Executive Summary: Bayesian Econometrics, Python, and AI for GDP per Capita Analysis

Executive Summary

This summary examines a Master's thesis focused on the application of **Bayesian econometrics**, implemented in **Python**, to analyze the determinants of **Gross Domestic Product (GDP) per capita**. The study rigorously compares this approach with classical econometric methods, notably **Ordinary Least Squares (OLS)** regression, and explores an extension to **artificial intelligence (AI) models** for predictive purposes.

Key findings indicate that **public spending on health and governance** are significant and positive drivers of GDP per capita growth. The effect of **education spending** is more complex, showing a negative effect in full models, potentially suggesting temporal lags or inefficiencies in resource allocation.

Comparative analysis shows that while OLS and Bayesian models converge on point estimates, the **Bayesian approach provides superior uncertainty quantification**. Bayesian credible intervals allow for direct probabilistic interpretation, more intuitive and informative for decision-making than classical confidence intervals. In terms of predictive performance, the Bayesian model shows slight superiority over OLS, with marginally lower errors (RMSE and MAE).

An extension using a **Random Forest (AI) model** demonstrated even better predictive performance. However, this advantage comes at the cost of interpretability, as the model operates as a "black box," reinforcing the value of Bayesian analysis for studies aimed at both prediction and causal explanation. In conclusion, the thesis establishes **Bayesian econometrics, supported by modern Python tools**, as a powerful, flexible framework particularly suited to the complexity and uncertainty inherent in macroeconomic data.

1. Research Problem and Context

The research is set in the context of modern economic analysis, where understanding the factors influencing GDP per capita is crucial for public policy. Traditional econometric methods (OLS, Maximum Likelihood) rely on strong assumptions (normality, homoscedasticity) and struggle to handle uncertainty, especially with small samples or noisy data.

Bayesian econometrics is presented as a relevant alternative. By treating parameters as random variables and updating prior knowledge with data via Bayes' theorem, this approach allows for **more flexible modeling** and better uncertainty representation.

The central research question is:

Can Bayesian econometrics provide a better understanding and forecasting of GDP per capita compared to classical methods?

To answer this, the thesis pursues three objectives:

1. Present Bayesian econometrics concepts and compare them with classical approaches.
2. Build a Bayesian model applied to GDP per capita analysis using macroeconomic variables.
3. Compare Bayesian and classical results to assess the added value of the Bayesian philosophy.

2. Theoretical Foundations and Methodological Framework

2.1 Limitations of Classical Econometrics

Frequentist methods such as OLS have structural limitations:

- **Small sample sizes:** undermine asymptotic properties and reliability of statistical tests.
- **Multicollinearity:** produces unstable and hard-to-interpret coefficient estimates.
- **Missing or outlier data:** challenging to handle without ad hoc techniques.
- **Limited uncertainty management:** confidence intervals have a non-probabilistic interpretation, often misused for decision-making.

2.2 Contributions of Bayesian Econometrics

Bayesian methods address these limitations through:

- **Uncertainty management:** parameters are probability distributions, allowing credible intervals with direct probabilistic interpretation.
- **Flexibility:** enables integration of prior information, useful with sparse or costly data.
- **Adaptation to complex models:** handles hierarchical structures, non-normal errors, and missing data effectively.
- **Computational advances:** MCMC methods and advanced algorithms (NUTS, HMC) via Python libraries (PyMC) allow practical large-scale application.

2.3 Synergies with Artificial Intelligence

The thesis highlights the growing synergy between Bayesian econometrics and AI. AI excels with massive, unstructured data, while Bayesian methods provide a rigorous probabilistic framework to quantify model uncertainty. Techniques such as **Bayesian deep learning** (e.g., Bayes by Backprop, Dropout) estimate probability distributions over neural network weights, producing robust predictions with uncertainty measures.

3. Data, Variables, and Exploratory Analysis

3.1 Data Source and Structure

Data are drawn from the **World Development Indicators (WDI)** database (World Bank). Panel data are annual for 30 countries, covering 2000–2023. WDI is chosen for reliability, international comparability, and relevance for economic development analysis.

