Financial Econometrics
Shanghai Jiao Tong University
Summer School 2026
Empirical Exercise 15.2
With Solutions

Fatih Kansoy

Reference: Stock, J.H. and Watson, M.W., Introduction to Econometrics (4th Edition, Global Edition)

Data: `us_macro_quarterly.csv`*. The PCE price index* `PCEP` *is the column* `PCECTPI` *and the CPI is* `CPIAUCSL`*. Use 1963:Q1–2012:Q4.*

Empirical Exercise 15.2. In the data file `us_macro_quarterly.csv`, you will find data on two aggregate price series for the United States: the price index for personal consumption expenditures (`PCEP`, the column `PCECTPI`) that you used in Empirical Exercise 14.1, and the Consumer Price Index (`CPI`, the column `CPIAUCSL`). These series are alternative measures of consumer prices in the United States. The CPI prices a basket of goods whose composition is updated every 5–10 years. `PCEP` uses chain-weighting to price a basket of goods whose composition changes from month to month. Economists have argued that the CPI will overstate inflation because it does not take into account the substitution that occurs when relative prices change. If this substitution bias is important, then average CPI inflation should be systematically higher than `PCEP` inflation. Let

$$\mathit{Infl}_t^{CPI} = 400\,[\ln(\texttt{CPI}_t) - \ln(\texttt{CPI}_{t-1})], \qquad \mathit{Infl}_t^{PCEP} = 400\,[\ln(\texttt{PCEP}_t) - \ln(\texttt{PCEP}_{t-1})],$$

and $Y_t = \mathit{Infl}_t^{CPI} - \mathit{Infl}_t^{PCEP}$, so Infl_t^{CPI} is the quarterly rate of price inflation (measured in percentage points at an annual rate) based on the CPI, Infl_t^{PCEP} is the quarterly rate of price inflation from the `PCEP`, and Y_t is their difference. Using data from 1963:Q1 through 2012:Q4, carry out the following exercises.

(a) Compute the sample means of $Infl_t^{CPI}$ and $Infl_t^{PCEP}$. Are these point estimates consistent with the presence of economically significant substitution bias in the CPI?

(b) Compute the sample mean of Y_t. Explain why it is numerically equal to the difference in the means computed in (a).

(c) Show that the population mean of Y is equal to the difference of the population means of the two inflation rates.

(d) Consider the "constant-term-only" regression: $Y_t = \beta_0 + u_t$. Show that $\beta_0 = E(Y)$. Do you think that u_t is serially correlated? Explain.

(e) Construct a 95% confidence interval for β_0. What value of the HAC standard error truncation parameter m did you choose? Why?

(f) Is there statistically significant evidence that the mean inflation rate for the CPI is greater than the rate for the `PCEP`?

(g) Is there evidence of instability in β_0? Carry out a QLR test. (Hint: Make sure you use HAC standard errors for the regressions in the QLR procedure.)

Solution:

Data Overview

The data file `us_macro_quarterly.csv` contains quarterly observations on two US consumer price indices. We use the column `CPIAUCSL` as the CPI and the column `PCECTPI` as the PCE price index (`PCEP`). Both series are levels (price indices), so we transform them into quarterly inflation rates, measured in percentage points at an annual rate, before any analysis.

Quantity	Definition	Role
$Infl_t^{CPI}$	$400\,[\ln(\texttt{CPI}_t) - \ln(\texttt{CPI}_{t-1})]$	CPI inflation (annualised %)
$Infl_t^{PCEP}$	$400\,[\ln(\texttt{PCEP}_t) - \ln(\texttt{PCEP}_{t-1})]$	PCEP inflation (annualised %)
Y_t	$Infl_t^{CPI} - Infl_t^{PCEP}$	Inflation gap (CPI less PCEP)

Sample: 1963:Q1 through 2012:Q4. The factor 400 turns a quarterly log change into an annualised percentage rate (multiply by 100 to make it a percentage, and by 4 to annualise).

