



UNIWERSYTET
WARSZAWSKI



UNIWERSYTET WARSZAWSKI
Wydział Nauk Ekonomicznych

Real GDP Estimator with Keynesian Regressors

Ege Güney Kıymaç

Index No. 447987

e.kiymac@student.uw.edu.pl

Sung Kwan Chiu

Index No. 444585

s.chiu@student.uw.edu.pl

November 26, 2025

Contents

1	Introduction	3
1.1	Overview	3
1.2	Hypothesis	3
1.3	Literature Review	3
2	Data	4
2.1	Series Characteristics	4
2.1.1	Real GDP	4
2.1.2	CPI	4
2.1.3	Unemployment Rate	4
3	Methodology	5
3.1	Data Preparation	5

1 Introduction

1.1 Overview

Real GDP is a prominent measure of output due to its simplicity, interpretability, and effectiveness in communicating the overall activity of an economy. Alongside academic settings, Real GDP has retained its relevance in politics as one of the go-to measures in public communications regarding economic growth and prosperity. As per its wide-spread adoption, the measure has accumulated a plethora of literature referencing it from a multitude of perspectives.

This paper aims to show that empirical observations of Real GDP in the US economy can be modelled optimally as BLUE through an $ARX(p, q)$ model where the exogenous contributors are theoretical descriptors of output.

1.2 Hypothesis

In the history of economic equilibria, multiple models have proposed mathematically defined relationships between Real GDP and many indicators. From the Keynesian approach, two of said indicators are:

- Price levels (CPI) through the **Aggregate Demand Curve**
- Unemployment rates through **Okun's Law**

The formal hypothesis of this paper is that an $ARX(p, q)$ representation of Real GDP, incorporating CPI and unemployment:

$$ARX_y(p, q) = C + \sum_{i=1}^p \beta_i y_{t-i} + \left[\sum_{j=1}^q \phi_j CPI_{t-j} + \gamma_j u_{t-j} \right] + \varepsilon_t$$

yields OLS parameter estimates that satisfy the **Gauss–Markov** conditions for time-series estimators.

1.3 Literature Review

...

2 Data

The proposed ARX model relies on two exogenous variables alongside the autoregressive target. All three variables are sourced from the Federal Reserve Bank of St. Louis database “FRED”. The implementation uses quarterly data from 1990Q1 to 2025Q1 and uses the FRED series:

- Real GDP (y): “GDPC1”
- CPI (CPI): “CPIAUCSL”
- Unemployment Rate (u): “UNRATE”

NOTE: All series are seasonally adjusted by FRED.

2.1 Series Characteristics

2.1.1 Real GDP

FRED’s Real GDP series is measured in chained 2017 dollars (Bn). The series is updated quarterly, which matches the frequency of the model natively.

2.1.2 CPI

The CPI series of choice for the model records the US-wide city average CPI for all urban consumers. The series is expressed as a chain index with base period 1982-1984=100. Entries are recorded monthly, and quarter-end observations are used to align with the Real GDP series.

2.1.3 Unemployment Rate

Unemployment rate in FRED is recorded as percentages with a monthly frequency. Similar to CPI, quarter-end observations are used to match the frequencies among the series.

3 Methodology

The proposed methodology of this paper can be split into four main components:

- Data Preparation
- Model Specification
- Model Estimation
- BLUE Evaluation

Each component is detailed in the following subsections.

3.1 Data Preparation

The data preparation for this project focuses on transforming the raw data for interpretability without effecting the underlying distributions. Rescaling the input matrix (X) allows better comparison of coefficients and ensures that $\sum X^T \varepsilon = 0$ must hold within very tight tolerances¹. A common approach to achieve this stability is standardization, which transforms each feature to have a mean of 0 and a standard deviation of 1. Standardization is performed column-wise on the input using the formula:

$$Z = \frac{X_{i,j} - \bar{X}_i}{S_i}$$

where $X_{i,j}$ is the j -th observation of the i -th column, \bar{X}_i is the mean of the i -th column, and S_i is the [unbiased] standard deviation of the i -th column.

¹Differences in magnitude between X columns can lead to numerical ambiguity in orthogonality checks. It is generally preferable to eliminate any differences in the order of magnitudes in X .