

DESCOMPOSICION DE DIFERENCIAS SALARIALES REGIONALES EN PRESENCIA DE SELECCION MUESTRAL

Enrique López-Bazo[♂]

Elisabet Motellón[♀]

(versión preliminar e incompleta)

[♂] AQR—IREA

Universitat de Barcelona

Avda Diagonal 690, 08034 Barcelona

Tel: +34 93 4037041 FAX: +34 93 4021821 Email: elopez@ub.edu

[♀] Universitat Oberta de Catalunya

Av. Tibidabo 39-43

Tel: + 34 932537501 FAX: +34 934176495 Email: emotellon@uoc.edu

&

AQR—IREA

Universitat de Barcelona

Tel: + 34 93 4021010 FAX: +34 93 4021821 Email: emotellon@ub.edu

Abstract: En este trabajo proponemos un método para descomponer las diferencias salariales regionales en presencia de selección muestral en las ecuaciones salariales. La no linealidad del término de corrección de la selección en las ecuaciones salariales impide la aplicación de los métodos de descomposición estándar basados en el método de Oaxaca-Blinder. Las descomposiciones sugeridas hasta el momento en la literatura para ese caso no incorporan el efecto de las variables que determinan el proceso de selección y, en consecuencia, solo proporcionan evidencia parcial acerca de la contribución de las diferencias en dotaciones y de la heterogeneidad en los rendimientos. El método propuesto es aplicado al análisis de las disparidades regionales en los niveles salariales medios en España, utilizando los microdatos procedentes de la ola ampliada del PHOGUE de 2000. Los resultados confirman que existen diferencias en la dotación de características que influyen directa e indirectamente (a través de la influencia en la participación en el mercado de trabajo) en las diferencias regionales de salarios, pero además sugieren que la contribución de la heterogeneidad regional en los rendimientos tiene una contribución incluso más importante. En particular, este fenómeno es especialmente significativo en el caso del capital humano, tanto educativo como general.

JEL: C24, J31, R23

1. INTRODUCCION

Al igual que sucede en el caso del producto per cápita y de la productividad del trabajo, se ha constatado que las diferencias salariales entre regiones de una economía son importantes y muestran una elevada persistencia. A modo de ejemplo, Webber (2002) ha mostrado que en el caso de las regiones de la Unión Europea las diferencias de salarios muestran un patrón muy estable, siendo sus principales características la estratificación y la persistencia. En el caso de la economía española tales diferencias han sido puestas de manifiesto en diversos trabajos (véase por ejemplo Serrano, 2002 y Villaverde y Maza, 2002). Las consecuencias de la heterogeneidad regional en los niveles salariales son diversas afectando, entre otras, a las diferencias en magnitudes clave del mercado de trabajo como el nivel de participación o de desempleo y, al menos teóricamente, a los movimientos migratorios interregionales. Asimismo, pueden ser consideradas, al menos parcialmente, como el reflejo de las diferencias en los niveles de productividad regionales. Por otra parte, recientemente se ha mostrado la conexión sugerida por diversos modelos teóricos de la geografía económica entre salarios y el potencial de mercado regional real (Head y Mayer, 2005), y entre aquellos y la localización de la actividad económica (Bernard et al, 2004).

Este artículo pretende contribuir al conocimiento de la magnitud y causas de las diferencias salariales entre las regiones españolas, complementando la evidencia existente hasta el momento que se ha basado, en gran medida, en un enfoque agregado. Para ello, utilizamos la información contenida en la muestra ampliada del año 2000 del Panel de Hogares de la UE (PHOGUE), la cual nos permite obtener muestras representativas para las comunidades autónomas, conteniendo información tanto de individuos empleados, percibiendo un salario, como de individuos no ocupados. Esta circunstancia es importante dado que es posible que los factores inobservables que influyen en el salario afecten también a la probabilidad de que el individuo esté ocupado, de forma que los resultados basados en la estimación de la ecuación salarial adolecerían del conocido sesgo de selección muestral. Por ello, en este trabajo proponemos una descomposición de la diferencia salarial regional en la contribución de las diferencias regionales en características y la de la heterogeneidad regional en el rendimiento o precio de las mismas, que es robusta a la existencia de selección muestral. En este sentido nuestro trabajo difiere del de García y Molina (2002), dado que estos autores estudian la contribución de las características y sus rendimientos a las diferencias en los salarios medios entre las regiones NUTS I españolas, pero sin considerar la posible presencia de selección muestral. También presenta diferencias respecto al análisis realizado por

Duranton y Monastiriotis (2002) para el caso de las regiones del Reino Unido, dado que estos autores centran su análisis únicamente en el efecto de las características observables de los trabajadores, fundamentalmente de su capital humano, no considerando el papel que pueden estar jugando las diferencias regionales en las características de los puestos de trabajo y de las empresas. En cualquier caso, y a partir de la evidencia reportada por ese último trabajo, prestaremos especial atención al papel de las diferencias regionales en el capital humano, tanto en la dotación como en su rendimiento, a la hora de explicar las disparidades salariales regionales.

Los resultados obtenidos confirman que las diferencias regionales en la dotación de características observadas que determinan los niveles salariales contribuyen a explicar una parte del gap salarial regional. Pero para muchas de las regiones, el distinto rendimiento a esas características tiene una contribución incluso más importante. Además, tanto en el caso de la dotación como en el del rendimiento, se considera la influencia que se produce a través del término de corrección de la selección muestral. Adicionalmente, el análisis de la descomposición detallada permite apreciar como en el caso del capital humano, la diferencia en el rendimiento entre las distintas regiones tiene un papel destacado para explicar las diferencias salariales entre ellas, con una intensidad incluso mayor a la que corresponde a la distinta dotación de ese tipo de capital.

El resto del trabajo se estructura como sigue. En el siguiente apartado se presenta brevemente la base de datos utilizada y se realiza un breve análisis descriptivo tanto de los niveles salariales regionales como de las diferencias regionales en las características observadas de individuos y de puestos de trabajo y empresas. En el tercer apartado se describe la metodología empleada para descomponer el efecto de características y rendimientos de éstas en el caso de que se controle por selección muestral en la estimación de las ecuaciones salariales. Los resultados obtenidos, tanto por lo que respecta a todo el conjunto de características como al papel específico del capital humano se presentan en el apartado quinto. Finalmente, en el último apartado se formulan las principales conclusiones.

2. BASE DE DATOS Y DESCRIPTIVO

Si bien para el caso español la Encuesta de Estructura Salarial representa la base de datos más completa sobre las remuneraciones de los trabajadores, el hecho de contener únicamente información sobre asalariados nos imposibilita la corrección de las ecuaciones salariales en

presencia de un proceso de selección. Por este motivo, se ha recurrido a los datos facilitados para España en el Panel de Hogares de la Unión Europea¹, concretamente en su muestra ampliada de 2000 especialmente diseñada para estudios transversales y que tiene como característica interesante para nuestro trabajo el ser representativa a nivel de comunidad autónoma. Esta base de datos nos facilita información detallada de las características personales de los individuos, incluyendo las particularidades de sus hogares, así como de las condiciones laborales para aquellos que están ocupados. En particular, nuestro análisis de las disparidades regionales en salarios se efectúa sobre una muestra de asalariados entre 16 y 65 años que trabajan más de 15 horas semanales² de todo el estado español, con excepción de Ceuta y Melilla.

La información contenida en la Tabla 1 nos permite realizar una primera valoración de las diferencias salariales entre regiones. En ella se muestra el salario bruto hora medio, la desviación estándar y el número de trabajadores contenido en la muestra para cada una de las diecisiete CCAA y para el conjunto del Estado. Se aprecia claramente la existencia de importantes diferencias regionales. De este modo, el salario medio de Extremadura, la región que tenía el nivel salarial más bajo para 2000, representa el 69.75% del salario medio de Madrid, la región con el nivel salarial más elevado. Estas diferencias no se aprecian solamente entre los casos de las regiones con posiciones extremas. Por ejemplo, el ratio entre las cinco regiones con salarios más altos y las cinco regiones con salarios más bajos es de 1.29, lo que nos lleva a concluir que la magnitud de las disparidades salariales a nivel regional es comparable a la existente en otras magnitudes clave como el PIB per cápita o la productividad del trabajo (véase por ejemplo Goerlich et al, 2002).

