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 29, 2025

Prepared under the supervision of Dr. A. W. Kolasa.

Contents

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 for 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 U.S. economy can be modelled optimally with a **Gauss–Markov** compliant $ARX(p, q)$ specification 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:

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

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 the following 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 three main components:

- Data Preparation and Examination
- Model Estimation
- BLUE Evaluation

Each component is detailed in the following subsections.

3.1 Data Preparation and Examination

Data preparation does not involve any steps that are specific to a given column of the dataset. Therefore, $X_{i,j}$ notation will be used to denote any given column/observation and β_i will be used to denote any given coefficient. In addition to formatting the X matrix to match the model structure, the preparation steps emphasize transforming the raw data for interpretability while preserving its underlying distributions.

3.1.1 Dataset Structure Validation

More often than not, the performance of a non-robust, linear estimator is governed by the quality of the input variables. Ideally, we look for a dataset that's **free of multicollinearity, extreme outliers, and missing values** while being **correlated to the target**. This general construction achieves a model where all variables have uniquely identifiable contributions to the output, often leading to lower variances and better predictive power.

Autoregressive models present a challenge in this regard since variables selected to be correlated to the target (y) are also correlated to the lagged representation of it within X. Acknowledging this issue, the endogenous portion of the model will only be checked to ensure there are no singularities by perfect correlation inside X. The presence/absence of multicollinearity will be displayed through a correlation heatmap.

Furthermore, data continuity will be prioritized over outlier removal since the model itself is time-dependent. Outliers, if any, will be addressed if they are found to be damaging to the model's performance after the initial estimation.

Finally, any missing values will be handled through linear interpolation of the observations prior and after the occurrence.

3.1.2 Design Matrix Formation

X is formed by converting the model equation specified in Equation 1 into matrix form. The inner product operation specified as $ARX(p,\ q)$ can be written as a matrix multiplication of observations and coefficients:

$$\begin{bmatrix} 1 & X_{1,1} & \cdots & X_{1,m} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n,1} & \cdots & X_{n,m} \end{bmatrix} \times \begin{bmatrix} C \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix} \tag{2}$$

Where n is the number of observations and m is the number of features (including the intercept) dictated by the $\{p,\ q\}$ orders of the model.

As shown in the above equation, a closed form matrix representation of the model requires p and q to be determined. In this paper, the selection of these hyperparameters will follow the simple and idealistic approach of selecting the lowest possible orders that lead to an efficient estimator.

3.1.3 Model Order Selection

First step of the design matrix creation is the identification of optimal $\{p,\ q\}$ parameters for the model. There are two distinct methods to select the autoregressive and exogenous lag order of the model:

- **Autoregressive Order** — "Autocorrelation Function" (ACF) and "Partial Autocorrelation Function" (PACF) plots of the dependent variable are examined to identify significant lags. The ACF plot helps to identify the overall correlation structure, while the PACF plot isolates the direct effect of each lag.
- **Exogenous Order** — Lagged correlation matrices and "Cross-Correlation Function" (CCF) plots between the dependent variable and each exogenous variable are analyzed to determine significant lags. The CCF plot reveals the correlation between the dependent variable and lagged values of the exogenous variables.

While ACF and PACF can directly be applied to the dependent variable, CCF results can be heavily distorted by nonstationary and the presence of AR structures in the variables being compared. Therefore, the **Box–Jenkins Prewhitening** procedure will be applied to expose a clearer signal of the underlying cross-correlations. The procedure transforms the exogenous variables into residuals of their AR representation, then uses the β coefficient to re-scale the dependent variable. (Hackhard and Romer, 2025) For our purposes, the process will be applied to normalized data, making the inclusion of an intercept redundant. Formally, the $AR(1)$ variant of the process can be shown as:

$$\begin{aligned} \hat{X}_i &= (L^1 X_i)\beta_i \\ \hat{y}_i &= y\beta_i \\ \rho_{y,i} &= \text{corr}(\hat{y}_i, X_i - \hat{X}_i) \end{aligned} \tag{3}$$

Where L^1 is the lag-1 operator.

