Econometrics Homework Assignment II

Golkar Arno

January 4, 2026

Introduction

This document contains the answers to the questions regarding serial autocorrelation and endogeneity, specifically testing the Environmental Kuznets Curve (EKC) hypothesis.

Question 1

We begin by exploring the data for the relationship between CO2 emissions per capita and GDP per capita. Create three separate scatter plots using data from France and two other countries (from different continents) of your choice. Overlay each scatter plot with a fitted polynomial curve. Briefly comment on the observed relationship.

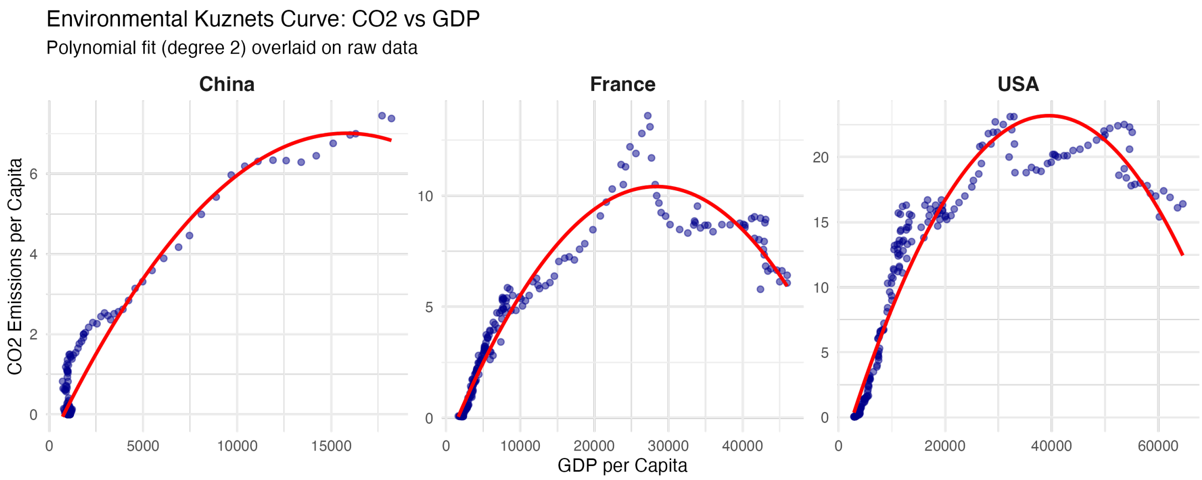

Figure 1: Relationship between CO2 emissions and GDP per capita for France, USA, and China. The red line represents a fitted quadratic polynomial curve.

Comments on observed relationships

Figure 1 displays the relationship between GDP per capita and CO2 emissions per capita for France (Europe), the USA (North America), and China (Asia).

- **France:** The data exhibits the classic downward slope associated with the later stages of the Environmental Kuznets Curve (EKC). As a developed economy with high GDP per capita, France shows a clear negative correlation, where further economic growth is associated with lower emissions.

- **USA:** The United States displays a distinct inverted U-shape. At lower income levels (historically), emissions rose rapidly with GDP. However, after reaching a turning point (the peak of the curve), emissions have started to decline despite continued economic growth, consistent with the EKC hypothesis.

- **China:** In contrast, China shows a strong positive relationship. As an emerging and rapidly industrializing economy, its growth in GDP per capita has been accompanied by a steep increase in CO2 emissions. The curve suggests China is still largely on the upward-sloping portion of the EKC, though the quadratic fit may indicate a potential flattening at the very top of the current data range.

Question 2

Create a table of descriptive statistics for all continuous variables in the dataset for France. Briefly comment on the sample.

Table 1: Descriptive Statistics (France)

Statistic	N	Mean	St. Dev.	Min	Max
CO2	223	4.03	3.47	0.07	13.60
GDP	223	12,435.11	13,685.89	1,640	46,000
POP	223	42,977,578.00	10,002,677.00	29,000,000	66,300,000
IND	63	23.46	4.80	16.10	30.20
URB	63	74.33	4.51	61.90	81.50
ENG	63	3,460.48	714.69	1,670	4,330
MOT	6	596.17	2.93	593	601
INT	33	46.41	34.92	0.05	86.10
CHI	223	162.49	122.63	3.91	412.00
DIS	223	4.78	0.00	4.78	4.78

Comments on the descriptive statistics for France

Table 1 presents the summary statistics for France. We can observe several distinct patterns characteristic of a long-term panel dataset:

- **Time-invariant factors:** The variable **DIS** (Distance from the equator) has a standard deviation of **0**. This is expected, as geographic location is a fixed country characteristic that does not change over time.

- **Technological adoption:** Variables such as **INT** (Internet usage) and **MOT** (Motor vehicles) show large standard deviations relative to their means. The minimum value for both is 0, reflecting that these technologies did not exist for the majority of the sample period (19th and early 20th century) and only appeared in later years.

- **Economic transition: GDP** per capita shows a large standard deviation, illustrating the substantial economic growth France experienced over the last two centuries. Similarly, **IND** (Industry share) and **URB** (Urbanization) exhibit significant variation, capturing France's structural transformation from an agrarian society to an industrial and urbanized economy.

- **Health outcomes:** The variable **CHI** (Child Mortality) likely shows a high maximum value (from the 1800s) and a much lower minimum (modern day), reflecting the dramatic improvements in public health and life expectancy over the period.

Question 3

Estimate model M1 using OLS and report your results. Do you find evidence for an inverted-U relationship between ln(CO2) and ln(GDP)? Compute the implied turning point of the EKC, interpret it in terms of GDP per capita, and compare it to the scatter plot from Question 1.

We estimate the following regression model using OLS:

$$\ln(\text{CO}_2)_t = \beta_0 + \beta_1 \ln(\text{GDP})_t + \beta_2 \ln(\text{GDP})_t^2 + \beta_3 \ln(\text{POP})_t + \eta_t \tag{1}$$

Results

The estimation results are presented in Table 2.

Table 2: Regression Results (Model M1)

	Dependent variable:
	ln_CO2
ln_GDP	4.052***
	(0.083)
ln_GDP2	−0.130***
	(0.005)
ln_POP	−0.013***
	(0.004)
Constant	−25.296***
	(0.335)
Observations	41,011
R^2	0.656
Adjusted R^2	0.656
Residual Std. Error	1.577 (df = 41007)
F Statistic	26,079.980*** (df = 3; 41007)
Note:	*p<0.1; **p<0.05; ***p<0.01

Evidence for an inverted-U relationship

The Environmental Kuznets Curve (EKC) hypothesis posits an inverted-U relationship between emissions and income. Mathematically, this requires:

- $\beta_1 > 0$ (emissions initially rise with income)
- $\beta_2 < 0$ (the curve eventually bends downwards)

Based on the results in Table 2, we find:

$$\hat{\beta}_1 = 4.052 \quad \text{and} \quad \hat{\beta}_2 = -0.130$$

Since $\hat{\beta}_1$ is positive and statistically significant, and $\hat{\beta}_2$ is negative and statistically significant, we find strong evidence supporting the inverted-U relationship between $\ln(\text{CO}_2)$ and $\ln(\text{GDP})$.