Para aislar el posible efecto de los diferenciales en precios entre unas regiones y otras, además de considerar el salario/hora nominal en cada región, se ha utilizado una estimación del nivel relativo de precios regionales³ para obtener lo que podemos definir como salario/hora real, es

¹ Son múltiples los estudios que han empleado esta base de datos para sus análisis salariales en nuestro mercado de trabajo. Véase a modo de ejemplo, Montuenga et al (2003), García-Pérez y Jimeno (2007), De la Rica et al (2007)

² La eliminación de los datos para los asalariados con una jornada laboral inferior a 15 horas está motivada a que el PHOGUE no facilita información para este colectivo de algunas de las variables relativas al mercado de trabajo más relevantes, como el sector de actividad o la antigüedad en la empresa.

³ El origen de esas cifras es una encuesta regional de precios realizada por el INE para 1989 que, aunque no ha sido oficializada por dicho instituto, es utilizada por otras entidades como EUROSTAT o el Institut d'Estadística de Catalunya. Los niveles regionales de precios que utilizamos en este trabajo para 2000 son el resultado de aplicar la inflación anual publicada por el INE desde 1989 hasta ese año para cada una de las CCAA. Nótese que de esta forma no se estarán recogiendo las modificaciones en la cesta representativa utilizada para el cálculo de

decir el nivel salarial descontando el diferencial de precios existentes en 2000 en cada una de las regiones consideradas. Se puede apreciar, en primer lugar, el cambio en el ranking en las comunidades, siendo el caso más significativo el de Extremadura que pasa a ocupar el octavo puesto con un salario medio real de más de diez puntos por encima de la comunidad con menor retribución media, Murcia. De este modo, si tenemos en cuenta el diferencial de precios, son Murcia y Madrid las comunidades con salarios extremos, más bajo y más alto respectivamente. En segundo lugar, se observa en general una modesta convergencia en los salarios regionales. Así, el salario medio real de las cinco regiones de bajo salario aumenta en torno al 2% debido a su menor nivel de precios relativos, mientras que el de las cinco regiones con niveles salariales elevados disminuye como consecuencia del mayor nivel de precios en esas regiones, en idéntico porcentaje. Pero a pesar de esta mínima convergencia en los niveles retributivos los diferenciales salariales en términos reales son todavía muy elevados: el salario de Murcia no representa ni el 75% del salario medio real percibido por los asalariados madrileños, por lo que no parece razonable imputar la mayor parte del diferencial salarial entre regiones a un efecto compensatorio de diferenciales de precios. Otra circunstancia que resulta interesante es el hecho de que existen también diferencias en el grado de dispersión de la distribución salarial entre las regiones con niveles salariales bajos y elevados, como se desprende de los valores que toma la desviación estándar en ambos tipos de regiones. Si bien es cierto que dicha medida está influida por el nivel medio de la distribución, en general la variabilidad salarial en el caso de las regiones de salarios elevados resulta notablemente mayor que aquella observada para el caso de las regiones de bajos.⁴

Una inmediata explicación al hecho de que los salarios difieran entre las regiones la podemos encontrar en lo que se puede calificar como efecto composición, es decir, que los salarios son distintos porque la composición de los trabajadores y de los puestos de trabajo que ocupan y en las empresas que lo hacen difieren entre unas regiones y otras. Si esa fuese la explicación, no encontraríamos diferencias en el salario entre trabajadores con iguales características ocupando puestos de trabajos similares en empresas con iguales características pero que estuviesen situadas en distintos territorios. O dicho de otra forma, las diferencias regionales en el salario medio serían ocasionadas por un fenómeno estadístico. Para valorar esta posibilidad, en la Tabla 2 se reproducen los valores medios tomados por las distintas características

los niveles relativos de precios en el año inicial, por lo que los precios relativos utilizados deben ser considerados únicamente como aproximados.

⁴ Para una mayor discusión acerca de esta cuestión véase El-Attar et al (2007).

observadas en nuestra muestra en el conjunto del Estado y en las dos regiones con los niveles salariales medios extremos.⁵ En cada uno de los casos se distingue entre la muestra de ocupados y la de los individuos no ocupados. Centrándonos en lo que nos resulta de interés para este trabajo, la información contenida en la citada tabla revela la existencia de importantes diferencias en la dotación media de capital humano y en las restantes características personales y familiares de los individuos entre las dos regiones, que se aprecian también en características del puesto de trabajo (como la ocupación) y de la empresa (como el tamaño y el sector de actividad). Es importante destacar que esas diferencias se observan tanto entre las muestras de ocupados como entre las de los no ocupados. Además esta circunstancia no es específica del caso de la comparación de estas dos regiones sino que el análisis detallado de la distribución de características en todas las regiones permite afirmar que existe una importante heterogeneidad regional en las mismas. De todas formas, si todo el diferencial salarial medio fuese debido a esa heterogeneidad regional en las características, cuando comparásemos los salarios medios para trabajadores con unas mismas características en distintas regiones no deberíamos apreciar notables diferencias en ellos. Pero ese no parece ser el caso para las regiones españolas, lo que sugiere que la hipótesis de que una parte importante del diferencial puede ser ocasionado por diferencias en el rendimiento de esas características aparece como un supuesto verosímil. El mismo será contrastado en los siguientes apartados.

3. METODOLOGIA

La presencia de selección muestral es una situación habitual en los modelos de determinación salarial ocasionando inconvenientes en su estimación y, sobre todo, impidiendo la aplicación del método de Oaxaca-Blinder para la obtención de la descomposición detallada de la diferencia salarial, cuando se aborda un análisis de discriminación en el mercado de trabajo de cualquier índole (por género, por tipo de contrato, por raza, por sindicalización, etc.). La asiduidad de este problema en este tipo de análisis radica en que frecuentemente el proceso de selección que hace que un individuo aparezca o no en la muestra no es aleatoria, estando además relacionada con los determinantes de la magnitud en la que supuestamente se manifiesta la discriminación.

⁵ Por motivos de espacio obvios no se reproduce esa información para el conjunto de las 17 regiones, aunque se encuentra disponible bajo requerimiento para el lector interesado.

Así, en un marco de este tipo, el salario para el individuo i del grupo g (en nuestro caso una de las regiones) se determina en un sistema de ecuaciones compuesto por la ecuación de determinación salarial (1) y el proceso de selección (2):

$$W_{ig} = X'_{ig} \beta_g + \varepsilon_{ig} \quad (1)$$

$$C_{ig}^* = Z'_{ig} \gamma_g + \mu_{ig} \quad (2)$$

donde W_{ig} denota el logaritmo del salario del individuo i perteneciente al grupo g , X_{ig} representa el conjunto de sus determinantes salariales y β_g es el vector de parámetros que recoge el rendimiento de esos determinantes. Por su parte, C_{ig}^* es un proceso latente e inobservable por el cual el individuo i del grupo g es seleccionado en la muestra, siendo Z_{ig} el vector de observaciones correspondiente a las variables que determinan la selección, y γ_g el vector de parámetros asociados a ese mecanismo. ε_{ig} y μ_{ig} son términos de error i.i.d. que siguen una distribución normal bivalente $(0, 0, \sigma_{\varepsilon\varepsilon}, \sigma_{\mu\mu}, \rho_g)$, con ρ_g el coeficiente de correlación de ambas perturbaciones para el grupo g .

Únicamente es observable el resultado del proceso de selección de la ecuación (2), la variable indicador C_{ig} , que toma valor 1 si $C_{ig}^* > 0$, y 0 en caso contrario. Así, la probabilidad de que el individuo i sea seleccionado en el grupo g se define como

$$C_{ig} = \text{Prob}(C_{ig}^* > 0) = \text{Prob}(\mu_{ig} > -Z'_{ig} \gamma_g) = \Phi(Z'_{ig} \gamma_g) \quad (3)$$

siendo $\Phi(\cdot)$ la función de distribución de la normal estándar.

La estimación de la regresión salarial en (1) obviando la ecuación de selección (2) proporciona estimaciones sesgadas e inconsistentes de sus parámetros. Por el contrario, estimaciones adecuadas de la ecuación de interés para cada grupo se pueden obtener maximizando la verosimilitud del sistema de ecuaciones, aunque ha sido más frecuente en la literatura la aplicación del método de estimación bietápico propuesto en Heckman (1976). En una primera etapa, la estimación del modelo probabilístico en (3) permite obtener para todos los individuos en cada grupo la estimación de la probabilidad de selección. Con esas probabilidades se construye la inversa del ratio de Mills, λ_{ig} :

$$\lambda_{ig} = \frac{\phi(Z'_{ig} \hat{\gamma}_g)}{\Phi(Z'_{ig} \hat{\gamma}_g)} \quad (4)$$

donde $\phi(\cdot)$ y $\Phi(\cdot)$ denotan respectivamente a las funciones de densidad y de distribución de la normal estándar. En la segunda etapa se estima la ecuación salarial incluyendo λ_{ig} como un regresor adicional.

