

## Temas metodológicos:

- Diferenciales de salarios, descomposición de Oaxaca y sesgo de selección
- Variables instrumentales
- Regresión de cuantilas
- Evaluación de impacto de políticas y programas

## **Diferenciales de salarios, descomposición de Oaxaca y sesgo de selección**

Para un análisis de diferenciales de salarios se debe estimar una ecuación de Mincer

$$\ln \omega_{ij} = X_{ij} \beta_j + u_{ij} \quad [1.1]$$

$i = 1, 2, \dots, N$  individuos

$j$  define el sector o grupo estudiado

$\ln \omega$  es el logaritmo natural del salario por hora

$X$  el vector de características de los individuos

$u$  error  $N(0, \sigma^2)$ .

Salario: ingreso de la ocupación principal.

Ej. variables explicativas ( $X$ ): educación, tiene título post secundario, experiencia potencial, experiencia en el actual trabajo, género, región donde vive, tiene contrato de trabajo (permanente o temporal).

Esta ecuación 1.1 se puede estimar en conjunto para ambos sectores, con una **dummy** de sector o grupo

⇒ los retornos al capital humano no difieren entre ambos sectores

Si las tasas de retorno a las características de capital humano difieren, se requiere estimar una ecuación 1.1 diferente para cada sector o grupo.

Una vez estimada la ecuación 1.1 para cada sector o grupo por MCO, se descomponen las diferencias de salarios entre los grupos de trabajadores según la metodología de Blinder-Oaxaca.

Objetivo descomposición: analizar qué parte del diferencial de salarios entre distintos sectores o grupos corresponde a:

- diferencias en la dotación de capital humano (diferentes características individuales) entre los trabajadores de ambos sectores/grupos
- diferencias en los retornos a ese capital humano entre los sectores/grupos.

La ecuación 1.2 descompone los diferenciales de salarios en: diferencias en las características ( $X$ ) y diferencias en los retornos ( $\beta$ ).

$$\ln \hat{\omega}_j - \ln \hat{\omega}_k = \bar{X}_j \beta_j - \bar{X}_k \beta_k$$

$$\ln \hat{\omega}_j - \ln \hat{\omega}_k = (\bar{X}_j - \bar{X}_k) [0.5(\beta_j + \beta_k)] + (\beta_j - \beta_k) [0.5(\bar{X}_j + \bar{X}_k)] \quad [1.2]$$

Los parámetros “gorro” ( $\hat{\phantom{x}}$ ) son aquellos estimados en las regresiones para cada sector o grupo  $j \neq k$ .

Los parámetros “barra” ( $\bar{\phantom{x}}$ ) son los promedios de las variables.

Se calcula la diferencia en las características de los trabajadores de ambos sectores evaluadas en el promedio de los retornos y la diferencia de los retornos evaluados en el promedio de las características de los individuos.

El modelo [1.1] es correcto siempre que no haya **sesgo de selección** entre sectores.

Cuando la asignación en el sector económico donde trabajan no es aleatoria, se debe estimar en dos etapas para corregir por este hecho.

1ª Etapa:

Se estima la probabilidad que un individuo, dadas ciertas características, trabaje en un determinado sector económico; i.e., que un individuo con ciertas características se autoseleccione en un sector económico (modelo probit/logit).

2ª Etapa

Se estima la versión corregida de la ecuación [1.1]:

$$E(\ln \omega_{ij} | I_i = j) = X_{ij} \beta_j + E(u_{ij} | I_i = j) = X_{ij} \beta_j + E(u_{ij} | \varepsilon_{ij} < Z_i \gamma_j) \quad [1.3]$$

Se analiza el valor esperado del salario dado que el individuo seleccionó el sector  $j$  para trabajar (la función  $I_i$  toma el valor 1 cuando  $i = j$ ).

¿Cual es la idea?

Se están omitiendo variables (no observables) que, al influir en la selección del sector de trabajo, afectan el salario y se manifiestan mediante errores que no se comportan bien.

La condición  $\varepsilon_{ij} < Z_i \gamma_j$  es equivalente a que la función  $I_i$  sea igual a 1 cuando  $i = j$  puesto que se deriva del modelo *probit* que estima la probabilidad de estar empleado en el sector  $j$ . Una expresión simplificada pero equivalente para [1.3] es:

$$\ln \omega_{ij} = X_{ij} \beta_j + \theta_j \lambda_{ij} + v_{ij}$$
$$E(\ln \omega_{ij} \mid I_i = j) = X_{ij} \beta_j + \theta_j \lambda_{ij} \quad [1.4]$$

Se ha sustituido el error  $u_{ij}$  por sus dos componentes: el término que muestra la corrección por sesgo de selección (el inverso de la razón de Mills y su coeficiente asociado) y el verdadero error, iid.

- $\theta$  es la correlación entre los términos de error de la ecuación de salarios y la de selección del sector.
- Para efectos de identificación es necesario contar con algunas variables en el modelo *probit* que no aparezcan en la regresión de salarios (equivalente a variables instrumentales).



