Financial Econometrics
Shanghai Jiao Tong University
Summer School – 2026
Empirical Exercise 15.1
With Solutions

Fatih Kansoy

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

Data: `us_macro_monthly.csv` (1955:M1–2017:M12). The Index of Industrial Production `IP` is the column `INDPRO`; the oil-price shock `Ot` (the net oil price increase) is constructed from the crude-oil price series `WPU0561` (PPI: crude petroleum). Use the sample 1960:M1–2012:M12.

Empirical Exercise 15.1. In this exercise you will estimate the effect of oil prices on macroeconomic activity, using monthly data on the Index of Industrial Production (`IP`) and the monthly oil-price shock `Ot` described below.

This exercise uses `us_macro_monthly.csv`, covering 1955:M1–2017:M12. Take `IP` to be the column `INDPRO` (the Index of Industrial Production). The oil-price shock `Ot` is the *net oil price increase*: it equals the greater of zero or the percentage point difference between the oil price at date t and its maximum value over the past 12 months, where the oil price is the crude-petroleum producer price series `WPU0561`. Throughout, restrict attention to the sample period 1960:M1–2012:M12.

(a) Compute the monthly growth rate in `IP`, expressed in percentage points, $ip\_growth_t = 100 \times [\ln(IP_t) - \ln(IP_{t-1})]$. What are the mean and standard deviation of `ip_growth` over the 1960:M1–2012:M12 sample period? What are the units for `ip_growth` (percent, percent per annum, percent per month, or something else)?

(b) Plot the value of `Ot`. Why are so many values of `Ot` equal to zero? Why aren't some values of `Ot` negative?

(c) Estimate a distributed lag model by regressing ip_growth onto the current value and 18 lagged values of Ot, including an intercept. What value of the HAC standard error truncation parameter m did you choose? Why?

(d) Taken as a group, are the coefficients on Ot statistically significantly different from zero?

(e) Construct graphs showing the estimated dynamic multipliers, cumulative multipliers, and 95% confidence intervals. Comment on the real-world size of the multipliers.

(f) Suppose that high demand in the United States (evidenced by large values of ip_growth) leads to increases in oil prices. Is Ot exogenous? Are the estimated multipliers shown in the graphs in (e) reliable? Explain.

Solution:

Data Overview

The analysis uses monthly United States macroeconomic data, us_macro_monthly.csv (1955:M1–2017:M12), restricted to the sample 1960:M1–2012:M12. Two series are central:

Variable	**Source column**	**Construction**
IP	INDPRO	Index of Industrial Production (level)
ip_growth	—	$100 \times [\ln(IP_t) - \ln(IP_{t-1})]$
Oil price	WPU0561	PPI: crude petroleum
Ot	—	Net oil price increase (see below)

Constructing the net oil price increase Ot. Following Hamilton, the regressor is not the raw change in oil prices but the *net* increase relative to the recent peak:

$$O_t = \max\Big(0,\ 100\ln(\mathit{Oilprice}_t) - \max_{1\le j\le 12} 100\ln(\mathit{Oilprice}_{t-j})\Big).$$

In words, Ot is the percentage point amount by which the current (log) oil price exceeds its highest value over the previous 12 months. When the oil price is at or below its trailing one-year peak, Ot $= 0$. The idea is that oil prices damage the economy only when they jump above their values in the recent past, so price recoveries after a fall do not count as new shocks.

Note: The standard errors throughout are heteroskedasticity- and autocorrelation-consistent (HAC), because the distributed lag error term is serially correlated.

Key Notation and Formulas

The distributed lag model with the current value and $r = 18$ lags of `Ot` is

$$ip\_growth_t = \beta_0 + \beta_1 O_t + \beta_2 O_{t-1} + \cdots + \beta_{19} O_{t-18} + u_t.$$

The estimated coefficients $\hat{\beta}_1, \ldots, \hat{\beta}_{19}$ are the **dynamic multipliers**: $\hat{\beta}_{h+1}$ is the effect on `ip_growth` of a one percentage point increase in the net oil price h months earlier. The **cumulative dynamic multiplier** at horizon h is the running sum

$$\widehat{\mathrm{Cum}}_h = \hat{\beta}_1 + \hat{\beta}_2 + \cdots + \hat{\beta}_{h+1},$$

which measures the total effect on the *level* of industrial production after h months of a permanent one percentage point rise in the net oil price.

(a) Mean and Standard Deviation of `ip_growth`

Over the 1960:M1–2012:M12 sample, the monthly growth rate of industrial production has:

Statistic	**Value**
Mean	0.21
Standard deviation	0.75

Units. Because $ip\_growth_t = 100 \times [\ln(IP_t) - \ln(IP_{t-1})]$ uses the change in the log index from one month to the next, the units are **percent per month**. A mean of 0.21 therefore says that industrial production grew on average by about 0.21% each month over this period, which is roughly $0.21 \times 12 \approx 2.5\%$ per year, in line with long-run United States output growth.

(b) Plotting the Net Oil Price Increase `Ot`

[See Figure: Time-series plot of the net oil price increase `Ot`, 1960:M1–2012:M12]

Why are so many values of `Ot` equal to zero? By construction, `Ot` equals zero whenever the current oil price is *not* above its maximum value over the previous 12 months. In most months the oil price sits at or below its recent one-year peak (prices are flat, falling, or merely recovering toward an earlier high), so `Ot` takes the value zero. Positive values appear only in the relatively few months when the oil price climbs to a genuinely new 12-month high, which is why the series is zero most of the time with occasional spikes.

