Empirical Asset Pricing: Problem Set 1

Your Name

February 26, 2025

1 Question 1

1.1 Simulation function

```
def routine(alpha, beta, theta, rho, var_u, var_v, cov_uv, T):
    # Generate correlated normal random variables
    errors = np.random.multivariate_normal(
        [0, 0],
        np.array([[var_u, cov_uv], [cov_uv, var_v]]),
        size=T
    )
    u, v = errors[:, 0], errors[:, 1]

    # Initialize x, y
    x, y = np.zeros(T), np.zeros(T)
    x[0] = theta / (1 - rho)  # x0 depends only on theta and rho
    y[0] = alpha + beta * x[0]
    for t in range(1, T):
        x[t] = theta + rho * x[t - 1] + v[t]
        y[t] = alpha + beta * x[t - 1] + u[t]

    return x, y
```

This function generates correlated time series $\{x_t, y_t\}$ using the specified parameters, starting at their unconditional means.

1.2 OLS Estimation

```
def ols_regression(x, y):
    # Regression 1: y[t+1] = alpha + beta * x[t] + u[t+1]
    Y = sm.add_constant(x[:-1])
    y_target = y[1:]
    model_y = OLS(y_target, Y).fit()

    X = sm.add_constant(x[:-1])
    x_target = x[1:]
    model_x = OLS(x_target, X).fit()

    return model_y, model_x
```

Plot of x and y

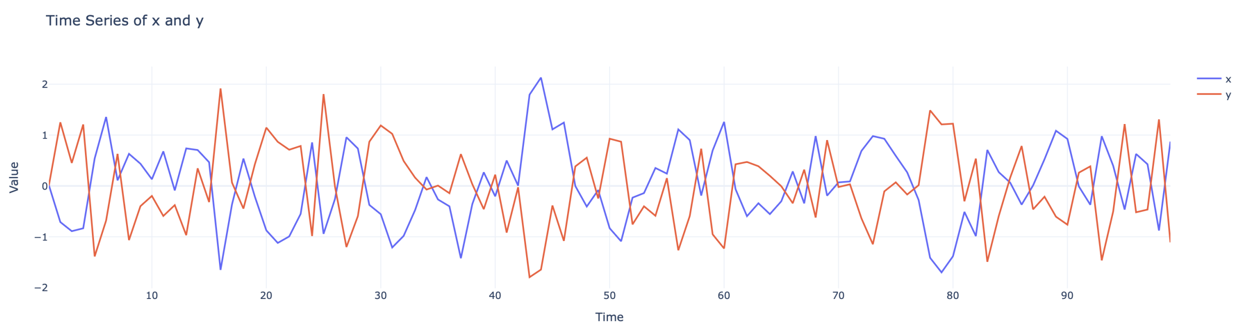

Figure 1: Simulated Time Series of x_t and y_t for $T = 100$. (Illustrative example from one random draw.)

2 Question 2: Single-Simulation OLS Results

After drawing one realization of $\{x_t, y_t\}$ with $T = 100$, we estimate:

- For the regression $y_{t+1} = \alpha + \beta x_t + u_{t+1}$:
  - $\hat{\alpha} \approx -0.0628$
  - $\hat{\beta} \approx 0.2151$
- For the regression $x_{t+1} = \theta + \rho x_t + \nu_{t+1}$ we also obtain numerical estimates for θ and ρ.
  - $\hat{\theta} \approx 0.0819$
  - $\hat{\rho} \approx 0.2302$

Figure 1 shows the time-series. Table 1 summarizes the OLS results for the regression for $\hat{\beta}$. We note that in this single run, $\hat{\beta}$ is clearly larger than the true $\beta = 0.05$; finite-sample bias is apparent.

Table 1: OLS results for y_{t+1} vs. x_t in one single draw ($T = 100$).