Para la estimación de la ecuación [1.4], la descomposición de Blinder-Oaxaca se modifica levemente en relación a la expresada en la ecuación [1.2]:

$$\hat{\ln \omega}_j - \hat{\ln \omega}_k = \bar{X}_j \hat{\beta}_j + \hat{\alpha}_j \bar{\lambda}_j - \bar{X}_k \hat{\beta}_k - \hat{\alpha}_k \bar{\lambda}_k$$

$$\hat{\ln \omega}_j - \hat{\ln \omega}_k = (\bar{X}_j - \bar{X}_k) [0.5(\hat{\beta}_j + \hat{\beta}_k)] + (\hat{\beta}_j - \hat{\beta}_k) [0.5(\bar{X}_j + \bar{X}_k) + (\hat{\alpha}_j \bar{\lambda}_j - \hat{\alpha}_k \bar{\lambda}_k)] [1.5]$$

La ecuación [1.5] incluye ahora la diferencia en el valor medio del parámetro de sesgo de selección.

En este caso, el diferencial de salarios se descompone en tres términos:

- (i) las diferencias en capital humano (primer término),
- (ii) las diferencias en los retornos a las características de capital humano en cada sector/grupo (segundo término)
- (iii) las diferencias en el sesgo de selección. En este caso el segundo término muestra las diferencias salariales luego de corregir por diferentes características de los agentes según el sector y por el sesgo de selección.

Si no hay sesgo de selección [1.5] se reduce a [1.2]

# Descomposición de Oaxaca para diferenciales de salarios publico/privado en Chile (resultados preliminares)

$$(\ln W_{pub} - \ln W_{pri})(\bar{X}_{pub} - \bar{X}_{pri}) \hat{\beta}_{prom} (\hat{\beta}_{pub} - \hat{\beta}_{pri}) \bar{X}_{prom} \quad \text{Dif Sesgo}$$

administración pública

MCO	0.1255	0.1690	-0.0435	
Heckman	0.1244	0.1619	0.4448	-0.4823

administración pública + empresas públicas

MCO	0.2161	0.2177	-0.0016	
Heckman	0.2140	0.2065	0.4039	-0.3964

Fuente: Mizala y Romaguera (2004) en base a CASEN 2000

**CUADRO 1. RESUMEN DE ALGUNOS TRABAJOS SOBRE DIFERENCIALES DE SALARIOS ENTRE SECTOR PÚBLICO VS PRIVADO**

ESTUDIO (AÑO)	AUTOR Y PAÍS	IMPORTANCIA DEL SECTOR PÚBLICO	METODOLOGÍA UTILIZADA	SIGNO DIFERENCIAL SALARIO PÚBLICO-PRIVADO	
Public/private wage differentials in Metropolitan Areas (1978)	Smith, S. (1981) Estados Unidos	-	Modelo Capital Humano (Ecuación de Mincer)	(+)	
An Analysis of Public- and Private- Sector Wages Allowing for Endogenous Choices of Both Government and Union Status (1988)	Gyourko, J. y Tracy, J. (1986) Estados Unidos	-	Modelo Capital Humano (Ecuación de Mincer) Corrige por sesgo de Selección	Hombre (-) <sup>2</sup>	Mujer (+)
Wage gaps between public and private sector in Spain (1990 – 1991)	Lassibille, G. España	-	-	(+)	
The Distribution of Public Sector Wage Premia: New Evidence Using Quantile Regression Methods (1994)	Poterba, J. y Rueben, K. (1994) Estados Unidos	-	Modelo Capital Humano (Ecuación de Mincer) Regresiones por cuantiles para hombres y mujeres No corrige por sesgo de selección	Hombre <sup>1</sup> (-) Poca educación (+)	Mujer <sup>1</sup> (+) Alta educación (-)
Wage structures in the private and public sectors in West Germany (1984 – 1993)	Dustmann, C. y Van Soest, A. Alemania Oriental	18% (1985)	Modelo de Capital Humano	Hombre (-)	Mujer (+)
Public-Private employment choice, wage differentials and gender in Turkey (1994)	Tansel, A. Turquía	30% de los ocupados	Modelo Capital Humano (Ecuación de Mincer)	(-)	
Diferencias Salariales entre Trabajadores del sector público y privado (1986 – 1999)	Amarante, V. Uruguay	16% (1999)	Modelo Capital Humano (Ecuación de Mincer) Descomposición de Oaxaca	(+)	
The Public-Private Wage Gap in Brazil (2000)	Foguel, M. <i>et. al.</i> Brasil (2000)	13.4% de la población ocupada (1995)	Modelo Capital Humano (Ecuación de Mincer)	(+)	
The strange case of public sector wage premium: evidence from Latin America	Panniza, U. América Latina y el Caribe (2001)	14% de la población ocupada – promedio para la región –	Modelo Capital Humano (Ecuación de Mincer) No corrige por sesgo de selección	(+)	
Are Public Sector workers underpaid? (1997 – 1998)	Bales, S. y Rama, M. (2002) Vietnam	5% de la población ocupada	Jobs Approach Workers Approach	(+)	

**Notas:**

<sup>1</sup> El estudio muestra que en el tiempo el castigo para los hombres ha disminuido (y en ciertos casos hasta se ha vuelto positivo) y se ha mantenido constante para las mujeres.

