Term Project in Introduction to Econometrics of Time Series

An analysis of the United Kingdom GDP, Trade Balance, and Exchange Rate, 1955-2024

Souleymane Faye & Gereon Staratschek

2025-05-24

1 Introduction

In this project, we aim to analyze the development of the Gross Domestic Product (GDP), exchange rates, and trade balance of the United Kingdom (UK). We collect our data from the OECD open data portal. We use data on a quarterly basis, starting latest in 1997. This period covers important financial events such as the Global Financial Crisis (GFC) in 2007, the Brexit referendum in 2016 and UK's final EU leave in 2020 as well as the COVID-19 pandemic from 2020-2022.

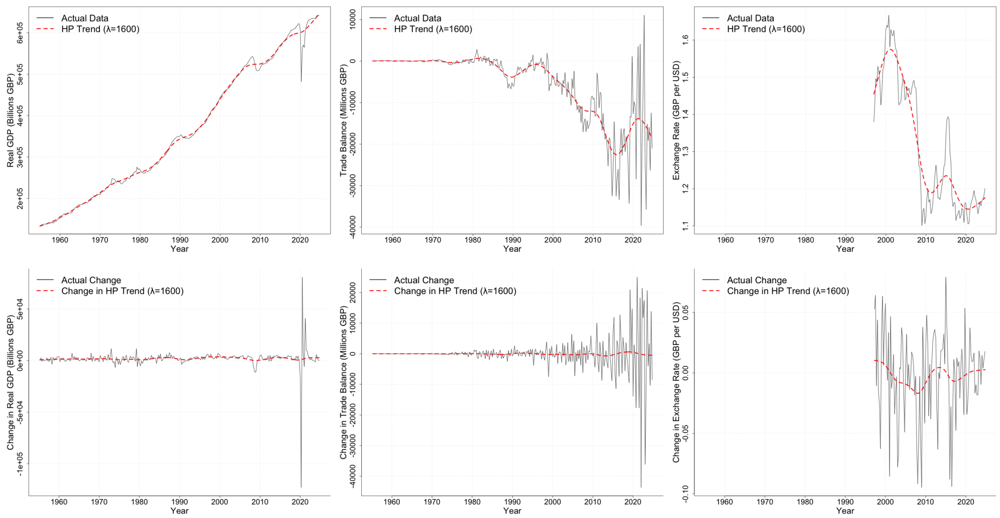
2 Univariate Analysis

2.1 Macro trends

Figure 1 illustrates the evolution of the UK's macroeconomic indicators—GDP, Trade Balance, and Exchange Rate—in both levels and first differences.

Trend Dynamics and Structural Breaks in Level Series. Data on UK's GDP, available on a quarterly basis since 1955, exhibits a clear, upwards trajectory with only a few shocks such as the 2007 financial crisis or the 2020 COVID-19 pandemic interrupting the general trend. Hence, the GDP of the UK is clearly not stationary. However, the trend seems to be linear, suggesting that the first-differences time series of the GDP might be stationary. Similarly, data on the trade balance is available on a quarterly basis starting in 1955.We see that the trade balance fluctuates around 0 until the 1990s where a negative trend seems to set in continueing until the 2010s, when trade balance starts fluctuating around a low, negative value. However, since the observed spikes are getting much bigger over time, we also see an increase in fluctuation around the respective stationary mean.

Figure 1: Trends in GDP, Trade Balance, and Exchange Rate



Notes: This figure presents levels (top row) and first differences (bottom row) of key UK macroeconomic indicators: real GDP (left), trade balance (center), and exchange rate against the US dollar (right), from 1955 to 2023. All series are shown alongside Hodrick-Prescott (HP) trends with a smoothing parameter of $\lambda = 1600$.
Source: Authors' computation from the UK Statistical Office.

Hence, the data seems to be stationary in the beginning and in the end with a negative trend being observed between the 1990s and the 2010s. Data on the exchange rate is available since 1997 on a quarterly basis. It exhibits stationarity between 1997 and 2007, as well as from 2007 onwards. In 2007, a shock seems to have shifted the mean of the stationary process downwards. The bottom row highlights first-differenced series, which strip away trends to reveal stationary fluctuations and cyclical patterns, supporting the idea that differencing mitigates non-stationarity. Across all panels, Hodrick-Prescott filtered trends ($\lambda = 1600$, standard for quarterly data) visually disentangle long-term trajectories from cyclical noise.