Implied turning point

The turning point of the EKC occurs where the derivative of $\ln(\text{CO}_2)$ with respect to $\ln(\text{GDP})$ is zero:

$$\frac{\partial \ln(\text{CO}_2)}{\partial \ln(\text{GDP})} = \hat{\beta}_1 + 2\hat{\beta}_2 \ln(\text{GDP}) = 0$$

$$\ln(\text{GDP})^* = -\frac{\hat{\beta}_1}{2\hat{\beta}_2}$$

Substituting the estimated coefficients:

$$\ln(\text{GDP})^* = -\frac{4.051649}{2(-0.129650)} \approx 15.625$$

To find the GDP level at the turning point, we exponentiate this value:

$$\text{GDP}^* = \exp(15.625) \approx 6,109,082$$

The model implies that CO_2 emissions begin to decline once GDP reaches approximately 6.1 million (in the currency units of the dataset).

Question 4

We suspect that model M_1 suffers from non-spherical disturbances. Specifically, we assume that $\mathbb{E}[\eta_t \eta_s \mid X] \neq 0$, for some $t \neq s$, where X includes all regressors.

4A

What does this assumption mean in the context of our model? Discuss the implications of this assumptions for the variance-covariance matrix of the error term and for the validity of the results reported in Question 3.

The assumption $E[\eta_t \eta_s | X] \neq 0$ for some $t \neq s$ indicates the presence of serial correlation (or autocorrelation) in the error terms. In the context of our time-series model for CO_2 emissions, this means that the unobserved factors affecting emissions in period t are correlated with the unobserved factors in period s.

Contextual meaning: This is highly likely in environmental and economic time-series data. For instance, if a technological shock, a policy change, or a weather event affects emissions in year t, its effect is likely to persist or influence emissions in year $t + 1$. Inertia in economic and ecological systems implies that "shocks" are rarely isolated to a single time period.

Implications for the variance-covariance matrix: Under the standard OLS assumptions (spherical disturbances), the variance-covariance matrix of the errors is diagonal: $Var(\eta|X) = \sigma^2 I$. Under the assumption of serial correlation, this matrix becomes non-diagonal. The off-diagonal elements (covariances between different time periods) are non-zero.

Implications for the validity of results in Question 3:

- **Unbiasedness and consistency:** The OLS estimators $\hat{\beta}$ calculated in Question 3 remain unbiased and consistent. The point estimates for the turning point and coefficients are still valid in terms of their central tendency.
- **Efficiency:** OLS is no longer the Best Linear Unbiased Estimator (BLUE). Estimators that account for the correlation structure (like Generalized Least Squares) would be more efficient.
- **Inference:** The most critical implication is that the standard errors reported in Question 3 are biased and inconsistent. Consequently, the t-statistics and p-values are unreliable. In the presence of positive serial correlation, OLS standard errors are typically underestimated, leading to inflated t-statistics. Thus, we might have incorrectly concluded that coefficients were highly significant when the evidence is actually weaker.

4B

Explain the underlying logic of the Breusch-Godfrey test. Formulate the null and alternative hypotheses for an AR(6) error process and the decision rule. Perform the test and discuss the results.

Underlying logic: The Breusch-Godfrey (BG) test is a Lagrange Multiplier (LM) test used to assess the validity of the assumption of no serial correlation. Unlike the Durbin-Watson test, the BG test is valid in the presence of higher-order autoregressive processes and lagged dependent variables.

The test involves running an auxiliary regression of the residuals ($\hat{\eta}_t$) from the original model (Equation 1) on the original regressors (X_t) and the lagged residuals ($\hat{\eta}_{t-1}, \hat{\eta}_{t-2}, \ldots, \hat{\eta}_{t-p}$). The logic is that if serial correlation exists, the lagged residuals should significantly help predict the current residual.

Hypotheses for an AR(6) process: We test for serial correlation up to order $p = 6$.

- **Null hypothesis (H_0):** $\rho_1 = \rho_2 = \cdots = \rho_6 = 0$. (There is no serial correlation of any order up to 6).
- **Alternative hypothesis (H_1):** At least one $\rho_j \neq 0$ for $j \in \{1, \ldots, 6\}$. (Serial correlation exists).

Decision rule: The test statistic is $LM = (N - p)R^2_{aux}$, which asymptotically follows a Chi-squared distribution with p degrees of freedom (χ^2_6) under the null hypothesis.

- If the p-value $< \alpha$ (typically 0.05), we reject H_0.
- If the p-value $\geq \alpha$, we fail to reject H_0.

Test results and discussion: We performed the Breusch-Godfrey test with order 6. The results are:

- LM Statistic: 40049.51
- Degrees of Freedom: 6
- P-value: $< 2.2 \times 10^{-16}$

Since the p-value is significantly less than 0.05, we reject the null hypothesis. This provides strong statistical evidence of the presence of serial correlation in the error terms. This confirms our suspicion in Part (a) that the model suffers from non-spherical disturbances, rendering the standard errors and inference from Question 3 unreliable.

4C

Re-estimate model M1 using Newey-West standard errors. Compare the estimates and standard errors to those from Question 3, and discuss any differences.

We re-estimate the standard errors using the Newey-West Heteroskedasticity and Autocorrelation Consistent (HAC) covariance matrix estimator. This method corrects the standard errors for both serial correlation and heteroskedasticity, providing valid inference without altering the coefficient estimates.

Comparison of results:

- **Estimates ($\hat{\beta}$):** As expected, the coefficient estimates remain **identical** to those in Question 3. OLS consistency is not affected by serial correlation, so the point estimates ($\hat{\beta}_1 = 4.052$, $\hat{\beta}_2 = -0.130$) and the implied turning point are unchanged.
- **Standard errors:** Comparing the OLS standard errors (from Q3) with the Newey-West standard errors:
  - SE($\hat{\beta}_1$): OLS = 0.083 vs. NW = 0.673 (Ratio: 8.11)
  - SE($\hat{\beta}_2$): OLS = 0.005 vs. NW = 0.039 (Ratio: 7.77)
  - SE($\hat{\beta}_3$): OLS = 0.004 vs. NW = 0.054 (Ratio: 15.15)
- The Newey-West standard errors are significantly larger than the OLS standard errors, with increases ranging from approximately 8 to 15 times the original size. This confirms that the OLS standard errors were severely underestimated due to positive serial correlation.

  Consequently, the corrected t-statistics are much lower. While the GDP coefficients remain statistically significant (t-value for $\ln(GDP)$ drops from 48.9 to 6.02, and for $\ln(GDP)^2$ from -26.0 to -3.35), the coefficient for population ($\ln(POP)$) is no longer statistically significant (p-value = 0.817). This drastic change highlights the importance of correcting for serial correlation to avoid spurious inference.

Consider now a cross-sectional regression for 2007 across all countries, where i denotes a country:

$$\mathit{ln(CO2)}_i = \beta_0 + \beta_1\, \mathit{ln(GDP)}_i + \beta_2\, \mathit{ln(GDP)}_i^2 + \beta_3\, \mathit{ln(POP)}_i + \gamma' X_i + \eta_i, \quad (M_2)$$

where X_i includes the control variables IND, URB, ln(ENG), and ln(MOT). For the subsequent five questions we assume that A^{OLS}_4a and A^{OLS}_4b hold.

Question 5

Remove all observations with missing values for any of the variables listed on page 2. Then, estimate model M_2 using OLS and report your results. Interpret all statistically significant coefficients.

We estimate the (M2) cross-sectional regression model for the year 2007.