<sup>2</sup> Para los hombres es negativo cuando se corrige por sesgo de selección, la estimación sin corrección por sesgo de selección indica un premio positivo para los hombres, pero inferior al de las mujeres.

## **Método de variables instrumentales**

$$Y = X\beta + \mu$$

X y  $\mu$  están correlacionadas, esto implica que los estimadores de  $\beta$  son sesgados e inconsistentes.

¿Por qué puede existir esta correlación?

- Variables omitidas correlacionadas con X, X medida con error, simultaneidad
- Necesitamos variables instrumentales (Z), ie, variables no correlacionadas con  $\mu$ , pero correlacionadas con X. En este caso:

$$\beta_{iv} = (Z'X)^{-1} Z'Y$$

Donde Z es una matriz de variables instrumentales en que cada una de sus columnas no está correlacionada con  $\mu$  y sí con X, algunas columnas de Z son columnas de X sin problemas de correlación, otros son instrumentos.

$\beta_{iv}$  es un estimador consistente de  $\beta$

Ejemplos en que se requiere usar variables instrumentales:

- Efecto de la competencia sobre la forma como se remunera y las características que se demandan de los profesores (Hoxby 2002, EEUU).

Las ecuaciones a estimar son:

$$Q_{im}^k = \alpha_0 + \alpha_1 C_m + X_m \alpha_2 + \varepsilon_m + \varepsilon_{im}$$

$$\ln(W_{im}) = \beta_0 + \beta_1 C_m + \beta_2 Q_{im}^k + \beta_3 C_m Q_{im}^k + \dots + X_m \beta_4 + \xi_m + \xi_{im}$$

donde  $C_m$  es uno menos el Índice de Herfindal, que varía entre 0 (sin elección) y 1 (máxima elección)

El test es determinar si  $\alpha_1$  y  $\beta_3$  son mayores que cero.

$C_m$  puede estar correlacionada con los errores

Variable instrumental: ríos y arroyos (afecta la formación de distritos escolares en una área metropolitana, pero no afecta la conducta de los distritos escolares).

Ejemplo Chile (Contreras, 2003)

**Analizar si la mayor competencia en determinadas comunas mejora resultados en la PAA. La ecuación a estimar es:**

$$PAA_{ij} = \alpha X_{ij} + \beta F_{ij} + \delta S_{ij} + \gamma C_j + \varepsilon_{ij}$$

X: características alumno

F: características Familia

S: características establecimiento

C: grado de competencia comuna

Variables instrumentales para C:

- población comunal
- área geográfica
- ingreso per cápita promedio
- disponibilidad de establecimientos PP
- disponibilidad establecimientos PS
- $C_{t-1}$

Variable dependiente: Puntaje matemáticas PAA 1998

<b>a) OLS</b>	<b>(1)</b>	<b>(2)</b>	<b>(3)</b>	<b>(4)</b>	<b>(6)</b>
<b>Indice de Herfindahl</b>	113.34 [8.5]	111.06 [12.2]	108.54 [12.0]	70.25 [9.5]	50.47 [4.5]
<b>b) 2SLS</b>	<b>(1)</b>	<b>(2)</b>	<b>(3)</b>	<b>(4)</b>	<b>(6)</b>
<b>Indice de Herfindahl</b>	126.98 [8.0]	123.67 [11.1]	119.88 [11.4]	78.98 [9.3]	58.01 [7.7]

Estadísticas t robustas, calculadas usando la corrección de Moulton

(Contreras, 2003)



## **Regresiones de Cuantilas**

El supuesto típico de texto es que, dadas las variables independientes, la variable dependiente se distribuye normal iid