	Coeff.	Std. Err.	t-stat	p-value
Constant	-0.0628	0.079	-0.79	0.429
x_t	0.2151	0.103	2.09	0.039
$R^2 = 0.043$			Adj. $R^2 = 0.033$	

Positive bias contradicts Stambaugh's formula. Likely due to random noise in single simulation. Now we verify with $N = 10000$ simulations. (It is important to note that the numbers in the table are not fixed. They may vary slightly with each run of the code due to the randomness in the simulations.)

3 Question 3: Distribution of $\hat{\beta}$

To quantify the bias across many realizations, we fix $T = 100$ with N independent simulations, and estimate $\hat{\beta}$ each time. Although the mean of the distribution is near the true value $\beta = 0.05$, we see a shift (bias) and a noticeable spread.

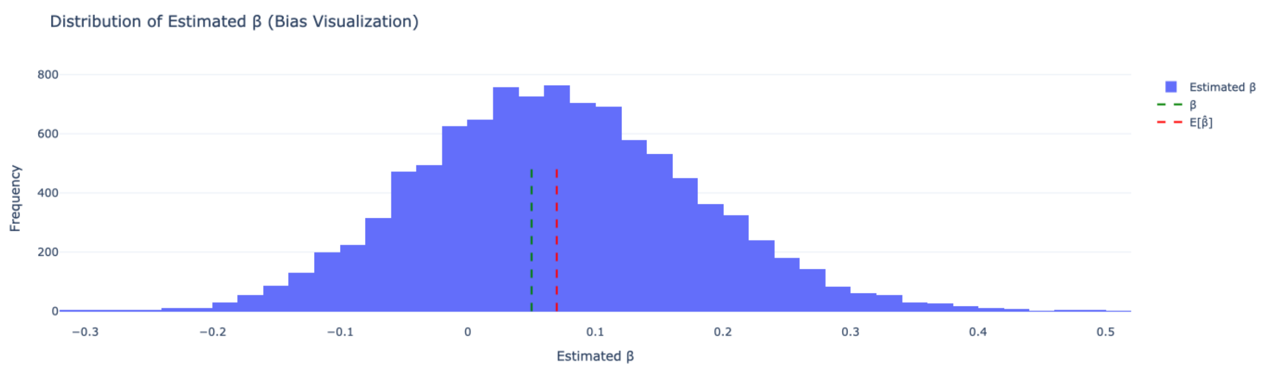

Figure 2: Empirical distribution of $\hat{\beta}$ for $N = 1000$ simulations at $T = 100$.

The persistent regressor and correlated errors drive this small-sample bias; as we increase T, the bias diminishes as expected.

4 Question 4: Bias of $\hat{\beta}$ as a Function of T

Here we vary T over $\{30, 50, 100, 200, 500\}$ (and so forth), generate N simulations for each T, and compute the bias. Figure 3 plots the resulting average bias. As predicted, the bias is the largest for small T and shrinks toward zero as T grows.

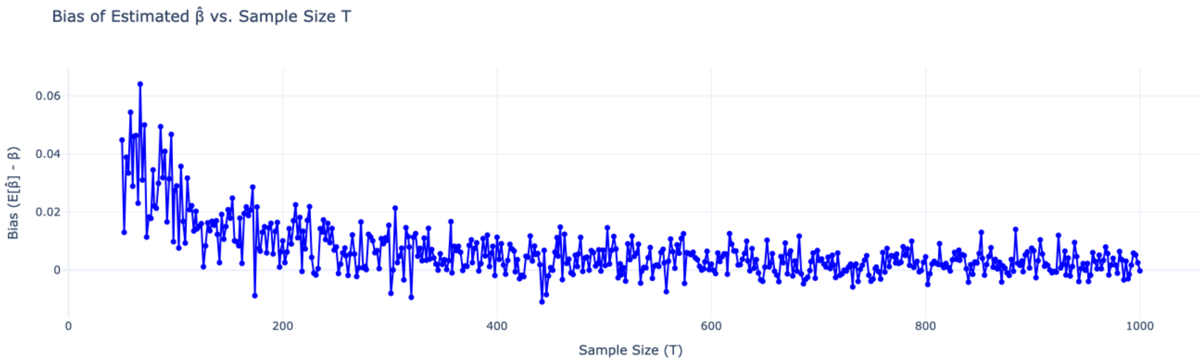

Figure 3: Average bias of $\hat{\beta}$ across simulations for varying T.