De esta forma, la ecuación salarial a estimar para los individuos de cada uno de los grupos se puede expresar como:

$$W_{ig} | C_{ig}^* > 0 = X'_{ig} \beta_g + \theta_g \lambda_{ig} + \varepsilon_{ig} \quad (5)$$

donde $\theta_g = \rho_g \sigma_{ug}$ es el parámetro asociado a la inversa del ratio de Mills y ε_{ig} es el componente inobservable.

Así, y a modo ilustrativo, si estamos interesados en la estimación de la discriminación salarial de los trabajadores del grupo B respecto a los trabajadores pertenecientes al grupo A, el salario ajustado para los individuos de cada grupo en una especificación como (5) viene dado por:⁶

$$\hat{W}_{iA} = X'_{iA} \hat{\beta}_A + \hat{\theta}_A \lambda_{iA} \quad (6)$$

$$\hat{W}_{iB} = X'_{iB} \hat{\beta}_B + \hat{\theta}_B \lambda_{iB} \quad (7)$$

La descomposición clásica de Oaxaca (1973) y Blinder (1973) permite descomponer el diferencial en el salario medio de trabajadores del grupo A y B en un doble efecto, el atribuible a las características observables y aquel asociado a la diferencia en los rendimientos de éstas. Este último componente es el que la literatura ha identificado con la existencia de discriminación. Pero la presencia del término asociado al proceso de selección en las ecuaciones salariales en (6) y (7) complica la realización de esa descomposición, dado que resulta obvio que una parte del gap salarial está vinculada con el proceso de selección. En presencia de éste, la descomposición del diferencial salarial entre los grupos A y B viene dada por:

$$\bar{W}_A - \bar{W}_B = (\bar{X}_A - \bar{X}_B) \hat{\beta}_A + \bar{X}_B (\hat{\beta}_A - \hat{\beta}_B) + (\hat{\theta}_A \bar{\lambda}_A - \hat{\theta}_B \bar{\lambda}_B) \quad (8)$$

donde la barra sobre las variables denota su media y $\bar{\lambda}_g$ es la media de la inversa del ratio de Mills para los individuos del grupo $g=A, B$. La expresión (8) nos indica que la media del gap salarial entre los grupos A y B puede explicarse a través de tres términos. Los primeros dos

⁶ Suponemos que la estructura salarial del grupo A es la no discriminatoria.

términos corresponden a los de la descomposición de Oaxaca-Blinder clásica en ausencia del proceso de selección: $(\bar{X}_A - \bar{X}_B)\hat{\beta}_A$ corresponde a la contribución de las diferencias en las variables que determinan el nivel salarial (capital humano, características de la empresa, etc.), mientras que $(\hat{\beta}_A - \hat{\beta}_B)\bar{X}_B$ hace referencia a la contribución de las diferencias en los rendimientos estimados para esas variables. El tercer término, $(\hat{\theta}_A\bar{\lambda}_A - \hat{\theta}_B\bar{\lambda}_B)$, es el debido a la existencia del proceso de selección y/o clasificación.

Partiendo de la expresión (8), que responde a la descomposición más intuitiva y más empleada en la literatura, Neuman y Oaxaca (2004) recogen distintas opciones para abordar la descomposición del diferencial salarial en presencia de sesgo de selección. La primera de ellas simplemente consiste en considerar el diferencial salarial neto de la contribución del término atribuible a la selección, de forma que es el diferencial neto el que se descompone en los términos habituales de la descomposición clásica (efecto de las dotaciones y discriminación). Esta propuesta, empleada entre otros por Duncan y Leigh (1980), Reimers (1983) y Boymond et al. (1994), se expresa como:

$$(\bar{W}_A - \bar{W}_B) - (\hat{\theta}_A\bar{\lambda}_A - \hat{\theta}_B\bar{\lambda}_B) = (\bar{X}_A - \bar{X}_B)\hat{\beta}_A + (\hat{\beta}_A - \hat{\beta}_B)\bar{X}_B \quad (9)$$

Las descomposiciones en (8) y (9) coinciden en tratar de forma global el efecto del sesgo en las diferencias salariales entre los grupos A y B. En contraste con esta opción, Neuman y Oaxaca (2004) proponen descomponer el efecto asociado a la selección, a partir de la configuración de la inversa del ratio de Mills en cada grupo. Así, tomando la definición de λ en (4), podemos expresar la componente asociada a la selección como:

$$\begin{aligned} (\hat{\theta}_A\lambda_A - \hat{\theta}_B\lambda_B) &= \hat{\theta}_A \frac{\phi(Z'_A \hat{\gamma}_A)}{\Phi(Z'_A \hat{\gamma}_A)} - \hat{\theta}_B \frac{\phi(Z'_B \hat{\gamma}_B)}{\Phi(Z'_B \hat{\gamma}_B)} = \\ &= \hat{\theta}_A \left(\frac{\phi(Z'_A \hat{\gamma}_A)}{\Phi(Z'_A \hat{\gamma}_A)} - \frac{\phi(Z'_B \hat{\gamma}_A)}{\Phi(Z'_B \hat{\gamma}_A)} \right) + \hat{\theta}_A \left(\frac{\phi(Z'_B \hat{\gamma}_A)}{\Phi(Z'_B \hat{\gamma}_A)} - \frac{\phi(Z'_B \hat{\gamma}_B)}{\Phi(Z'_B \hat{\gamma}_B)} \right) + (\hat{\theta}_A - \hat{\theta}_B) \frac{\phi(Z'_B \hat{\gamma}_B)}{\Phi(Z'_B \hat{\gamma}_B)} \end{aligned} \quad (10)$$

Definiendo $\lambda_B^A \equiv \frac{\phi(Z'_B \hat{\gamma}_A)}{\Phi(Z'_B \hat{\gamma}_A)}$ como la contrafactual de la inversa del ratio de Mills resultante de evaluar las funciones de densidad y de distribución para los individuos del grupo B en los parámetros del proceso clasificatorio del grupo A, podemos reescribir la igualdad anterior como:

$$(\hat{\theta}_A \lambda_A - \hat{\theta}_B \lambda_B) = \hat{\theta}_A (\lambda_B^A - \lambda_B) + \hat{\theta}_A (\lambda_A - \lambda_B^A) + (\hat{\theta}_A - \hat{\theta}_B) \lambda_B \quad (11)$$

A partir de la descomposición de la contribución de la selección en (11) Neuman y Oaxaca (2004) proponen la siguiente agrupación de términos en las componentes atribuibles a diferencias en características, a discriminación y a un remanente de selección:

$$\bar{W}_A - \bar{W}_B = \underbrace{(\bar{X}_A - \bar{X}_B) \hat{\beta}_A + \hat{\theta}_A (\bar{\lambda}_A - \bar{\lambda}_B^A)}_{\text{Diferencia en características}} + \underbrace{\bar{X}_B (\hat{\beta}_A - \hat{\beta}_B) + \hat{\theta}_A (\bar{\lambda}_B^A - \bar{\lambda}_B)}_{\text{Discriminación}} + \underbrace{(\hat{\theta}_A - \hat{\theta}_B) \bar{\lambda}_B}_{\text{Selección}} \quad (12)$$

En la descomposición en (12) se asume que la desigualdad salarial entre los grupos se origina por: i) un primer componente que recoge las diferencias en las características observables, tanto de los determinantes salariales (X) como de aquellas que determinan la probabilidad del proceso de selección (Z), ii) una componente identificada con discriminación salarial, y que se conforma a través de las diferencias en los rendimientos de las variables de la ecuación salarial (β) y de las diferencias en los parámetros del modelo de selección (γ), y iii) un último componente que tiene que ver con la diferencia en el efecto de la selección sobre los salarios en cada uno de los grupos (θ).