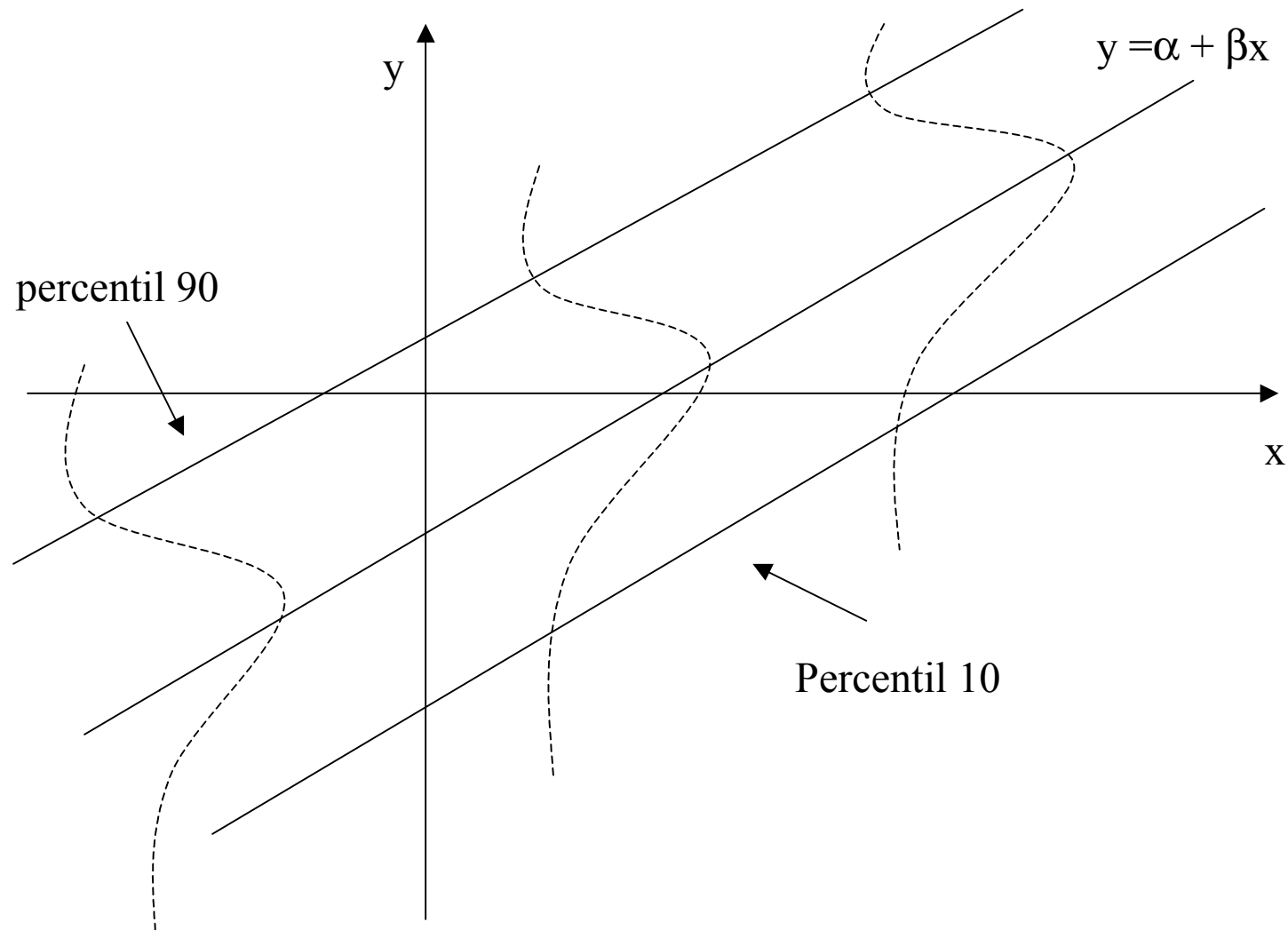
Pero, es difícil suponer que, por ejemplo, datos de encuestas de hogares se distribuyan normal iid, ellos son heterocedásticos (ie, no idénticamente distribuidos)

Heterocedasticidad implica que OLS es ineficiente (errores estándar incorrectos)

-

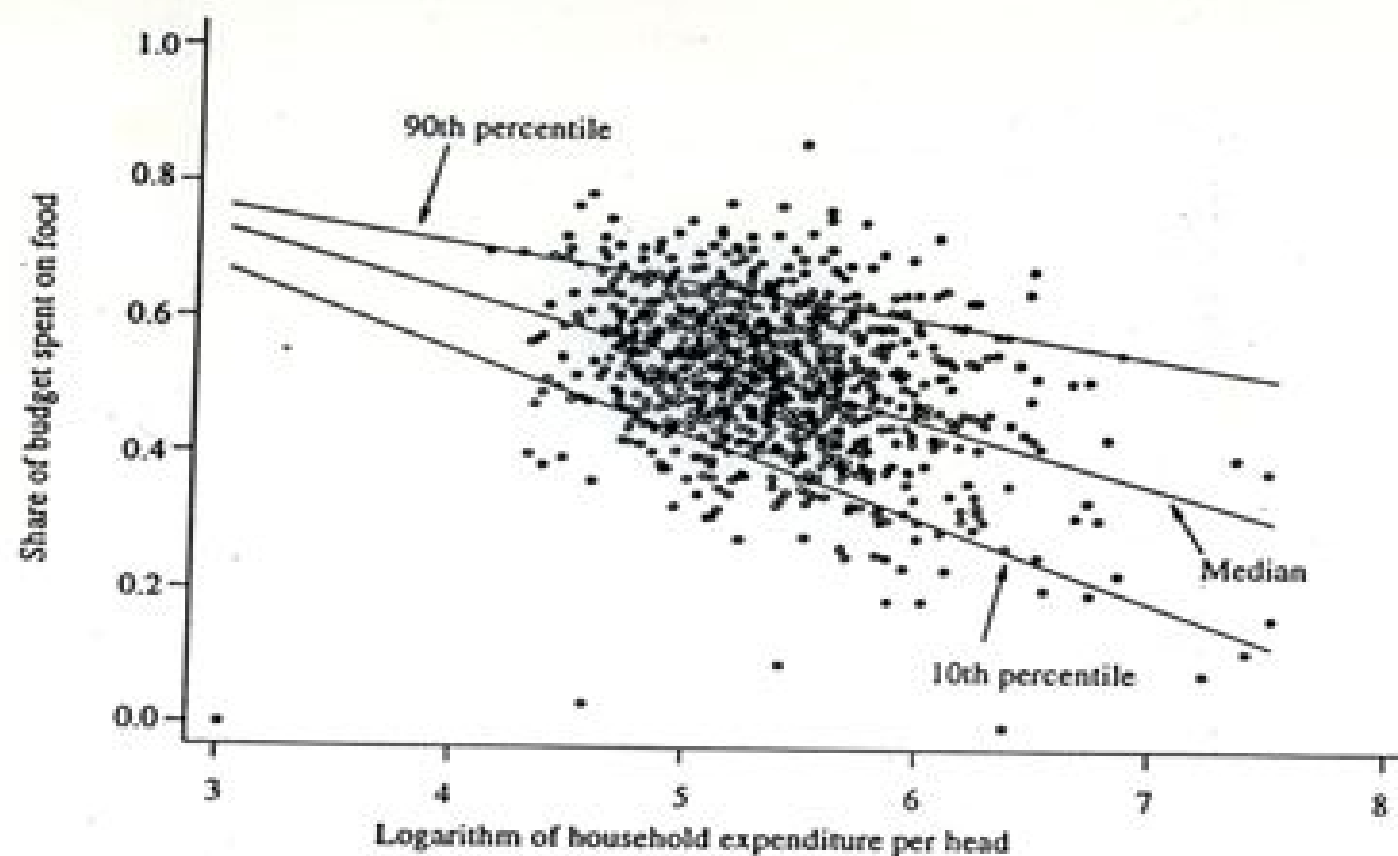
- Para enfrentar el problema de heterocedasticidad recurrimos a las regresiones de cuantilas (gráficos)
- Regresiones para diferentes cuantilas también permite explorar la forma de la distribución condicional
- Además, la regresión de la mediana está menos afectada por *outliers* que la OLS
- Resultados (perfiles de ingreso, insumos educacionales, salarios públicos y privados)

