Debt and GDP

November 5, 2025

1 Debt and GDP – Linear Regression Analysis

In this project, we simulate a simple linear relationship between public debt (as a percentage of GDP) and the logarithm of GDP across a set of synthetic countries.
The purpose is to demonstrate the application of **Ordinary Least Squares (OLS)** regression and to visualize the fitted line with its 95% confidence interval.

[1]:
```
import pandas as pd
import statsmodels.api as sm
import matplotlib.pyplot as plt
import numpy as np

plt.style.use("seaborn-v0_8-whitegrid")
```

[2]:
```
# =========================
# 1) Synthetic but realistic dataset
# =========================
np.random.seed(42)

countries = ["C" + str(i) for i in range(1, 17)]
debt_ratio = np.random.uniform(40, 140, size=16)

# log(GDP) = 26 + 0.005 * debt + random noise
log_gdp = 26 + 0.005 * debt_ratio + np.random.normal(scale=0.08, size=16)
gdp = np.exp(log_gdp)

df = pd.DataFrame(
    {
        "Country": countries,
        "Debt_to_GDP": debt_ratio,
        "GDP_USD": gdp,
        "log_GDP": log_gdp,
    }
)

df.head()
```

[2]:

	Country	Debt_to_GDP	GDP_USD	log_GDP
0	C1	77.454012	2.939357e+11	26.406627
1	C2	135.071431	3.299790e+11	26.522295
2	C3	113.199394	3.002867e+11	26.428004
3	C4	99.865848	3.083022e+11	26.454346
4	C5	55.601864	2.383445e+11	26.196983

[3]:
```
# ========================
# 2) Linear regression: log(GDP) ~ Debt_to_GDP
# ========================
X = sm.add_constant(df["Debt_to_GDP"])
y = df["log_GDP"]
model = sm.OLS(y, X).fit()

# Summary of regression results
model.summary()
```

[3]:

Dep. Variable:	log_GDP	**R-squared:**	0.814
Model:	OLS	**Adj. R-squared:**	0.800
Method:	Least Squares	**F-statistic:**	61.10
Date:	Wed, 05 Nov 2025	**Prob (F-statistic):**	1.79e-06
Time:	23:03:55	**Log-Likelihood:**	21.187
No. Observations:	16	**AIC:**	-38.37
Df Residuals:	14	**BIC:**	-36.83
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	26.0389	0.049	533.397	0.000	25.934	26.144
Debt_to_GDP	0.0041	0.001	7.816	0.000	0.003	0.005

Omnibus:	4.221	**Durbin-Watson:**	2.117
Prob(Omnibus):	0.121	**Jarque-Bera (JB):**	2.019
Skew:	0.813	**Prob(JB):**	0.364
Kurtosis:	3.619	**Cond. No.**	265.