Finalmente, Neuman y Oaxaca plantean la posibilidad de incorporar el término $(\hat{\theta}_A - \hat{\theta}_B) \hat{\lambda}_B$ a la componente de discriminación, dado que este elemento no es más que la diferencia en rendimientos de la inversa del ratio de Mills:

$$\bar{W}_A - \bar{W}_B = \underbrace{(\bar{X}_A - \bar{X}_B) \hat{\beta}_A + \hat{\theta}_A (\bar{\lambda}_A - \bar{\lambda}_B^A)}_{\text{Diferencia en características}} + \underbrace{\bar{X}_B (\hat{\beta}_A - \hat{\beta}_B) + \hat{\theta}_A (\bar{\lambda}_B^A - \bar{\lambda}_B) + (\hat{\theta}_A - \hat{\theta}_B) \bar{\lambda}_B}_{\text{Discriminación}} \quad (13)$$

Las descomposiciones en (8) hasta (13) nos permiten realizar una valoración global de la contribución del proceso de selección o clasificatorio en la desigualdad salarial entre los grupos A y B, pero únicamente posibilitan el análisis de la contribución detallada de las variables o grupos de variables, contenidas en X y de sus rendimientos, es decir de aquéllas que directamente intervienen en la ecuación de determinación salarial. Por tanto, las citadas descomposiciones no permiten conocer la contribución indirecta de las diferencias en esas variables y en su rendimiento a través de su influencia sobre el proceso de selección. Y tampoco de la influencia de variables que determinan el proceso clasificatorio sin afectar directamente al nivel salarial. Esto es debido a la no linealidad de la inversa del ratio de Mills.

A continuación planteamos una propuesta para estimar la contribución total de las diferencias en todas las variables que intervienen en el sistema y en las diferencias en sus rendimientos. Para ello extendemos la metodología de descomposición de diferencias en el primer momento de funciones no lineales propuesta en Yun (2004) al caso de una especificación con un proceso de selección o clasificación.

Una propuesta de descomposición salarial detallada en presencia de selección muestral

Como se desprende de (5), en presencia de un mecanismo de selección, el salario se determina a través de una función con una componente no lineal que tiene como argumento una función lineal de las variables determinantes del proceso de selección ($Z'\gamma$). El método generalizado de descomposición de las diferencias en el primer momento propuesto en Yun (2004) proporciona una metodología válida para aislar la contribución de la diferencia entre cada una de las variables y entre cada uno de los parámetros en los dos grupos considerados, para cualquier forma funcional (lineal o no) que tenga como argumento una combinación lineal de las variables. La propuesta que realizamos en este trabajo se basa en la aplicación del principio del método general en Yun (2004) a la función salarial en presencia de un mecanismo de selección.

Supongamos que la variable de interés, Y , es una función lineal o no lineal de una combinación lineal de variables independientes, $M\psi$, donde ψ es el vector de coeficientes:

$$Y_{ig} = F(M'_{ig}\psi_g) \tag{14}$$

donde F es una función diferenciable de primer orden.

La diferencia en las medias de Y entre los grupos A y B puede descomponerse como:

$$\bar{Y}_A - \bar{Y}_B = \left[\overline{F(M_A\psi_A)} - \overline{F(M_B\psi_A)} \right] + \left[\overline{F(M_B\psi_A)} - \overline{F(M_B\psi_B)} \right] \tag{15}$$

donde $\overline{F(\cdot)}$ representa la media muestral de la función F .

La descomposición en (15) representa la forma más genérica de expresar las diferencias en la media de la variable dependiente entre los grupos A y B en las diferencias atribuibles a las características observables $\left[\overline{F(M_A\psi_A)} - \overline{F(M_B\psi_A)} \right]$ y en las diferencias en los coeficientes $\left[\overline{F(M_B\psi_A)} - \overline{F(M_B\psi_B)} \right]$. Pero en el caso de estar interesados en la contribución de cada variables a la diferencia total, es decir en la descomposición detallada, debemos ir un paso

más allá. En el caso en que la función F sea lineal, la solución es inmediata, correspondiendo con la tradicional descomposición detallada de Oaxaca-Blinder. Pero en el caso en el que F es no lineal, no es posible la aplicación de la aproximación clásica.

Como señala Yun (2004), el problema es cómo ponderar adecuadamente la contribución de cada variable a los efectos de las características y de los coeficientes. Su propuesta consiste en aproximar $\overline{F(M_g \psi_g)}$ mediante $F(\overline{M}_g \psi_g)$, y a través de una expansión de Taylor de primer orden en $\overline{M}_A \psi_A$ y $\overline{M}_B \psi_B$ para linealizar los efectos de características y coeficientes respectivamente.

En el caso específico del modelo con mecanismo de selección, sea $F(H_g, \tau_g) \equiv X_g \beta_g + \theta_g \lambda_g$, $g = A, B$, y $F(H_B, \tau_A) \equiv X_B \beta_A + \theta_A \lambda_B^A$, de forma que la descomposición global equivalente a (VI.15) se puede denotar como:

$$\overline{W}_A - \overline{W}_B = \left[\overline{F(H_A, \tau_A)} - \overline{F(H_B, \tau_A)} \right] + \left[\overline{F(H_B, \tau_A)} - \overline{F(H_B, \tau_B)} \right] \quad (16)$$

y aproximando la media de la función a través de la función en la media de las variables en H :

$$\overline{W}_A - \overline{W}_B = \left[F(\overline{H}_A, \tau_A) - F(\overline{H}_B, \tau_A) \right] + \left[F(\overline{H}_B, \tau_A) - F(\overline{H}_B, \tau_B) \right] + R_M \quad (17)$$

donde R_M es el término de error al aproximar la media de la función a través de la función en la media de las variables en H .

De forma equivalente a la descomposición en (12), la expresión anterior se puede escribir como:

$$\overline{W}_A - \overline{W}_B = \left[(\overline{X}_A - \overline{X}_B) \hat{\beta}_A + \hat{\theta}_A (\hat{\lambda}_A - \hat{\lambda}_B^A) \right] + \left[\overline{X}_B (\hat{\beta}_A - \hat{\beta}_B) + \hat{\theta}_A (\hat{\lambda}_B^A - \hat{\lambda}_B) \right] + \left(\hat{\theta}_A - \hat{\theta}_B \right) \hat{\lambda}_B + R_M \quad (18)$$

donde ahora $\hat{\lambda}_g = \left(\phi(\overline{Z}_g \hat{\gamma}_g) / \Phi(\overline{Z}_g \hat{\gamma}_g) \right)$, $g = A, B$, y $\hat{\lambda}_B^A = \left(\phi(\overline{Z}_B \hat{\gamma}_A) / \Phi(\overline{Z}_B \hat{\gamma}_A) \right)$. El término de error de aproximar las funciones a la media de las variables, R_M , se define como:

$$R_M = R_{MA} + R_{MB} + R_{MB}^A \quad (19)$$

siendo:

$$R_{Mg} = \bar{\lambda}_g - \hat{\lambda}_g = \frac{\overline{\phi(Z'_{ig} \hat{\gamma}_g)}}{\Phi(Z'_{ig} \hat{\gamma}_g)} - \frac{\phi(\bar{Z}_g \hat{\gamma}_g)}{\Phi(\bar{Z}_g \hat{\gamma}_g)} ; \quad g = A, B$$

$$R_{MB}^A = \bar{\lambda}_B^A - \hat{\lambda}_B^A = \frac{\overline{\phi(Z'_{iB} \hat{\gamma}_A)}}{\Phi(Z'_{iB} \hat{\gamma}_A)} - \frac{\phi(\bar{Z}_B \hat{\gamma}_A)}{\Phi(\bar{Z}_B \hat{\gamma}_A)}$$

Utilizando una expansión de Taylor de primer orden para linealizar los efectos asociados a los dos primeros elementos de (18), diferencias en características y en los coeficientes β y γ , en torno a $\bar{Z}_A \hat{\gamma}_A$ y $\bar{Z}_B \hat{\gamma}_B$ respectivamente:

$$\left[(\bar{X}_A - \bar{X}_B) \hat{\beta}_A + \hat{\theta}_A (\hat{\lambda}_A - \hat{\lambda}_B^A) \right] = (\bar{X}_A - \bar{X}_B) \hat{\beta}_A + \hat{\theta}_A \bar{f}_A (\bar{Z}_A - \bar{Z}_B) \hat{\gamma}_A + R_{T1}$$

$$\left[\bar{X}_B (\hat{\beta}_A - \hat{\beta}_B) + \hat{\theta}_A (\hat{\lambda}_B^A - \hat{\lambda}_B) \right] = \bar{X}_B (\hat{\beta}_A - \hat{\beta}_B) + \hat{\theta}_A \bar{f}_B \bar{Z}_B (\hat{\gamma}_A - \hat{\gamma}_B) + R_{T2}$$
(20)

con $\bar{f}_g = \partial(\lambda_g) / \partial(\hat{\alpha}_g) = -\lambda_g^2 + \hat{\alpha}_g \lambda_g$, $g = A, B$, $\hat{\alpha}_g = \bar{Z}_g \hat{\gamma}_g$, y R_{T1}, R_{T2} los errores de aproximación.