## Gráfico esquemático de una regresión lineal homocedástica



Nota: la línea sólida muestra la función de regresión de  $y$  en  $x$ , asumiendo linealidad. Las líneas cortadas muestran la distribución condicional de  $y$  dado  $x$ .

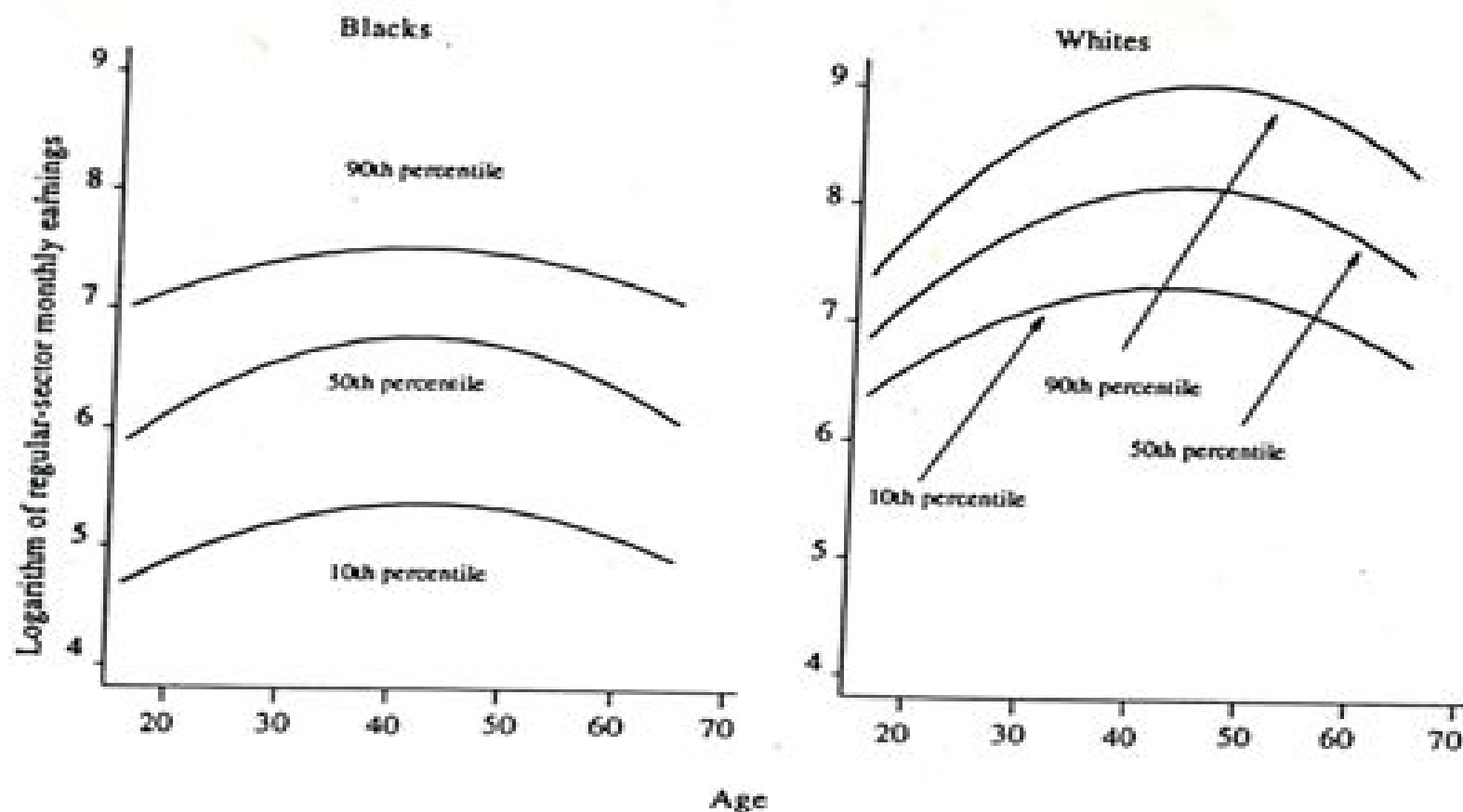
**Figure 2.2. Scatter diagram and quantile regressions for food share and total expenditure, Pakistan, 1984–85**



