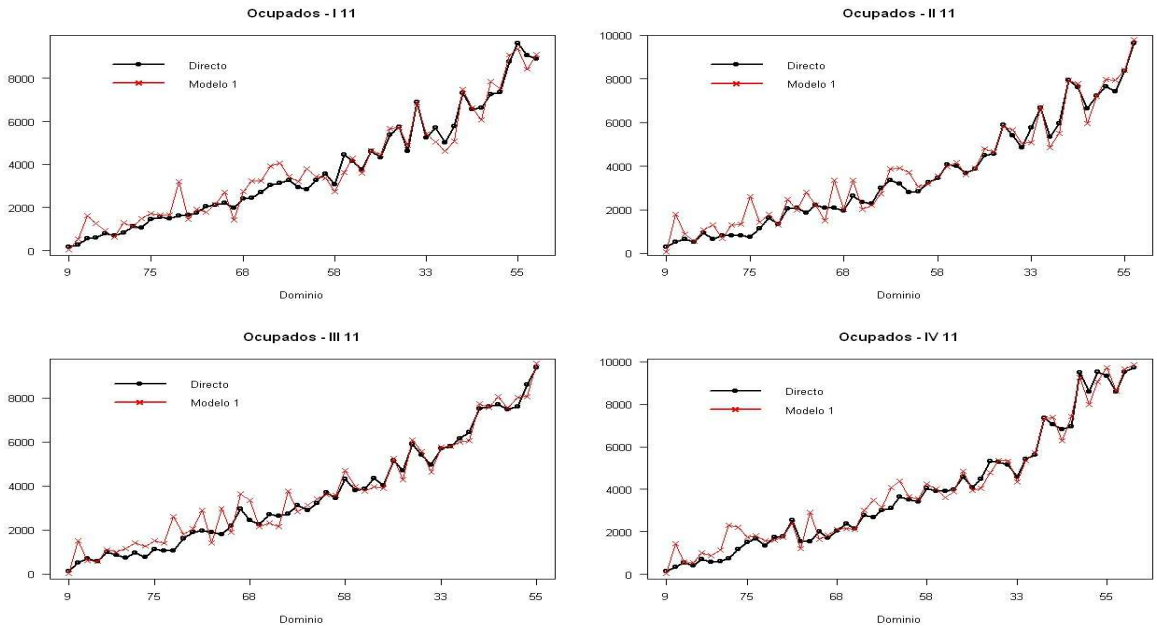


**Figura 4.5. Comparación del estimador directo y el obtenido por el Modelo 1 para un tamaño del número de ocupados <10000 en los cuatro trimestres del 2011.**

En el análisis exploratorio se observaban unos coeficientes de variación muy altos para determinadas observaciones en el estimador directo. Es conveniente comprobar que el modelo ha sido capaz de reducir el valor de estos coeficientes. Para ello se representan en la Figura 4.6 los cuatro últimos trimestres objeto de estudio, comparando el coeficiente de variación del estimador directo (3) y el coeficiente de variación obtenido con el Modelo1, que se obtiene a través de la siguiente expresión:

$$\widehat{CV} = \frac{\sqrt{MSE\left(\hat{Y}_{dt}^{EBLUP}\right)}}{\hat{Y}_{dt}^{EBLUP}} \qquad (6)$$

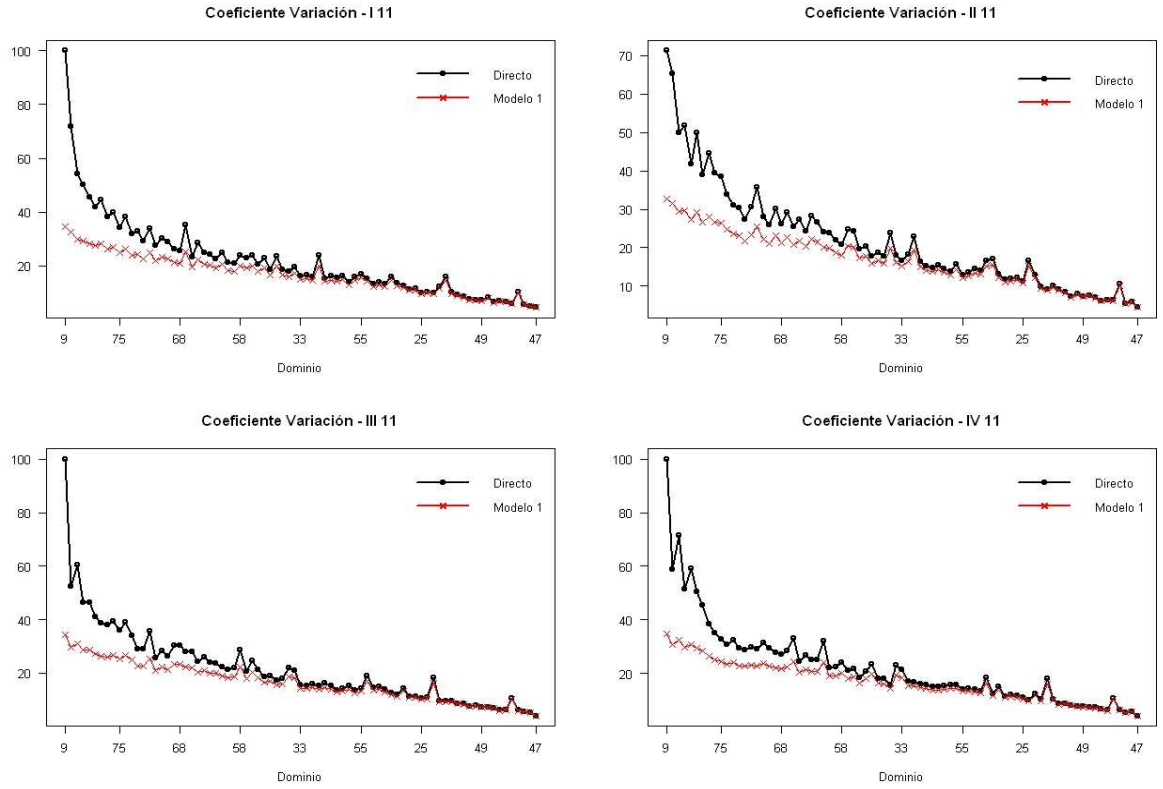


**Figura 4.6. Coeficiente de variación original frente al estimado para los cuatro últimos trimestres.**

Tal como muestra la Figura 4.6, los coeficientes de variación han experimentado una gran disminución, sobre todo para aquellos datos que contienen pocas observaciones. El Modelo 1 ha reducido coeficientes de variación de más de un 100% en coeficientes que no llegan a superar el 40%, lo cual supone un gran avance en la estimación.

Para concluir con el análisis de los resultados del modelo se han escogido una serie de actividades económicas representativas, o dominios en este caso, para ver la evolución del estimador directo y el obtenido a partir del Modelo 1 entre el tercer trimestre de 2009 y el cuarto trimestre de 2011. El criterio de elección se presenta en la Tabla 4.2 (Véase en el Anexo I la descripción de las mismas).

| | |
|---|---|
| **Fila 1** | Actividades con un tamaño muestral inferior a 10. |
| **Fila 2** | Actividades con un tamaño muestral de entre 11 y 20. |
| **Fila 3** | Actividades con un tamaño muestral de entre 21 y 70. |
| **Fila 4** | Actividades con un tamaño muestral de entre 71 y 200. |
| **Fila 5** | Actividades con un tamaño muestral mayor que 200. |

**Tabla 4.2. Criterio empleado para la elección de los dominios en la Figura 4.7.**

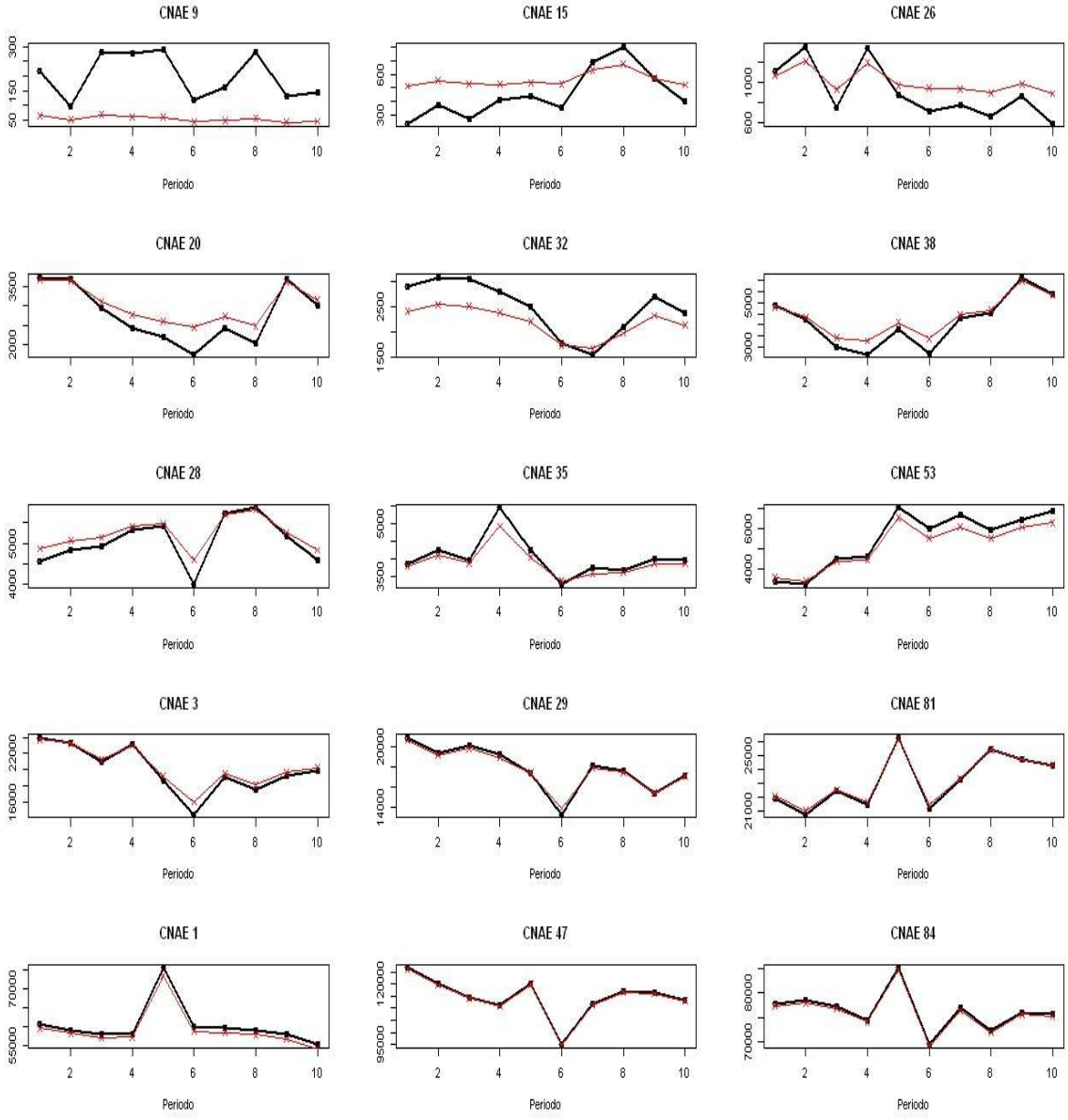


**Figura 4.7. Evolución del estimador directo (negro) y del estimador obtenido con el Modelo 1 (rojo) entre el tercer trimestre del 2009 y el cuarto trimestre de 2011 para algunas actividades económicas.**

Figura 4.7 muestra la evolución a lo largo de los periodos, siendo el primero el tercer trimestre de 2009 y el décimo, y último, el cuarto trimestre de 2011. Se observa que la estimación del modelo suaviza la evolución de aquéllos dominios que presentan pocas observaciones mientras que se aproxima a la de los dominios que presentan muchas observaciones. Se van a representar también los coeficientes de variación de

estas actividades económicas para ver cómo ha sido el cambio que ha provocado el modelo en ellas.

La Figura 4.8 pone de manifiesto que la disminución en el coeficiente de variación es tanto más grande cuantas menos observaciones tiene el dominio. Observamos que el dominio 9, que posee tan sólo una observación experimenta una disminución en su coeficiente de variación de noventa puntos porcentuales, aunque todavía se queda algo lejos del límite del 20% estipulado por la ONS (2004).