3.2 Variables

Full Variable Name	Short Name	Description
GDP per capita (current USD)	log_gdp	Dependent variable, log-transformed to reduce skewness.
Total public expenditure (% GDP)	gov_spend	Measures state involvement in the economy.
Education expenditure (% GDP)	edu_spend	Proxy for human capital investment.
Health expenditure (% GDP)	health_spend	Proxy for population well-being and productivity.
Unemployment rate (%)	unemp	Reflects underemployment of labor resources.
Education × Health	edu_health	Derived variable to study the combined effect of human capital investment.

3.3 Descriptive Analysis

- **Univariate statistics:** High heterogeneity, e.g., gov_spend ranges from 4.26% to 83.61% of GDP; unemp up to 34.15%.
- **Correlation matrix:** log_gdp positively correlated with health (0.57), gov_spend (0.54), and edu_spend (0.32). Unemployment shows weak correlation with other variables.
- **PCA:** First two components explain 71.8% of total variance; PC1 (54.8%) is a socio-economic axis linking GDP to public, health, and education spending; PC2 (17.0%) is primarily driven by unemployment.

4. Modeling, Results, and Comparison

Three research hypotheses were tested by comparing OLS and Bayesian models.

Hypothesis	Status
H1: Public spending in education and health positively affects GDP per capita	Validated
H2: Bayesian approach provides better uncertainty quantification	Validated
H3: Bayesian model offers better predictive performance than OLS	Partially validated

4.1 OLS Results

- **Restricted model (education & health):** Explains 33.4% of GDP variance ($R^2 = 0.334$). Education spending increases GDP by 9.7%, health by 28.1%. Residuals show strong autocorrelation (Durbin-Watson = 0.173) and non-normality.
- **Expanded model:** Adjusted $R^2 = 0.427$. Public and health spending positive; education spending negative; unemployment negative, consistent with theory.

4.2 Bayesian Modeling Results

Estimated using weakly informative priors via MCMC (NUTS). Convergence diagnostics excellent (R-hat = 1.0).

- **Restricted model:** Confirms positive effect of education and health with narrow credible intervals.
- **Expanded model:**
  - Health spending: $\beta = 0.451$, 95% CI [0.305, 0.598]
  - Total public spending: $\beta = 0.554$, 95% CI [0.502, 0.604]
  - Education spending: $\beta = -0.150$, 95% CI [–0.266, –0.031]
  - Unemployment: $\beta = -0.168$, 95% CI [–0.211, –0.127]

Bayesian results broadly confirm OLS findings but with **more rigorous uncertainty quantification**.

4.3 Predictive Performance and AI Extension

Method	Mean RMSE	Mean MAE
OLS	1.0993	0.8116
Bayesian (MCMC)	1.0954	0.8095
Random Forest (AI)	0.7311	0.4890

Bayesian slightly outperforms OLS, capturing structural uncertainty better. Random Forest provides higher predictive accuracy but lower interpretability, reinforcing the Bayesian model's value for prediction **and causal explanation**.

5. Conclusion, Limitations, and Future Research

5.1 Contributions

Bayesian econometrics adds value to macroeconomic analysis by:

1. Rigorously quantifying uncertainty.
2. Building flexible models adapting to real-world economic data constraints.
3. Achieving robust predictive performance, slightly exceeding classical methods.

Synergy with AI is noted: AI for pure prediction, Bayesian methods for balance between prediction, interpretability, and inferential rigor.

5.2 Limitations

- Cross-sectional snapshot: Temporal dynamics not fully captured.
- Computational cost: MCMC can be slow for large datasets.

- Variable selection: Model may omit important structural factors.

5.3 Future Research

- Use **variational inference** to accelerate Bayesian computations.
- Develop **dynamic Bayesian models** for temporal effects.
- Integrate **unstructured data** (text, images) via AI for richer economic modeling.