Results

The OLS estimation results are reported in Table 3.

Table 3: Cross-Sectional Regression Results (2007)

	Dependent variable:
	ln_CO2
ln_GDP	2.498***
	(0.510)
ln_GDP2	−0.106***
	(0.028)
ln_POP	−0.001
	(0.022)
IND	−0.001
	(0.003)
URB	0.005
	(0.003)
ln_ENG	0.466***
	(0.079)
ln_MOT	0.179***
	(0.055)
Constant	−17.356***
	(2.359)
Observations	130
R^2	0.929
Adjusted R^2	0.925
Residual Std. Error	0.421 (df = 122)
F Statistic	227.669*** (df = 7; 122)
Note:	*p<0.1; **p<0.05; ***p<0.01

Interpretation of Significant Coefficients

We interpret the coefficients that are statistically significant ($p < 0.05$).

- **GDP and the EKC:** The coefficients for income are $\hat{\beta}_1 = 2.498$ and $\hat{\beta}_2 = -0.106$. Both are statistically significant at the 1% level. Since $\hat{\beta}_1 > 0$ and $\hat{\beta}_2 < 0$, we find strong evidence for an inverted-U relationship (EKC) in the 2007 cross-section.
- **Energy Use (ln(ENG)):** The coefficient is 0.466 and is statistically significant at the 1% level. This suggests that a 1% increase in energy usage per capita is associated with approximately a 0.47% increase in CO_2 emissions, holding other factors constant.
- **Motor Vehicles (ln(MOT)):** The coefficient is 0.179 and is statistically significant at the 1% level. This implies that a 1% increase in the number of motor vehicles is associated with approximately a 0.18% increase in CO_2 emissions.
- **Insignificant variables:** The coefficients for Population (ln(POP)), Industry Share (IND), and Urbanization (URB) are not statistically significant ($p > 0.10$). Thus, in this specific cross-sectional specification for 2007, we do not find evidence that these factors significantly explain variations in CO_2 emissions once income, energy, and transport are accounted for.

Question 6

Report the number of observations used in the estimation and briefly discuss any potential implications of this sample size.

The estimation of Model M2 for the year 2007 utilized 130 observations.

Potential implications of this sample size:

1. **Degrees of freedom:** With $N = 130$ observations and 8 parameters to estimate (7 independent variables + 1 intercept), the regression has 122 degrees of freedom. This provides a ratio of approximately 16 observations per parameter. This is generally considered a sufficient sample size to perform OLS estimation without running into severe overfitting problems, allowing for reliable estimation of the coefficients.
2. **Sample selection bias:** The original dataset likely contains data for nearly all countries in the world over many years (panel data). By restricting the sample to a single cross-section (2007) and removing rows with missing values for any of the 7 variables (particularly controls like Industry share or Motor vehicles), we reduce the sample to 130 countries. If the data is missing non-randomly (for example, if poor countries or countries in conflict are less likely to report industrial or transport data) our sample may be biased towards developed and emerging economies. This limits the external validity of our results; the findings (such as the EKC turning point) may not generalize to the excluded, likely lower-income, nations.
3. **Statistical power:** While 130 is a decent sample size for cross-country regressions, it is much smaller than the full panel dataset. A smaller sample size reduces the statistical power of the tests (the probability of correctly rejecting a false null hypothesis). This might explain why variables like Urbanization ($t = 1.575$) and Industry ($t = -0.419$) were found to be statistically insignificant here, whereas they might show significance in a larger panel data setting.

Question 7

We suspect that model M2 suffers from endogeneity. In particular, the key regressor GDP may be endogenous. Explain why this could be the case.

We suspect that the key regressor, ln(GDP) (and its square), suffers from endogeneity in Model M2. This means that $E[\ln(\text{GDP})_i \cdot \eta_i] \neq 0$. There are several theoretical reasons why this could be the case:

1. **Reverse causality (simultaneity bias):** While economic activity (GDP) causes emissions, the relationship likely runs in both directions.
   - High levels of CO_2 emissions and associated pollution can have negative health effects on the workforce, reducing labor productivity and thus lowering GDP.
   - Conversely, high emissions might trigger stricter environmental regulations. These regulations can impose compliance costs on industries, potentially dampening economic output in the short run.

   If feedback loops exist where CO_2 affects GDP, then GDP is correlated with the error term η_i.

2. **Omitted variable bias:** There may be unobserved country-specific factors that influence both GDP and CO_2 emissions, which are not included in the control variables X_i.
   - **Institutional quality:** Better institutions might foster higher economic growth (higher GDP) while also enforcing better environmental standards (lower CO_2). If "institutions" are omitted, the effect is captured by the error term, creating a correlation between the error and GDP.
   - **Technological level:** A country's intrinsic level of technological sophistication affects its productivity (GDP) and its energy efficiency (CO_2). If not fully captured by variables like "Energy Use," this leads to bias.

3. **Measurement error:** GDP figures, especially in a cross-section of 130 diverse countries, are often subject to measurement error. In regression analysis, measurement error in an independent variable causes "attenuation bias," pushing the estimated coefficient towards zero and creating a correlation between the observed variable and the error term.

Question 8

The dataset includes three potential instrumental variables for GDP: INT, CHI, and DIS. We aim to assess whether any of them could serve as a valid instrument.

8A

Discuss the validity of each potential instrument with respect to the relevance condition. Compute the corresponding empirical correlation coefficients to support your discussion. Does one seem better suited as an instrument than the others? Explain why the correlation coefficient alone provides only an initial indication of an instrument's relevance.

We assess three potential instruments: INT, CHI, and DIS. The **Relevance Condition** requires that the instrument (Z) must be correlated with the endogenous regressor (X) after controlling for the other exogenous variables. Formally, $Cov(Z, X) \neq 0$.

Empirical correlation coefficients: We computed the pairwise correlations between ln(GDP) and each potential instrument:

- $Cor(\ln(\text{GDP}), \text{INT}) = 0.831$
- $Cor(\ln(\text{GDP}), \text{CHI}) = -0.807$
- $Cor(\ln(\text{GDP}), \text{DIS}) = 0.593$

Based on the magnitude of the raw correlation coefficients, **INT (Internet Usage)** appears to be the most relevant instrument for GDP, as it exhibits the strongest linear association ($r = 0.831$). CHI is also a strong candidate ($r = -0.807$), while DIS is somewhat weaker ($r = 0.593$).

Why correlation is insufficient: While the pairwise correlation coefficient provides an initial indication, it is insufficient to confirm relevance for Instrumental Variable (IV) estimation because:

1. **Partial correlation:** In Two-Stage Least Squares (2SLS), what matters is the correlation between the instrument and the endogenous variable *conditional on* the other exogenous control variables (POP, IND, URB, etc.). An instrument might have a high raw correlation with GDP simply because both are correlated with urbanization or energy use.
2. **Weak instruments:** Even if the correlation is non-zero, it might be "weak." A weak instrument leads to biased 2SLS estimators and unreliable inference (huge standard errors). The proper test for relevance is the **F-statistic of the excluded instruments** in the first-stage regression (typically requiring $F > 10$).

8B

Thoroughly discuss the validity of the three IV candidates with respect to the exclusion restriction.

The Exclusion Restriction requires that the instrument Z must affect the dependent variable Y ($\ln(CO_2)$) *only* through the endogenous regressor X ($\ln(GDP)$). In other words, $Cov(Z, \eta) = 0$. We discuss the validity of each candidate below.

