

# Эконометрика — Совбак ВШЭ и РЭШ, 2021 final

Совбак ВШЭ и РЭШ

Эконометрика

2021

final

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

**QUESTION 1**

**Part I: The trend toward later retirement — 20 points**

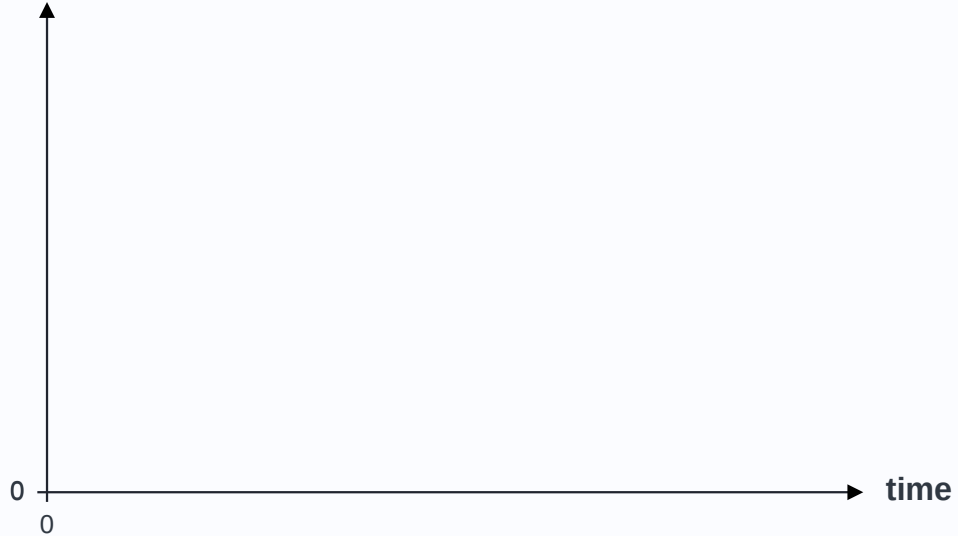
The labor-force participation rate, LFPR, is the fraction of people who are employed or looking for work. From the mid-1990s through approximately 2009, the LFPR of Americans aged 55 and over,  $LFPR_{55}$ , rose from roughly 30% to more than 40%. Since 2010 it has plateaued.

The four-quarter change is

$$\Delta_4 LFPR_{55_t} = LFPR_{55_t} - LFPR_{55_{t-4}}$$

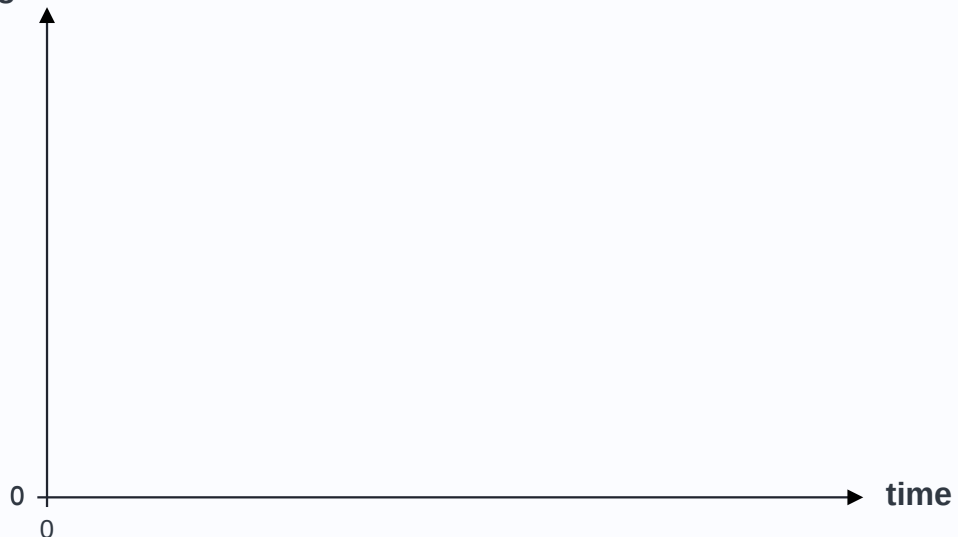
The part examines whether the slowdown was caused by the 2008-2009 recession using unemployment, real disposable personal-income growth, and housing starts.