*Note:* The scatter as shown is a ten percent random sample of the points used in the regressions. The regression lines shown were obtained using the "qreg" command in STATA and correspond to the 10th, 50th, and 90th percentiles.

*Source:* Author's calculations based on Household Income and Expenditure Survey.

**Figure 2.3. Quantile regressions of the logarithm of earnings on age by race, South Africa, 1993**  
(log of rand per month)



Source: Author's calculations using the South African Living Standards Survey, 1993.

**Table 5. Quantile Regressions for Language Achievement in Chile, Tenth Grade, 1998\***

<i>Variable</i>	<i>OLS</i>	<i>Q10</i>	<i>Q25</i>	<i>Q50</i>	<i>Q75</i>	<i>Q90</i>
Dummy private	18.974	24.204	21.838	18.709	17.114	15.162
subsidized	(0.425)*	(0.959)*	(0.685)*	(0.448)*	(0.383)*	(0.544)*
Dummy private	26.251	37.090	31.313	25.332	22.428	19.219
fee-paying	(0.887)*	(2.061)*	(1.146)*	(1.032)*	(0.926)*	(0.900)*
H&S schools	16.127	15.679	16.125	16.541	17.273	17.614
	(0.438)*	(0.728)*	(0.603)*	(0.557)*	(0.636)*	(0.600)*
H&S +	2.152	4.071	2.924	1.488	2.029	1.919
Technical	(0.505)*	(1.245)*	(0.955)*	(0.555)*	(0.593)*	(0.505)*
Total	0.007	0.007	0.005	0.007	0.008	0.010
enrollment	(0.0009)*	(0.001)*	(0.0008)*	(0.001)*	(0.001)*	(0.001)*
Total	-1.39E-6	-1.62E-6	-9.60E-7	-1.31E-6	-1.58E-6	-1.95E-6
enrollment <sup>2</sup>	(2.69E-7)*	(4.16E-7)*	(2.69E-7)*	(3.39E-7)*	(3.18E-7)*	(4.22E-7)*
Gender	5.452	9.246	6.053	4.760	3.611	3.676
	(0.329)*	(0.813)*	(0.476)*	(0.473)*	(0.375)*	(0.441)*
Student/teacher	-0.201	-0.209	-0.193	-0.215	-0.183	-0.169
ratio	(0.023)*	(0.034)*	(0.032)*	(0.027)*	(0.027)*	(0.033)*
Teacher	0.703	0.889	0.885	0.727	0.632	0.500
experience	(0.039)*	(0.902)*	(0.071)*	(0.061)*	(0.037)*	(0.044)*
Student	17.868	17.443	19.142	19.029	17.802	16.130
socioeconomic	(0.254)*	(0.533)*	(0.343)*	(0.258)*	(0.280)*	(0.368)*
level						
Student	-2.407	-1.582	-2.509	-2.751	-2.824	-2.518
socioeconomic	(0.114)*	(0.208)*	(0.180)*	(0.106)*	(0.106)*	(0.125)*
level squared						
Achievement	0.388	0.669	0.630	0.319	0.145	0.154
peer effect	(0.028)*	(0.046)*	(0.031)*	(0.028)*	(0.028)*	(0.028)*
Socioeconomic	0.097	0.136	0.128	0.092	0.053	0.027
peer effect	(0.022)*	(0.050)*	(0.036)*	(0.025)*	(0.023)*	(0.031)
Constant	206.705	130.512	166.032	213.329	248.793	273.054
	(1.599)*	(3.743)*	(2.496)*	(1.923)*	(1.479)*	(2.029)*
(Pseudo) $R^{2b}$	0.2509	0.1080	0.1346	0.1503	0.1572	0.1493
No.						
observations	67549					

\* Statistically significant at 1 percent.

a. Excluded dummy variables: municipal schools, technical schools, coeducational schools. Asymptotic standard errors in parenthesis. Heteroskedasticity robust for OLS, bootstrapped for quantiles.

b. Pseudo  $R^2 = (1 - \frac{\text{sum of weighted deviations around estimated quantile}}{\text{sum of weighted deviations around raw quantile}})$