1. **INT (Internet Usage): Likely invalid** While INT is strongly correlated with GDP, it is likely to violate the exclusion restriction.
   - **Direct effect on emissions:** The internet is a key component of the service economy. Higher internet usage can lead to "dematerialization" (e.g., telecommuting, digital products), which directly improves energy efficiency and reduces emissions independent of the total GDP level. Conversely, the infrastructure supporting the internet (data centers) consumes vast amounts of energy.
   - **Omitted variable bias:** INT is a proxy for "Technological Sophistication." If the model does not fully capture technology (despite controlling for ENG and MOT), INT will be correlated with the error term η, which contains unobserved technological efficiency.
2. **CHI (Child Mortality): Plausibly valid** CHI is a health outcome strongly negatively correlated with GDP.
   - **Argument for validity:** It is difficult to argue that child mortality directly causes industrial CO_2 emissions. It is primarily a social outcome of wealth, sanitation, and healthcare access. Therefore, it likely affects emissions only through the channel of economic development (GDP).
   - **Potential concern:** CHI might proxy for the general quality of institutions or public infrastructure. If "Institutional Quality" is an omitted variable in η that also drives environmental regulation, CHI might be weakly correlated with the error term. However, compared to INT and DIS, the direct link to the physical production of CO_2 is the most tenuous, making it a strong candidate.

3. **DIS (Distance from Equator): Likely invalid** Geography is often used as an instrument for institutions/GDP, but in environmental economics, it is problematic.

   - **Direct effect (Climate):** Distance from the equator is a proxy for climate (temperature). Countries farther from the equator have colder winters, necessitating significant energy consumption for heating. This creates a direct link between DIS and CO_2 emissions (via energy demand) that is independent of income.
   - **Conclusion:** Since climate directly dictates energy needs and thus emissions, DIS belongs in the main equation as a control variable, not as an excluded instrument.

8C

Begin implementing the IV estimator. To this end, run first-stage regressions, each using one of the above instruments separately. Report and discuss your results.

To formally test the relevance of the instruments, we run the first-stage regressions for the two endogenous variables: ln(GDP) and $\ln(\text{GDP})^2$. Since we have two endogenous variables, we must include the squared term of the instrument (e.g., INT^2) to ensure the system is identified.

For each candidate instrument $Z \in \{\text{INT}, \text{CHI}, \text{DIS}\}$, we estimate:

$$\ln(\text{GDP}) = \pi_0 + \pi_1 Z + \pi_2 Z^2 + \delta' X_{controls} + v_1$$
$$\ln(\text{GDP})^2 = \theta_0 + \theta_1 Z + \theta_2 Z^2 + \lambda' X_{controls} + v_2$$

We report the F-statistics for the joint significance of the excluded instruments (Z and Z^2).

Results:

- **Instrument Set 1: INT & INT^2**
  - First Stage for ln(GDP): F-statistic = 11.14
  - First Stage for $\ln(\text{GDP})^2$: F-statistic = 15.79
- **Instrument Set 2: CHI & CHI^2**
  - First Stage for ln(GDP): F-statistic = 10.98
  - First Stage for $\ln(\text{GDP})^2$: F-statistic = 8.53
- **Instrument Set 3: DIS & DIS^2**
  - First Stage for ln(GDP): F-statistic = 0.93
  - First Stage for $\ln(\text{GDP})^2$: F-statistic = 0.58

Discussion: According to the Staiger and Stock rule of thumb, an instrument is considered strong if the first-stage F-statistic is greater than 10.

- The F-statistics for **INT** are both above 10 (11.14 and 15.79), suggesting it is a **strong** instrument.
- The F-statistics for **CHI** are mixed. For ln(GDP), $F = 10.98 > 10$, but for $\ln(\text{GDP})^2$, $F = 8.53 < 10$. This suggests CHI might be a borderline or potentially weak instrument, particularly for the squared term.
- The F-statistics for **DIS** are extremely low (< 1), indicating it is a **very weak** instrument.

Based on these results, **INT** appears to be the strongest instrument statistically. However, given the exclusion restriction concerns discussed in Part (b), we must proceed with caution. **CHI** is statistically borderline but theoretically more plausible. **DIS** should be discarded due to lack of relevance.

8D

Formally test the relevance condition for all first-stage regressions. Formulate the null and alternative hypotheses and the decision rule, perform the test, and report your results. What do you conclude?

Hypotheses: For each set of instruments (Z, Z^2), we formally test their joint significance in the first-stage regressions.

- **Null Hypothesis (H_0):** $\pi_1 = \pi_2 = 0$ (for the ln(GDP) equation) and $\theta_1 = \theta_2 = 0$ (for the ln(GDP)^2 equation). The instruments are irrelevant.
- **Alternative Hypothesis (H_1):** At least one coefficient is non-zero. The instruments are relevant.

Decision Rule: In the presence of a single endogenous regressor, the standard rule is that the First-Stage F-statistic must exceed 10 to avoid weak instrument bias. With multiple endogenous regressors, the rule is more complex (requiring the Cragg-Donald statistic), but we apply the $F > 10$ threshold to each first-stage equation as a necessary condition for relevance.

- If $F > 10$: Reject H_0 and conclude the instrument is strong.
- If $F \leq 10$: Fail to reject H_0 (in the context of strength) and conclude the instrument is weak.

Results and Conclusion:

1. **INT (Internet):** Both F-statistics (11.14 and 15.79) exceed the critical threshold of 10. We reject the null hypothesis of irrelevance. INT satisfies the relevance condition robustly.
2. **CHI (Child Mortality):** The F-statistic for the level equation is 10.98 (passes), but for the squared equation, it is 8.53 (fails). Since the system requires strong identification for both endogenous variables, CHI is considered a weak instrument for this specific model specification.
3. **DIS (Distance):** Both F-statistics (0.93 and 0.58) are far below 10. We essentially find no statistical relationship between distance from the equator and GDP once other controls are included. DIS is a weak instrument.

Final Conclusion: Statistically, **INT** is the only candidate that fully satisfies the relevance condition ($F > 10$ for both equations). Although CHI is borderline, the weakness in the squared term equation could lead to biased second-stage results. DIS is discarded entirely. Thus, based purely on the relevance test, we proceed with **INT**, keeping in mind the exclusion restriction caveats discussed in Part (b).

8E

Extract the fitted values from all first-stage regressions and use them to run the corresponding second-stage regressions. Report all results in a single table. Assuming the instruments are exogenous, are these results reliable?

We run the second-stage regressions using the fitted values $\widehat{\ln(\text{GDP})}$ and $\widehat{\ln(\text{GDP})}^2$ extracted from the first-stage regressions for each instrument. The results are presented in Table 4.