De esta forma, la descomposición en (18) se puede escribir como:

$$\bar{W}_A - \bar{W}_B = \left[(\bar{X}_A - \bar{X}_B) \hat{\beta}_A + \hat{\theta}_A \bar{f}_A (\bar{Z}_A - \bar{Z}_B) \hat{\gamma}_A \right] + \left[\bar{X}_B (\hat{\beta}_A - \hat{\beta}_B) + \hat{\theta}_A \bar{f}_B \bar{Z}_B (\hat{\gamma}_A - \hat{\gamma}_B) \right] + (\hat{\theta}_A - \hat{\theta}_B) \bar{\lambda}_B + R_M + R_{T1} + R_{T2}$$
(21)

Por lo que una descomposición detallada del diferencial salarial entre los grupos A y B se obtiene a través de:

$$\bar{W}_A - \bar{W}_B = \sum_{i=1}^l P_{\Delta H}^i \left[(\bar{X}_A - \bar{X}_B) \hat{\beta}_A + \hat{\theta}_A (\lambda_A - \lambda_B^A) \right] + \sum_{i=1}^l P_{\Delta \tau}^i \left[\bar{X}_B (\hat{\beta}_A - \hat{\beta}_B) + \hat{\theta}_A (\lambda_B^A - \lambda_B) \right] + (\hat{\theta}_A - \hat{\theta}_B) \bar{\lambda}_B$$
(22)

donde

$$P_{\Delta H}^i = \frac{(\bar{X}_A^i - \bar{X}_B^i) \hat{\beta}_A^i + \hat{\theta}_A \bar{f}_A (\bar{Z}_A^i - \bar{Z}_B^i) \hat{\gamma}_A^i}{(\bar{X}_A - \bar{X}_B) \hat{\beta}_A + \hat{\theta}_A \bar{f}_A (\bar{Z}_A - \bar{Z}_B) \hat{\gamma}_A}$$

$$P_{\Delta \tau}^i = \frac{\bar{X}_B^i (\hat{\beta}_A^i - \hat{\beta}_B^i) + \hat{\theta}_A \bar{f}_B \bar{Z}_B^i (\hat{\gamma}_A^i - \hat{\gamma}_B^i)}{\bar{X}_B (\hat{\beta}_A - \hat{\beta}_B) + \hat{\theta}_A \bar{f}_B \bar{Z}_B (\hat{\gamma}_A - \hat{\gamma}_B)}$$

$$\sum_{i=1}^l P_{\Delta H}^i = \sum_{i=1}^l P_{\Delta \tau}^i = 1$$

son las ponderaciones que permiten asignar la contribución de cada variable en H a las diferencias atribuibles a características ($P_{\Delta H}^i$) y a sus rendimientos ($P_{\Delta \tau}^i$).

4. RESULTADOS

En este apartado se presentan los resultados de la estimación de los efectos de las variables que determinan tanto directa como indirectamente los salarios de los individuos en cada una de las regiones, y los de la descomposición de las diferencias salariales regionales, siguiendo la metodología expuesta en el apartado previo. Por motivos de espacio, no se muestran los resultados de la estimación de los parámetros de las ecuaciones salariales con selección muestral para cada una de las regiones, pero sí se reproduce el valor estimado para el parámetro asociado a la IRM en cada una de las regiones y en el conjunto del Estado, así como su significación (Tabla 3). Se puede apreciar como dicho coeficiente únicamente resulta significativo en 3 regiones (en una de ellas de forma marginal) y en el conjunto del Estado. En consecuencia, en las descomposiciones en las que estuviesen implicadas esas regiones sería necesaria aplicar la descomposición en presencia de selección muestral derivada en el apartado anterior. Pero además, por el tipo de aproximación que aplicamos en este trabajo, el que existan indicios de selección muestral para el conjunto del Estado nos conduce a utilizar las estimaciones Heckit en todos los casos y aplicar la citada descomposición para analizar el gap salarial en cada una de las regiones.

Así, y como paso previo, en la Tabla 4 comparamos los efectos marginales asociados a todas las variables implicadas en el modelo de determinación salarial para el conjunto del Estado y, por motivos de espacio, únicamente para las dos regiones con valores salariales extremos, Madrid y Murcia.⁷ En la tabla se presenta el efecto marginal condicional a estar ocupado y el efecto marginal incondicional, es decir el impacto que tendría la característica correspondiente no únicamente sobre el salario de los ocupados sino sobre el del conjunto de la población (es decir incorporando el efecto indirecto a través del impacto sobre la probabilidad de ocupación).⁸ Dado que nuestro objetivo no es el análisis detallado del efecto de cada una de las características, sino simplemente mostrar como las regiones no difieren únicamente en su dotación de capital humano y en características personales de los individuos y de los puestos de trabajo y de las empresas, nos interesa poner el acento en las diferencias

⁷ Los resultados completos para todas las regiones se encuentra a disposición del lector interesado.

⁸ Véase Hoffman y Kassouf (2005).

que se observan entre las dos regiones y entre cada una de ellas y el conjunto del Estado. Se aprecia claramente como existen notables diferencias regionales en el efecto que tienen las características observadas sobre el salario. Esas diferencias se observan tanto en el rendimiento del capital humano (con mayor intensidad en experiencia y antigüedad en el caso de las regiones mostradas en la tabla) como en las características del puesto de trabajo y la empresa (por ejemplo, el rendimiento a la ocupación y al sector de actividad), y de las personales y familiares. Ello no impide que para algunas características concretas no se aprecien discrepancias sustanciales en su rendimiento en las dos regiones con niveles salariales medios extremos (por ejemplo, en el rendimiento a la educación, en el género o en los ingresos del hogar).

Para completar esos resultados parciales referidos a dos regiones, y dado nuestro interés especial en el efecto del capital humano, en la Tabla 5 se muestran los efectos marginales estimados para la educación en cada una de las regiones. En ese caso se ilustra claramente la heterogeneidad regional en el efecto sobre el salario de un incremento en los años de escolarización, con valores en algunas regiones que superan el 2% y que en otras no resultan significativamente distintos de cero para el rendimiento condicional.⁹ Para el rendimiento incondicional, mucho más elevado en general debido al importante efecto de la educación sobre la ocupación, las diferencias regionales son también muy pronunciadas, con valores por encima del 10% en algunas regiones y alrededor del 7% en otras. En resumen, estos resultados nos confirman que las regiones españolas no sólo difieren de forma sustancial en la dotación de algunas de las características que determinan los niveles salariales, sino que también lo hacen en el rendimiento que los trabajadores obtienen por esas características. Y que esas diferencias regionales se dan tanto en los factores que directamente afectan al salario como a través de los que afectan a éste de forma indirecta mediante su efecto sobre la ocupación. Esta circunstancia motiva la descomposición del gap salarial regional para poder valorar la contribución al mismo de la diferencia en dotación y en rendimiento de las características observadas, que afectan al salario tanto directa como indirectamente.

Dado el elevado número de regiones implicadas y lo limitado del espacio, para cada una de las regiones se ha efectuado la descomposición del gap salarial medio del conjunto del Estado,

⁹ Recuérdese que al estar estimando el rendimiento a la educación en una especificación que incluye controles como la ocupación y el sector de actividad, el valor obtenido debe estar subestimando el verdadero rendimiento a la educación. La estimación basada en una ecuación salarial simplificada que no incorpora esos controles proporciona estimaciones del rendimiento condicional de la educación bastante más elevadas.

excluyendo esa región, y la región. Es decir, se ha descompuesto $\bar{W}_{ESP-CAi} - \bar{W}_{CAi}$, donde $\bar{W}_{ESP-CAi}$ es la media del logaritmo del salario para la muestra de todos los individuos salvo los residentes en la región i y \bar{W}_{CAi} es la media correspondiente a esa región. Los resultados se han sintetizado en la Tabla 6. La primera columna muestra la diferencia salarial total entre el resto del Estado y cada región, de forma que si la magnitud es positiva implica que la región en cuestión tiene un salario medio inferior al del resto del Estado, mientras que si es negativa es porque el salario en la región es superior. En las dos columnas siguientes se muestra la diferencia salarial atribuible a la distinta dotación de características y a la diferencia en el rendimiento de éstas. En este caso, el signo positivo indica que la dotación o el rendimiento favorece al resto del Estado, mientras que será favorable a la región cuando la magnitud presente signo negativo. Finalmente, la última columna recoge la contribución del elemento residual asociado a la selección muestral en la expresión VI.22.