3.1.4 Rescaling

Rescaling the input matrix (X) allows better comparison of coefficients and ensures that $\sum X^T \varepsilon = 0$ must hold within very tight tolerances[1]. 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:

$$Z = \frac{X_{i,j} - \bar{X}_i}{S_i} \tag{4}$$

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

This transformation has no effect on the error profile of the fitted model as the distribution shapes and inter-variable interactions are perfectly preserved.[2] However, the transformation brings all variables to the same scale and units. As a result, the **Kolmogorov–Smirnov (KS) test** which is sensitive to such differences now becomes available.

3.1.5 Normalized Input Distributions

With each regressor standardized to zero mean and unit variance, a one unit change corresponds to a 1σ shift in the original scales. However, this unit equality does not make the "likelihood" (or "extremeness") of different variables comparable. The "Z-scores" are comparable in likelihood only if the variables subject to comparison are identically distributed. This part of the methodology works towards identifying such equalities in X.

Concretely, the aforementioned KS test can be used to check for equality of distributions. However, the KS' assumption of **i.i.d. samples** is virtually impossible to satisfy in time-series data. To address this, a modification of the **Bootstrap KS Test** (Præstgaard, 1995) can be applied.

[1] Differences 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.

[2] The statement holds for any non-penalized estimator of the form $\hat{\beta} = \arg\min_\beta L(y, X\beta)$, since rescaling X can be absorbed into β.

The specific test statistic is computed by calculating

$$D_{i,j} = \sup_x |F_{X_i}(x) - F_{X_j}(x)|$$

Where F_{X_i} and F_{X_j} are the Empirical CDFs of the respective regressors. Afterward, a comparison of the bootstrapped distribution of $D^*_{i,j}$ to the observed $D_{i,j}$ is performed.

The empirical p-value is estimated as the proportion defined by:

$$p_{i,j} = \frac{1}{B+1}\left(\sum_{b=1}^{B}(\mathbb{1}_{\{D^{*(b)}_{i,j} \geq D_{i,j}\}}) + 1\right) \tag{5}$$

Where B is the number of bootstrap samples. To preserve the in-block serial dependence of observations, we employ a **Block Bootstrap** approach instead of the standard bootstrap resampling.[3]

3.2 Model Estimation

...

3.3 BLUE Evaluation

...

[3]See Section 5.1 for further discussion on the time-series application of the KS test.

4 Results

4.1 Design Matrix

First step of the design matrix creation is the identification of optimal $\{p,\ q\}$ parameters for the model.

4.1.1 Autoregressive Order

Application of ACF yielded a very common pattern of monotonically decreasing correlations, not outlining any specific lag order as superior. Looking for more conclusive insights, PACF was applied.

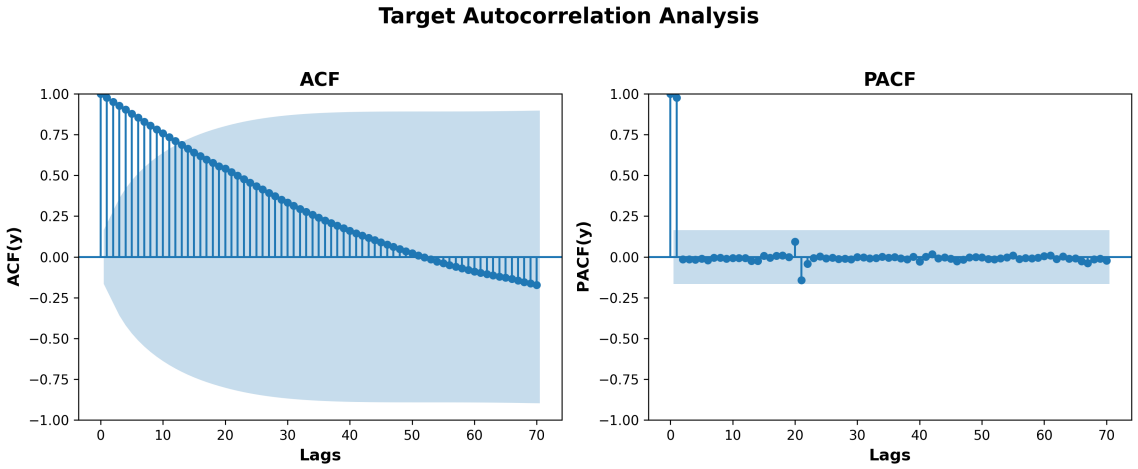

Figure 1: ACF and PACF plots of the dependent (target) variable.