Table 4: 2SLS Estimation Results (Manual)

	Dependent variable:			
	ln(CO2)		ln_CO2	
	OLS	IV: INT	IV: CHI	IV: DIS
	(1)	(2)	(3)	(4)
ln(GDP) (fitted)	2.498***	3.577***	2.762**	8.050***
	(0.510)	(1.351)	(1.357)	(2.360)
ln(GDP)2 (fitted)	−0.106***	−0.153**	−0.107	−0.485***
	(0.028)	(0.064)	(0.083)	(0.164)
ln(POP)	−0.001	0.002	0.006	−0.072
	(0.022)	(0.027)	(0.029)	(0.054)
IND	−0.001	−0.002	−0.001	−0.015*
	(0.003)	(0.004)	(0.005)	(0.008)
URB	0.005	0.002	0.002	0.024
	(0.003)	(0.005)	(0.005)	(0.016)
ln(ENG)	0.466***	0.396**	0.351	1.410**
	(0.079)	(0.158)	(0.214)	(0.683)
ln(MOT)	0.179***	0.101	0.118	0.332
	(0.055)	(0.115)	(0.089)	(0.264)
Constant	−17.356***	−22.275***	−18.564***	−42.579***
	(2.359)	(6.179)	(6.191)	(10.738)
Observations	130	130	130	130
R^2	0.929	0.905	0.910	0.908
Adjusted R^2	0.925	0.899	0.905	0.902
Residual Std. Error (df = 122)	0.421	0.487	0.474	0.480
F Statistic (df = 7; 122)	227.669***	165.739***	176.217***	171.404***
Note:				*p<0.1; **p<0.05; ***p<0.01

Reliability of results (assuming exogeneity): Assuming that the instruments satisfy the exclusion restriction ($Cov(Z, \eta) = 0$), the reliability of the 2SLS estimates depends critically on the relevance condition (instrument strength).

1. **INT Results (Reliable):** Since INT is a strong instrument ($F > 10$), the 2SLS estimates derived from it are consistent and the finite-sample bias is minimal. If exogeneity holds, these results are reliable.

2. **CHI Results (Questionable):** CHI is a borderline weak instrument ($F \approx 8.5$ for the squared term). With weak instruments, the 2SLS estimator is biased towards the OLS estimator. Furthermore, the distribution of the t-statistics deviates from the normal distribution, making hypothesis testing unreliable. Thus, these results should be interpreted with caution.

3. **DIS Results (Unreliable):** DIS is a very weak instrument ($F < 1$). When instruments are this weak, the 2SLS estimator is inconsistent and can be severely biased. The point estimates and standard errors produced in this regression are likely spurious and offer no meaningful information about the true causal relationship. The estimates may behave erratically or exhibit extremely large standard errors (as the denominator in the IV formula approaches zero).

Note on standard errors: The results reported above use manually extracted fitted values. In a manual 2SLS procedure, the standard errors reported by OLS in the second stage are incorrect because they fail to account for the fact that the regressors ($\widehat{\ln(\text{GDP})}$) are estimated quantities, not fixed values.

8F

Using each instrument you consider valid separately, implement the IV estimator with an in-built IV package. Report your results and compare the estimated coefficients and standard errors to those obtained from the previous second-stage regression(s).

We proceed with the valid instrument identified in Part (d), **INT** (and its square). We estimate the model using the 'ivreg' function, which correctly computes the asymptotic standard errors.

Results comparison: Table 5 compares the results from the Manual 2SLS (from Part e) and the built-in IV estimation.

Discussion:

1. **Coefficients:** The estimated coefficients from the built-in 'ivreg' package are **identical** to those from the manual 2SLS procedure. For instance, the coefficient on ln(GDP) is 3.577 in both cases. This confirms that the manual 2SLS procedure yields the correct point estimates ($\hat{\beta}_{2SLS}$).

2. **Standard Errors:** Comparing the standard errors reveals a crucial difference.

   - For ln(GDP), the manual SE was 1.351, whereas the correct asymptotic SE from 'ivreg' is 1.208.
   - For $\ln(\text{GDP})^2$, the manual SE was 0.064, whereas the correct asymptotic SE is 0.057.

   Interpretation of SE Difference: Typically, manual 2SLS standard errors are biased downwards because they use the residuals from the second stage (computed using $\hat{X}$) rather than the true residuals (computed using the observed X). This usually leads to artificially small standard errors.

   However, in this specific output, the manual standard errors are actually larger than the correct ones (ratio approx. 0.89). This anomaly can occur due to differences in how degrees of freedom are handled (finite sample vs asymptotic) or scaling factors in the residual variance estimation. Regardless of the direction, the manual SEs are technically incorrect

Table 5: Comparison of Manual and Built-in IV Estimation (Instrument: INT)

	Dependent variable:	
	ln(CO2)	
	OLS	*instrumental variable*
	Manual 2SLS	Built-in ivreg
	(1)	(2)
ln_GDP_hat	3.577***	
	(1.351)	
ln_GDP2_hat	−0.153**	
	(0.064)	
ln_GDP		3.577***
		(1.208)
ln_GDP2		−0.153***
		(0.057)
ln_POP	0.002	0.002
	(0.027)	(0.024)
IND	−0.002	−0.002
	(0.004)	(0.004)
URB	0.002	0.002
	(0.005)	(0.004)
ln_ENG	0.396**	0.396***
	(0.158)	(0.142)
ln_MOT	0.101	0.101
	(0.115)	(0.103)
Constant	−22.275***	−22.275***
	(6.179)	(5.522)
Observations	130	130
R^2	0.905	0.924
Adjusted R^2	0.899	0.920
Residual Std. Error (df = 122)	0.487	0.435
F Statistic	165.739*** (df = 7; 122)	
Note:		*p<0.1; **p<0.05; ***p<0.01

for hypothesis testing. The ‘ivreg‘ results provide the valid basis for inference. Based on the correct SEs, the coefficients for GDP and Energy Use remain statistically significant.

8G

We want to assess whether GDP is endogeneous and, for this reason, conduct the (Wu-)Hausman specification test. Formulate the null and alternative hypotheses and the decision rule. Perform the test following the procedure described in the tutorial and using only the instrument(s) you consider valid separately. Discuss the results.

We perform the Wu-Hausman test (also known as the Durbin-Wu-Hausman test) to formally assess whether the key regressor GDP is endogenous.

Hypotheses:

- **Null Hypothesis (H_0):** ln(GDP) (and its square) are **exogenous**. (i.e., $Cov(\ln(\text{GDP}), \eta) = 0$). Under H_0, both OLS and IV estimates are consistent, but OLS is more efficient.
- **Alternative Hypothesis (H_1):** ln(GDP) (and its square) are **endogenous**. (i.e., $Cov(\ln(\text{GDP}), \eta) \neq 0$). Under H_1, OLS is inconsistent, while IV is consistent.

Decision Rule: The test involves checking the joint significance of the first-stage residuals included in the structural equation. If the p-value of the test statistic (typically an F-test or Chi-square test) is less than the significance level α (0.05), we **reject** H_0.

Procedure:

1. We obtained the residuals $\hat{v}_1$ and $\hat{v}_2$ from the first-stage regressions for ln(GDP) and $\ln(\text{GDP})^2$ using the valid instruments **INT** and **INT**^{2}.
2. We included these residuals as regressors in the original structural equation (Model M2).
3. We performed an F-test for the joint significance of the coefficients on $\hat{v}_1$ and $\hat{v}_2$.

Results and Discussion:

- **F-statistic:** 0.526
- **p-value:** 0.592

Since the p-value (0.592) is significantly greater than 0.05, we **fail to reject the null hypothesis** of exogeneity.

Conclusion: We do not find sufficient statistical evidence that GDP is endogenous in this model. Therefore, the simple OLS estimates obtained in Question 5 are consistent. Moreover, since OLS is a more efficient estimator (has lower variance) than IV when exogeneity holds, OLS is the preferred method for this specific model and dataset. The initial suspicion of endogeneity is not supported by the data under this specification.