Se aprecia como tanto la mejor dotación de características como el mayor rendimiento de éstas contribuye a explicar el diferencial salarial en aquellas regiones con salario medio superior al resto de regiones. Lo mismo no se puede decir en el caso de las regiones con niveles medios inferiores a los del resto del Estado, dado que en ese caso la mitad de las regiones presentan una contribución de la diferencia en rendimientos que contrarresta la contribución de su peor dotación de características. Pero incluso más interesante es el hecho de que la contribución de los rendimientos llega a superar la de las características en un gran número de regiones españolas. De hecho, se puede concluir que en algunas de ellas no se habrían apreciado diferencias significativas con el salario medio del resto del Estado si no hubiesen existido diferencias en el rendimiento de las características. Y en las restantes las diferencias con las otras regiones hubiesen sido mucho menores de haber existido homogeneidad regional en los rendimientos. Por último, señalar que la componente residual tiene una contribución poco relevante en todas las regiones salvo en Extremadura donde, debido a la cancelación del efecto de características y rendimientos, la contribución del término residual asociado a la selección muestral es elevado. En consecuencia, los resultados de la descomposición del gap salarial regional, robustos a la existencia de selección muestral, sugieren que las importantes diferencias en los niveles salariales medios de las distintas regiones no tienen su origen principal en diferencias en la dotación de características personales y del puesto de trabajo y las empresas, sino que se deben en gran parte a heterogeneidad regional en la retribución de esas características.

Para finalizar el análisis, y dado nuestro interés en el papel del capital humano en la explicación de las diferencias salariales regionales, en la Tabla 7 presentamos la contribución detallada al gap salarial de las diferencias regionales en el nivel educativo y en el conjunto de características consideradas para aproximar el capital humano (educación y experiencia genérica y específica). Para cada una de las regiones españolas, la citada tabla muestra la contribución de la diferencia en la dotación de educación y capital humano, y de la diferencia en su rendimiento al diferencial salarial entre el resto de regiones y la región considerada. Se puede apreciar como, salvo en contadas excepciones, las regiones con mejor dotación educativa y, en general, de capital humano tienen salarios por encima de las regiones con peores dotaciones. Es decir, que la contribución de la diferencia en dotación aparece como negativa para las regiones con salarios por encima de los del resto de regiones y como positiva para las que los tienen inferiores. En consecuencia, la diferencia en la dotación de capital humano en general y del educativo en particular contribuye a explicar el diferencial regional en salarios. Pero además, en muchas regiones las diferencias en el rendimiento al capital humano contribuyen a acentuar las disparidades salariales. Así, salvo en País Vasco y Navarra, todas las regiones de salario superior al del resto del Estado presentan mayor rendimiento a esa característica, lo que contribuye a ensanchar el diferencial salarial a su favor. Una situación simétrica encontramos en la mayoría de regiones de nivel salarial inferior al resto del Estado, en las que el menor rendimiento al capital humano se une al efecto de la menor dotación. Las excepciones en este caso son Galicia en el caso del nivel educativo (Murcia presenta también una contribución negativa de esta componente pero de una dimensión muy reducida) y de Extremadura y Galicia en el del capital humano general (en este caso Baleares también presenta un efecto negativo pero de escasa dimensión). Por último, y no menos importante para una de las hipótesis planteadas en este trabajo, para muchas de las regiones, la contribución de la diferencia en el rendimiento es como mínimo de igual importancia a la correspondiente a la diferencia regional en la dotación de capital humano, tanto educativo como general. De hecho, en bastantes regiones el rendimiento tiene un efecto mucho más importante que la dotación. De lo que se desprende que la homogeneización de los niveles educativos de las distintas regiones no aseguraría, ceteris paribus, la igualación de los salarios regionales medios.

5. CONCLUSIONES

En este trabajo hemos derivado un método de descomposición detallada de las diferencias salariales regionales en presencia de selección muestral y lo hemos aplicado al caso de las regiones españolas para mostrar cómo la heterogeneidad regional en los rendimientos de las características tiene un papel primordial en la explicación de las disparidades entre regiones en los niveles salariales. Además, hemos mostrado cómo la no consideración del efecto de las características a través de su influencia en el proceso de selección da lugar a estimaciones inadecuadas de la contribución de las características y los rendimientos.

En el caso concreto de las regiones españolas, existen importantes diferencias en la distribución regional de las características, pero además existe una intensa disparidad en su rendimiento, tanto entre aquellos individuos ocupados como en el conjunto de la población. De hecho, la descomposición llevada a cabo confirma que la contribución de los rendimientos es más importante que la de la dotación de características en un gran número de regiones. El análisis detallado de la contribución de las distintas características nos ha permitido adicionalmente valorar el efecto de la dotación de capital humano en las diferencias salariales regionales. De nuevo en este caso concreto, el distinto rendimiento al capital humano, tanto educativo como genérico, tiene una contribución al gap salarial regional que es incluso más importante que la correspondiente a la diferente dotación del mismo. En consecuencia, la simple homogeneización del capital humano en el territorio no necesariamente se traduciría en una sustancial reducción de las diferencias salariales entre las regiones españolas.

Referencias

- Bernard A, Redding S, Schott P y Simpson H (2004) Relative wage variation and industry location, *CEPR Discussion Paper 4213*
- Blinder A (1973) Wage discrimination: reduced form and structural estimates, *Journal of Human Resources* 8: 436-455
- Boymond et al. (1994): Wage discrimination and occupational segregation by gender: Some evidence from Swiss data. Mimeo, Geneva.
- De la Rica S, Dolado JJ y Llorens V (2007): Ceilings or floors? Gender wage gaps by education in Spain, *Journal of Population Economics*
- Duncan, G.M. and Leigh, D.E. (1980): Wage determination in the union and non-union sectors: A sample selectivity approach, *Industrial and Labor Relations Review* 34:24-34.
- Duranton G y Monastiriotis V (2002) Mind the gaps: The evolution of regional earnings inequalities in the U.K., 1982-1997, *Journal of Regional Science* 42, 219-256
- El-Attar M, López-Bazo E y Motellón E (2007) Heterogeneity in regional wage distributions, Universidad de Barcelona, mimeo
- García I y Molina A (2002) Inter-regional wage differentials in Spain, *Applied Economics Letters* 9: 209-215
- García-Pérez JI y Jimeno JF (2007): Public sector wage gaps in Spanish regions, *The Manchester School* Vol 75 No. 4, 501-531.
- Gardeazabal J y Ugidos A (2004): More on identification in detailed wage decompositions. *The Review of Economics and Statistics* 86:1034-1036
- Head K y Mayer T (2005) Regional wage and employment responses to market potential in the UK, *CEPR Discussion Paper 4908*
- Heckman, J. (1976): The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and a Simple Estimator for Such Models, *Annals of Economic and Social Measurement* 5:475-492.
- Hoffmann R y Kassouf A (2005) Deriving conditional and unconditional marginal effects in log earnings equations estimated by Heckman's procedure, *Applied Economics* 37: 1303-1311
- Montuenga V, García I y Fernandez M (2003): Wage flexibility: Evidence from five EU countries based on the wage curve, *Economics letters* 78, 169-174.
- Neuman, S. and Oaxaca, R. (2004): Wage decompositions with selectivity-corrected wage equations: A methodological note, *Journal of Economic Inequality* 2:3-10.
- Oaxaca R (1973) Male-Female wage differentials in urban labor markets, *International Economic Review* 14: 693-709
- Reimers, C. (1983): Labor market discrimination against hispanic and black men, *Economics and Statistics* 65:570-579.

- Serrano L (2002) Salarios regionales y dotaciones de capital humano, *Revista de Economía Aplicada* 28: 23-38
- Villaverde J y Maza A (2002) Salarios y desempleo en las regiones españolas, *Papeles de Economía Española*, 93:182-194
- Webber DJ (2002) Labour's reward across regions of the EU: a distributional dynamic approach, *Applied Economics* 34: 385-394
- Yun, M. (2004): Decomposing differences in the first moment, *Economics letters* 82: 275-280.

