# Эконометрика — Совбак ВШЭ и РЭШ, 2022 final

Совбак ВШЭ и РЭШ | Эконометрика | 2022 | final

---

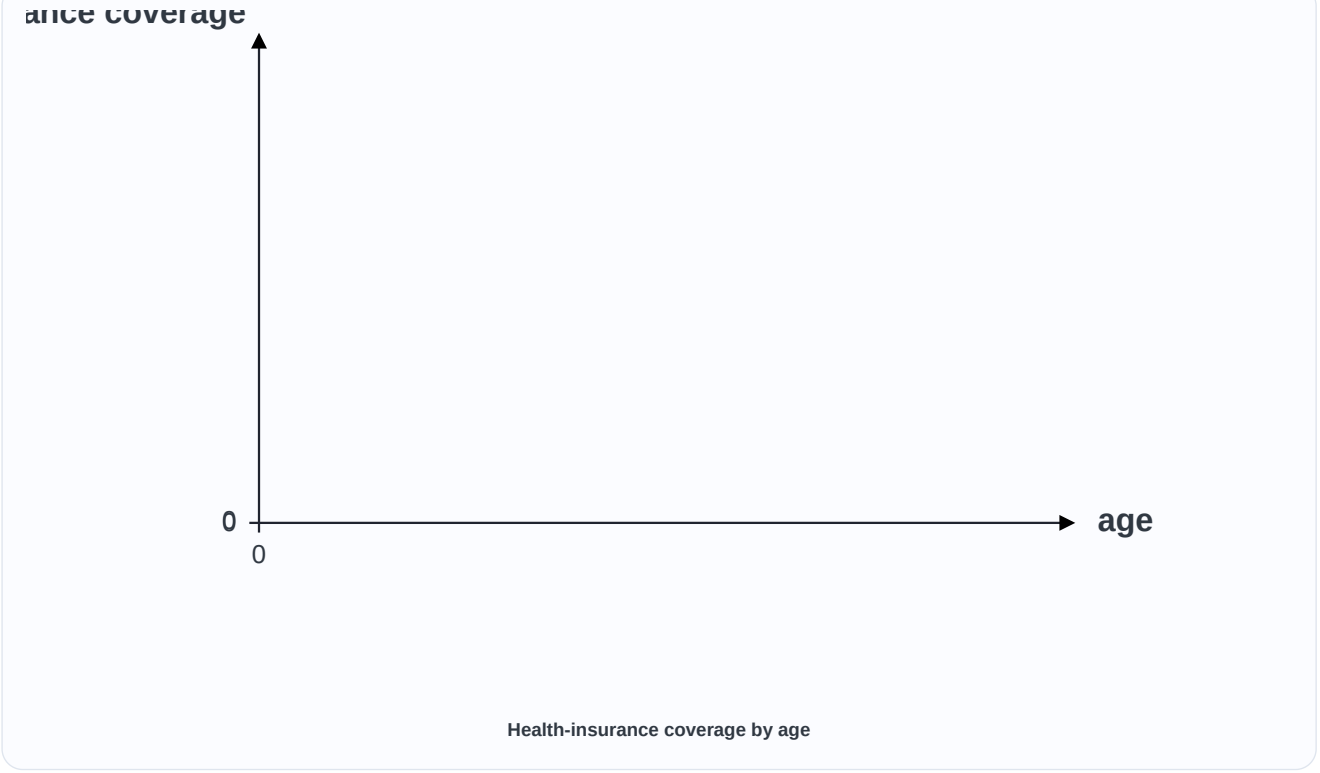
**Рисунки пока рендерятся в тестовом режиме и могут отличаться от исходных материалов.**

QUESTION 1

## Part I: Health insurance and medical-care use — 32 points

The fraction of U.S. adults with medical insurance rose substantially after implementation of the Affordable Care Act. This part studies whether insurance causes greater use of medical services by exploiting Medicare eligibility at age 65.

The data are from the 2006 U.S. Health and Retirement Survey and are restricted to working, nonretired people aged 48-75.




**Health-insurance coverage by age**




**Average number of doctor visits by age**

Table 1. Variable definitions

| Variable | Definition | Mean | Std. dev. |
|---|---|---|---|
| HI | 1 if the individual has any health insurance, 0 otherwise | 0.87 | 0.34 |
| Dr visits | Number of doctor visits during the year | 7.31 | 12.81 |
| hospital stays | Number of hospital stays during the year | 0.22 | 0.67 |
| prescription drugs | Number of prescription drugs currently taken | 0.68 | 0.47 |
| Age | Age in years | 59.62 | 6.71 |
| Age65+ | 1 if age is at least 65, 0 otherwise | 0.24 | 0.67 |
| assets | Total assets, dollars | 319,435 | 1,603,809 |
| income | Total income, dollars | 90,251 | 123,143 |
| married | 1 if married and spouse is alive | 0.77 | 0.42 |
| female | 1 if female | 0.52 | 0.50 |
| Hispanic | 1 if Hispanic | 0.09 | 0.29 |
| black | 1 if Black | 0.14 | 0.35 |

$N = 7{,}036$.