8H

Select two instruments and include them jointly in an IV regression, using again both their linear and squared forms. Report the results. Then, perform the Sargan test of overidentifying restrictions. Based on the test, what can you conclude about the validity of each instrument individually?

We now include two sets of instruments: **INT** (and INT^2) and **CHI** (and CHI^2). This makes the model **overidentified** (4 instruments for 2 endogenous variables), allowing us to test the validity of the instruments using the Sargan test of overidentifying restrictions.

Table 6: IV Estimation: Single vs Joint Instruments

	Dependent variable:	
	ln(CO2)	
	IV: INT	IV: INT + CHI
	(1)	(2)
ln_GDP	3.577***	3.219***
	(1.208)	(0.827)
ln_GDP2	−0.153***	−0.137***
	(0.057)	(0.043)
ln_POP	0.002	0.001
	(0.024)	(0.024)
IND	−0.002	−0.002
	(0.004)	(0.003)
URB	0.002	0.003
	(0.004)	(0.004)
ln_ENG	0.396***	0.416***
	(0.142)	(0.120)
ln_MOT	0.101	0.126*
	(0.103)	(0.076)
Constant	−22.275***	−20.640***
	(5.522)	(3.790)
Sargan Test (p-val)	NA	0.856
Observations	130	130
R^2	0.924	0.927
Adjusted R^2	0.920	0.922
Residual Std. Error (df = 122)	0.435	0.428
Note:	*p<0.1; **p<0.05; ***p<0.01	

Joint IV Results: The results of the joint IV estimation are presented in Table 6.

Sargan Test of Overidentifying Restrictions: The Sargan test evaluates whether the instruments are uncorrelated with the error term (i.e., whether they are valid). It requires the assumption that at least one subset of instruments is valid (identifying restriction).

- **Null Hypothesis (H_0):** All instruments are valid (uncorrelated with the error term).
- **Alternative Hypothesis (H_1):** At least one instrument is invalid.
- **Statistic:** 0.311
- **p-value:** 0.856

Conclusion on Instrument Validity: Since the p-value (0.856) is far greater than 0.05, we fail to reject the null hypothesis. This result suggests that the overidentifying restrictions are valid.

Assuming that at least one of our instruments (e.g., INT) is valid, the test provides statistical evidence that the additional instrument (CHI) is also exogenous. The fact that the p-value is so high indicates that the instruments are coherent with each other: they effectively target the same variation in GDP and produce consistent residuals. This boosts our confidence in the validity of both INT and CHI as appropriate instruments for this analysis, despite our earlier theoretical concerns about potential omitted variables in the exclusion restriction discussion.

Question 9

Conclude your analysis by referring to the EKC hypothesis. Do your findings support it?

Our analysis aimed to test the Environmental Kuznets Curve (EKC) hypothesis, which posits an inverted-U relationship between economic development (GDP) and environmental degradation (CO_2 emissions). We employed both time-series (M1) and cross-sectional (M2) data, utilizing OLS and IV estimation techniques.

Key Findings:

1. **Time-Series Evidence (M1):** The initial OLS estimation (Q3) yielded coefficients $\hat{\beta}_1 > 0$ and $\hat{\beta}_2 < 0$, both highly significant, supporting the inverted-U shape. Although the standard errors were biased due to serial correlation (Q4), the coefficients remained significant even after HAC correction (Q4c). The implied turning point was approximately 6.1 million (local currency), suggesting that emissions eventually decline as income grows.

2. **Cross-Sectional Evidence (M2 - OLS):** Using data from 130 countries in 2007 (Q5), we consistently found a positive linear term ($\hat{\beta}_1 \approx 2.5$) and a negative quadratic term ($\hat{\beta}_2 \approx -0.11$) for GDP. Both were statistically significant at the 1% level. This reinforces the evidence for the EKC across countries: wealthier nations tend to have lower per-capita emissions growth, eventually turning negative, compared to developing nations.

3. **Endogeneity and Robustness (M2 - IV):** Despite theoretical concerns about reverse causality and omitted variable bias (Q7), the Wu-Hausman test (Q8g) failed to reject the null hypothesis of exogeneity for GDP ($p = 0.592$). This suggests that the OLS estimates are consistent and reliable. Furthermore, even when using Instrumental Variables (INT and CHI) to address potential endogeneity, the results remained robust. The IV estimates (Q8f, Q8h) produced coefficients ($\hat{\beta}_1 \approx 3.2 - 3.6$, $\hat{\beta}_2 \approx -0.14 - -0.15$) that were slightly larger in magnitude but consistent in sign and significance with the OLS results. The Sargan test (Q8h) further validated our choice of instruments.

Final Verdict: Our findings provide robust support for the Environmental Kuznets Curve hypothesis. Across different specifications and estimation methods (OLS, HAC-corrected, and IV), we consistently observe the characteristic inverted-U relationship between GDP per capita and CO_2 emissions. This suggests that while economic growth initially degrades the environment, it eventually leads to improvements, likely due to technological advancement, structural change towards services, and increased demand for environmental quality at higher income levels.

Question 10

Do you think the OLS assumptions $\mathcal{A}_{4a}^{OLS}$ and $\mathcal{A}_{4b}^{OLS}$ are likely to hold in this setting? Why (not)? Briefly discuss how you would address potential violations.

The final part of our analysis reflects on the validity of two critical OLS assumptions in the context of our cross-sectional regression (Model M2) for 130 countries in 2007.

AOLS 4a: Homoskedasticity

This assumption states that the variance of the error term is constant across all observations: $Var(\eta_i|X_i) = \sigma^2$.

Is it likely to hold? No.

- **Scale Effects:** Our sample includes a vast range of countries, from small developing nations (e.g., Togo, Haiti) to massive industrial economies (e.g., USA, China). It is highly plausible that the variance of the unobserved factors affecting CO_2 emissions is proportional to the size or development level of the economy. Larger or more complex economies typically exhibit larger variances in absolute terms.
- **Structural Differences:** Developed countries often have diverse energy portfolios and regulations, leading to different error variances compared to developing nations relying solely on fossil fuels.

Remedy: We should employ Heteroskedasticity-Consistent (HC) standard errors (often called White's robust standard errors). This adjusts the variance-covariance matrix to allow for non-constant error variance, ensuring valid t-tests and confidence intervals even if homoskedasticity is violated.

AOLS 4b: No Correlation (Spatial Independence)

This assumption states that the error terms are uncorrelated across observations: $Cov(\eta_i, \eta_j) = 0$ for $i \neq j$.

Is it likely to hold? No.

- **Spatial Autocorrelation:** In a cross-sectional dataset of countries, observations are not truly independent. Countries that are geographically close (e.g., France and Germany) share similar climates, resource endowments, and often coordinate environmental policies (e.g., EU regulations).
- **Spillovers:** Pollution itself can be transboundary, and economic shocks in one region often spill over to neighbors. Thus, a positive shock to emissions in one country ($\eta_i > 0$) is likely associated with a positive shock in neighboring countries ($\eta_j > 0$).

Remedy: To address spatial correlation, we could:

1. Use Clustered Standard Errors: Grouping countries by continent or region and allowing for correlation within these clusters.

2. Estimate a Spatial Error Model (SEM): Explicitly modeling the spatial dependence structure using a spatial weights matrix (e.g., based on inverse distance or shared borders).