**LFPR age 55+**



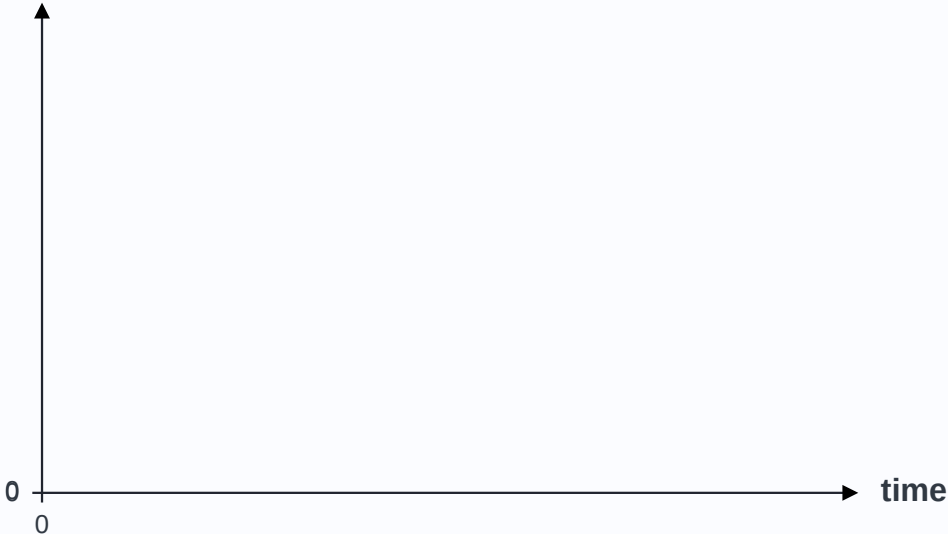
Labor-force participation rate, ages 55+, quarterly

**quarter change**



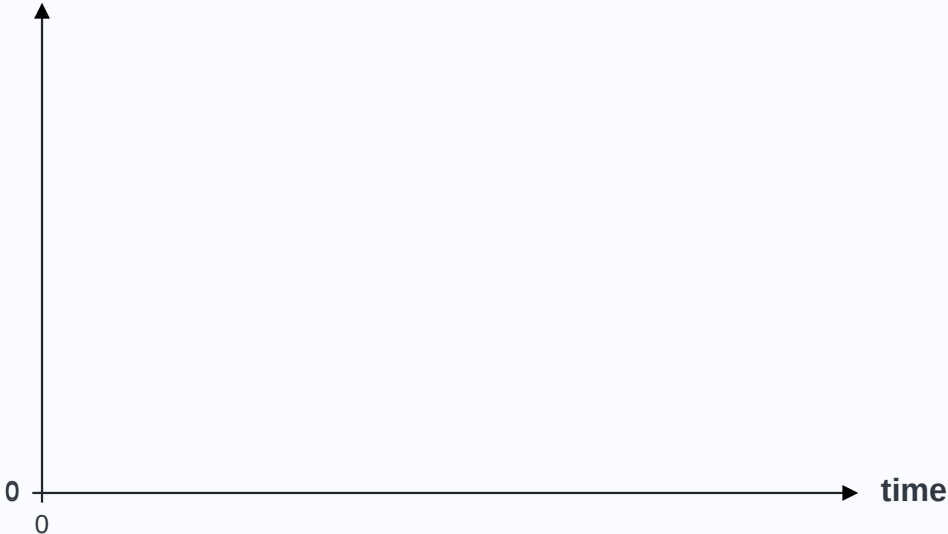
Four-quarter change in LFPR55

employment rate



Unemployment rate

inflation growth



Real disposable personal-income growth



Table 1. Variable definitions and summary statistics

Unit: U.S. quarterly data, 1995Q1-2015Q3,  $T = 83$ .

Variable	Definition	Mean	Std. dev.
$LFPR_{55}$	LFPR for U.S. residents aged 55 and over	36.2	3.8
$\Delta LFPR_{55}$	$LFPR_{55_t} - LFPR_{55_{t-1}}$	0.12	0.22
$\Delta_4 LFPR_{55}$	$LFPR_{55_t} - LFPR_{55_{t-4}}$	0.47	0.48
Unemployment Rate	Unemployed as percentage of labor force	6.0	1.7
Real Income	After-tax personal income, billions of 2010 dollars	9,989	1,616
$\Delta \ln(\text{RealIncome})$	Quarterly log change	0.0069	0.0099
Housing Starts	New home construction started in the quarter, millions	1.334	0.471
$\Delta \text{HousingStarts}$	First difference	-0.0038	0.084

Table 2. Forecasting regressions for  $\Delta_4 LFPR_{55}$

All regressions use quarterly data from 1995Q1 through 2007Q4. Standard errors are in parentheses. HR denotes heteroskedasticity-robust; HAC denotes heteroskedasticity- and autocorrelation-consistent.