Table 2. Medical utilization and health insurance

Standard errors are in parentheses. All standard errors and $F$ tests are heteroskedasticity-robust. Regressions (4)-(6) use TSLS with Age65+ as the instrument for HI.

| Regressor/statistic | (1) OLS: Dr visits | (2) OLS: HI | (3) OLS: Dr visits | (4) TSL |
|---|---|---|---|---|
| HI | 2.01**<br>(0.69) | — | — | 9.62 |
| Age65+ | — | 0.153**<br>(0.006) | 1.47**<br>(0.34) | |
| Age | 1.06<br>(5.18) | — | — | |
| $Age^2$ | -0.02<br>(0.08) | — | — | |
| $Age^3$ | 0.00013<br>(0.00046) | — | — | |
| Other controls: income, assets, female, Black, Hispanic, married | Yes | No | No | |
| Constant | -15.7<br>(104.3) | 0.834**<br>(0.005) | 6.96**<br>(0.18) | -1.0 |
| $R^2$ | 0.010 | 0.038 | 0.002 | |
| $F$: Age, $Age^2$, $Age^3$ | 5.99, $p = .001$ | — | — | |
| $F$: other controls | 6.44, $p = .000$ | — | — | |

1

Using regression (1):

1. **(2 points)** Interpret the coefficient on HI.
2. **(2 points)** Construct a 95% confidence interval.

2

Using regression (3):

1. **(2 points)** Interpret the intercept.

2. **(2 points)** Interpret the coefficient on Age65+.

**3 (3 points)** In regression (1), the individual $t$ statistics for Age, $Age^2$, and $Age^3$ are below one, yet their joint $F$ statistic is 5.99 and rejects at 1%. How is this possible?

**4 (3 points)** People expecting high medical costs may be more likely to buy insurance. Explain why this produces simultaneous-causality bias in the coefficient on HI in regression (1).

**5 (3 points)** The coefficient on HI in regression (4) is the ratio of the Age65+ coefficients in regressions (3) and (2):

$$\frac{1.47}{0.153} = 9.62.$$

Explain why, intuitively or mathematically.

## 6. IV regression (4)

1. **(2 points)** What is a weak instrument?
2. **(3 points)** Is Age65+ weak or strong here? Explain.
3. **(2 points)** What does instrument exogeneity mean?
4. **(3 points)** Is Age65+ plausibly exogenous in regression (4)? Use judgment.
5. **(5 points)** Regressions (1), (4), (5), and (6) estimate the effect of insurance on doctor visits. Which has the greatest internal validity? Explain.

QUESTION 2

## Part II: Performance-linked pay — 18 points

Table IV from Lemieux, MacLeod, and Parent (Quarterly Journal of Economics, 2009) reports regressions of log wages on a performance-pay dummy and other variables using worker panel data. The regressions also contain industry, occupation, and year dummies; county unemployment; marital status; race dummies; and union status. Standard errors are in parentheses.

Experience and tenure enter as quadratics. “Potential experience, effect at 20 years” and “Tenure, effect at ten years” report marginal effects evaluated at those values. Interaction rows report corresponding differences for performance-pay jobs.

Table IV. Skills-related wage differentials and performance-pay jobs

| Variable | PP jobs OLS (1) | Non-PP jobs OLS (2) | Sample OLS (3) | Sample FE (4) |
|---|---|---|---|---|
| Performance-pay job dummy | — | — | -0.4526<br>(0.1019) | -0.2061<br>(0.0723) |
| Years of education | 0.0929<br>(0.0071) | 0.0665<br>(0.0039) | 0.0637<br>(0.0040) | 0.0167<br>(0.0091) |
| Education × performance-pay job | — | — | 0.0365<br>(0.0071) | 0.0169<br>(0.0048) |
| Education × 1990-1993 | — | — | — | — |
| Education × performance pay × 1990-1993 | — | — | — | — |
| Potential experience, effect at 20 years | 0.4259<br>(0.0635) | 0.2882<br>(0.0288) | 0.3010<br>(0.0294) | 0.4545<br>(0.1258) |
| Experience × performance-pay job | — | — | 0.1162<br>(0.0584) | 0.0149<br>(0.0501) |
| Tenure, effect at ten years | 0.1670<br>(0.0268) | 0.2197<br>(0.0154) | 0.2262<br>(0.0154) | 0.1158<br>(0.0128) |
| Tenure × performance-pay job | — | — | -0.0666<br>(0.0301) | 0.0278<br>(0.0237) |
| Observations | 9,680 | 16,466 | 26,146 | 26,146 |

Models (5)-(6) include interactions between performance pay and education, potential experience, and tenure. Models (5)-(6) include the full set of interactions for 1980-1984, 1985-1989, and 1990-1993, but only 1990-1993 estimates are reported. FE is the fixed-effect method at the individual-job level.