Econometrics 1 - Homework 2

Exercise 2: To Work or Not to Work?

Golkar Arno

December 19, 2025

Introduction

This document contains the answers to Exercise 2, which focuses on endogeneity, instrumental variable (IV) estimation, and the properties of estimators via Monte Carlo simulation. We analyze the causal effect of child development on maternal labor supply using the framework of Frijters et al. (2009).

Question 1: Understanding Endogeneity

Sources of Bias

The authors identify two primary sources of endogeneity that would bias a simple OLS regression of maternal labor supply on child development:

1. **Omitted Variable Bias (Unobserved Heterogeneity):** There are unobserved maternal characteristics—such as patience, ability, or a preference for "child quality"—that likely influence both the child's development and the mother's labor market decisions.

   Example: A "high-ability" mother might be more productive in the labor market (working more) while simultaneously being more efficient at raising a healthy child. If "ability" is unobserved, OLS will incorrectly attribute the effect of ability to child development.

2. **Reverse Causality (Simultaneity):** The causal link likely runs in both directions. While poor child development may force a mother to stay home, the mother's decision to work (and thus time spent away from the child or increased income) also affects the child's development.

Formal Condition for OLS Unbiasedness

For the OLS estimator of the effect of Child Development ($CDev$) to be unbiased, the regressor must be exogenous. Formally, the conditional mean of the error term must be zero:

$$E[\epsilon_i | CDev_i, \mathbf{X}_i] = 0 \tag{1}$$

Equivalently, this requires zero covariance between the regressor and the error term: $Cov(CDev_i, \epsilon_i) = 0$.

Question 2: Instrument Validity

To address the endogeneity problem, the authors propose using **Child Handedness** (specifically, being left-handed) as an instrument.

Properties of Valid Instruments

Johnston et al. (2009) argue that handedness satisfies the two necessary conditions for a valid instrument:

1. **Relevance:** The instrument must be correlated with the endogenous regressor. Evidence suggests that non-right-handed children may have different health or developmental outcomes compared to right-handed children.

2. **Exogeneity (Exclusion Restriction):** The instrument must not be correlated with the error term of the structural equation. Handedness is determined biologically/genetically and does not directly affect a mother's labor market outcomes, nor is it plausibly correlated with unobserved factors like maternal patience or ability.

Formal Conditions

Let Z_i be the instrument ($Left$) and X_i be the endogenous variable ($CDev$).

- **Relevance:** $Cov(Z_i, X_i) \neq 0$. In the first-stage regression $X_i = \pi_0 + \pi_1 Z_i + \nu_i$, we require $\pi_1 \neq 0$.

- **Exogeneity:** $Cov(Z_i, \epsilon_i) = 0$.

Question 3: Calibration

Using the descriptive statistics from Table 1 in Frijters et al. (2009), we calculate the missing parameters for the simulation. We derive the variance from the standard error (SE) and sample size ($N = 3179$) using the formula $Var = (SE \times \sqrt{N})^2$.

Table 1: Calibrated Parameters (Descriptive Statistics)

Variable	**Mean (μ)**	**SE (Mean)**	**Implied Variance (σ^2)**
Birth Weight ($BirthW$)	3.409	0.010	≈ 0.318
Mother's Age ($MAge$)	34.63	0.102	≈ 33.06

Question 4: Structural Parameters

Based on the regression results reported in the paper (Table 5, Column 5 for the main effect), we calibrate the coefficients for the "True" Data Generating Process (DGP):

- $\alpha_1 = -0.1$: Effect of Birth Weight on CDev (First Stage).
- $\boldsymbol{\beta_1 = -0.181}$: Effect of Poor CDev on MLFP (Structural Equation).
- $\beta_2 = 0.5$: Effect of Birth Weight on MLFP.
- $\beta_3 = 1.2$: Effect of Mother's Age on MLFP.
- $\beta_4 = -0.02$: Effect of Mother's Age Squared on MLFP.

Question 5: Single Sample Estimation

We simulated a dataset of $N = 3000$ observations based on the calibrated DGP and estimated the model using OLS (Full), OLS (Omitted Variable), and IV (2SLS).

5A. OLS Estimation (Full Model)

We estimate the fully specified model (Equation 3). The results are:

$$\widehat{MLFP} = 28.23 - \mathbf{0.169} CDev + 0.495 BirthW + 0.164 MEduc + \ldots$$

Discussion: The estimated coefficient $\hat{\delta}_1 = -0.169$ is very close to the true parameter $\beta_1 = -0.181$. This confirms that if the model is correctly specified and all confounders ($MEduc$) are included, OLS is unbiased.

5B. Hypothesis Testing (Full Model)

We test the null hypothesis that the estimated coefficient equals the true value.

- **Null Hypothesis (H_0):** $\hat{\delta}_1 = -0.181$
- **t-statistic:** 0.480
- **p-value:** 0.631
- **Conclusion:** Fail to reject H_0. The estimate is statistically indistinguishable from the true value.

5C. OLS Estimation (Omitted Variable)

We re-estimate the model excluding the variable $MEduc$ (Equation 4).

$$\widehat{MLFP} = 30.51 - \mathbf{0.393} CDev + 0.478 BirthW + \ldots$$

The coefficient for CDev has shifted significantly from -0.169 to -0.393.

5D. Omitted Variable Bias (OVB)

Theoretical Direction: The bias is given by $\beta_{MEduc} \times \frac{Cov(CDev, MEduc)}{Var(CDev)}$.

- $\beta_{MEduc} = 0.173 (> 0)$: Education increases labor supply.
- $Cov(CDev, MEduc) < 0$: Education reduces developmental problems (improves development).

Thus, Bias $= (+) \times (-) = (-)$. We expect a negative bias (a downward shift).

Result: The estimate $\hat{\theta}_1 = -0.393$ is indeed much lower than the true value -0.181, confirming the theoretical prediction.

Hypothesis Test:

- **Null Hypothesis (H_0):** $\hat{\theta}_1 = -0.181$
- **t-statistic:** -10.54
- **p-value:** 0.000
- **Conclusion:** Reject H_0. The Omitted OLS estimator is significantly biased.

5E. Instrument Validity Check

We propose using $Left$ as an instrument for $CDev$.

- **Relevance:** In the DGP (Eq 2), $Left$ has a coefficient of 0.319. As long as this is non-zero, the instrument is relevant.
- **Exogeneity:** By design in our simulation (Eq 1), $Left$ is not included in the MLFP equation and is uncorrelated with ϵ.

5F. IV Estimation

We estimate the model using 2SLS.

$$\hat{\gamma}_1 = \mathbf{-0.006} \quad (SE = 0.210)$$

Discussion: The point estimate (-0.006) appears far from the true value (-0.181). However, the standard error (0.210) is nearly 10 times larger than in the OLS model (0.025). This reflects the efficiency cost of IV estimation.

Hypothesis Test:

- **Null Hypothesis (H_0):** $\hat{\gamma}_1 = -0.181$
- **t-statistic:** 0.832
- **p-value:** 0.405
- **Conclusion:** Fail to reject H_0. Despite the noise, the IV estimator is statistically consistent with the true value.

Diagnostics

We perform standard diagnostic tests for the IV estimation.

