MSCFE 610 - Financial Econometrics

Group Work Project #1 Report for Group 12188

David Olutunde Daniel, Pratyush Mathur

December 18, 2025

Problem 1: Omitted Variable Bias

1a) Independence of Error Terms

Question: If the error terms $\epsilon(i)$ satisfy the standard assumptions, does it mean that the error terms $\mu(i)$ do as well?

Answer: No. The true model is $Y(i) = \alpha + \beta x(i) + \gamma w(i) + \delta z(i) + \epsilon(i)$, while the estimated model is $Y(i) = \alpha + \beta x(i) + \gamma w(i) + \mu(i)$. The new error term is effectively $\mu(i) = \delta z(i) + \epsilon(i)$. Even if $\epsilon(i)$ satisfies standard assumptions, $\mu(i)$ contains the omitted variable $z(i)$. If $z(i)$ is correlated with the included regressors, $\mu(i)$ will be correlated with them as well. This violates the assumption of Strict Exogeneity ($E[\mu|x, w] \neq 0$), rendering the error terms non-compliant.

1b) Consequences for Parameter Estimates

If the omitted variable z is correlated with x or w, the OLS estimators for α, β, and γ from model (2) will be **biased** and **inconsistent**. The coefficients will capture both the direct effect of the included variables and the indirect effect of the omitted variable. This Omitted Variable Bias (OVB) means the expected value of the estimator does not equal the true parameter.

1c) Condition for Unbiased Estimates

The estimates would be identical (unbiased) only if the omitted variable z is **orthogonal** (uncorrelated) to the included explanatory variables x and w. In this case, the covariance between the regressor and the error term μ remains zero. However, the omission of a relevant variable still increases the error variance, inflating the standard errors of the estimates.

1d) Simulation Analysis

We simulated a model $Y = \alpha + \beta X + \gamma Z + \epsilon$ with known $\beta = 0.5$ and introduced a correlation between X and Z.

- **Small Sample ($n = 50$):** Estimating without Z yielded $\hat{\beta} \approx 0.76$, a significant bias of $+0.26$.
- **Large Sample ($n = 500$):** The estimated $\hat{\beta}$ remained approximately 0.77.

Conclusion: Increasing sample size does not cure Omitted Variable Bias; the estimator remains inconsistent because the bias is structural, not a result of sampling noise.

Problem 2: Outlier Analysis

2a) Sensitivity to Outliers

Ordinary Least Squares (OLS) minimizes the sum of squared residuals ($\sum e_i^2$). By squaring errors, the objective function heavily penalizes large deviations. A single outlier can exert massive "leverage," pulling the regression line toward itself to minimize the squared distance. This makes OLS non-robust compared to methods like Least Absolute Deviations.

2b) Simulation of Outlier Effects

We simulated a linear trend and introduced outliers to observe their impact.

- **Clean Data:** The regression yielded $\hat{\beta} \approx 0.49$, matching the true $\beta = 0.5$.
- **With Outliers:** Introducing outliers shifted the slope to $\hat{\beta} \approx 0.82$, a 64% deviation.
- **Magnitude Sensitivity:** The bias in β scales linearly with the outlier magnitude (in σ). A 10σ outlier caused triple the bias of a 3σ outlier.

Problem 3: Model Selection

3a) Methodology and Results

We performed model selection on the assigned dataset to predict Y using predictors X_1 through X_5 via Stepwise Selection (BIC), Lasso (CV), and Best Subsets.

Method	**Selected Features**	**BIC**	**Test R^2**
Stepwise (BIC)	X_2, X_3, X_4	216.63	0.403
Lasso (CV)	X_1, X_2, X_3, X_4, X_5	218.12	0.462
Best Subsets (Adj R^2)	X_2, X_3, X_4, X_5	218.12	0.462

Conclusion: We selected the model with features $[X_2, X_3, X_4]$. We prioritized the **BIC criterion** (216.63) for parsimony. The 3-variable model prevents overfitting and excludes variables (X_1, X_5) that add complexity with minimal explanatory gain.

Problem 4: Elasticity

4a) Elasticity Calculations

The elasticity of y with respect to x is $E = \frac{dy}{dx}\frac{x}{y}$.

1. **Model (a):** $y = 2 + 0.8x$

$$E = 0.8\left(\frac{x}{2 + 0.8x}\right) \quad \text{(Variable)}$$

2. **Model (b):** $\ln(y) = 0.1 + 0.4x$

$$\frac{d\ln y}{dx} = 0.4 \implies E = \frac{dy}{dx}\frac{x}{y} = (0.4y)\frac{x}{y} = 0.4x \quad \text{(Linear in } x\text{)}$$

3. **Model (c):** $\ln(y) = 0.1 + 0.25\ln(x)$

$$E = \frac{d\ln y}{d\ln x} = 0.25 \quad \text{(Constant)}$$

4. **Model (d):** $y = 0.15 + 1.2\ln(x)$

$$\frac{dy}{dx} = \frac{1.2}{x} \implies E = \frac{1.2}{x}\frac{x}{y} = \frac{1.2}{y} \quad \text{(Inverse to } y\text{)}$$

4b) Identification

The correct regression for identifying constant elasticity is **Model (c):** $\ln(y) = 0.1 + 0.25\ln(x)$. In this log-log specification, the slope coefficient (0.25) is the elasticity itself: a 1% change in x leads to a 0.25% change in y.

Problem 5: Stationarity and Time Series

5a) Stationarity and Testing

Stationarity implies constant statistical properties (mean, variance) over time. We use the **Augmented Dickey-Fuller (ADF)** test (H_0: Unit Root). If we fail to reject H_0, the series is non-stationary, and we must difference it (ΔY_t).

5b) Real Economic Data Analysis

We analyzed S&P 500 prices (2020-2022).

- **Price Levels:** ADF p-value = 0.41. Fail to reject null. Prices are **Non-Stationary**.
- **Log Returns:** ADF p-value = 0.00. Reject null. Returns are **Stationary**.

5c) Unit Root vs. Explosive Root

A unit root ($\phi = 1$) implies a Random Walk where shocks have permanent effects. A root > 1 implies an **Explosive Process** that diverges exponentially. While mathematically possible, explosive roots are economically unrealistic over long periods (except during bubbles). The primary challenge is distinguishing between stable (stationary) and wandering (unit root) processes.

Problem 6: Structural Break

6a) Testing with Dummy Variables

To test for a structural break in slope at $t = 10$ using a single regression, we define a dummy variable $D_t = 1$ if $t > 10$ and 0 otherwise. We estimate:

$$Y(t) = \alpha + \beta_1 X(t) + \delta(D_t \cdot X(t)) + \epsilon(t) \tag{1}$$

Here, δ represents the *change* in slope. We perform a t-test on δ ($H_0 : \delta = 0$). If significant, we conclude a structural break exists.