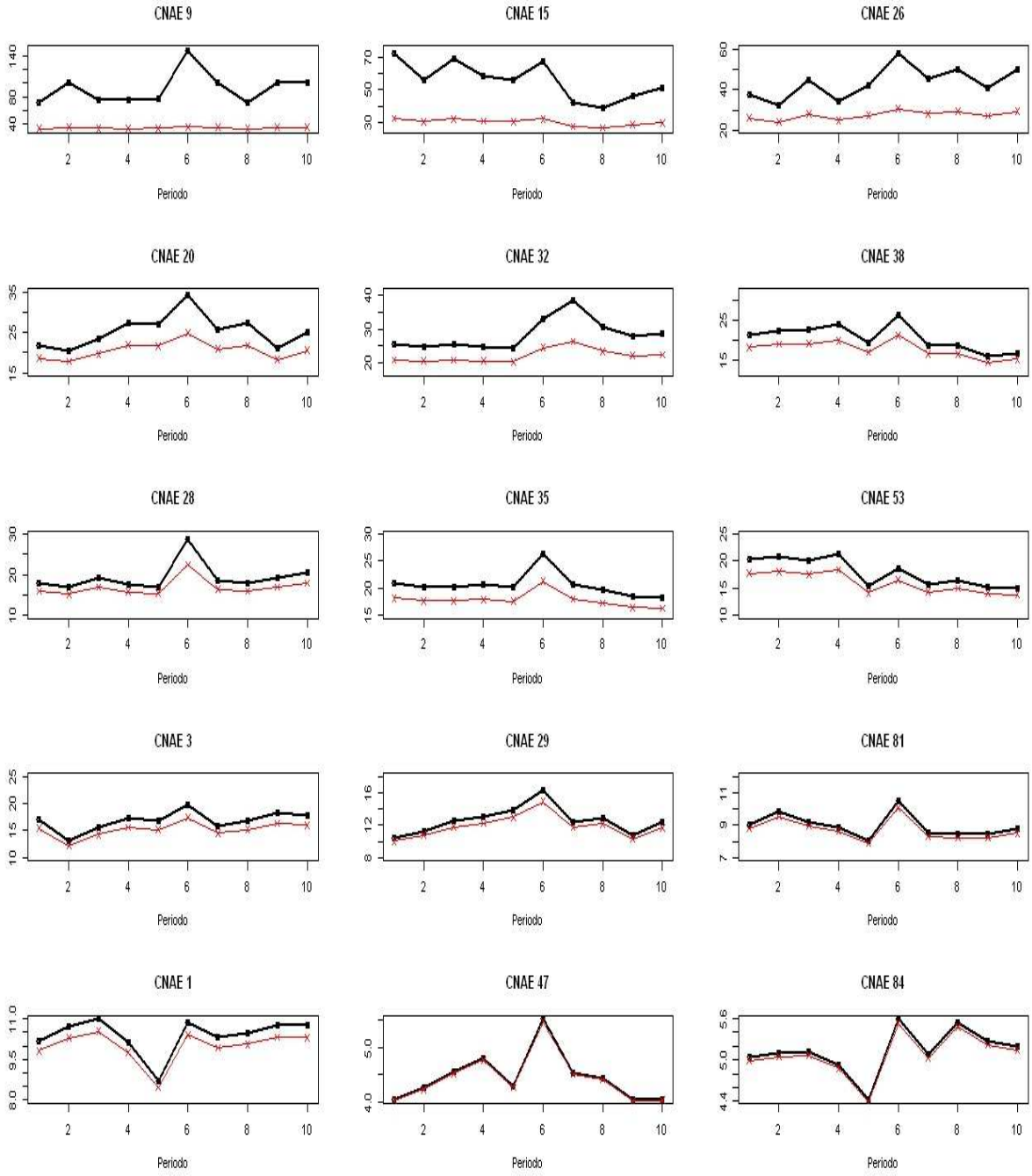


**Figura 4.8. Evolución del coeficiente de variación del estimador directo (negro) y del obtenido con el Modelo 1 (rojo) entre el tercer trimestre del 2009 y el cuarto trimestre de 2011 para algunas actividades económicas.**

En este apartado solo se presentan resultados gráficos, los resultados numéricos correspondientes el 3º y 4º trimestre del 2011 se pueden consultar en los anexos II a V.

Es conveniente por tanto, profundizar en el análisis e incluir otros supuestos para tratar de disminuir un poco más este coeficiente y llevarlo a niveles aceptables. El modelo 2 que se explica a continuación considera que los efectos aleatorios temporales están correlados y siguen un proceso autorregresivo de orden 1, AR(1), es decir, lo que sucede en un instante de tiempo $t$ tiene que ver con lo que sucede en el instante de tiempo anterior $t-1$, por ejemplo, el número de ocupados en la construcción de edificios en el cuarto trimestre de 2010 depende del número de ocupados en la construcción en el tercer trimestre de ese año.

## 4.2 Modelo 2. Modelo a Nivel de Área con Efectos de Tiempo Correlados.

Se considera el siguiente modelo (Modelo 2):

$$Y_{dt} = X_{dt}\beta + u_{dt} + e_{dt}$$

$$d = 1, \dots, D$$

$$t = 1, \dots, m_d$$

$Y_{dt}$ es el estimador directo de los ocupados EPA para el dominio d, $d = 1, \dots, D$ en el instante t, $t = 1, \dots, m_d$.

$X_{dt}$ es el vector que contiene los valores de las p variables auxiliares para el dominio d, $d = 1, \dots, D$ en el instante t, $t = 1, \dots, m_d$.

$u_{dt}$ son los efectos aleatorios temporales tales que $u_{dt}$ $i.i.d.$ $AR(1)$ con parámetros de varianza y autocorrelación $\sigma_u^2$ y $\rho$ respectivamente.

$e_{dt}$ son los errores $e_{dt} \sim N(0, \sigma_{dt}^2)$ $\sigma_{dt}^2$ son conocidos

$u_{dt}$ independientes de $e_{dt}$.

En notación matricial el modelo es $Y = X\beta + Zu + e$

donde $Y = \underset{1\le d\le D}{col}(Y_d)$ , $Y_d = \underset{1\le t\le m_d}{col}(Y_{dt})$ , $u = \underset{1\le d\le D}{col}(u_d)$ , $u_d = \underset{1\le t\le m_d}{col}(u_{dt})$ , $e = \underset{1\le d\le D}{col}(e_d)$ , $e_d = \underset{1\le t\le m_d}{col}(e_{dt})$ , $X = \underset{1\le d\le D}{col}(X_d)$ , $X_d = \underset{1\le t\le m_d}{col}(x_{dt})$ , $x_{dt} = \underset{1\le i\le p}{col}'(x_{dti})$ , $\beta = \underset{1\le i\le p}{col}(\beta_i)$ , $Z = I_{MxM}$ y $M = \sum_{d=1}^{D} m_d$ . En esta notación $u \sim N(0, V_u)$ y $e \sim N(0, V_e)$ son independientes con matrices de varianza – covarianza:

$$V_u = \sigma_u^2 \Omega(\rho) \quad , \quad \Omega(\rho) = \underset{1\le d\le D}{diag}\,(\Omega_d(\rho)) \quad , \quad V_e = \underset{1\le d\le D}{diag}\,(\mathrm{V}_{ed}) \quad , \quad V_{ed} = \underset{1\le t\le m_d}{diag}\,(\sigma_{dt}^2)$$

donde $\sigma_{dt}^2$ son conocidas y

$$\Omega_d = \Omega_d(\rho) = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & \rho & \dots & \rho^{m_d-2} & \rho^{m_d-1} \\ \rho & 1 & \ddots & & \rho^{m_d-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \rho^{m_d-2} & & \ddots & 1 & \rho \\ \rho^{m_d-1} & \rho^{m_d-2} & \dots & \rho & 1 \end{pmatrix}$$

Si las componentes de la varianza son conocidas, entonces los estimadores BLUE de β y BLUP de u son:

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}Y \qquad \hat{u} = V_u Z'V^{-1}(Y - X\hat{\beta})$$

donde $var(Y) = V = \sigma_u^2 \underset{1\le d\le D}{diag}\,(\Omega_d(\rho)) + V_e = \underset{1\le d\le D}{diag}\,(\sigma_u^2\Omega_d(\rho) + V_{ed}) = \underset{1\le d\le D}{diag}\,(V_d)$

## 4.2.1 Estimación del Modelo.

Para el cálculo de $\hat{\beta}$ y $\hat{u}$ se han aplicado las siguientes fórmulas:

$$\hat{\beta} = \left(\sum_{d=1}^{D} X'_d V_d^{-1} X_d\right)^{-1}\left(\sum_{d=1}^{D} X'_d V_d^{-1} Y_d\right) \qquad \hat{u} = \sigma_u^2 \underset{1\le d\le D}{col}\left(\Omega_d(\rho)V_d^{-1}\left(Y_d - X_d\hat{\beta}\right)\right)$$

Al ser los componentes de la varianza desconocidos, sus estimadores REML se calculan mediante el algoritmo de puntuación de Fisher con la siguiente fórmula.

$$\theta^{k+1} = \theta^k + F^{-1}(\theta^k)S(\theta^k)$$

donde $\theta = (\theta_1, \theta_2) = (\sigma_u^2, \rho)$. Como semillas, se utilizan $\rho = 0$ y $\sigma_u^{2(0)} = \hat{\sigma}_{uH}^2$, donde $\hat{\sigma}_{uH}^2$ es el estimador Henderson 3 de $\sigma_u^2$ bajo el modelo restringido a $\rho = 0$. Las puntuaciones REML y las componentes de la matriz de información de Fisher son:

$$S_a = -\frac{1}{2}tr(PV_a) + \frac{1}{2}Y'PV_aPY, \quad F_{ab} = \frac{1}{2}tr(PV_aPV_b), \qquad a, b = 1,2$$

donde

$$V_1 = \frac{\partial V}{\partial \sigma_u^2} = \underset{1\le d\le D}{diag}\left(\Omega_d(\rho)\right),\ V_2 = \frac{\partial V}{\partial \rho} = \sigma_u^2 \underset{1\le d\le D}{diag}\left(\dot{\Omega}_d(\rho)\right),\ Q = (\textstyle\sum_{d=1}^{D} X'_d V_d^{-1} X_d)^{-1}$$

$$P = \underset{1\le d\le D}{diag}\,(V_d^{-1}) - \underset{1\le d\le D}{col}\,(V_d^{-1}X_d)Q\ \underset{1\le d\le D}{col'}\,(X'_d V_d^{-1})$$

$$PV_a = \underset{1\le d\le D}{diag}\,(V_d^{-1}V_{ad}) - \underset{1\le d\le D}{col}\,(V_d^{-1}X_d)Q\ \underset{1\le d\le D}{col'}\,(X'_d V_d^{-1}V_{ad})$$

$$tr(PV_a) = \sum_{d=1}^{D} tr(V_d^{-1}V_{ad}) - \sum_{d=1}^{D} tr(X'_d V_d^{-1} V_{ad} V_d^{-1} X_d Q)$$

$$tr(PV_aPV_b) = \sum_{d=1}^{D} tr(V_d^{-1}V_{ad}V_d^{-1}V_{bd}) - 2\sum_{d=1}^{D} tr(X'_d V_d^{-1} V_{ad} V_d^{-1} V_{bd} V_d^{-1} X_d Q) + tr\left\{\left(\sum_{d=1}^{D} X'_d V_d^{-1} V_{ad} V_d^{-1} X_d\right) Q \left(\sum_{d=1}^{D} X'_d V_d^{-1} V_{bd} V_d^{-1} X_d\right) Q\right\}$$

$$Y'PV_aPY = \sum_{d=1}^{D} Y'_d V_d^{-1} V_{ad} V_d^{-1} Y_d - \left(\sum_{d=1}^{D} Y'_d V_d^{-1} V_{ad} V_d^{-1} X_d\right) Q \left(\sum_{d=1}^{D} Y'_d V_d^{-1} X_d\right)' - \left(\sum_{d=1}^{D} Y'_d V_d^{-1} X_d\right) Q \left(\sum_{d=1}^{D} X'_d V_d^{-1} V_{ad} V_d^{-1} Y_d\right) + \left(\sum_{d=1}^{D} Y'_d V_d^{-1} X_d\right) Q \left(\sum_{d=1}^{D} X'_d V_d^{-1} V_{ad} V_d^{-1} X_d\right) Q \left(\sum_{d=1}^{D} Y'_d V_d^{-1} X_d\right)'$$

Finalmente, la derivada de la matriz $\Omega_d(\rho)$ con respecto a $\rho$ es:

$$\dot{\Omega}_d(\rho) = \frac{1}{1-\rho^2}\begin{pmatrix} 0 & 1 & \dots & \dots & (m_d-1)\rho^{m_d-2} \\ 1 & 0 & \ddots & & (m_d-2)\rho^{m_d-3} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ (m_d-2)\rho^{m_d-3} & & \ddots & 0 & 1 \\ (m_d-1)\rho^{m_d-2} & \dots & \dots & 1 & 0 \end{pmatrix} + \frac{2\rho\Omega_d(\rho)}{(1-\rho^2)^2}$$

El estimador REML de $\beta$ se calcula aplicando la fórmula:

$\hat{\beta}_{REML} = \left(X'\hat{V}^{-1}X\right)^{-1}X'\hat{V}^{-1}Y$

Las distribuciones asintóticas de los estimadores REML de $\theta$ y $\beta$ son:

$$\hat{\theta} \sim N_2\left(\theta, F^{-1}(\theta)\right), \quad \hat{\beta} \sim N_p(\beta, (X'V^{-1}X)^{-1})$$

Y los intervalos de confianza asintóticos a un nivel $(1-\alpha)$ para $\theta_a$ y $\beta_i$ son:

$$\theta_a \mp z_{\alpha/2} v_{aa}^{1/2}, a = 1,2 \quad \hat{\beta}_i \mp z_{\alpha/2} q_{ii}^{1/2}, \qquad i = 1, \dots, p$$

Donde $\hat{\theta} = \theta^{\kappa}$, $F^{-1}(\theta^{\kappa}) = (v_{ab})_{a,b=1,2}$, $(X'V^{-1}(\theta^{\kappa})X)^{-1} = \left(q_{ij}\right)_{i,j=1,\dots,p}$, $\kappa$ es la iteración final del algoritmo de puntuación de Fisher y $z_{\alpha}$ es el cuantil $\alpha$ de la distribución normal estándar $N(0,1)$.