Figure 4.1.1 clearly shows the aforementioned decay in ACF. Conversely, PACF shows lag-1 is the only entry with actual partial contribution into the descriptive structure of y.[4] Interestingly, the small spikes around lag-20 of the PACF plot (corresponding to 20 quarters or 5 years) indicate that the COVID-19 shock still has a minor effect on the 2025 Real GDP.

In any case, the visualization of the autocorrelation structure presents strong evidence towards the $p = 1$ selection, which is used for the model generation in this paper.

4.1.2 Exogenous Order

The exogenous order selection process started with the visualization of the simple correlation vector y and the lags of exogenous variables.

[4]The first spike in both ACF and PACF refers to lag-0 (the variable itself) which is equal to 1 by definition.

$\mathbf{i_{lag}}$	$\mathbf{corr(y,\ CPI_{t-i})}$	$\mathbf{corr(y,\ u_{t-i})}$
0	0.989	-0.221
1	0.988	-0.255
2	0.988	-0.221
3	0.988	-0.190
4	0.988	-0.236
5	0.988	-0.205

Table 1: Pearson correlation coefficients between the dependent variable and lags of exogenous variables.

Inspecting the table of coefficients, we see that CPI consistently stays at a coefficient of 0.988. This indicates that the autoregressive process of CPI is likely very strong, and related to the autoregressive process explaining y. On the other hand, u shows a more reasonable set of coefficients. Although coefficients of u are also relatively stable, they reside in the reasonable range of $[-0.255,\ -0.190]$. However, a lack of decay or noticeable regime changes in the matrices leaves this simple test inconclusive. This inconclusiveness was expected, and the CCF plots of whitened variables were examined for clearer insights.

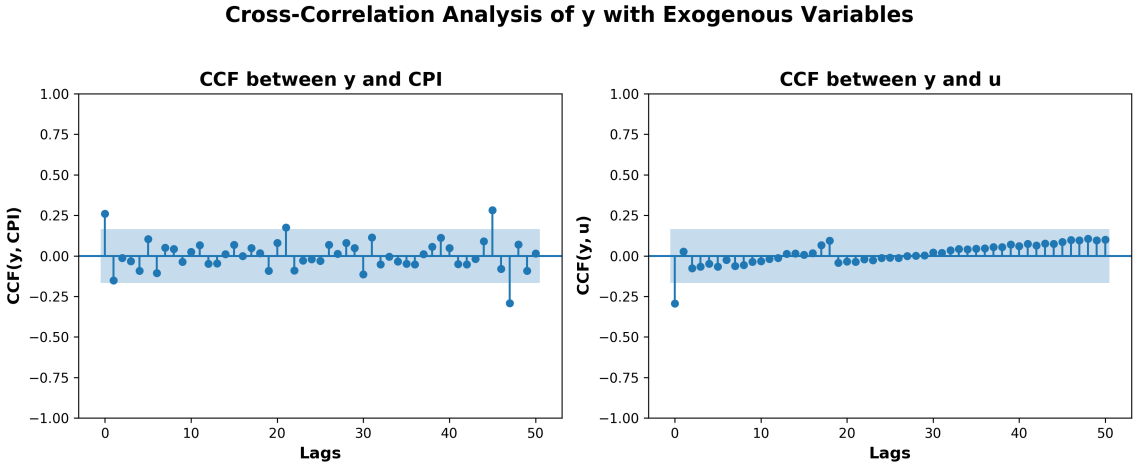

Figure 2: CCF plots of the whitened exogenous variables and the dependent variable.

Although not as clear as the PACF results, Figure 4.1.2 carries much stronger insights into the cross-correlation structure compared to raw correlation coefficients. CPI shows several spikes above the significance bounds with notable lags being $\{0, 45, 47\}$. For u, we observe that lag-0 is the only significant entry. Based on the observations, lag-0 is the only selection that ensures significant lags of both exogenous variables are included. Therefore, the cross-correlation analysis yields $q = 0$ to be the optimal selection.

Using $q = 0$ introduces the problem of **data leakage**, as current quarter values of exogenous variables would not be available at the time of estimation. This results in a design choice of either using the closest available selection in time (i.e. lag-1) or in significance (selection of lag-45 or lag-47 for CPI). Since introducing 45 or 47 more coefficients per variable would essentially guarantee overfitting, the final decision was in favor of using $q = 1$.[5]

4.1.3 ARX Equation

With the order specification finalized, we were left with an $ARX(1,\ 1)$ model. The objective equation is as follows:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 CPI_{t-1} + \beta_3 u_{t-1} + \varepsilon_t \tag{6}$$

At this stage, all model variables were known, and rescaling was performed to standardize the data.