## Retorno a la educación y a tener título universitario. Sector público versus privado. Chile

Variable	OLS	10	25	50	75	90
S. público						
años educación	0.083	0.052	0.063	0.077	0.094	0.115
título	0.499	0.654	0.644	0.584	0.399	0.244
S. privado						
años educación	0.080	0.065	0.059	0.068	0.086	0.105
título	0.832	0.661	0.863	0.933	0.903	0.917

(1) Sector público: administración central, empresas públicas, salud y educación pública

(2) Todas las variables son significativas al 1%

Dos ecuaciones separadas para S. público y S. privado:

$\ln W = f(\text{cte}, \text{exp}, \text{exp}^2, \text{antigüedad actual trabajo}, \text{antig}^2, \text{educ}, \text{genero}, \text{título}, \text{contrato perm}, \text{urbano})$

Fuente: Mizala y Romaguera (2004)

## **Evaluación de impacto de políticas y programas**

¿Qué habría pasado si el programa no hubiera existido?

- Se compara un grupo que ha sido beneficiario del programa con un grupo de control que no ha sido beneficiario
- Hay diseños experimentales y cuasi experimentales  
X Caso experimental beneficiarios y no beneficiarios son escogidos aleatoriamente de un conjunto de individuos elegibles
- Caso cuasi experimental se construye un grupo de control similar al grupo de beneficiarios a través de técnicas econométricas



## Ejemplos de experimentos en educación

- (1) Kenia: Asignación aleatoria de 7 colegios al tratamiento y 7 como grupo de control
  - Tratamiento: uniformes, textos y construcción de salas
  - Resultados: estudiantes en colegios tratados alcanzaron alrededor de 15% más escolaridad que el grupo de control (Kremer, 2002)
  
- (2) Kenia: asignación aleatoria de 25 jardines escolares a tratamiento y 25 grupo de control
  - Tratamiento: desayuno gratis en el jardín a los niños tratados
  - Resultados: la participación escolar fue 30% mayor en los jardines tratados que en el grupo de control (Vermeersch y Kremer 2004)

Métodos paramétricos: Se observan diferencias entre resultados de grupo afectado y grupo de comparación (diferencias en diferencias)

Se controla vía regresión por las diferencias entre las características iniciales observadas entre ambos grupos y por cambios en las variables exógenas (ej. cambio legislación laboral en Brasil)

Condiciones que debe cumplir el grupo de comparación:

No debe ser afectado por la política o programa

Impacto de otros cambios debe ser similar para grupo de tratamiento y control

## Efectos aumentos costos de despido en Brasil

(aumento de la multa desde 10% a 40% (1988), y luego a 50% (2001))

$$d_{it} = \alpha + \beta X_{it} + \gamma_o Treat_{it} + \gamma_1 8901_{it} + \gamma_2 Treat_{it} * 8901_{it} + \gamma_3 Post01_{it} + \gamma_4 Treat_{it} * Post01_{it} + \delta Y_t + u_{it}$$

d duración del empleo

i individuo

t meses

X características observadas (género, edad, educación, sector, región)

Treat: dummy para grupo de tratamiento

8901 dummy cambio constitucional, 1 entre enero 1989 y septiembre 2001

Post 01 dummy periodo posterior sept 2001, 1 a partir de octubre 2001

Y vector de variables agregadas (desempleo, inflación, apertura de la economía)

Grupos de control: informales, trabajadores contratados a prueba (menos de 3 meses)

Test:  $\gamma_2$  y  $\gamma_4 > 0$  y  $\gamma_4 > \gamma_2$ .

Métodos para seleccionar un grupo de comparación (Regalia 1999; Heckman, Ichimura y Todd 1997; Baker, 1999).

Se trata de emular un experimento

Matching: se construye el grupo de comparación seleccionando individuos similares (vecinos cercanos) a los beneficiarios a partir de datos de encuestas. *Propensity scores*.

Pasos a seguir:

- calcular propensity score  $P(X)$ , usando un modelo probit o logit,  $X$  características de los individuos
- formar la distancia  $|P(X_i) - P(X_j)|$  para cada  $i$  con tratamiento y  $j$  comparación
- determinar número de vecinos cercanos
- construir el match

El parámetro de interés es el impacto medio del tratamiento en los tratados

$$\Delta_{D=1}(X) = E(Y_1 - Y_0 | X, D=1)$$

- Corte transversal: compara resultados en el mismo periodo de tiempo después del programa.

$$\alpha^{CS} = \left( \frac{1}{n_1} \right) \sum_{i=1}^{n_1} (Y_{1i} - E[Y_{0i} | P(X_i)])$$