El p-valor para contrastar la hipótesis nula $H_0: \beta_i = 0$ es

$$p = 2P_{H_0}\left(\hat{\beta}_i > |\beta_0|\right) = 2P\left(N(0,1) > |\beta_0|/\sqrt{q_{ii}}\right)$$

Nosotros estamos interesados en predecir $\mu_{dt} = X_{dt}\beta + u_{dt}$ con el estimador EBLUP $\hat{\mu}_{dt} = X_{dt}\hat{\beta} + \hat{u}_{dt}$ . Si no se tiene en cuenta el error $e_{dt}$ esto es equivalente a predecir $Y_{dt} = a'Y$ donde *a* es un vector con "1" en la celda $t + \sum_{l=1}^{d-1} m_l$ y "0" en las restantes celdas. Por tanto para estima el total $Y_{dt}$ se estima con $\hat{Y}_{dt}^{EBLUP} = \hat{\mu}_{dt}$.

## 4.2.2 Estimación del MSE.

Utilizando la metodología de Prasad y Rao (1990) el error cuadrático medio de $\hat{\bar{Y}}_{dt}^{EBLUP}$ se aproxima por :

$$MSE\left(\hat{\bar{Y}}_{dt}^{EBLUP}\right) \approx g_1(\theta) + g_2(\theta) + g_3(\theta)$$

por tanto, el estimador de $MSE\left(\hat{\bar{Y}}_{dt}^{EBLUP}\right)$ es:

$$mse\left(\hat{\bar{Y}}_{dt}^{EBLUP}\right) = g_1(\hat{\theta}) + g_2(\hat{\theta}) + 2g_3(\hat{\theta})$$

donde $\theta = (\sigma_u^2, \rho)$ , $a_d = \underbrace{col}_{1\le k\le m_d}(\delta_{tk})$ y las expresiones para $g_1$, $g_2$, $g_3$ son:

$$g_1(\theta) = \sigma_u^2 a'_d \Omega_d a_d - \sigma_u^4 a'_d \Omega_d V_d^{-1} \Omega_d a_d$$

$$g_2(\theta) = [a'_d X_d - \sigma_u^2 a'_d \Omega_d V_{ed}^{-1} X_d + \sigma_u^4 a'_d \Omega_d V_d^{-1} \Omega_d V_{ed}^{-1} X_d] Q [X'_d a_d - \sigma_u^2 X'_d V_{ed}^{-1} \Omega_d a_d + \sigma_u^4 X'_d V_{ed}^{-1} \Omega_d V_d^{-1} \Omega_d a_d]$$

$$g_3(\theta) \approx tr\left\{ \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix} \begin{pmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{pmatrix}^{-1} \right\}$$

donde $F_{ab}$ es el elemento de la matriz de información de Fisher calculado por REML y:

$$q_{11} = a'_d \Omega_d V_d^{-1} \Omega_d a_d - 2\sigma_u^2 a'_d \Omega_d V_d^{-1} \Omega_d V_d^{-1} \Omega_d a_d + \sigma_u^4 a'_d \Omega_d V_d^{-1} \Omega_d V_d^{-1} \Omega_d V_d^{-1} \Omega_d a_d$$

$$q_{12} = \sigma_u^2 a'_d \Omega_d V_d^{-1} \dot{\Omega}_d a_d - \sigma_u^4 a'_d \Omega_d V_d^{-1} \dot{\Omega}_d V_d^{-1} \Omega_d a_d - \sigma_u^4 a'_d \Omega_d V_d^{-1} \Omega_d V_d^{-1} \dot{\Omega}_d a_d + \sigma_u^6 a'_d \Omega_d V_d^{-1} \Omega_d V_d^{-1} \dot{\Omega}_d V_d^{-1} \Omega_d a_d$$

$$q_{22} = \sigma_u^4 a'_d \dot{\Omega}_d V_d^{-1} \dot{\Omega}_d a_d - 2\sigma_u^6 a'_d \Omega_d V_d^{-1} \dot{\Omega}_d V_d^{-1} \dot{\Omega}_d a_d + \sigma_u^8 a'_d \Omega_d V_d^{-1} \dot{\Omega}_d V_d^{-1} \dot{\Omega}_d V_d^{-1} \Omega_d a_d$$

### 4.2.3 Aplicación a Datos Reales.

En este apartado se aplica el modelo temporal con efectos de tiempo correlados a los datos descritos en el Capítulo 2. En esta aplicación contamos con $D = 79$ dominios y $m_d = 10$ para $d = 1, \dots, 79$.

La Tabla 4.3 presenta las estimaciones de los parámetros de regresión y las componentes de la varianza y sus errores estándar. También presenta los p-valores resultantes del contraste de significatividad para los parámetros $\beta_i$: $\begin{matrix} H_0: \beta_i = 0 \\ H_1: \beta_i \neq 0 \end{matrix}$.

| Parámetros | Valor | Error estándar | p-valor |
|---|---|---|---|
| $\boldsymbol{\beta_o}$ | 1.729 | 0.608 | 0.000 |
| $\boldsymbol{\beta_1}$ | 0.803 | 0.084 | 0.000 |
| $\boldsymbol{\beta_2}$ | 0.021 | 0.045 | 0.352 |
| $\boldsymbol{\sigma_u}$ | 0.002 | 0.001 | |
| $\boldsymbol{\rho}$ | 0.994 | 0.003 | |

**Tabla 4.3. Estimación de los parámetros del Modelo 2 y p-valor asociado a cada uno de ellos.**

La Tabla 4.3 pone de manifiesto que el parámetro asociado a la variable contratos no se muestra como significativo dentro del modelo. Aunque convendría eliminar esta variable y realizar de nuevo la estimación, ésta se va a mantener debido a la importancia económica que tiene para el IGE el uso de esta información.

De igual forma que para el modelo Fay-Herriot con efectos de tiempo independientes, se presentan los diagramas de dispersión correspondientes al estimador directo respecto al estimador proporcionado por el modelo tanto para el conjunto total de los datos como para aquéllas observaciones de la EPA que tienen menos de 10.000 ocupados.

Si bien no se aprecia gran diferencia entre la Figura 4.9 y la Figura 4.1 en la que se presentaba el mismo diagrama de dispersión para el Modelo 1, lo cierto es que a priori parece que se observa algo más de dispersión en las observaciones, el ajuste a la recta no es tan exacto como lo era antes. Esta dispersión es probable que se vea mejor realizando un zoom sobre aquellas áreas que presentan pocas observaciones (Figura 4.10).

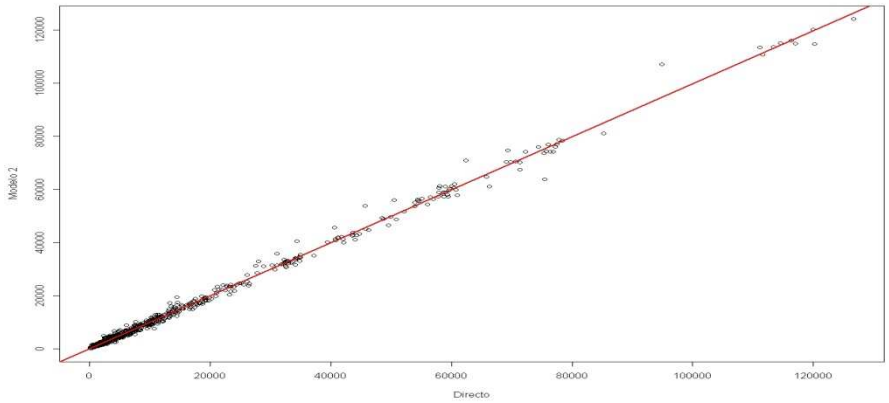


**Figura 4.9. Diagrama de dispersión del estimador directo frente al obtenido con el Modelo 2.**

La Figura 4.10 muestra mayor dispersión de los datos respecto a la recta en la que se igualan estimador directo y estimador del modelo en comparación con la dispersión obtenida en el primer modelo (véase Figura 4.2).

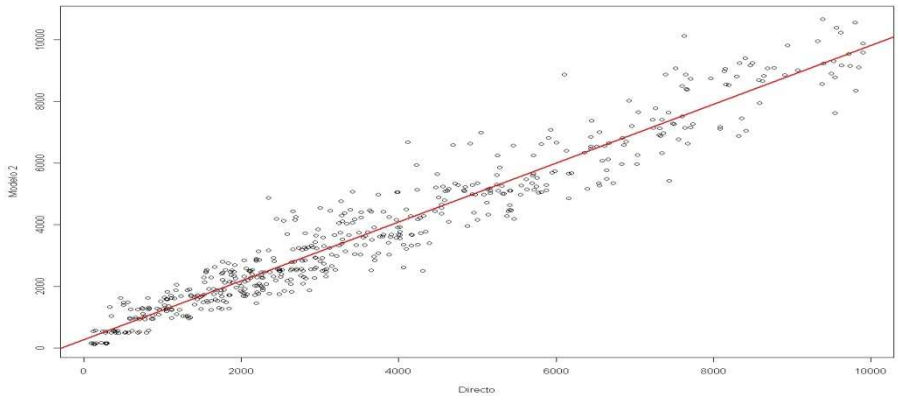


**Figura 4.10. Diagrama de dispersión del estimador directo frente al obtenido con el modelo 2 para un tamaño de ocupados <10000.**

Es conveniente observar también los residuos del modelo en frente al tamaño de muestra de los dominios calculados empleando la expresión (5).

La Figura 4.11 muestra una mayor dispersión en los residuos del Modelo 2. En el modelo con efectos de tiempo independientes se observaban residuos que se acercaban al cero muy rápidamente (véase Figura 4.3), y en cambio en éste, los residuos presentan una mayor dispersión a lo largo de las observaciones.

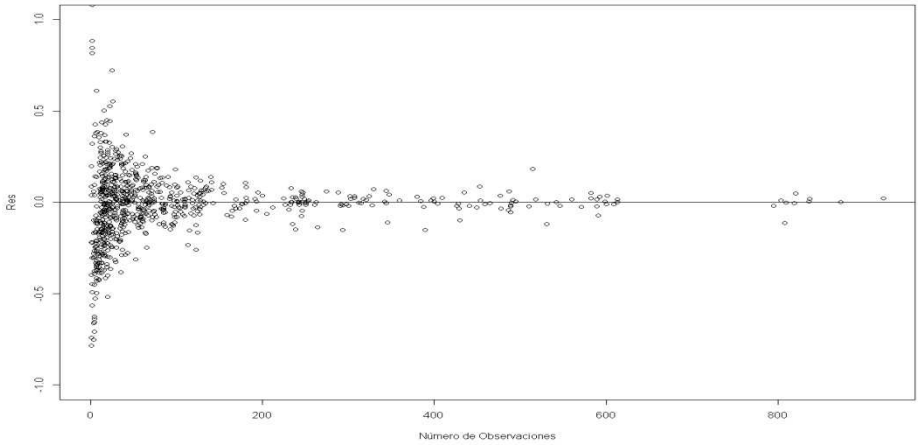


**Figura 4.11. Residuos del Modelo 2 en frente al tamaño de muestra.**

Se presentan en la Figura 4.12 los gráficos para los cuatro últimos trimestres que muestran el estimador directo y los estimadores obtenidos con los modelos con efectos de tiempo independientes y con efectos de tiempo correlados, una vez ordenados los dominios por el tamaño de muestra. No se observan demasiado bien las diferencias entre el estimador directo y los obtenidos con los modelos para estos cuatro trimestres, por lo que se va a volver a representar pero para aquellos dominios con un tamaño de ocupados inferior a 10.000 personas en la Figura 4.13.