5 Question 5: Regression of Bias on $1/T$ and $1/T^2$

Finally, we take the bias estimates Bias_i at each of five sample sizes T_i and run the cross-sectional OLS:

$$\text{Bias}_i \;=\; \gamma_0 \;+\; \gamma_1 \frac{1}{T_i} \;+\; \gamma_2 \frac{1}{T_i^2} \;+\; \varepsilon_i.$$

The regression output below show a good fit ($R^2 = 0.530$) but also the limited degrees of freedom (only five T_i values):

Table 2: OLS regression of bias on $1/T$ and $1/T^2$.

	$\hat{\gamma}$	Std. Err.	t-stat	p-value
Constant	0.0003	0.001	0.126	0.900
$1/T$	1.8212	0.250	7.274	0.000
$1/T^2$	8.7301	15.959	0.547	0.585
$R^2 = 0.530$, Adj.$R^2 = 0.528$.				

- The $1/T$ and $1/T^2$ terms capture the leading small-sample expansion of the bias predicted by Stambaugh (1999).
- Despite that, the overall regression "shape" tracks the bias well (high R^2).

The regression results provide key insights into the small-sample bias structure:

- **Significance of $1/T$**: The coefficient estimate $\hat{\gamma}_1 = 1.8212$ is highly significant ($t = 7.274$, $p = 0.000$), confirming that the leading-order term in the bias follows an $O(1/T)$ structure, consistent with Stambaugh (1999).
- **Insignificance of $1/T^2$**: The coefficient $\hat{\gamma}_2 = 8.7301$ is not statistically significant ($t = 0.547$, $p = 0.585$). This suggests that the higher-order term $O(1/T^2)$ may not be as relevant in explaining the bias for the given sample sizes.
- **Interpretation of the constant term**: The estimated intercept $\hat{\gamma}_0 = 0.0003$ is statistically insignificant ($p = 0.900$), indicating no meaningful bias independent of $1/T$ and $1/T^2$.
- **Overall model fit**: The relatively high R^2 value (0.530) suggests that the regression effectively captures the bias structure, even though only five sample sizes are used.

Stambaugh (1999) derives the small-sample bias in the estimation of the first-order autoregressive coefficient ρ when the regressor is persistent. The bias in the OLS estimate $\hat{\rho}$ is approximated by:

$$E[\hat{\beta} - \beta] = \frac{\sigma_{uv}}{\sigma_v^2} E[\hat{\rho} - \rho] = -\frac{\sigma_{uv}}{\sigma_v^2} \left(\frac{1 + 3\rho}{T}\right) + O\left(\frac{1}{T^2}\right).$$

For the given parameters:

- $\sigma_{uv} = -0.5$, $\sigma_v^2 = 0.5$, and $\rho = 0.3$,
- The theoretical coefficient for $1/T$ is computed as:

$$\gamma_1 = \frac{-\sigma_{uv}}{\sigma_v^2}\left(\frac{1+3\rho}{T}\right) = \frac{-(-0.5)}{0.5}\left(\frac{1+3(0.3)}{T}\right) = \frac{1.9}{T}.$$

- This aligns closely with the estimated $\hat{\gamma}_1 = 1.8212$, validating Stambaugh's prediction that bias is dominated by an $O(1/T)$ term.

The insignificance of γ_2 ($p = 0.585$) suggests that the higher-order correction $O(1/T^2)$ is less influential. This confirms Stambaugh's assertion that the leading bias term is primarily $O(1/T)$, with the $O(1/T^2)$ term contributing only marginally in finite samples. The high p-value for γ_2 reinforces the idea that practical applications may not need additional higher-order bias corrections in small samples.