Regressor/statistic	(1)	(2)	(3)	(4)	(5)
Standard errors/tests	HR	HAC	HAC	HAC	HAC
$\Delta LFPR_{55_{t-4}}$	0.138 (0.245)	0.138 (0.211)	-0.059 (0.149)	0.169 (0.244)	0.024 (0.313)
$\Delta LFPR_{55_{t-5}}$	0.380 (0.289)	0.380 (0.234)	0.216 (0.148)	0.409 (0.245)	0.212 (0.367)
$\Delta LFPR_{55_{t-6}}$	0.430+ (0.255)	0.430* (0.166)	0.269* (0.135)	0.458** (0.154)	0.270 (0.224)
Unemployment Rate $_{t-4}$	—	—	0.756** (0.217)	—	—
Unemployment Rate $_{t-5}$	—	—	-0.236 (0.332)	—	—

Regressor/statistic	(1)	(2)	(3)	(4)	(5)
Unemployment Rate <sub>t-6</sub>	—	—	-0.675** (0.215)	—	—
$\Delta \ln RealIncome_{t-4}$	—	—	—	-3.08 (3.16)	—
$\Delta \ln RealIncome_{t-5}$	—	—	—	-0.205 (6.79)	—
$\Delta \ln RealIncome_{t-6}$	—	—	—	7.65 (9.64)	—
$\Delta HousingStarts_{t-4}$	—	—	—	—	1.01** (0.30)
$\Delta HousingStarts_{t-5}$	—	—	—	—	-0.89 (0.61)
$\Delta HousingStarts_{t-6}$	—	—	—	—	0.51 (0.38)
Constant	0.497** (0.088)	0.497** (0.112)	1.447** (0.412)	0.444 (0.291)	-0.460 (0.712)
BIC	-1.641	-1.641	-2.039	-1.433	-1.537
Adjusted $R^2$	0.021	0.021	0.441	0.002	0.077
RMSFE, 2008Q1-2015Q3	0.558	0.558	0.685	0.554	0.381
Observations	52	52	52	52	52
$F$ : three lagged LFPR changes	1.51, $p = .224$	4.55, $p = .007$	3.33, $p = .028$	6.40, $p = .001$	2.33, $p = .087$
$F$ : three unemployment terms	—	—	11.76, $p = .000$	—	—
$F$ : three income-growth terms	—	—	—	1.20, $p = .320$	—
$F$ : three housing-start terms	—	—	—	—	4.21, $p = .010$

+, \*, and \*\* denote significance at 10%, 5%, and 1%.



1 — 4 points

Regressions (1) and (2) differ only in how standard errors are computed. Which method is preferred here, or does it not matter in theory? Explain.

2

You must forecast the change in LFPR55 from 2015Q3 to 2016Q3 using one of regressions (2)-(5).

1. **(4 points)** Which specification would you choose, and why?
2. **(4 points)** Give the standard error of the forecast. You need not compute the forecast itself.

3 — 8 points

Assess the hypothesis that the post-2010 plateau in LFPR55 was caused by the recession and slow recovery. Absent the recession, older workers might have continued postponing retirement; because of job losses and difficulty finding work, many may instead have retired earlier.

Using Figures 1-6 and Table 2, write one or two paragraphs evaluating whether the evidence supports, contradicts, or is insufficient to assess this hypothesis.

## QUESTION 2

### Part II: Bank lending and small-business profitability — 21 points

The Bank of Russia wants to estimate the effect of bank lending on small-business profitability. It has data on one hundred businesses over five years, described in the source as 50,000 observations:

- $P_{it}$ : profitability;
- $B_{it}$ : total bank loans received;
- $G_{it}$ : total government loans received;
- $R_{ji}$ : region dummies,  $j = 1, \dots, 85$ ;
- $L_{ki}$ : line-of-business dummies,  $k = 1, \dots, 20$ ;
- $U_t$ : unemployment;
- $D_i$ : distance to the nearest bank branch;
- $C_t$ : number of business closures in the city;
- $O_t$ : number of business openings in the city.

Business and year dummies and a time trend can also be created. The baseline regression is

$$P_{it} = \beta_0 + \beta_1 B_{it} + \beta_2 G_{it} + \sum_{j=2}^{85} \beta_{3,j} R_{ji} + \sum_{k=2}^{20} \beta_{4,k} L_{ki} + \beta_5 t + \varepsilon_{it}.$$

For each concern, explain briefly how you would address it.