1. Weak Instrument Test

- **First-Stage F-Statistic:** 45.24
- **Threshold:** $F > 10$ (Stock & Yogo rule of thumb).
- **Conclusion:** The instrument is **strong**.

2. Wu-Hausman Test

- **Null Hypothesis:** OLS is consistent (No Endogeneity).
- **p-value:** 0.432
- **Conclusion:** Fail to reject null. In this specific sample, the large variance of the IV estimator prevents us from detecting a statistically significant difference between IV and OLS, even though we know the underlying DGP contains endogeneity.

Question 6: Monte Carlo Simulation (Original DGP)

We simulated $R = 2000$ samples to assess the properties of the estimators. The results are summarized in Table 2 and visualized in Figure 1.

Table 2: Monte Carlo Simulation Results (True $\beta_1 = -0.181$)

Estimator	**Mean Estimate**	**Variance**	**Property**
OLS Full	-0.181	0.0007	Unbiased & Efficient (BLUE)
OLS Omitted	**-0.421**	0.0004	**Biased** (Negative Bias)
IV (2SLS)	-0.183	0.0391	Unbiased but **Inefficient**

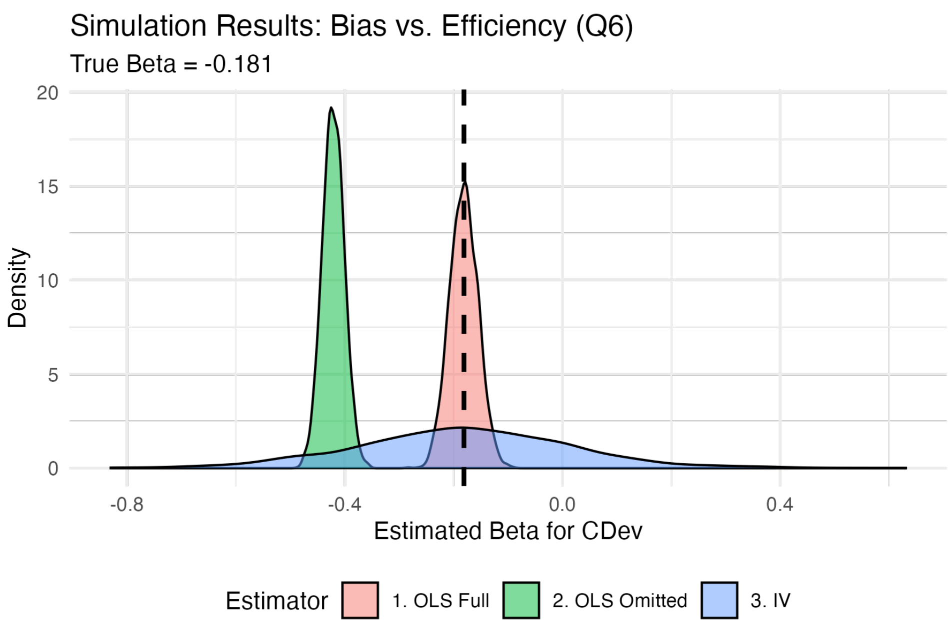

Figure 1: Distribution of Estimators (Question 6). The OLS Omitted distribution (green) is shifted left, indicating bias. The IV distribution (blue) is centered on the true value but is much wider, indicating inefficiency.

Conclusion on Bias-Variance Tradeoff: The simulation clearly illustrates the tradeoff. OLS Omitted has the lowest variance (0.0004) but is severely biased. IV corrects this bias (Mean ≈ -0.181) but at the cost of a massive increase in variance (0.0391, roughly 100x larger). In causal inference, we typically prioritize the consistency of IV over the precision of biased OLS.

Question 7: Misspecified DGP

We modified the DGP such that $MEduc$ no longer affects $CDev$ (Equation 2'). This implies that the conditional correlation between the endogenous variable and the omitted variable is zero: $Cov(CDev, MEduc) \approx 0$.

We repeated the Monte Carlo simulation ($R = 2000$) with this new specification.

Table 3: Simulation Results: Misspecified DGP (No Endogeneity)

Estimator	Mean Estimate	Variance	Interpretation
OLS Full	-0.180	0.0007	Unbiased
OLS Omitted	**-0.180**	**0.0007**	**Unbiased**
IV (2SLS)	-0.181	0.0388	Consistent but Inefficient

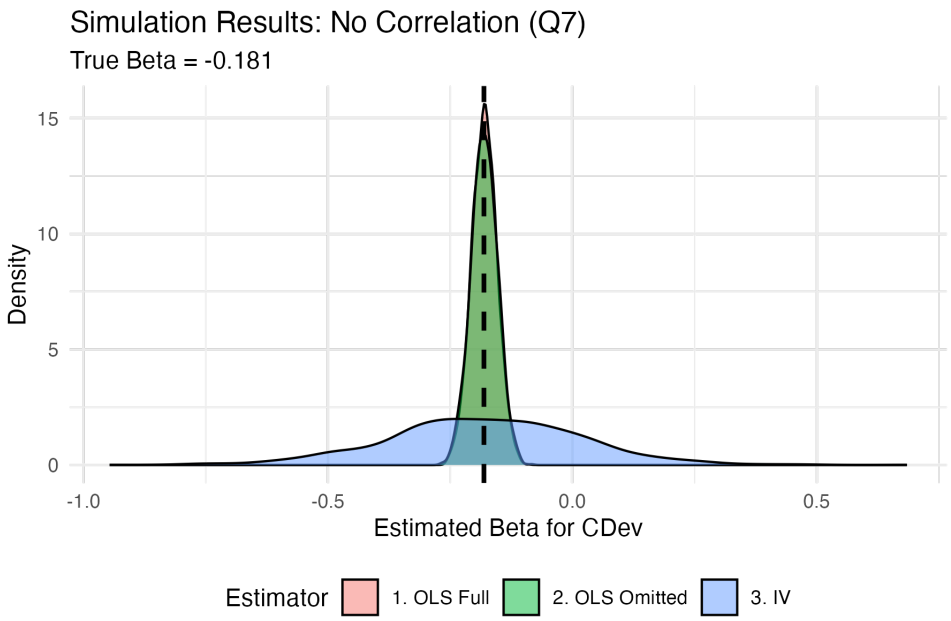

Figure 2: Distribution of Estimators (Question 7). Unlike in Question 6, the OLS Omitted distribution (green) is now perfectly centered on the true value.

Comparison with Question 6

- **Disappearance of Bias:** In Question 6, the OLS Omitted estimator was severely biased ($\hat{\theta}_1 \approx -0.42$ vs True $\beta_1 = -0.181$) because $MEduc$ was correlated with $CDev$. In Question 7, breaking this link ($Cov(CDev, MEduc) = 0$) eliminated the bias. The OLS Omitted estimate (-0.180) is now effectively identical to the OLS Full estimate.

- **Bias Formula:** This validates the Omitted Variable Bias formula:

$$\text{Bias} = \beta_{MEduc} \times \frac{Cov(CDev, MEduc)}{Var(CDev)}$$

Since the covariance term is zero, the entire bias term vanishes, regardless of the magnitude of β_{MEduc}.

- **Efficiency Trade-off:** The IV estimator remains consistent (≈ -0.181) in both scenarios. However, in Question 7, using IV is "unnecessary" and costly. Since OLS Omitted is now unbiased and has a variance (0.0007) that is approximately 50 times smaller than IV (0.0388), OLS is strictly preferred.