Why aren't some values of `Ot` negative? The definition takes the *greater* of zero and the gap between the current price and the trailing peak:

$$O_t = \max\big(0,\ \text{price}_t - \text{12-month peak}\big) \geq 0.$$

The outer $\max(0, \cdot)$ censors all negative values at zero. When the price falls below its recent peak the raw gap is negative, but `Ot` is set to zero rather than recording a fall. This encodes Hamilton's idea that only *net increases* above recent experience disturb the economy, so price declines are deliberately not treated as shocks.

(c) Distributed Lag Regression and the Truncation Parameter

Model. Regress `ip_growth` on the current value and 18 lags of `Ot`, with an intercept, giving 19 oil-price coefficients plus the constant.

Choice of the HAC truncation parameter m. Because the error term in a distributed lag regression is serially correlated, the standard errors must be HAC (Newey–West). The truncation parameter m is chosen using the benchmark rule

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

rounded up to the nearest integer. With the sample used here this rule gives

$$m = 7.$$

This value balances the two failure modes of HAC estimation: too few autocovariances ignore genuine serial correlation, while too many introduce so much estimation noise that the variance estimator becomes inconsistent. The rule scales m gently with the sample size, here delivering $m = 7$.

Estimated coefficients. The 19 estimated dynamic multipliers (the coefficients on `Ot` and its 18 lags), together with their 95% confidence intervals, are displayed graphically in part (e).

(d) Joint Significance of the Oil Price Coefficients

We test the null hypothesis that all 19 coefficients on `Ot` and its lags are jointly equal to zero, using a HAC F-test:

Test	**Value**
HAC F-statistic (19 coefficients = 0)	2.74
p-value	0.00

Conclusion. The p-value is essentially zero, so the coefficients are jointly **statistically significant at the 5% level**. They are, however, **not significant at the 1% level** once the conventional critical values for this test are applied. Taken as a group the net oil price increase therefore has a detectable, though not overwhelmingly strong, association with subsequent industrial production growth.

(e) Dynamic and Cumulative Multipliers

[See Figure: Estimated dynamic multipliers with 95% confidence intervals]

[See Figure: Estimated cumulative multipliers with 95% confidence intervals]

What the multipliers show. The cumulative multipliers reveal a **persistent and large decline in industrial production** following an increase in oil prices above their previous 12-month peak. The effect builds up over many months rather than appearing all at once, and it does not reverse within the horizon considered.

Real-world size. Reading the cumulative multiplier at the 18-month horizon:

A 100% increase in oil prices is associated with about a 21% decline in industrial production after 18

This is an economically substantial effect. A doubling of oil prices above the recent peak is followed, over the course of roughly a year and a half, by a fall in the level of industrial production of around one fifth. The size and persistence of this response are consistent with the long-standing view that sharp oil price increases have contributed to several post-war recessions in oil-importing economies.

Summary: The dynamic multipliers accumulate into a deep, slow-building, and persistent contraction in industrial production after a net oil price increase, peaking at roughly a 21% decline (per 100% oil price rise) about 18 months out.

(f) Is `Ot` Exogenous? Are the Multipliers Reliable?

The exogeneity concern. The distributed lag estimator delivers consistent estimates of dynamic causal effects only if `Ot` is exogenous, that is, if the regression error u_t has conditional mean zero given current and past values of `Ot`. The scenario in the question breaks exactly this condition: if strong United States demand (high `ip_growth`) itself pushes up oil prices, then causation runs in *both* directions. Oil prices affect output, but output also affects oil prices.

Why this matters here. When buoyant domestic activity raises the oil price, the net oil price increase `Ot` becomes correlated with the very demand factors contained in the error term u_t. The regressor is then **endogenous** rather than exogenous, because high-demand months tend to coincide with both high `ip_growth` and rising oil prices. This is a simultaneous-causality (reverse-causation) problem.

> **Conclusion:** In this case `Ot` is **not exogenous**, so the OLS distributed lag estimates are not consistent estimates of the dynamic causal effect of oil prices on output. The multipliers summarised in part (e), and the conclusions drawn from them, are therefore **not reliable** as causal statements.

Summary: Key Takeaways

Part	Result	Interpretation
(a) `ip_growth` stats	Mean 0.21, SD 0.75	Percent per month
(b) Shape of `Ot`	Many zeros, no negatives	Censored net increase
(c) Truncation parameter	$m = 0.75\,T^{1/3} = 7$	HAC (Newey–West)
(d) Joint test	$F = 2.74$, $p = 0.00$	Significant at 5%, not 1%
(e) Cumulative effect	-21% at 18 months	Large, persistent decline
(f) Exogeneity	`Ot` not exogenous	Multipliers not reliable

The Big Picture

This exercise illustrates several core ideas in the estimation of dynamic causal effects:

1. **Distributed lag models capture timing.** The response of output to an oil price shock is spread over many months. A single contemporaneous coefficient would miss the slow build-up that the 18 lags reveal.

2. **Cumulative multipliers measure total effect.** Summing the dynamic multipliers gives the effect on the *level* of industrial production. Here that cumulative effect reaches roughly -21% (per 100% oil price rise) after 18 months.

3. **HAC standard errors are essential.** Because the error term is serially correlated, ordinary OLS standard errors would be misleading. The Newey–West truncation rule $m = 0.75\,T^{1/3}$ (here $m = 7$) provides valid inference.

4. **Exogeneity is the decisive assumption.** If high domestic demand drives oil prices, then `Ot` is endogenous and the estimated multipliers no longer have a causal interpretation. Statistical significance does not rescue a regression whose key regressor is correlated with the error term.

Calculations performed using Stata. See accompanying ***`E15_1_solution.do`*** *file.*