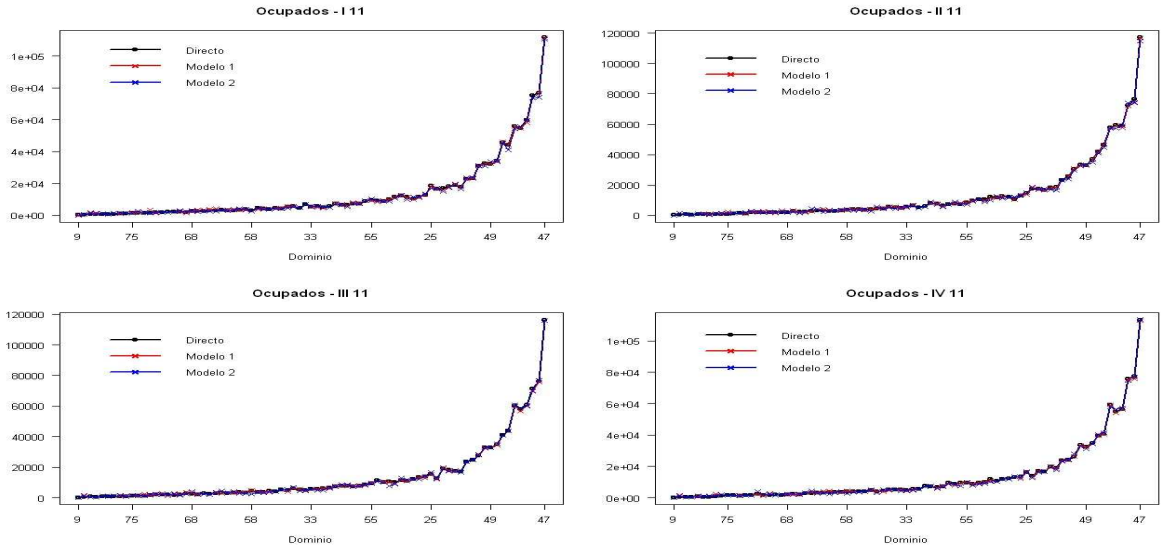


**Figura 4.12. Estimador directo y estimadores obtenidos en los modelos 1 y 2 ordenados por tamaño de muestra para los cuatro trimestres de 2011.**

La Figura 4.13 no presenta grandes diferencias entre los dos modelos en cuanto a resultados para tamaño muestral pequeño pero sí parece que el modelo con efectos de tiempo correlados tiene un comportamiento distinto al modelo con efectos de tiempo independientes tal como muestra la Figura 4.5. Las diferencias entre el estimador directo y el modelo con efectos correlados son cuantitativamente más grandes para los dominios que presentan pocas observaciones.

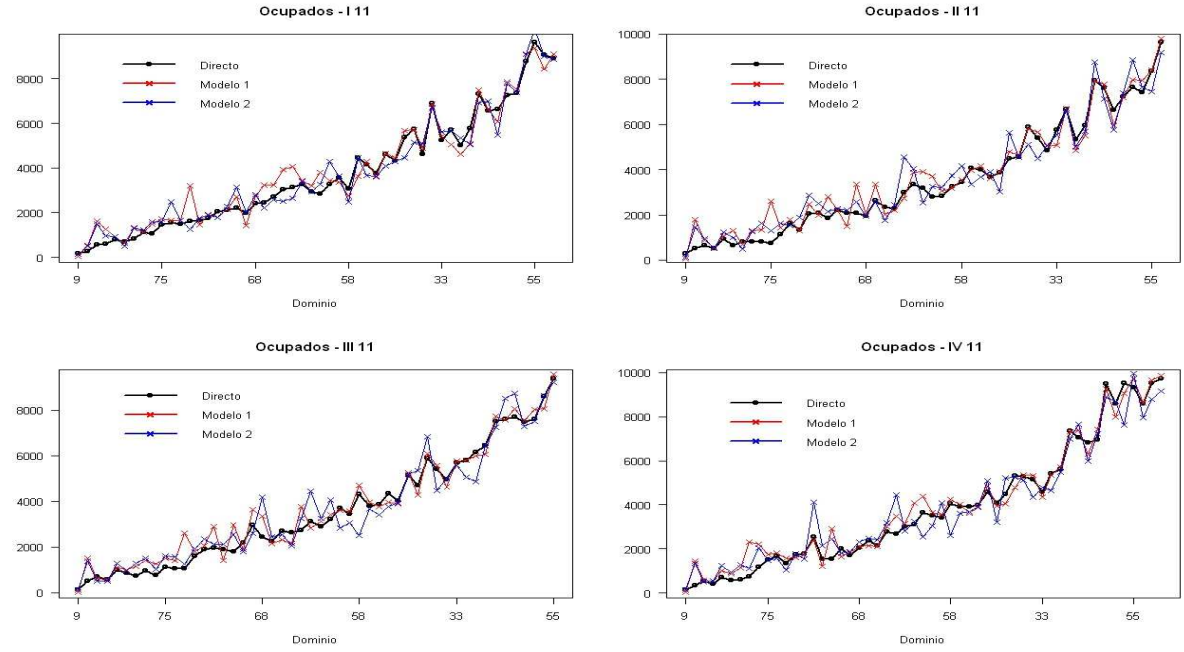


**Figura 4.13. Comparación del estimador directo y los obtenidos en los modelos para un tamaño del número de ocupados <10000 en los cuatro trimestres de 2011.**

Para ver si la evolución en cuanto a modelos es la correcta, y se está avanzando en una línea de mejora de los resultados, se pueden observar los coeficientes de variación obtenidos a través de (6), y ver si mejoran los que daba el modelo previo.

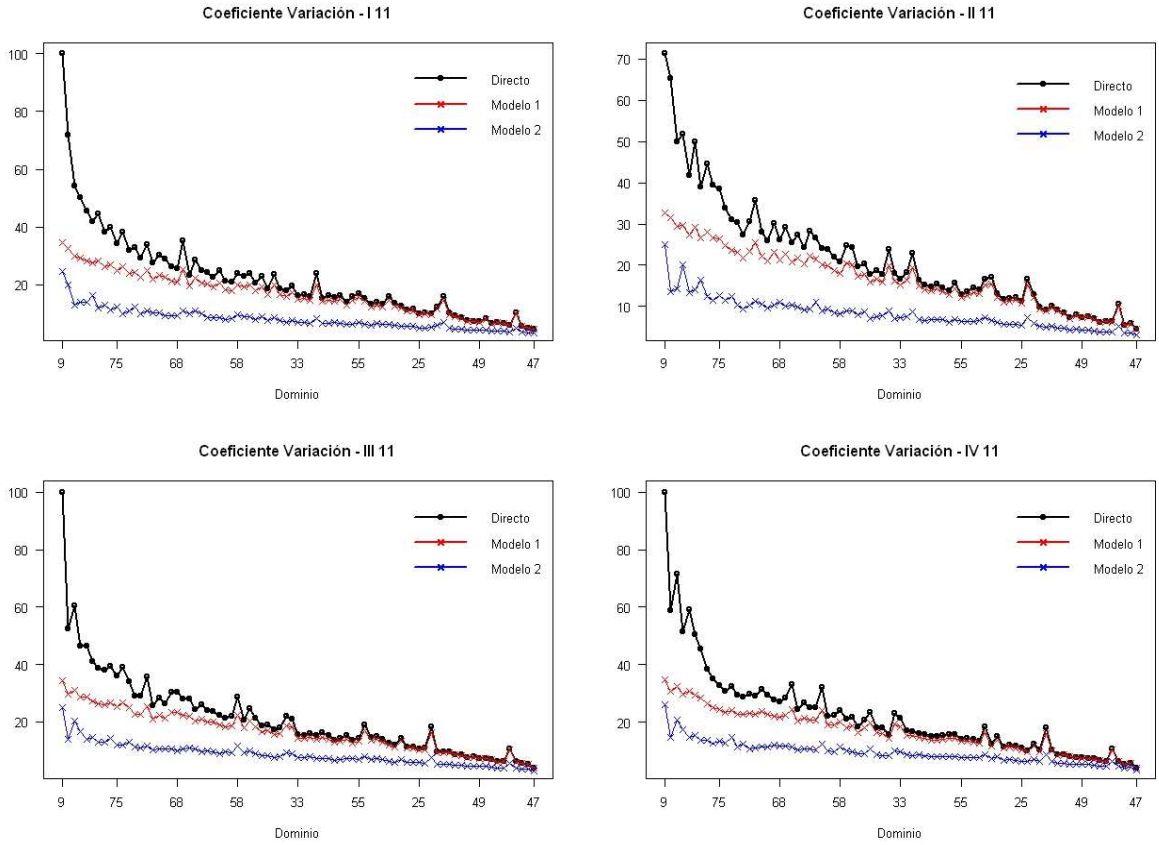


**Figura 4.14. Coeficiente de variación del estimador directo frente a los estimadores basados en los modelos 1 y 2 para los cuatro trimestres de 2011.**

La Figura 4.14 pone de manifiesto una mejoría notable entre los coeficientes de variación de los dos modelos estudiados. Existe una diferencia de cerca de diez puntos porcentuales en la práctica totalidad de los dominios, disminuyendo esta diferencia en aquéllos que poseen un número más grande de observaciones. Además, excepto el dominio perteneciente a las actividades de apoyo a la industria extractiva (actividad 09) que tiene un coeficiente de variación del 30%, el resto de actividades no supera el límite del 20% que se comentaba al principio del estudio, por lo que el resultado es muy bueno.

Por último se presenta la comparativa para una serie de actividades escogidas mediante el criterio previamente explicado en la Tabla 4.2. La Figura 4.15 presenta la evolución del estimador directo y los estimadores de los modelos a lo largo del periodo estudiado en las actividades seleccionadas. Muestra que la evolución que presenta el segundo modelo es mucho más suave que la que presentaba el primer modelo, incluso se muestra suave para actividades, o dominios, que tienen un tamaño muy grande de observaciones.