Tabla 1. Descriptivo del nivel salarial en las regiones españolas.

	Obs	WBH		WBHR	
		Mean	Std. Dev.	Mean	Std. Dev.
Galicia	795	6.42	3.830	6.51	3.888
Asturias	396	7.06	4.152	7.00	4.111
Cantabria	455	6.55	3.963	6.65	4.023
País Vasco	618	8.75	4.362	8.29	4.134
Navarra	496	7.91	3.543	7.36	3.296
La Rioja	358	6.67	2.932	6.57	2.886
Aragón	576	8.00	4.507	8.29	4.671
Madrid	1174	8.49	4.800	8.51	4.809
Castilla-León	683	7.78	4.659	8.15	4.879
Castilla La Mancha	613	6.53	3.545	7.06	3.836
Extremadura	482	6.10	3.496	7.05	4.037
Catalunya	1513	7.92	4.490	7.44	4.216
C. Valenciana	886	6.37	2.922	6.46	2.961
Baleares	379	6.80	3.291	6.50	3.143
Andalucía	1336	6.60	3.256	6.91	3.409
Murcia	558	6.23	3.498	6.37	3.580
Canarias	848	6.49	4.224	6.65	4.327
Estado	12166	7.19	4.062	7.24	4.063

Tabla 2. Descriptivo de las variables utilizadas en el análisis para las regiones con salarios extremos.

	Estado		Madrid		Murcia	
	Ocupados	No ocupados	Ocupados	No ocupados	Ocupados	No ocupados
CAPITAL HUMANO Y CARÁCTERÍSTICAS DE PUESTO DE TRABAJO Y EMPRESA						
VARIABLES NO CATEGÓRICAS. MEDIA Y DESVIACIÓN ESTÁNDAR						
Edad	37,259 (10,684)	44,522 (11,876)	37,383 (10,456)	45,783 (11,207)	36,159 (11,342)	41,855 (12,168)
Educación (en años)	8,960 (5,370)	4,943 (4,195)	10,931 (5,156)	6,077 (4,608)	7,889 (5,516)	4,820 (4,372)
Experiencia Potencial	18,206 (12,075)	- -	18,045 (11,921)	- -	17,972 (12,513)	- -
VARIABLES CATEGÓRICAS. CATEGORÍAS EN PORCENTAJES						
Nivel de estudios						
EDUC1	2,41%	14,12%	0,93%	8,95%	7,61%	17,64%
EDUC2	13,90%	32,78%	6,59%	29,43%	14,51%	30,23%
EDUC3	26,88%	30,81%	18,92%	28,23%	29,91%	30,62%
EDUC4	9,05%	5,21%	6,33%	7,40%	10,62%	5,81%
EDUC5	11,91%	4,20%	10,81%	7,75%	8,50%	4,07%
EDUC6	12,57%	7,97%	22,04%	11,36%	9,56%	5,43%
EDUC7	11,30%	3,04%	11,66%	2,75%	9,91%	4,84%
EDUC8	11,98%	1,88%	22,72%	4,13%	9,38%	1,36%
Antigüedad						
Antigüedad ≤ 1	29,41%	-	24,07%	-	29,38%	-
Antigüedad 2-4	20,88%	-	22,38%	-	23,54%	-
Antigüedad 5-9	9,53%	-	11,23%	-	9,56%	-
Antigüedad 10-14	10,95%	-	10,98%	-	11,68%	-
Antigüedad ≥ 15	28,31%	-	30,49%	-	24,60%	-
Tipo de contrato						
Indefinido	72,99%	-	80,66%	-	63,89%	-
Temporal	27,01%	-	19,34%	-	36,11%	-
Ocupación						
OCUP1	1,91%	-	1,94%	-	0,53%	-
OCUP2	14,26%	-	21,37%	-	12,74%	-
OCUP3	10,86%	-	14,10%	-	7,79%	-
OCUP4	12,99%	-	22,72%	-	12,04%	-
OCUP5	15,58%	-	13,43%	-	14,87%	-
OCUP6	1,04%	-	0,00%	-	1,06%	-
OCUP7	19,03%	-	12,50%	-	25,31%	-
OCUP8	10,66%	-	7,35%	-	6,90%	-
OCUP9	13,67%	-	6,59%	-	18,76%	-
Tamaño empresarial						
Tamaño 1-4	13,91%	-	9,71%	-	12,92%	-
Tamaño 5-19	27,55%	-	18,16%	-	29,20%	-
Tamaño 20-49	17,94%	-	10,90%	-	20,00%	-
Tamaño 50-99	11,34%	-	10,05%	-	9,03%	-
Tamaño 100-499	15,82%	-	17,91%	-	15,40%	-
Tamaño >500	13,43%	-	33,28%	-	13,45%	-

Tabla 2. Descriptivo de las variables utilizadas en el análisis para las regiones con salarios extremos. (cont.)

	Estado		Madrid		Murcia	
	Ocupados	No ocupados	Ocupados	No ocupados	Ocupados	No ocupados
Propiedad						
Privada	74,99%	-	72,13%	-	79,47%	-
Pública	25,01%	-	27,87%	-	20,53%	-
Supervisión						
SUPERV	23,47%	-	31,93%	-	16,11%	-
SUPERVNO	76,53%	-	68,07%	-	83,89%	-
Sector de actividad						
STOR1	3,25%	-	0,17%	-	7,26%	-
STOR2	1,50%	-	1,77%	-	1,24%	-
STOR3	3,80%	-	1,35%	-	8,14%	-
STOR4	4,79%	-	2,96%	-	4,25%	-
STOR5	3,52%	-	3,21%	-	4,07%	-
STOR6	5,13%	-	2,20%	-	3,01%	-
STOR7	4,71%	-	2,87%	-	3,19%	-
STOR8	11,31%	-	6,93%	-	14,34%	-
STOR9	13,21%	-	12,67%	-	14,87%	-
STOR10	4,79%	-	3,80%	-	3,89%	-
STOR11	6,34%	-	14,53%	-	5,49%	-
STOR12	8,92%	-	18,92%	-	5,84%	-
STOR13	8,96%	-	9,54%	-	7,26%	-
STOR14	8,09%	-	8,87%	-	7,43%	-
STOR15	7,85%	-	5,74%	-	6,90%	-
STOR16	3,85%	-	4,48%	-	2,83%	-

CARACTERÍSTICAS PERSONALES Y FAMILIARES

Variables no categóricas. Media y Desviación estándar

Nº miembros hogar	3,697 (1,276)	3,813 (1,382)	3,542 (1,205)	3,676 (1,168)	3,933 (1,320)	4,000 (1,501)
Ingresos netos hogar	1081,951 (972,951)	1465,931 (913,266)	1280,170 (1127,170)	1774,817 (1089,941)	1024,620 (1025,685)	1349,367 (865,200)
Nº niños ≤ 15	0,753 (0,892)	0,928 (1,106)	0,693 (0,881)	0,914 (1,081)	0,922 (0,990)	1,275 (1,380)

Variables categóricas. Categorías en porcentajes

Género						
Hombre	59,99%	12,54%	53,29%	7,57%	63,72%	12,60%
Mujer	40,01%	87,46%	46,71%	92,43%	36,28%	87,40%
Nacionalidad						
Española	98,88%	98,34%	98,31%	96,21%	99,47%	99,42%
Extranjera	1,10%	1,66%	1,69%	3,79%	0,53%	0,58%
Estado civil						
Casado	65,65%	84,49%	65,79%	88,12%	66,02%	85,08%
Otro	34,35%	15,51%	34,21%	11,88%	33,98%	14,92%
Hijos entre 0-6	24,58%	29,10%	22,80%	27,54%	30,97%	36,24%
Enfermedad crónica	8,75%	25,16%	9,04%	22,72%	10,27%	24,61%

Tabla 3. Estimación del parámetro asociado a la IRM y su significación

Región	Coficiente de IRM	Región	Coficiente de IRM
Galicia	.0009	Castilla-La Mancha	.0455
Asturias	.0586	Extremadura	-.0273
Cantabria	.0246	Cataluña	.0126
País Vasco	.1049 **	Comunidad Valenciana	-.0426
Navarra	-.0250	Baleares	-.0266
La Rioja	.1312 **	Andalucía	.0197
Aragón	.0276	Murcia	-.0388
Madrid	.0710 *	Canarias	-.0600
Castilla y León	.0384	Estado	.0236 **

Tabla 4. Efectos marginales para las regiones con salarios extremos.