donde  $E[Y_{0i} | P(X)]$  se estima a través de método de vecinos cercanos

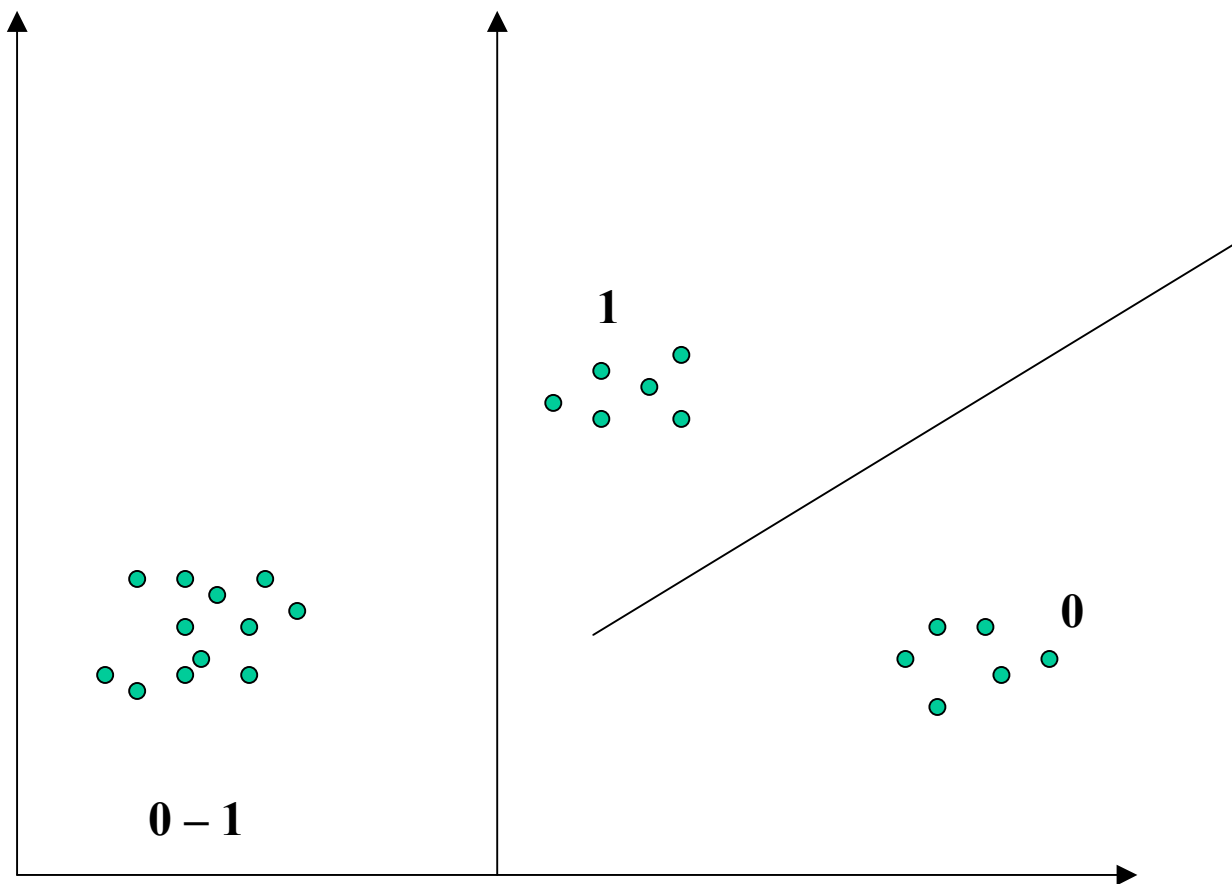
Supuestos:

- No hay sesgo de selección (misma distribución de resultados)
- Existen beneficiarios y no beneficiarios para cada X

No hay componentes privilegiadas, cada individuo tiene posibilidad de recibir tratamiento. No hay una característica que haga que un individuo sea tratado.

$$(Y_0, Y_1) \perp D \mid x,$$

$$0 < \text{Prob}(D=1 \mid x) < 1$$



- Diferencias en diferencias: compara cambios en beneficiarios con respecto al grupo comparación, respecto a un periodo de referencia antes del programa/política

Puede ser utilizado para reducir el potencial sesgo de selección y el impacto de otros factores exógenos al programa

$$a^{DD} = \left( \frac{1}{n_1} \right) \sum_{i=1}^{n_1} \{ (Y_{1i} - E[Y_{0i} | P(X_i)]) \} - \left( \frac{1}{n_1} \right) \sum_{i=1}^{n_1} \{ (Y'_{1i} - E[Y'_{0i} | P(X_I)]) \}$$

Supuestos:

- no hay sesgo de selección

$$E(Y_{0t} - Y_{0t'} / P(X), D = 1) = E(Y_{0t} - Y_{0t'} / P(X), D = 0)$$

- existen beneficiarios y no beneficiarios para cada X

$$0 < \Pr(D = 1 / X) < 1$$

Donde:

- $Y_0$  resultado sin tratamiento
- $Y_1$  resultado con tratamiento
- $t$  y  $t'$  momentos previo y posterior a la aplicación del programa/política
- $X$  característica de individuos
- $D=1$  individuo fue tratado



-Comparaciones reflexivas, compara mismos beneficiarios antes y después del programa

- Resultados año base conforman el grupo de control
- Evaluación compara el cambio promedio en los indicadores de resultados antes y después de la intervención.

- Regresión basada en variables instrumentales, caso en que hay sesgo de selección.

Cualquier variable correlacionada con la participación del individuo en el programa, pero no correlacionada con resultados individuales, dada su participación, puede ser usada como VI.

El método tiene dos etapas

- (i) Se predice la participación en el programa usando VI
- (ii) Se comparan indicadores promedio, condicional a la participación o no participación predicha.