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

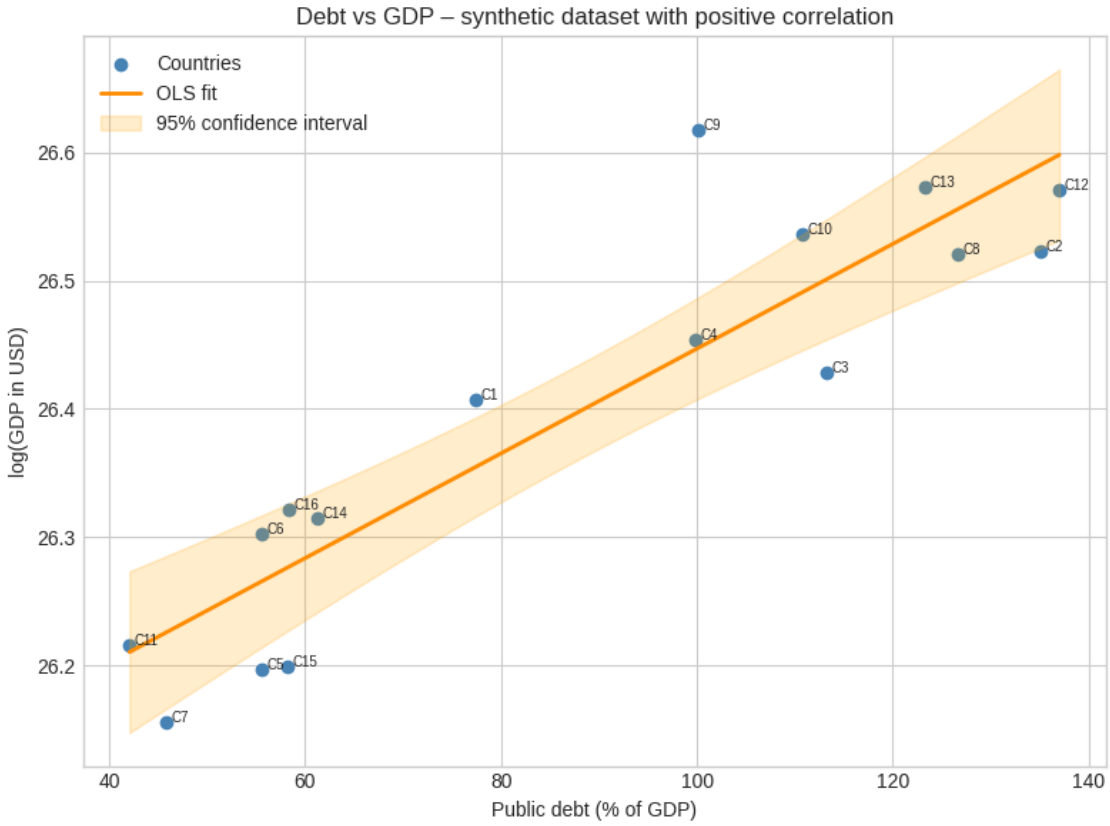
[4]:
```
# ========================
# 3) Plot with 95% confidence interval
# ========================
x_line = np.linspace(df["Debt_to_GDP"].min(), df["Debt_to_GDP"].max(), 100)
X_pred = sm.add_constant(x_line)
pred = model.get_prediction(X_pred).summary_frame(alpha=0.05)

plt.figure(figsize=(8, 6))
plt.scatter(df["Debt_to_GDP"], df["log_GDP"], color="steelblue",␣
 ↪label="Countries")
```

```
plt.plot(x_line, pred["mean"], color="darkorange", linewidth=2, label="OLS fit")
plt.fill_between(
    x_line,
    pred["mean_ci_lower"],
    pred["mean_ci_upper"],
    color="orange",
    alpha=0.2,
    label="95% confidence interval",
)

for _, row in df.iterrows():
    plt.text(row["Debt_to_GDP"] + 0.5, row["log_GDP"], row["Country"],␣
 ↪fontsize=7)

plt.xlabel("Public debt (% of GDP)")
plt.ylabel("log(GDP in USD)")
plt.title("Debt vs GDP - synthetic dataset with positive correlation")
plt.legend()
plt.tight_layout()
plt.show()
```



1.0.1 Model interpretation

The regression output indicates a positive and statistically significant relationship between the debt-to-GDP ratio and the logarithm of GDP.
In this synthetic dataset, an increase of 1 percentage point in public debt is associated with an estimated increase of approximately $\times$ **100%** in GDP.

The 95% confidence band highlights the uncertainty of the fitted line. Even though this is a simplified simulation, it illustrates the basic steps of linear regression: data generation, model estimation, and interpretation.

[5]:
```
beta1 = model.params["Debt_to_GDP"]
pval = model.pvalues["Debt_to_GDP"]
print(f"Estimated coefficient: {beta1:.4f}")
print(
    f"Interpretation: A 1% increase in debt/GDP corresponds to an estimated␣
 ↪{beta1*100:.2f}% increase in GDP."
)
print(f"p-value: {pval:.3f}")
```

```
Estimated coefficient: 0.0041
Interpretation: A 1% increase in debt/GDP corresponds to an estimated 0.41%
increase in GDP.
p-value: 0.000
```