1. **(3 points)** Only businesses surviving all five years are included.
2. **(3 points)** Larger bank loans are usually granted to businesses more likely to be profitable.
3. **(3 points)** A loan may have a positive long-run effect but initially be a burden because repayments begin before benefits are realized.
4. **(3 points)** Government-loan recipients may use bank loans more effectively because the government-loan program includes business education.
5. **(3 points)** Only businesses that received loans are in the data set.
6. **(3 points)** Doubling loan size more than doubles the expected benefit.
7. **(3 points)** Unobserved, time-invariant business characteristics, such as management quality, may affect both loan access and profitability.

### QUESTION 3

#### Part III: True, false, or uncertain — 24 points

For each statement, answer **true**, **false**, or **uncertain**, with an explanation.

**(a) (4 points)** In a probit or logit model, marginal effects on the probability that the dependent variable equals one are functions of  $X$ . Therefore, it is difficult to give  $\hat{\beta}$  alone a useful interpretation.

**(b) (4 points)** You want to estimate the supply equation for snow boots

$$Q_i = \gamma + \delta P_i + \psi C_i + u_i,$$

where  $Q_i$  is quantity sold,  $P_i$  is price, and  $C_i$  is transportation cost from a central production facility. Annual snowfall is proposed as an instrument for  $P_i$ . The fact that snowfall is correlated with  $Q_i$  proves that the instrument is invalid.

**(c) (4 points)** In a panel model where regressors are correlated with individual fixed effects, the fixed effects should be omitted to avoid multicollinearity.

**(d) (4 points)** The same instrument can be used for two different endogenous regressors in one regression.

**(e) (4 points)** OLS with cross-sectional data is bad because heteroskedasticity biases both slope estimates and standard errors.

**(f) (4 points)** It is possible to construct standard errors robust to unknown heteroskedasticity, but not standard errors robust to both heteroskedasticity and autocorrelation.

**QUESTION 4****Part IV: Miscellaneous problems — 35 points****1. Omitted variables and instruments — 15 points**

Suppose

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon,$$

and the main parameter of interest is  $\beta_1$ . Assume  $X_1$  and  $X_2$  are uncorrelated and, in fact,

$$\text{plim} \frac{1}{n} X_1' X_2 = 0.$$

- (5 points)** Without data on  $X_2$ , can you consistently estimate  $\beta_1$ ? Explain.
- (5 points)** Suppose  $X_1$  is endogenous and  $X_2$  remains unobserved. A proposed instrument  $W_1$  is a noisy linear combination of  $X_1$  and  $X_2$ , for example

$$W_1 = X_1 + 0.1X_2 + \eta, \quad \eta \sim N(0,1).$$

Discuss whether  $W_1$  is a suitable instrument and the properties of the IV estimator.

- (5 points)** If  $X_2$  is observed, is there a way to obtain a consistent estimate of  $\beta_1$ ? Give a specific procedure or explain why not.

**2. Cigarette-demand IV estimates — 20 points**

The observations are the 48 continental U.S. states. Let  $q_i$  denote cigarette sales,  $p_i$  a state price index, and  $I_i$  average income.

Using log state sales tax as an instrument for log price gives

$$\widehat{\ln q_i} = 9.43 - 1.14 \ln p_i + 0.21 \ln I_i,$$

with standard errors

$$(0.37) \quad (0.31)$$

under the coefficients on  $\ln p_i$  and  $\ln I_i$ .

- (2 points)** Construct a 95% confidence interval for the price elasticity.
- (2 points)** Test the significance of the income coefficient.

Adding log state cigarette tax as a second instrument gives

$$\widehat{\ln q_i} = 9.89 - 1.28 \ln p_i + 0.28 \ln I_i,$$

with standard errors

$$(0.25) \quad (0.25).$$

- (1 point)** Construct a 95% confidence interval for the price elasticity.
- (5 points)** Compare the  $\ln p_i$  coefficients across the two specifications to test instrument validity. You may use the fact that the variance of their difference can be estimated as the difference of their estimated variances, with the second specification having the lower variance.

Using a second year of data, first differences, and the differences of the same two instruments gives

$$\widehat{\ln(q_i/q_{i1})} = 0.37 - 1.20 \ln(p_i/p_{i1}) + 0.46 \ln(I_i/I_{i1}),$$

with standard errors

$$(0.25) \quad (0.25).$$

5. **(5 points)** Write a model in which this differenced IV estimator is consistent but the previous level estimators are not.
6. **(5 points)** Regressing  $\ln(p_i/p_{i1})$  on a constant,  $\ln(I_i/I_{i1})$ , and the two instruments gives an  $F$  statistic of 88.6, with 48 observations. Is it significant? Which of the two key IV conditions does this test confirm?