4.2 Model Fit

The fit was performed using Ordinary Least Squares (OLS) on the matrix expansion of Equation 6.

4.3 Residual Analysis

...

4.4 Gauss–Markov Conditions

...

4.5 Residual Distribution

...

[5] Further regularization techniques, or using sparse lagged estimators are indeed possible but are deemed outside the scope of this paper.

5 Remarks

5.1 Remark 1: Time-Series Application of KS Test

The i.i.d. assumption of the classic KS test was addressed by Præstgaard with a bootstrapping approach. Yet, the examples, lemmas, and theorems presented in the paper were derived from row-exchangeable samples.

With row-exchangeability, Præstgaard was able to apply standard bootstrap resampling while preserving the continuity of the empirical distribution. However, time-series data inherently requires the row-structure to be preserved; hence rendering the standard bootstrap inappropriate. To preserve the time-dependent dynamics, our methodology elected to use a **Block Bootstrap** approach. This specific adaptation is not documented nor theoretically justified in Præstgaard's work; consequently, while we regard it as a reasonable and practically reliable extension for the purposes of this study, it remains an empirical choice that invites further theoretical investigation.

A block bootstrap 1-sample KS test was derived in a recent preprint paper (Chandy et al., 2025) and applications of bootstrap techniques in KS style tests were demonstrated in multiple studies. (Psaradakis, 2003 & Wyłupek, 2023) These works place bootstrap KS tests on a more solid theoretical footing, even though our specific use case is undocumented.

6 Bibliography

References in the Paper

Chandy, M., Schifano, E., Yan, J., & Zhang, X. (2025, November). Nonparametric Block Bootstrap Kolmogorov-Smirnov Goodness-of-Fit Test [arXiv:2511.05733 [stat]]. https://doi.org/10.48550/arXiv.2511.05733

Hackhard, B., & Romer, M. (2025, July). 9 Prewhitening; Intervention Analysis – STAT 510 — Applied Time Series Analysis. Retrieved November 28, 2025, from https://online.stat.psu.edu/stat510/Lesson09

Præstgaard, J. (1995). Permutation and Bootstrap Kolmogorov-Smirnov Tests for the Equality of Two Distributions. *Scandinavian Journal of Statistics*, *22*(3), 305–322. Retrieved November 28, 2025, from https://www.jstor.org/stable/4616362

Psaradakis, Z. (2003). A Bootstrap Test for Symmetry of Dependent Data Based on a Kolmogorov–Smirnov Type Statistic [Publisher: Taylor & Francis _eprint: https://doi.org/10.1081/SAC-120013116]. *Communications in Statistics - Simulation and Computation*, *32*(1), 113–126. https://doi.org/10.1081/SAC-120013116

Wyłupek, G. (2023). A nonparametric test for paired data [Publisher: Elsevier]. *Journal of Multivariate Analysis*, *198*(C). https://doi.org/10.1016/j.jmva.2023.105229

References in the Source Code

David, D., & Wayne, F. (1979). Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association*, *74*(366), 427–431. https://doi.org/10.2307/2286348

MacKinnon, J. (1996). Numerical Distribution Functions for Unit Root and Cointegration Tests. *Journal of Applied Econometrics*, *11*(6), 601–618. Retrieved November 21, 2025, from https://www.jstor.org/stable/2285154

Præstgaard, J. (1995). Permutation and Bootstrap Kolmogorov-Smirnov Tests for the Equality of Two Distributions. *Scandinavian Journal of Statistics*, *22*(3), 305–322. Retrieved November 28, 2025, from https://www.jstor.org/stable/4616362

Royston, J. P. (1982). An Extension of Shapiro and Wilk's W Test for Normality to Large Samples. *Journal of the Royal Statistical Society*, *31*(2), 115–124. https://doi.org/10.2307/2347973

Royston, J. P. (1995). Remark AS R94: A Remark on Algorithm AS 181: The W-test for Normality. *Journal of the Royal Statistical Society*, *44*(4), 547–551. https://doi.org/10.2307/2986146

Shapiro, S. S., & Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrica*, *52*(3/4), 591–611. https://doi.org/10.2307/2333709