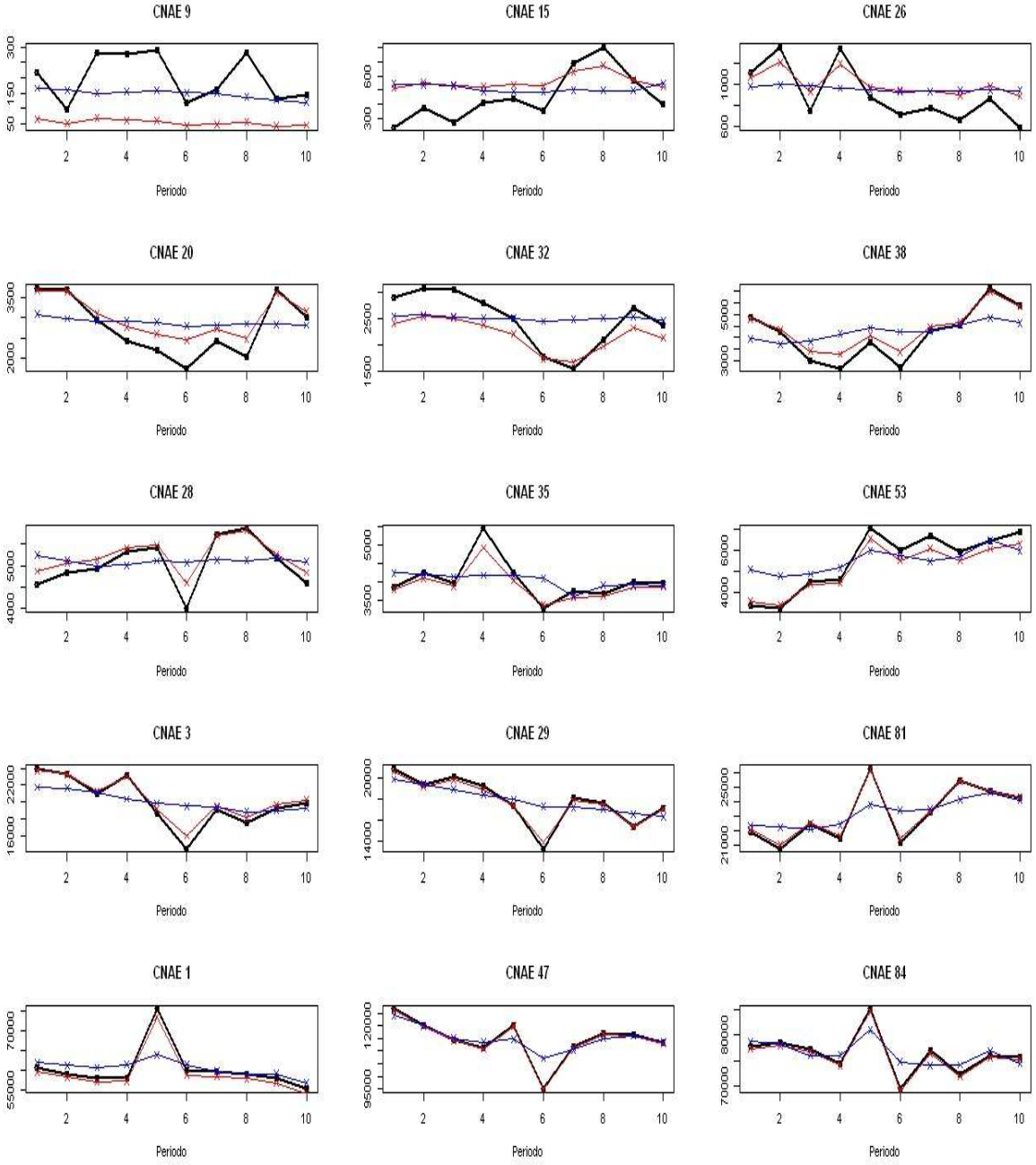


**Figura 4.15. Evolución del estimador directo (negro) y de los estimadores obtenidos con los modelos 1 (rojo) y 2 (azul) desde el tercer trimestre del 2009 hasta el cuarto trimestre de 2011 para algunas actividades económicas.**

La Figura 4.16 muestra la evolución del coeficiente de variación para todos los estimadores considerados a lo largo del periodo de estudio para las actividades económicas seleccionadas según el criterio de la Tabla 4.2. Se observa una mejora en los coeficientes de variación de todas las actividades, y siempre en niveles inferiores al 20% excepto en la actividad del margen superior izquierdo (actividades de apoyo a la industria extracticva). Se puede concluir que el modelo introduce mejoras significativas en la estimación.

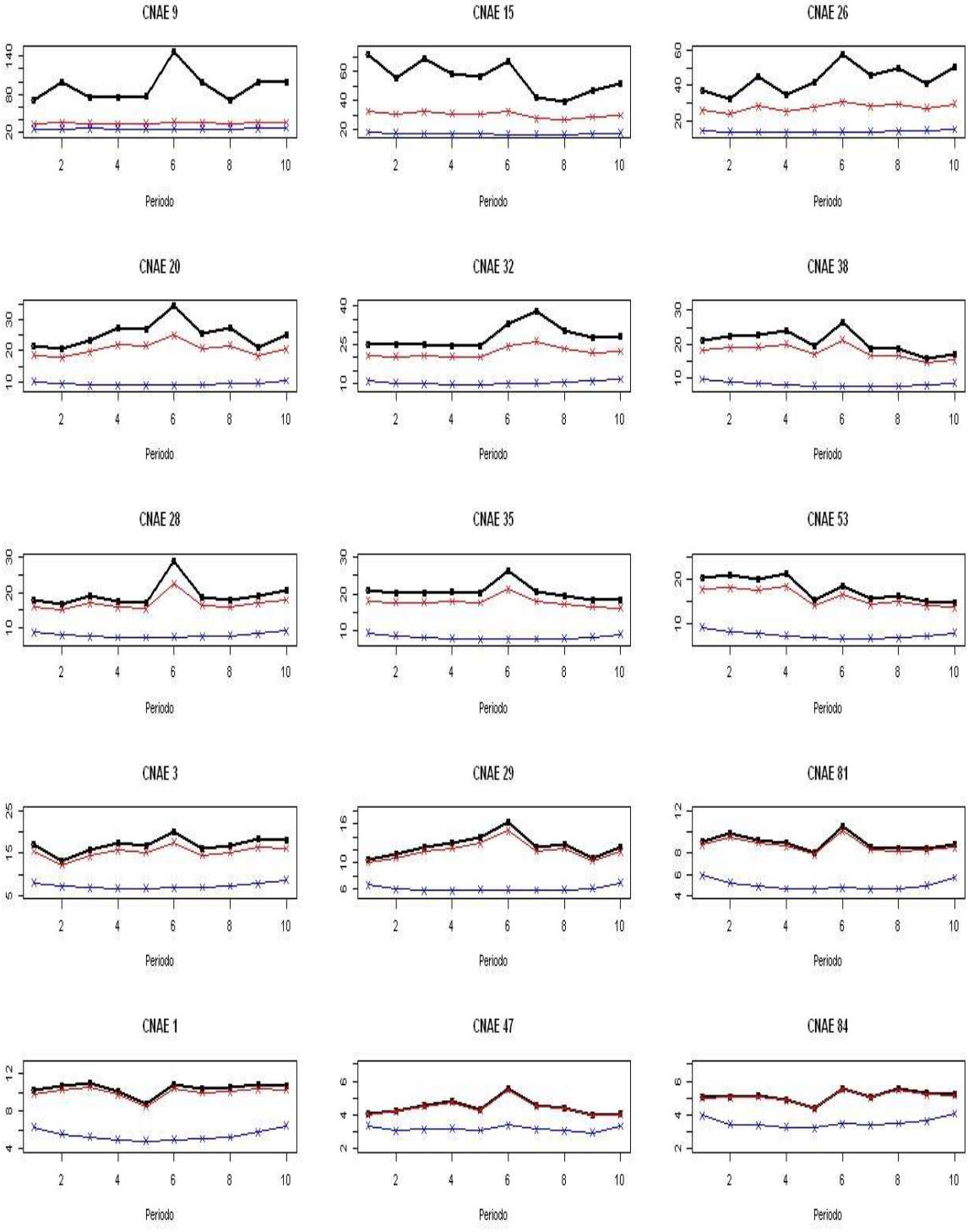


**Figura 4.16. Evolución del coeficiente de variación directo (negro) y de los obtenidos con los modelos 1 (rojo) y 2 (azul) desde el tercer trimestre del 2009 hasta el cuarto trimestre de 2011 para algunas actividades económicas.**

En este apartado solo se presentan resultados gráficos, los resultados numéricos correspondientes el 3º y 4º trimestre del 2011 se pueden consultar en los anexos II a V.

# Capítulo 5

# Conclusiones y Futuras Investigaciones

En este trabajo se proporcionan dos estimaciones basadas en modelos para los totales de ocupados de la EPA por actividad económica desde el tercer trimestre del 2009 hasta el cuarto trimestre del 2011. Estas dos estimaciones se proponen como alternativa al estimador directo que se obtiene directamente de la EPA. Se proporcionan los coeficientes de variación para los tres estimadores, utilizando para el estimador directo el paquete *"survey"* de R (T.Lumley 2004-2010), y para los basados en modelos la aproximación analítica de Prasad y Rao (1990). Los resultados indican que el estimador basado en el modelo con efectos de tiempo correlados es más eficiente para estimar los ocupados por actividad económica, constituyéndose como un buen competidor del estimador directo proporcionado por la EPA.

En efecto, las estimaciones obtenidas a partir del modelo 2 mejoran en algunos casos los coeficientes de variación en más de 100 puntos porcentuales, dejando finalmente coeficientes de variación inferiores al 30% , límite muy cercano al propuesto por la ONS para considerar un dato publicable y por tanto, oficial.

Si bien es cierto que la calidad en la estimación ha sufrido una considerable mejora, hay que tener en cuenta que los modelos utilizados en este trabajo suponen que hay una relación lineal entre las variables explicativas y la variable respuesta a nivel de actividad económica. En el análisis exploratorio se mostró que esta relación lineal se podía suponer para todas las actividades económicas y por sectores económicos (agrupaciones de actividades) pero puede no ser cierta para algunas de las actividades económicas consideradas.

La Figura 5.1 muestra algunas actividades económicas objeto de estudio que no se muestran como lineales a lo largo de los períodos examinados. La primera fila enfrenta los ocupados EPA $(Y)$ y las afiliaciones a la SS $(X_1)$ en dos actividades que muestran un comportamiento un poco alejado de la linealidad. De igual forma sucede en la segunda fila pero enfrentando a los ocupados EPA $Y$ y los contratos registrados $(X_2)$.

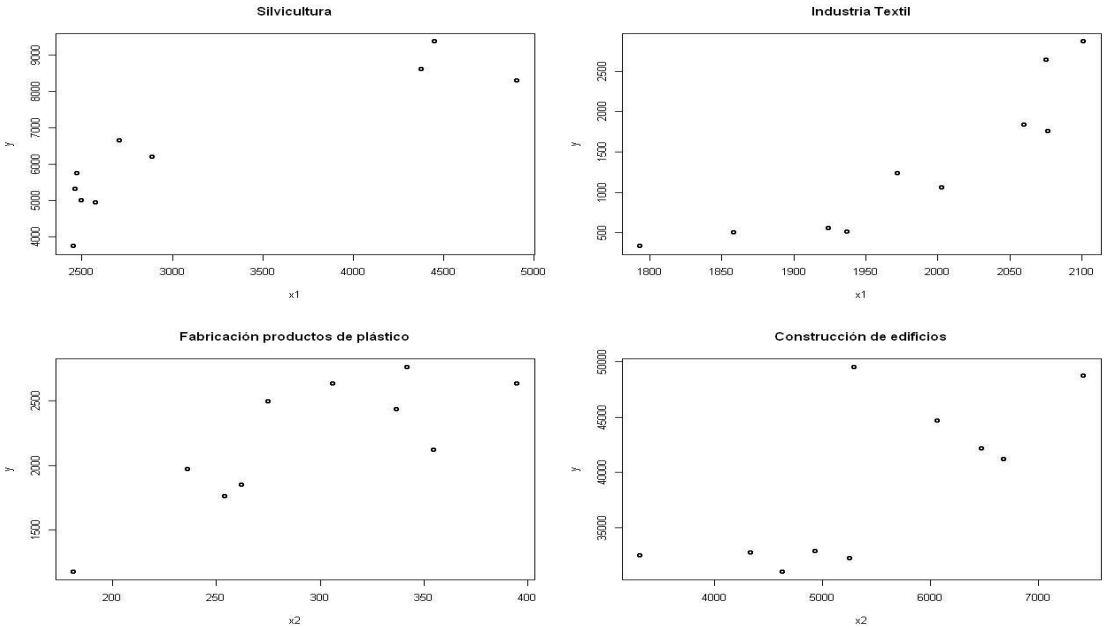


**Figura 5.1 Nube de puntos de los ocupados EPA en algunas actividades frente los asegurados a la SS $X_1$(fila superior) y los contratos registrados $X_2$(fila inferior) en dichas actividades.**

Es por ello que, para tener en cuenta la no linealidad en algunas actividades económicas una posible línea de estudio futura podría ser el uso del modelo propuesto por Rueda y Lombardía (2012). En el artículo de Rueda y Lombardía (2012) se propone el empleo de modelos mixtos semiparamétricos, monótonos y aditivos como alternativa a los modelos lineales y no paramétricos.

Los modelos monótonos semiparamétricos están definidos en un contexto de áreas pequeñas, asumiendo que algunas de las variables auxiliares tienen una relación monótona con la respuesta y con la incorporación de términos lineales para modelar otras variables auxiliares como las variables dummy.

La monotonía es una propiedad simple e intuitiva que establece que cuanto más grande (o más pequeña) es una información auxiliar, entonces más grande será la respuesta, siendo las otras variables iguales y preservando el orden de medias. En nuestro ejemplo, la monotonía es una propiedad que se cumple, en efecto, cuantos más afiliados a la SS hay, mayor es el número de ocupados recogidos en la EPA, y también cuantos más contratos se registran, mayor es el número de ocupados EPA.