(a) Sample means of CPI and PCEP inflation

Computing the sample averages of the two annualised inflation series over 1963:Q1–2012:Q4 gives:

Series	Sample mean (% at annual rate)
Infl_t^{CPI}	3.81
Infl_t^{PCEP}	3.35
Difference	0.46

Interpretation: On average, the CPI records inflation that is about 0.5 percentage points per year higher than the PCEP. This gap is large enough to matter economically. Compounded over a twenty-year horizon, the price level rises by

$$(1.0381)^{20} \approx 2.1 \text{ times}$$

using the CPI inflation rate, but only by

$$(1.0335)^{20} \approx 1.9 \text{ times}$$

using the PCEP inflation rate. A difference of this size over two decades changes cost-of-living adjustments, indexed contracts, and real-income calculations in a material way. The point estimates are therefore consistent with an economically significant substitution bias in the CPI.

(b) Sample mean of Y_t

The sample mean of the inflation gap is

$$\boxed{\overline{Y} = 0.47}$$

(measured in percentage points at an annual rate). This is, up to rounding, exactly the difference between the two means in part (a), $3.81 - 3.35 = 0.46$.

Why this must hold: The sample mean is a linear operator. Because $Y_t = \mathit{Infl}_t^{CPI} - \mathit{Infl}_t^{PCEP}$ is defined observation by observation,

$$\overline{Y} = \frac{1}{T}\sum_{t=1}^{T} Y_t = \frac{1}{T}\sum_{t=1}^{T}\left(\mathit{Infl}_t^{CPI} - \mathit{Infl}_t^{PCEP}\right) = \frac{1}{T}\sum_{t=1}^{T}\mathit{Infl}_t^{CPI} - \frac{1}{T}\sum_{t=1}^{T}\mathit{Infl}_t^{PCEP} = \overline{\mathit{Infl}^{CPI}} - \overline{\mathit{Infl}^{PCEP}}.$$

The mean of a difference equals the difference of the means, so the two ways of computing the gap must give the same number. (The small discrepancy between 0.47 and 0.46 is only rounding of the individual means.)

(c) Population mean of Y

The same linearity holds for the population (expectation) operator. Since $Y = \mathit{Infl}^{CPI} - \mathit{Infl}^{PCEP}$,

$$\boxed{E(Y) = E\big(\mathit{Infl}^{CPI} - \mathit{Infl}^{PCEP}\big) = E\big(\mathit{Infl}^{CPI}\big) - E\big(\mathit{Infl}^{PCEP}\big)\,.}$$

The expectation is a linear operator, so the population mean of the gap is exactly the difference of the population means of the two inflation rates. Part (b) is the sample analogue of this population statement.

(d) The constant-term-only regression

Consider the regression with only an intercept:

$$Y_t = \beta_0 + u_t.$$

Take expectations of both sides. By the least-squares assumption $E(u_t) = 0$,

$$E(Y_t) = \beta_0 + E(u_t) = \beta_0 + 0 = \beta_0, \qquad \Longrightarrow \qquad \boxed{\beta_0 = E(Y).}$$

So the intercept in the constant-only regression is just the population mean of Y, and its OLS estimator $\hat{\beta}_0$ is the sample mean $\overline{Y}$.

Is u_t serially correlated? Here $u_t = Y_t - E(Y)$ is the deviation of the inflation gap from its mean. The factors that drive a wedge between CPI and PCEP inflation (the substitution bias, differences in basket composition and weighting, and the way each index treats specific categories such as housing and medical care) do not switch on and off from one quarter to the next. They persist, so Y_t and therefore u_t are serially correlated. This matters for inference: with serial correlation we must use heteroskedasticity- and autocorrelation-consistent (HAC) standard errors rather than the usual homoskedasticity-only or heteroskedasticity-robust ones.

(e) 95% confidence interval for β_0 and choice of m

The OLS estimate of the intercept is $\hat{\beta}_0 = 0.47$ with a HAC standard error of $\mathrm{SE}(\hat{\beta}_0) = 0.09$. A 95% confidence interval is

$$\hat{\beta}_0 \pm 1.96 \times \mathrm{SE}(\hat{\beta}_0) = 0.47 \pm 1.96 \times 0.09,$$

that is,

$$\boxed{0.30 \leq \beta_0 \leq 0.64 \text{ (percentage points at an annual rate).}}$$

Choice of the truncation parameter m: The HAC (Newey–West) standard error needs a bandwidth (truncation parameter) m that says how many lags of autocovariance to include. A common rule of thumb is

$$m = 0.75\, T^{1/3},$$

rounded up to the nearest integer. With T quarterly observations from 1963:Q1 to 2012:Q4, this rule gives

$$m = 5.$$

Because u_t is serially correlated (part (d)), using HAC errors with this bandwidth gives a confidence interval that is valid in the presence of that autocorrelation, whereas the homoskedasticity-only interval would understate the true sampling uncertainty.