	Efecto Marginal Condicional			Efecto Marginal Incondicional		
	Estado	Madrid	Murcia	Estado	Madrid	Murcia
Educación	0,0147	0,0131	0,0190	0,0819	0,0762	0,0732
Experiencia Pot.	0,0152	0,0146	0,0067	0,0100	0,0117	0,0038
Experiencia Pot. ²	-0,0002	-0,0002	-0,0001	-0,0002	-0,0002	0,0000
Antigüedad 2-4	0,0231	0,0569	0,0252	0,0152	0,0457	0,0142
Antigüedad 5-9	0,0534	0,0987	-0,0179	0,0352	0,0794	-0,0101
Antigüedad 10-14	0,1096	0,1953	0,1021	0,0722	0,1571	0,0575
Antigüedad ≥ 15	0,1504	0,2229	0,1855	0,0992	0,1792	0,1045
Contrato Indefinido	0,0898	0,0987	0,1021	0,0592	0,0793	0,0575
Ocupación 1	0,6188	0,8041	0,5334	0,4079	0,6467	0,3004
Ocupación 2	0,4545	0,5915	0,3318	0,2996	0,4757	0,1869
Ocupación 3	0,2402	0,3246	0,2191	0,1583	0,2611	0,1234
Ocupación 4	0,1485	0,2467	0,0903	0,0979	0,1984	0,0509
Ocupación 5	0,0804	0,1059	0,0293	0,0530	0,0852	0,0165
Ocupación 6	0,0892	0,2250	0,0362	0,0588	0,1810	0,0204
Ocupación 7	0,0813	0,0862	0,1364	0,0536	0,0693	0,0768
Tamaño 1-4	-0,2404	-0,2115	-0,2981	-0,1585	-0,1701	-0,1679
Tamaño 5-19	-0,1612	-0,1320	-0,1801	-0,1063	-0,1062	-0,1014
Tamaño 20-49	-0,1085	-0,1076	-0,1635	-0,0715	-0,0865	-0,0921
Tamaño 50-99	-0,0768	0,0338	-0,1238	-0,0506	0,0272	-0,0697
Tamaño 100-499	-0,0273	0,0220	-0,0347	-0,0180	0,0177	-0,0195
Prop. Privada	-0,0833	-0,0437	-0,1191	-0,0549	-0,0351	-0,0671
Supervisión	0,1035	0,1215	0,0900	0,0682	0,0977	0,0507
Sector 1	-0,0142	0,4179	0,1046	-0,0094	0,3361	0,0589
Sector 2	0,2728	0,1042	0,4572	0,1798	0,0838	0,2575
Sector 3	0,0615	0,0234	0,1040	0,0405	0,0188	0,0586
Sector 4	0,0403	-0,0974	0,2312	0,0266	-0,0783	0,1302
Sector 5	0,1767	-0,1162	0,2752	0,1165	-0,0934	0,1550
Sector 6	0,1705	0,0341	0,2000	0,1124	0,0274	0,1126
Sector 7	0,1316	0,0375	0,1254	0,0867	0,0302	0,0706
Sector 8	0,1716	0,0055	0,2608	0,1131	0,0044	0,1469
Sector 9	0,0502	0,0168	0,0742	0,0331	0,0135	0,0418
Sector 10	0,0345	-0,0237	0,0705	0,0227	-0,0190	0,0397
Sector 11	0,1542	0,1295	0,0851	0,1016	0,1041	0,0479
Sector 12	0,1515	0,0341	0,2443	0,0998	0,0274	0,1376
Sector 13	0,1301	0,0868	0,0995	0,0858	0,0698	0,0560
Sector 14	0,2132	0,0470	0,3814	0,1405	0,0378	0,2148
Sector 15	0,0747	-0,1064	0,1424	0,0492	-0,0856	0,0802
Hombre	0,1251	0,0813	0,0889	0,9351	0,7748	0,9504
Nacionalidad Esp.	0,0275	0,0679	0,0132	0,2362	0,2783	0,6112
Casado	0,0040	0,0120	-0,0094	-0,1905	-0,1908	-0,2636
Enfermedad Crónica	-0,0118	-0,0340	0,0238	0,5324	0,5625	0,5704
Edad	-0,0007	-0,0008	0,0009	0,0314	0,0126	0,0258
Edad ²	0,0000	0,0000	0,0000	-0,0006	-0,0004	-0,0004
Nº hijos ≤ 15 años	0,0029	0,0121	-0,0084	-0,1404	-0,1975	-0,2311
Hijos entre 0-6 años	0,0036	0,0044	-0,0007	-0,1697	-0,0727	-0,0194
Nº miembros hogar	-0,0014	-0,0036	0,0041	0,0653	0,0588	0,1137
Ingresos hogar	4,24E-06	9,10E-06	-8,97E-06	-0,0002	-0,0001	-0,0002

Tabla 5. Efectos marginales para los años de educación en las regiones españolas.

Región	Ef. M. Cond.	Ef. M. Incond.	Región	Ef. M. Cond.	Ef. M. Incond.
Galicia	0.0229	0.0799	Castilla-La Mancha	0.0106	0.1104
Asturias	0.0006	0.0804	Extremadura	0.0156	0.1050
Cantabria	0.0128	0.0623	Cataluña	0.0177	0.0687
País Vasco	0.0054	0.0811	Valencia	0.0084	0.0898
Navarra	0.0113	0.0487	Baleares	0.0164	0.0637
La Rioja	0.0038	0.0683	Andalucía	0.0103	0.0836
Aragón	0.0211	0.1051	Murcia	0.0190	0.0732
Madrid	0.0131	0.0762	Canarias	0.0144	0.0696
Castilla y León	0.0170	0.0960	Estado	0.0146	0.0819

Tabla 6. Descomposición del gap salarial regional

	Diferencia Total	Diferencia en Características	Diferencia en Rendimientos	Componente Residual
Galicia	0.1201	0.0341	0.0778	0.0083
Asturias	0.0414	0.0086	0.0489	-0.0160
Cantabria	0.0993	-0.0046	0.1048	-0.0009
País Vasco	-0.1666	-0.0611	-0.0778	-0.0277
Navarra	-0.0517	-0.0282	-0.0378	0.0142
Aragón	-0.1352	-0.0855	-0.0478	-0.0020
Madrid	-0.1699	-0.1418	-0.0150	-0.0134
Castilla y León	-0.1054	-0.0675	-0.0312	-0.0068
Castilla-La Mancha	0.0259	0.0772	-0.0432	-0.0080
Extremadura	0.0307	0.0396	-0.0300	0.0294
Cataluña	-0.0456	-0.0250	-0.0236	0.0030
Valencia	0.0896	0.0429	0.0205	0.0262
Baleares	0.0845	0.0780	-0.0130	0.0193
Andalucía	0.0340	0.0680	-0.0355	0.0017
Murcia	0.1366	0.0709	0.0362	0.0295
Canarias	0.1146	0.0840	-0.0080	0.0388

Tabla 7. Contribución del nivel educativo y del capital humano al gap salarial regional

	Nivel Educativo		Capital Humano	
	Diferencia en Dotación	Diferencia en Rendimiento	Diferencia en Dotación	Diferencia en Rendimiento
Galicia	0.0047	-0.0638	0.0113	-0.1022
Asturias	-0.0034	0.0924	-0.0031	0.1579
Cantabria	0.0010	0.0189	-0.0041	0.0456
País Vasco	-0.0254	0.0398	-0.0321	0.1011
Navarra	-0.0067	0.0534	-0.0194	0.0792
Aragón	-0.0209	-0.0783	-0.0434	-0.0710
Madrid	-0.0361	-0.0083	-0.0408	-0.0008
Castilla y León	-0.0108	-0.0393	-0.0335	-0.1253
Castilla-La Mancha	0.0215	0.0081	0.0331	0.0172
Extremadura	0.0137	0.0272	0.0226	-0.0310
Cataluña	-0.0020	-0.0181	-0.0070	-0.0518
Valencia	0.0098	0.1007	0.0130	0.1279
Baleares	0.0118	0.0087	0.0217	-0.0082
Andalucía	0.0181	0.0428	0.0307	0.0221
Murcia	0.0176	-0.0048	0.0258	0.0590
Canarias	0.0217	0.0422	0.0406	0.0666