El modelo monótono aditivo es de la forma:

$$y_d = \alpha + \sum_{j=1}^{p} \beta_j X_{jd} + \sum_{j=p+1}^{p} h_j(x_{jd}) + u_d + \varepsilon_d$$

$$f(x_{1d}, \dots, x_{pd}) = \alpha + \sum_{j=1}^{p} \beta_j x_{jd} + \sum_{j=p+1}^{p+q} h_j(x_{jd})$$

donde $h_j(\,)$ son funciones monótonas.

Este modelo presenta una serie de ventajas:

- Es un enfoque flexible que incluye varias formas o regresiones (no solo la relación lineal).
- El problema de optimización para derivar los estimadores es fácilmente resuelto.
- La restricción isotónica evita el problema de definir opciones específicas del usuario como la definición de la ventana, parámetros de suavización, número de knots, …, y otra serie de elementos no paramétricos.
- Los estimadores correspondientes para las funciones $h_j$ en el modelo aditivo tienen la propiedad oráculo: la tasa de convergencia $\hat{h}_j$ es independiente del número de compenentes aditivas en el modelo.

Otra posible mejora de este trabajo puede venir en la estimación de los MSE. En este trabajo solo se han empleado expresiones analíticas para su cálculo. Estas fórmulas son aproximaciones que asumen hipótesis fuertes y que son específicas del modelo, en el sentido de que se obtienen bajo el modelo particular considerado. Es deseable introducir métodos que permitan la comparación de estimadores de áreas pequeñas obtenidos a partir de diferentes modelos y que además respete las propiedades del diseño muestral. Los métodos de remuestreo pueden aplicarse de manera similar bajo cualquier modelo estadístico, hacen comparables los resultados obtenidos por los distintos estimadores y, generalmente se basan en condiciones más débiles que las de las aproximaciones analíticas. En numerosos artículos disponibles en la literatura se muestra que con técnicas de remuestreo, como los métodos Jackknife y Bootstrap introducidos en Herrador y otros (2008), se obtienen mejores resultados que con las aproximaciones analíticas a la expresión proporcionada por Prasad y Rao (1990).

# Bibliografía


[1] Brackstone, G.J. (1987). Small Area Data: Policy Issues and Technical Challenges. En R. Platek, J.N.K. Rao, C.E. Sarndal y M. Singh eds., *Small Area Statistics*, pp. 3-20. Wiley, New York.

[2] Battese, G.E., Harter, R.M., Fuller, W.A. (1988). An Error Component Model for Prediction of County Crop Areas Using Survey and Satelite Date. *Journal of the American Statistical Association*, **83**, 28-36.

[3] Belsley, D.A., Kuh, E., Welsch, R.E. (1980). Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. *Wiley Series in Probability and Statistics*.

[4] Chow, G., Lin, A.L. (1971). Best Linear Unbiased Distribution and Extrapolation of Economic Time Series by Related Series, *Review of Economic and Statistics*, ***53***, n. 4, 372-375.

[5] Esteban, M.D., Morales, D., Pérez, A., Santamaría, L. (2009). Small area estimation of poverty indicators under area-level time models.

[6] Esteban, M.D., Morales, D., Pérez, A., Molina, I., Santamaría, L., Marhuenda, Y., et.al. (2010). Project SAMPLE (Small Area Methods for Poverty and Living Conditions Estimates).

[7] Fay, R.E., Herriot, R.A. (1979). Estimation of Income from Small Places: An Application of James-Stein Procedures to Census Data. *Journal of the American Statistical Association*, **74**, 269-277.

[8] Ghosh, M., Nangia, N., Kim, D.H. (1996). Estimation of Median Income of Four-Person Families: a Bayesian Time Series Approach. *Journal of the American Statistical Association* **91**, 1423-1431.

[9] Henderson, C.R. (1948). Estimation of General, Specific and Maternal Combining Abilities in Crosses Among Inbred Lines of Swine. *Ph.D. Thesis*, Iowa State University, Ames, Iowa.

[10] Henderson, C.R. (1949). Estimation of Changes in Herd Environment (Abstract). *J. Dairy Sci.* 32:706.

[11] Henderson, C.R. (1963). Selection Index and Expected Genetic Advance. En *Statistical Genetics and Plant Breeding* (W.D. Hanson and H.F. Robinson, eds.), 141-163. National
Academy of Sciences and National Research Council Publication No. 982,Washington, D.C.

[12] Henderson, C.R. (1973). Sire Evaluation and Genetic Trends. *Proceedings of the Animal*
*Breeding and Genetics Symposium in Honour of Dr. Jay L. Lush* 10-41. Amer. Soc. Animal Sci.-Amer. Dairy Sci. Assn.-Poultry Sci. Assn. Champaigne, Illinois.

[13] Henderson, C.R. (1975). Best Linear Unbiased Estimation and Prediction Under Selection Model. *Biometrics* **31**, 423-447.

[14] Henderson, C.R., Kempthorne, O., Searle, S.R., von Krosigk, C.M. (1959). The
Estimation of Environmental and Genetic Trends from Records Subject to Culling. *Biometrics* **15**, 192-218.

[15] Harville, D.A. (1991). Discussion on Robinson paper, That BLUP is a good thing: the
Estimation of Random Effects. *Statistical Science*. **6**, 15-51.

[16] Harville, D.A., Jeske, D.R. (1992). Mean Squared Error of Estimation or Prediction Under a General Linear Model. *Journal of the American Statistical Association* **87**, 724-731.

[17] Herrador, M., Morales, D., Esteban, M.D., Sánchez, A., Santamaría, L.,Marhuenda, Y., Pérez, A. (2008). Sampling Design Variance Estimation of Small Area Estimators in the Spanish Labour Force Survey. *SORT*, **32(2)**, 177-198.

[18] Herrador, M., Morales, D., Esteban, M.D., Sánchez, A., Santamaría, L.,Marhuenda, Y., Pérez, A., Molina, I. (2009). Estimadores de Áreas Pequeñas Basados en Modelos para la Encuesta de Población Activa, *Estadística Española*, **51**, 133-172.

[19] INE., IGE. (2005). Metodología Encuesta de Población Activa (EPA). http://www.ine.es/daco/daco43/resumetepa.pdf

[20] Jiang, J., Lahiri, P. (2006). Mixed Model Prediction and Small Area Estimation, *Sociedad de Estadística e Investigación Operativa*, **15**, 1-96.

[21] Lumley, T. (2004). Analysis of Complex Survey Samples. *Journal of statistical software* 9(1): 1-19.

[22] Lumley, T. (2010). Survey: Analysis of Complex Survey Samples. *R package* version 3.26.

[23] Malec, D., Davis, W.W., Cao, X. (1999). Model-Based Small Area Estimates of Overweight Prevalence Using Sample Selection Adjustment. *Statistics in Medicine*, **18**, 3189-3200.

[24] ONS. (2004). Labour Force Survey User Guide. vol 6.

[25] Pfeffermann, D., Burck, L. (1990). Robust Small Area Estimation Combining Time Series and Cross-Sectional Data. *Survey Methodology* **16**, 217-237.

[26] Prasad, N.G.N., Rao, J.N.K. (1990). The Estimation of the Mean Squared Error of Small-Area Estimators. *Journal of the American Statistical Association*, **85**, 163-171.

[27] Quilis, E.M. (2001). Notas Sobre Desagregación Temporal de Series Económicas. Instituto de Estudios Fiscales. Papeles de Trabajo n.1/01.

[28] Rao, J.N.K., Yu, M. (1992). Small Area Estimation Combining Time Series and Crosssectional Data. *Proceeding Survey Research Methods Section. American Statistical association*, 1-9.

[29] Rao, J. N. K., Yu, M. (1994). Small-Area Estimation by Combining Time-Series and Cross-Sectional Data", *The Canadian Journal of Statistics* , **22**, 511-528.

[30] Rao, J.N.K.(1999). Some Recent Advances in Model-Based Small Area Estimation. *Survey Methodology*, **25**, 175-186.

[31] Rao, J.N.K. (2003). Small Area Estimation. *John Wiley*. New York.

[32] Robinson, G.K. (1991). That BLUP is a Good Thing: the Estimation of Random Effects.
*Statistical Science*. **6**, 15-51.

[33] Rueda, C., Lombardía, M.J. (2011). Small Area Semiparametric Additive Monotone Models. Working Paper.

[34] Saei, A., Chambers, R. (2003). Small Area Estimation: A Review of Methods Base on the Application of Mixed Models. *Southampton Statistical Sciences Research Institute*. University of Southamptom, U.K.

[35] Scott, A. J., Smith, T.M.F. (1974). Analysis of Repeated Surveys Using Time Series Methods. *Journal of the American Statistical Association*, **69**, 674-678.

[36] Singh, A.C., Mantel, H.J., Thomas, B.W. (1994). Time Series EBLUPs for Small Areas
Using Survey Data. *Survey Methodology*, **20**, 33-43.

[37] Tiller, R.B. (1991). Times Series Modelling of Sample Survey Data from the U.S. Current
Population Survey. *Journal of Official Statistics* **8**, 149-166.

[38] You, Y., Rao, J.N.K. (2000). Hierarchical Bayes Estimation of Small Area Means Using
Multi-Level Models. *Survey Methodology*, **26**, 173–181.

[58] You, Y., Rao, J.N.K., Gambino, J. (2001). Model-Based Unemployment Rate Estimation for the Canadian Labour Force Survey: a Hierarchical Approach. *Technical report, Household Survey Method Division.* Statistics Canada.

# Anexo I. Clasificación Nacional de Actividades Económicas (CNAE 09)

| Sección | Sector | Código | Descripción de la actividad |
|---|---|---|---|
| A | Primario | 1 | Agricultura, ganadería, caza y servicios relacionados con ellas |
| A | Primario | 2 | Silvicultura y explotación forestal |
| A | Primario | 3 | Pesca y acuicultura |
| B | Industria | 5 | Extracción de antracita, hulla y lignito |
| B | Industria | 6 | Extracción de crudo de petróleo y gas natural |
| B | Industria | 7 | Extracción de minerales metálicos |
| B | Industria | 8 | Otras industrias extractivas |
| B | Industria | 9 | Actividades de apoyo a las industrias extractivas |
| C | Industria | 10 | Industria de la alimentación |
| C | Industria | 11 | Fabricación de bebidas |
| C | Industria | 12 | Industria del tabaco |
| C | Industria | 13 | Industria textil |
| C | Industria | 14 | Confección de ropa de vestir |
| C | Industria | 15 | Industria del cuero y del calzado |
| C | Industria | 16 | Industria de la madera y del corcho, excepto muebles; cestería e espartería |
| C | Industria | 17 | Industria del papel |
| C | Industria | 18 | Artes gráficas y reproducción de soportes grabados |
| C | Industria | 19 | Coquerías y refinación del petróleo |
| C | Industria | 20 | Industria química |
| C | Industria | 21 | Fabricación de productos farmacéuticos |
| C | Industria | 22 | Fabricación de productos de caucho y plásticos |
| C | Industria | 23 | Fabricación de otros productos minerales no metálicos |
| C | Industria | 24 | Metalurgia; fabricación de productos de hierro, acero y aleaciones de hierro |
| C | Industria | 25 | Fabricación de productos metálicos, excepto maquinaria y equipamiento |
| C | Industria | 26 | Fabricación de productos informáticos, electrónicos y ópticos |
| C | Industria | 27 | Fabricación de material y equipamiento eléctrico |
| C | Industria | 28 | Fabricación de maquinaria y equipamiento n.c.n. |
| C | Industria | 29 | Fabricación de vehículos de motor, remolques y semirremolques |
| C | Industria | 30 | Fabricación de otro material de transporte |
| C | Industria | 31 | Fabricación de muebles |
| C | Industria | 32 | Otras industrias manufactureras |
| C | Industria | 33 | Reparación e instalación de maquinaria y equipamiento |
| D | Industria | 35 | Suministro de energía eléctrica, gas, vapor y aire acondicionado |
| E | Industria | 36 | Captación, depuración y distribución de agua |
| E | Industria | 37 | Recogida y tratamiento de aguas residuales |
| E | Industria | 38 | Recogida, tratamiento y eliminación de residuos; valorización |
| E | Industria | 39 | Actividades de descontaminación y otros servicios de gestión de residuos |
| F | Construcción | 41 | Construcción de edificios |
| F | Construcción | 42 | Ingeniería civil |
| F | Construcción | 43 | Actividades de construcción especializada |
| G | Servicios | 45 | Venta y reparación de vehículos de motor y motocicletas |
| G | Servicios | 46 | Comercio por junto e intermediarios del comercio, salvo de vehículos de motor y motocicletas |
| G | Servicios | 47 | Comercio al por menor, salvo de vehículos de motor y motocicletas |
| H | Servicios | 49 | Transporte terrestre y por tubo |
| H | Servicios | 50 | Transporte marítimo y por vías navegables interiores |
| H | Servicios | 51 | Transporte aéreo |