(f) Is CPI inflation significantly greater than PCEP inflation?

Yes. The parameter β_0 is the difference in the mean inflation rates (parts (c) and (d)). The confidence interval from part (e) runs from about 0.30 to 0.64 percentage points and lies entirely above zero. Equivalently, the t-statistic for $H_0 : \beta_0 = 0$ is

$$t = \frac{\hat{\beta}_0}{\mathrm{SE}(\hat{\beta}_0)} = \frac{0.47}{0.09} \approx 5.2,$$

which is far beyond the 5% critical value of 1.96. We therefore reject the null that the two indices have the same mean inflation rate. There is statistically significant evidence that average CPI inflation exceeds average PCEP inflation, by somewhere between roughly 0.3 and 0.6 percentage points per year. This is the empirical signature of the substitution bias discussed in the question.

(g) QLR test for instability in β_0

To check whether the mean inflation gap is stable over the sample, we run a Quandt likelihood ratio (QLR) test on the constant-only regression, using HAC standard errors in the underlying regressions as the hint instructs.

Quantity	**Value**
QLR statistic	7.88
Date of maximum	2008:Q2
10% critical value	7.12
5% critical value	8.68

Conclusion: The QLR statistic of 7.88 exceeds the 10% critical value (7.12) but falls short of the 5% critical value (8.68). There is therefore statistically significant evidence of a break in β_0 at the 10% level, but not at the 5% level. The evidence for instability is suggestive rather than decisive.

Reading the timing: The largest value of the test sequence occurs around 2008:Q2, close to the inflation outlier in 2008:Q4 that was uncovered in Empirical Exercise 15.1. The global financial crisis and the sharp swing in commodity (especially energy) prices in 2008 affected the CPI and the PCEP differently, which plausibly nudged the mean gap during that episode. So the modest instability the QLR test detects is tied to a recognisable economic event rather than to a smooth, long-run drift.

Summary: Key Takeaways

Part	Result	Interpretation
(a)	$\overline{Infl^{CPI}} = 3.81$, $\overline{Infl^{PCEP}} = 3.35$	Gap $\approx$ 0.5 pp; economically large
(b)	$\overline{Y} = 0.47$	Mean of difference = difference of means
(c)	$E(Y) = E(Infl^{CPI}) - E(Infl^{PCEP})$	Expectation is linear
(d)	$\beta_0 = E(Y)$; u_t serially correlated	Persistent wedge $\Rightarrow$ use HAC
(e)	$0.47 \pm 1.96(0.09)$; $m = 5$	95% CI: 0.30 to 0.64
(f)	$t \approx 5.2$; reject H_0	CPI inflation significantly higher
(g)	QLR = 7.88 at 2008:Q2	Break at 10% only; tied to 2008

The Big Picture

This exercise illustrates several themes in the analysis of time-series data:

1. **Linearity of means:** The sample mean and the expectation are linear operators, so the mean of a difference is the difference of the means (parts (b) and (c)). The constant-only regression simply re-expresses the sample mean as a regression intercept (part (d)).

2. **Serial correlation demands HAC inference:** When the regression error is serially correlated, conventional standard errors are wrong. The Newey–West HAC standard error, with a bandwidth chosen by the $0.75\,T^{1/3}$ rule, delivers valid confidence intervals and tests (parts (d) and (e)).

3. **Economic significance, not just statistical:** The CPI–PCEP gap of about half a percentage point a year is both statistically significant (part (f)) and economically meaningful, since it compounds into a sizeable difference in the price level over two decades (part (a)). This is the measurable footprint of substitution bias.

4. **Stability is not automatic:** The QLR test (part (g)) finds borderline evidence of a break around 2008, reminding us to check whether a relationship that we treat as constant actually holds across the whole sample, especially around major macroeconomic episodes.

Calculations performed using Stata. See accompanying `E15_2_solution.do` *file.*