**1 (3 points)** Using column (3), is the return to education higher in performance-pay or non-performance-pay jobs? What is the difference, and is it statistically significant?

**2 (3 points)** Using column (3), what is the return to having a performance-pay job for a worker with 16 years of education, 20 years of experience, and 10 years of tenure?

**3 (3 points)** Column (4) includes worker fixed effects. The education coefficient falls from 0.0637 in column (3) to 0.0167 in column (4). Is this a large economic change? Explain.

**4 (5 points)** Give a concrete explanation for the difference between 0.0637 and 0.0167.

**5 (4 points)** Choose among homoskedasticity-only, heteroskedasticity-robust, and individual-job clustered standard errors. Which is most appropriate, and why?

QUESTION 3

## Part III: Miscellaneous questions — 50 points

**1 (8 points)** Measurement error

The true model is

$$Y^* = \beta_0 + \beta_1 X^* + \varepsilon,$$

but only

$$Y = Y^* + \eta, \qquad X = X^* + \xi$$

are observed. The variables $\varepsilon$, $\eta$, and $\xi$ have zero means and are mutually independent and independent of $X^*$.

1. **(4 points)** Show that the OLS slope from regressing $Y$ on $X$ is inconsistent.
2. **(4 points)** Briefly describe one way to obtain a consistent estimator of $\beta_1$.

**2 (13 points)** Linear probability model

Consider

$$E(y_i \mid x_i) = \beta_0 + \beta_1 x_i,$$

where $y_i = 1$ if student $i$ receives an A in econometrics and 0 otherwise, and $x_i$ is hours spent studying.

1. **(3 points)** In

   $$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,$$

   find $E(\varepsilon_i \mid x_i)$ and $\mathrm{Var}(\varepsilon_i \mid x_i)$.
2. **(2 points)** Is OLS consistent for $\beta_1$?
3. **(4 points)** What is the correct way to estimate standard errors of $\hat{\beta}_1$?
4. **(4 points)** Suggest a method that may produce narrower confidence intervals than OLS.

**3 (14 points)** Expectations hypothesis for Treasury bills

Let $hy6_t$ be the three-month holding yield, in percent, from buying a six-month Treasury bill at $t-1$ and selling it at $t$ as a three-month bill. Let $hy3_{t-1}$ be the known three-month holding yield from buying a three-month bill at $t-1$.

The expectations hypothesis says

$$E(hy6_t \mid I_{t-1}) = hy3_{t-1}.$$

This suggests

$$hy6_t = \beta_0 + \beta_1 hy3_{t-1} + u_t$$

and the null $H_0 : \beta_1 = 1$.

Using 123 observations,

$$\widehat{hy6_t} = -0.058 + 1.104 hy3_{t-1},$$

with heteroskedasticity-robust standard error 0.039 for the slope.

1. **(3 points)** Test $H_0 : \beta_1 = 1$ against a two-sided alternative at 1%. Is the estimate significantly different from 1?

Another implication is that no variable dated $t-1$ or earlier should help once $hy3_{t-1}$ is controlled for. Let $d_{t-1}$ be the price difference between six- and three-month bills. The expanded regression is

$$\widehat{hy6_t} = -0.123 + 1.053 hy3_{t-1} + 0.480 d_{t-1},$$

with robust standard errors 0.039 and 0.109 for the two slopes.

2. **(3 points)** Is the coefficient on $hy3_{t-1}$ significantly different from 1? Is the coefficient on $d_{t-1}$ significant?
3. **(3 points)** Explain how to test jointly that the coefficient on $hy3_{t-1}$ is 1 and the coefficient on $d_{t-1}$ is zero.
4. **(5 points)** A regression of $hy3_t$ on $hy3_{t-1}$ gives a coefficient of 0.914. Why may this be concerning? How would you modify the analysis?

**4 (15 points)** Dynamic causal effects

1. **(1 point)** What specification is appropriate for estimating the dynamic causal effect of $X$ on $Y$?
2. **(2 points)** What does it mean for $X$ to be exogenous? Strictly exogenous?
3. For each example, assess whether exogeneity and strict exogeneity are plausible:
   - **(4 points)** $Y$: U.S. inflation; $X$: percentage change in world oil prices set by OPEC.
   - **(4 points)** $Y$: GDP growth; $X$: Federal Funds rate.
   - **(4 points)** $Y$: change in inflation; $X$: unemployment rate, as in a Phillips curve.