| | | | |
|---|---|---|---|
| **H** | **Servicios** | **52** | Almacenamiento y actividades anexas al transporte |
| **H** | **Servicios** | **53** | Actividades postales y de correos |
| **I** | **Servicios** | **55** | Servicios de alojamiento |
| **I** | **Servicios** | **56** | Servicios de comidas y bebidas |
| **J** | **Servicios** | **58** | Edición |
| **J** | **Servicios** | **59** | Actividades cinematográficas, de vídeo y de programas de televisión, grabación de sonido y edición musical |
| **J** | **Servicios** | **60** | Actividades de programación y emisión de radio y televisión |
| **J** | **Servicios** | **61** | Telecomunicaciones |
| **J** | **Servicios** | **62** | Programación, consultoría y otras actividades relacionadas con la informática |
| **J** | **Servicios** | **63** | Servicios de información |
| **K** | **Servicios** | **64** | Servicios financieros, excepto seguros y fondos de pensiones |
| **K** | **Servicios** | **65** | Seguros, reaseguros y fondos de pensiones, excepto seguridad social obligatoria |
| **K** | **Servicios** | **66** | Actividades auxiliares a los servicios financieros y a los seguros |
| **L** | **Servicios** | **68** | Actividades inmobiliarias |
| **M** | **Servicios** | **69** | Actividades jurídicas y de contabilidad |
| **M** | **Servicios** | **70** | Actividades de las sedes centrales; actividades de consultoría de gestión empresarial |
| **M** | **Servicios** | **71** | Servicios técnicos de arquitectura e ingeniería; ensayos y análisis técnicas |
| **M** | **Servicios** | **72** | Investigación y desenvolvimiento |
| **M** | **Servicios** | **73** | Publicidad y estudios de mercado |
| **M** | **Servicios** | **74** | Otras actividades profesionales, científicas y técnicas |
| **M** | **Servicios** | **75** | Actividades veterinarias |
| **N** | **Servicios** | **77** | Actividades de alquiler |
| **N** | **Servicios** | **78** | Actividades relacionadas con el empleo |
| **N** | **Servicios** | **79** | Actividades de agencias de viajes, operadores turísticos, servicios de reservas y actividades relacionadas con ellas |
| **N** | **Servicios** | **80** | Actividades de seguridad e investigación |
| **N** | **Servicios** | **81** | Servicios a edificios y actividades de jardinería |
| **N** | **Servicios** | **82** | Actividades administrativas de oficina y otras actividades auxiliares a las empresas |
| **O** | **Servicios** | **84** | Administración pública y defensa; seguridad social obligatoria |
| **P** | **Servicios** | **85** | Educación |
| **Q** | **Servicios** | **86** | Actividades sanitarias |
| **Q** | **Servicios** | **87** | Asistencia en establecimientos residenciales |
| **Q** | **Servicios** | **88** | Actividades de servicios sociales sin alojamiento |
| **R** | **Servicios** | **90** | Actividades de creación, artísticas y espectáculos |
| **R** | **Servicios** | **91** | Actividades de bibliotecas, archivos, museos y otras actividades culturales |
| **R** | **Servicios** | **92** | Actividades de juegos de azar y apuestas |
| **R** | **Servicios** | **93** | Actividades deportivas, recreativas y de entretenimiento |
| **S** | **Servicios** | **94** | Actividades asociativas |
| **S** | **Servicios** | **95** | Reparación de ordenadores, efectos personales y artículos de uso doméstico |
| **S** | **Servicios** | **96** | Otros servicios personales |
| **T** | **Servicios** | **97** | Actividades de los hogares como empleadores de personal doméstico |
| **T** | **Servicios** | **98** | Actividades de los hogares como productores de bienes y servicios para uso propio |
| **U** | **Servicios** | **99** | Actividades de organizaciones y organismos extraterritoriales |

# Anexo II. Estimación de los ocupados en el tercer trimestre de 2011.

| Dominio | Estimador Directo | Estimador Modelo 1 | Estimador Modelo 2 |
|---|---|---|---|
| **1** | 57.900,82 | 56.506,01 | 58.927,34 |
| **2** | 8.623,31 | 8.063,08 | 8.656,94 |
| **3** | 19.189,53 | 19.679,30 | 19.006,41 |
| **8** | 4.968,94 | 4.629,00 | 4.985,09 |
| **9** | 130,94 | 40,46 | 125,88 |
| **10** | 34.931,95 | 34.584,30 | 35.414,19 |
| **11** | 3.870,45 | 3.785,71 | 3.414,73 |
| **13** | 504,27 | 1.519,13 | 1.390,70 |
| **14** | 12.133,04 | 12.075,34 | 12.209,04 |
| **15** | 566,49 | 569,80 | 495,97 |
| **16** | 9.391,91 | 9.569,81 | 9.225,23 |
| **17** | 746,62 | 1.170,31 | 1.274,26 |
| **18** | 3.808,05 | 3.964,95 | 3.680,28 |
| **19** | 1.004,38 | 1.107,36 | 1.273,18 |
| **20** | 3.699,92 | 3.618,67 | 2.840,17 |
| **21** | 2.204,44 | 1.899,49 | 1.804,76 |
| **22** | 1.968,24 | 2.897,36 | 2.109,03 |
| **23** | 7.599,54 | 8.043,50 | 7.519,88 |
| **24** | 5.431,60 | 5.589,34 | 4.466,13 |
| **25** | 18.056,06 | 17.986,65 | 17.234,78 |
| **26** | 856,59 | 981,07 | 946,49 |
| **27** | 975,13 | 1.428,49 | 1.519,24 |
| **28** | 5.155,90 | 5.266,72 | 5.180,82 |
| **29** | 15.309,21 | 15.398,57 | 16.550,12 |
| **30** | 11.274,48 | 10.490,52 | 10.437,24 |
| **31** | 10.063,42 | 9.299,84 | 9.033,45 |
| **32** | 2.704,53 | 2.321,71 | 2.527,91 |
| **33** | 10.676,27 | 10.023,75 | 7.706,31 |
| **35** | 4.015,76 | 3.883,38 | 3.932,31 |
| **36** | 1.116,20 | 1.506,84 | 1.621,71 |
| **38** | 6.155,46 | 5.986,98 | 4.853,03 |
| **41** | 32.738,22 | 32.779,35 | 32.867,57 |
| **42** | 13.862,87 | 13.041,00 | 14.267,24 |
| **43** | 43.599,83 | 43.751,17 | 43.758,48 |
| **45** | 23.375,39 | 23.380,73 | 23.454,50 |
| **46** | 40.867,52 | 41.025,11 | 40.952,71 |
| **47** | 116.295,96 | 115.947,36 | 115.952,37 |

| | | | |
|---|---|---|---|
| **49** | 27.747,25 | 278.654,02 | 28.514,39 |
| **50** | 2.633,03 | 2.147,94 | 2.066,41 |
| **51** | 706,69 | 620,24 | 518,52 |
| **52** | 7.712,79 | 8.053,59 | 8.733,56 |
| **53** | 6.443,28 | 6.045,90 | 6.460,15 |
| **55** | 11.444,90 | 11.398,55 | 12.850,49 |
| **56** | 60.330,51 | 60.104,93 | 60.739,78 |
| **58** | 2.912,01 | 3.139,71 | 3.264,16 |
| **59** | 1.056,25 | 1.398,11 | 15.837,71 |
| **60** | 1.907,46 | 2.046,19 | 2.338,24 |
| **61** | 5.815,26 | 5.817,44 | 5.051,86 |
| **62** | 5.901,16 | 6.102,43 | 6.824,50 |
| **64** | 13.353,94 | 13.354,19 | 12.292,75 |
| **65** | 4.719,44 | 4.272,57 | 5.340,75 |
| **66** | 2.726,00 | 3.768,03 | 3.254,43 |
| **68** | 3.440,15 | 3.579,88 | 3.047,72 |
| **69** | 16.845,64 | 16.669,81 | 16.775,47 |
| **70** | 3.235,64 | 3.405,69 | 4.057,55 |
| **71** | 10.070,40 | 10.431,18 | 102.084,98 |
| **72** | 4.344,44 | 3.955,95 | 3.775,39 |
| **73** | 2.948,56 | 3.637,71 | 2.611,73 |
| **74** | 2.443,94 | 3.334,92 | 4.189,02 |
| **75** | 777,10 | 1.268,49 | 1.020,84 |
| **77** | 4.309,59 | 4.708,84 | 2.501,66 |
| **78** | 1.046,51 | 2.621,01 | 1.254,89 |
| **79** | 1.591,70 | 1.810,04 | 1.916,75 |
| **80** | 5.705,27 | 5.785,44 | 5.592,33 |
| **81** | 24.749,24 | 24.739,08 | 24.659,89 |
| **82** | 7.501,70 | 7.730,42 | 7.261,97 |
| **84** | 76.027,60 | 75.557,99 | 76.939,30 |
| **85** | 71.353,69 | 69.654,88 | 70.209,00 |
| **86** | 60.718,81 | 60.499,07 | 59.873,55 |
| **87** | 12.582,83 | 12.340,44 | 12.754,86 |
| **88** | 11.046,84 | 10.919,55 | 11.717,27 |
| **90** | 3.126,51 | 2.824,76 | 4.461,23 |
| **91** | 1.890,94 | 1.408,79 | 2.127,67 |
| **92** | 2.252,84 | 2.170,27 | 2.451,46 |
| **93** | 7.600,05 | 7.591,20 | 8.501,54 |
| **94** | 7.477,42 | 7.559,34 | 7.289,42 |
| **95** | 1.796,02 | 2.947,34 | 2.528,98 |
| **96** | 17.260,03 | 17.346,58 | 17.728,13 |
| **97** | 32.850,58 | 32.365,36 | 32.818,85 |

# Anexo III. Estimación de los ocupados en el cuarto trimestre de 2011.

| Dominio | Estimador Directo | Estimador Modelo 1 | Estimador Modelo 2 |
|---|---|---|---|
| **1** | 55032,02 | 53973,74 | 56586,34 |
| **2** | 5307,45 | 4787,67 | 5263,48 |
| **3** | 19833,86 | 20232,81 | 19252,60 |
| **8** | 4584,24 | 4343,19 | 4792,01 |
| **9** | 142,64 | 44,12 | 119,04 |
| **10** | 34650,39 | 34249,76 | 34280,70 |
| **11** | 4093,36 | 3952,39 | 3209,53 |
| **13** | 330,98 | 1440,21 | 1334,21 |
| **14** | 11822,76 | 11791,16 | 11908,93 |
| **15** | 400,47 | 521,16 | 545,32 |
| **16** | 9733,21 | 9857,17 | 9148,37 |
| **17** | 593,08 | 1152,28 | 1256,99 |
| **18** | 3898,51 | 4025,46 | 3600,11 |
| **19** | 686,54 | 1006,10 | 1247,13 |
| **20** | 3003,18 | 3146,55 | 2817,34 |
| **21** | 1731,42 | 1611,20 | 1782,74 |
| **22** | 1177,58 | 2199,26 | 2054,41 |
| **23** | 6963,45 | 7426,93 | 7206,26 |
| **24** | 5137,81 | 5332,13 | 4330,34 |
| **25** | 16092,26 | 16141,28 | 16580,77 |
| **26** | 584,94 | 885,36 | 938,37 |
| **27** | 1520,18 | 1726,80 | 1478,87 |
| **28** | 4587,09 | 4834,87 | 5090,20 |
| **29** | 17114,77 | 17013,53 | 16323,21 |
| **30** | 12001,07 | 10995,90 | 10099,64 |
| **31** | 8578,99 | 7975,07 | 8693,82 |
| **32** | 2362,81 | 2127,94 | 2460,77 |
| **33** | 9542,29 | 9049,61 | 7625,63 |
| **35** | 3983,79 | 3861,80 | 3909,08 |
| **36** | 1669,00 | 1805,50 | 1573,40 |
| **38** | 5414,01 | 5332,72 | 4632,48 |
| **41** | 32469,08 | 32434,63 | 31137,05 |
| **42** | 13275,09 | 12429,94 | 13215,24 |
| **43** | 40648,62 | 40850,88 | 41329,49 |
| **45** | 23848,68 | 23820,07 | 23266,52 |
| **46** | 39451,80 | 39649,63 | 40210,06 |
| **47** | 113340,54 | 113019,09 | 113634,65 |
| **49** | 26126,93 | 26346,39 | 27861,32 |

| | | | |
|---|---|---|---|
| **50** | 2004,47 | 1637,24 | 1809,40 |
| **51** | 523,16 | 560,62 | 493,40 |
| **52** | 9541,47 | 9675,37 | 8788,19 |
| **53** | 6832,25 | 6287,01 | 5970,98 |
| **55** | 10581,22 | 10386,77 | 10338,19 |
| **56** | 59332,16 | 59082,57 | 58007,55 |
| **58** | 2771,89 | 3015,40 | 3183,19 |
| **59** | 1769,75 | 1743,76 | 1528,87 |
| **60** | 2038,16 | 2114,86 | 2322,30 |
| **61** | 5268,26 | 5359,93 | 5100,41 |
| **62** | 7363,05 | 7334,20 | 6971,92 |
| **64** | 13138,20 | 13277,57 | 13413,28 |
| **65** | 4468,04 | 4056,47 | 5209,55 |
| **66** | 3122,11 | 4084,97 | 3241,91 |
| **68** | 3511,02 | 3635,17 | 3031,71 |
| **69** | 16585,09 | 16422,29 | 16796,70 |
| **70** | 3412,05 | 3537,67 | 4062,93 |
| **71** | 9322,52 | 9711,72 | 9963,85 |
| **72** | 3902,32 | 3616,84 | 3679,11 |
| **73** | 4058,98 | 4258,58 | 2612,23 |
| **74** | 2658,69 | 3464,14 | 4433,51 |
| **75** | 1326,90 | 1568,91 | 1037,87 |
| **77** | 3649,94 | 4362,99 | 2527,44 |
| **78** | 750,99 | 2321,99 | 1108,79 |
| **79** | 1690,60 | 1835,54 | 1870,61 |
| **80** | 5620,25 | 5714,82 | 5480,98 |
| **81** | 24285,84 | 24302,81 | 24138,29 |
| **82** | 7041,71 | 7399,26 | 7654,63 |
| **84** | 75677,30 | 75230,63 | 74406,91 |
| **85** | 77440,88 | 75727,42 | 77026,14 |
| **86** | 56453,69 | 56301,83 | 57058,69 |
| **87** | 13806,90 | 13469,21 | 12723,50 |
| **88** | 12076,56 | 11852,17 | 11702,96 |
| **90** | 2534,93 | 2399,24 | 4123,63 |
| **91** | 1533,64 | 1188,07 | 2129,18 |
| **92** | 2134,79 | 2090,23 | 2404,35 |
| **93** | 9500,53 | 9265,98 | 8901,65 |
| **94** | 8585,45 | 8609,03 | 7943,91 |
| **95** | 1551,18 | 2892,21 | 2455,12 |
| **96** | 18785,13 | 18756,35 | 17641,94 |
| **97** | 33650,07 | 33138,90 | 3279,34 |

# Anexo IV. Estimación del coeficiente de variación(%) en el tercer trimestre de 2011.

| Dominio | CV Directo | CV Modelo1 | CV Modelo 2 |
|---|---|---|---|
| **1** | 10,77 | 10,32 | 5,77 |
| **2** | 13,60 | 12,73 | 7,31 |
| **3** | 18,29 | 16,32 | 7,78 |
| **8** | 21,15 | 18,26 | 9,04 |
| **9** | 100,00 | 34,51 | 25,10 |
| **10** | 7,31 | 7,17 | 4,45 |
| **11** | 24,56 | 20,32 | 10,06 |
| **13** | 52,45 | 29,74 | 14,02 |
| **14** | 14,33 | 13,32 | 7,05 |
| **15** | 46,51 | 28,61 | 16,66 |
| **16** | 14,28 | 13,28 | 68,23 |
| **17** | 38,86 | 26,48 | 12,91 |
| **18** | 20,73 | 17,98 | 9,35 |
| **19** | 46,50 | 28,55 | 13,96 |
| **20** | 21,21 | 18,30 | 9,78 |
| **21** | 26,22 | 21,24 | 10,74 |
| **22** | 35,82 | 25,43 | 11,79 |
| **23** | 15,22 | 14,03 | 71,58 |
| **24** | 21,90 | 18,73 | 94,13 |
| **25** | 9,62 | 9,30 | 5,40 |
| **26** | 41,15 | 27,19 | 14,63 |
| **27** | 37,96 | 26,18 | 12,85 |
| **28** | 19,16 | 16,93 | 8,32 |
| **29** | 10,79 | 10,34 | 6,07 |
| **30** | 19,09 | 16,88 | 7,88 |
| **31** | 13,85 | 12,94 | 6,98 |
| **32** | 27,89 | 22,09 | 10,84 |
| **33** | 15,14 | 13,96 | 7,20 |
| **35** | 18,51 | 16,49 | 8,35 |
| **36** | 35,93 | 25,49 | 11,96 |
| **38** | 15,98 | 14,62 | 7,93 |
| **41** | 7,43 | 7,28 | 4,45 |
| **42** | 11,42 | 10,89 | 5,87 |
| **43** | 6,20 | 6,11 | 3,98 |
| **45** | 8,71 | 8,47 | 4,97 |
| **46** | 7,02 | 6,90 | 4,30 |
| **47** | 4,03 | 4,00 | 2,97 |
| **49** | 7,75 | 7,58 | 4,57 |

| | | | |
|---|---|---|---|
| **50** | 24,43 | 20,26 | 10,71 |
| **51** | 60,51 | 31,11 | 20,25 |
| **52** | 13,79 | 12,89 | 6,67 |
| **53** | 15,22 | 14,04 | 7,29 |
| **55** | 12,76 | 12,03 | 6,45 |
| **56** | 6,35 | 6,26 | 4,02 |
| **58** | 23,75 | 19,85 | 9,54 |
| **59** | 38,98 | 26,56 | 11,87 |
| **60** | 28,93 | 22,59 | 11,01 |
| **61** | 15,43 | 14,19 | 7,48 |
| **62** | 18,00 | 16,11 | 7,97 |
| **64** | 11,35 | 10,83 | 6,05 |
| **65** | 17,23 | 15,56 | 7,74 |
| **66** | 25,92 | 21,07 | 9,60 |
| **68** | 21,96 | 18,77 | 9,21 |
| **69** | 9,80 | 9,45 | 5,42 |
| **70** | 22,46 | 19,07 | 8,83 |
| **71** | 14,67 | 13,60 | 6,85 |
| **72** | 21,21 | 18,30 | 8,98 |
| **73** | 30,53 | 23,33 | 10,62 |
| **74** | 30,43 | 23,27 | 10,03 |
| **75** | 39,35 | 26,63 | 14,32 |
| **77** | 28,67 | 22,46 | 11,51 |
| **78** | 34,10 | 24,89 | 13,06 |
| **79** | 29,10 | 22,68 | 11,36 |
| **80** | 15,68 | 14,39 | 7,52 |
| **81** | 8,47 | 8,25 | 4,92 |
| **82** | 16,22 | 14,80 | 7,36 |
| **84** | 5,27 | 5,21 | 3,63 |
| **85** | 5,77 | 5,69 | 3,76 |
| **86** | 6,16 | 6,08 | 3,94 |
| **87** | 10,90 | 10,44 | 5,76 |
| **88** | 11,89 | 11,29 | 6,00 |
| **90** | 23,94 | 20,02 | 9,90 |
| **91** | 25,83 | 21,04 | 10,18 |
| **92** | 28,12 | 22,20 | 10,49 |
| **93** | 15,16 | 13,99 | 7,18 |
| **94** | 14,34 | 13,33 | 6,79 |
| **95** | 28,22 | 22,25 | 10,74 |
| **96** | 9,72 | 9,39 | 5,41 |
| **97** | 7,81 | 7,64 | 4,62 |

# Anexo V. Estimación del coeficiente de variación(%) en el cuarto trimestre de 2011.

| Dominio | CV Directo | CV Modelo 1 | CV Modelo 2 |
|---|---|---|---|
| **1** | 10,76 | 10,31 | 6,45 |
| **2** | 17,86 | 16,01 | 8,27 |
| **3** | 17,90 | 16,04 | 8,60 |
| **8** | 21,31 | 18,36 | 9,78 |
| **9** | 100,00 | 34,73 | 25,88 |
| **10** | 7,47 | 7,32 | 5,14 |
| **11** | 23,29 | 19,58 | 10,60 |
| **13** | 58,72 | 30,76 | 14,66 |
| **14** | 14,89 | 13,77 | 7,85 |
| **15** | 51,55 | 29,64 | 17,22 |
| **16** | 13,48 | 12,63 | 7,55 |
| **17** | 45,55 | 28,34 | 13,60 |
| **18** | 20,95 | 18,13 | 10,02 |
| **19** | 59,19 | 30,85 | 14,60 |
| **20** | 25,02 | 20,58 | 10,53 |
| **21** | 29,40 | 22,83 | 11,41 |
| **22** | 35,03 | 25,15 | 12,49 |
| **23** | 14,89 | 13,77 | 7,94 |
| **24** | 22,90 | 19,35 | 10,14 |
| **25** | 10,86 | 10,40 | 6,31 |
| **26** | 50,41 | 29,42 | 15,21 |
| **27** | 32,75 | 24,28 | 13,42 |
| **28** | 20,60 | 17,90 | 9,13 |
| **29** | 12,36 | 11,69 | 6,93 |
| **30** | 18,50 | 16,47 | 8,71 |
| **31** | 15,61 | 14,34 | 7,87 |
| **32** | 28,49 | 22,39 | 11,54 |
| **33** | 15,71 | 14,41 | 8,03 |
| **35** | 18,24 | 16,30 | 9,07 |
| **36** | 30,61 | 23,37 | 12,60 |
| **38** | 16,92 | 15,33 | 8,64 |
| **41** | 7,62 | 7,46 | 5,19 |
| **42** | 11,55 | 11,00 | 6,69 |
| **43** | 6,60 | 6,49 | 4,70 |
| **45** | 8,72 | 8,47 | 5,70 |
| **46** | 7,21 | 7,07 | 5,00 |
| **47** | 4,05 | 4,02 | 3,35 |
| **49** | 8,02 | 7,83 | 5,34 |

| | | | |
|---|---|---|---|
| **50** | 29,26 | 22,76 | 11,42 |
| **51** | 71,58 | 32,36 | 20,66 |
| **52** | 13,93 | 13,00 | 7,49 |
| **53** | 14,92 | 13,79 | 8,01 |
| **55** | 12,48 | 11,80 | 7,17 |
| **56** | 6,19 | 6,10 | 4,57 |
| **58** | 24,52 | 20,29 | 10,29 |
| **59** | 28,84 | 22,58 | 12,44 |
| **60** | 27,06 | 21,67 | 11,64 |
| **61** | 15,78 | 14,46 | 8,23 |
| **62** | 16,07 | 14,68 | 8,62 |
| **64** | 11,83 | 11,24 | 6,85 |
| **65** | 18,10 | 16,19 | 8,57 |
| **66** | 25,13 | 20,64 | 10,37 |
| **68** | 22,18 | 18,91 | 9,95 |
| **69** | 10,18 | 9,80 | 6,21 |
| **70** | 22,46 | 19,08 | 9,62 |
| **71** | 14,15 | 13,17 | 7,64 |
| **72** | 21,81 | 18,68 | 9,74 |
| **73** | 24,02 | 20,01 | 11,21 |
| **74** | 26,54 | 21,39 | 10,79 |
| **75** | 32,24 | 24,07 | 14,73 |
| **77** | 32,02 | 23,96 | 12,16 |
| **78** | 38,24 | 26,37 | 13,67 |
| **79** | 27,64 | 21,98 | 11,97 |
| **80** | 16,71 | 15,17 | 8,33 |
| **81** | 8,76 | 8,51 | 5,66 |
| **82** | 15,73 | 14,43 | 8,14 |
| **84** | 5,20 | 5,15 | 4,08 |
| **85** | 5,62 | 5,56 | 4,27 |
| **86** | 6,26 | 6,17 | 4,57 |
| **87** | 10,10 | 9,73 | 6,40 |
| **88** | 11,45 | 10,91 | 6,76 |
| **90** | 29,70 | 23,00 | 10,66 |
| **91** | 28,99 | 22,64 | 10,96 |
| **92** | 33,14 | 24,43 | 11,31 |
| **93** | 15,25 | 14,05 | 7,94 |
| **94** | 14,31 | 13,31 | 7,61 |
| **95** | 31,37 | 23,70 | 11,48 |
| **96** | 10,37 | 9,97 | 6,24 |
| **97** | 7,73 | 7,56 | 5,27 |