2.2 Unit roots and stationarity tests

In this section, we aim to formally conduct stationarity tests for the series. As explained in the last section, we have reason to doubt that our series are entirely stationary. However, for our analysis, we are relying on stationarity properties of the series. Hence, after identifying the non-stationary series formally, we will conduct the first-difference transformation to obtain stationary series for our analysis.

Table 1 presents the results of unit root and stationarity tests for our variables of interest, analyzed in both levels and first differences. The Augmented Dickey-Fuller (ADF), Phillips-Perron (PP), and Elliott, Rothenberg and Stock (ERS DF-GLS) tests evaluate the null hypothesis of a unit root (non-stationarity), while the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test assesses the null of stationarity. For GDP, the level series exhibits non-stationarity across all tests (e.g., ADF p-value = 0.353, KPSS statistic = 4.732*), but its first-differenced series shows strong evidence of stationarity (ADF statistic = -8.020*, KPSS = 0.086). Similarly, the trade balance in levels displays persistent non-stationarity (KPSS = 3.395*), with stationarity achieved after first differencing (ADF = -8.171*). The exchange rate shows mixed results in levels (e.g., ADF p-value = 0.418), but clear stationarity in differences (PP = -82.885***)

	Test Statistics			
Series/Test	ADF	PP	ERS DF-GLS	KPSS
Gross Domestic Product				
Levels	-2.531 (0.353)	-21.533^{***} (0.048)	3.074	4.732^{***}
Differences	-8.020^{***} (0.010)	-321.613^{***} (0.010)	-7.941^{***}	0.086
Trade Balance				
Levels	-2.225 (0.481)	-157.415^{***} (0.010)	-0.672	3.395^{***}
Differences	-8.171^{***} (0.010)	-300.088^{***} (0.010)	-12.912^{***}	0.035
Exchange Rate				
Levels	-2.382 (0.418)	-13.158 (0.354)	-1.415	1.754^{***}
Differences	-5.041^{***} (0.010)	-82.885^{***} (0.010)	-2.188	0.084

Notes: Null hypotheses—ADF/PP/ERS: series has a unit root (non-stationary); KPSS: series is stationary. To establish stationarity: reject ADF/PP/ERS null (significant ***/**) *and* fail to reject KPSS null (statistic < critical value). P-values in parentheses. Critical values (1% level): ADF/PP = -3.43, ERS DF-GLS = -2.57, KPSS = 0.739. $^{***}p < 0.01$, $^{**}p < 0.05$. First differences calculated as $\Delta y_t = y_t - y_{t-1}$.
Source: Author's calculations using data from the UK Statistical Office.

Table 1: Unit Root and Stationnary Test Results

We highlight that first-differencing effectively mitigates non-stationarity, as evidenced by statistically significant rejections of unit root hypotheses (***p<0.01) and failure to reject KPSS stationarity for differenced series.These results justify the use of differenced series for subsequent analysis, ensuring compliance with the stationarity assumptions underlying any further econometric treatment. Critical values and p-values are reported to validate the robustness of conclusions.

2.3 Model Estimation

Next, we proceed to the model estimation.

Statistical Models. Denote $\nabla y_t = y_t - y_{t-1}$ the first-difference operator. We write down three statistical models to guide our analysis. The univariate trade balance MA (4) model writes

$$\nabla tb_t = \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \theta_3 \epsilon_{t-3} + \theta_4 \epsilon_{t-4} + \epsilon_t \tag{1}$$

where tb_t denotes trade balance and $(\epsilon_t)_{1,\dots,T}$ is the innovation process.

The exchange rate AR (1) model is

$$\nabla e_t = \phi_1 \nabla y_{t-1} + \nu_t \tag{2}$$

where e_t represents exchange rate and $(\nu_t)_{1,\dots,T}$ is the innovation process The GDP univariate ARIMA (1,1,4) model with drift writes

$$\nabla y_t = c + \phi_1 \nabla y_{t-1} + \sum_{i=1}^{4} \theta_i \gamma_{t-i} + \gamma_t \tag{3}$$

where y_t is GDP and $(\gamma_t)_{1,\dots,T}$ is the innovation process.

Estimation. Table 2 shows the results of our estimations. We find that the Trade Balance has significant MA terms at the 1% level ($\theta_1 = -0.775^{***}$, $\theta_4 = 0.278^{***}$) indicate strong short- and long-lagged shock persistence. As for the exchange rate, The autoregressive term ($\phi_1 = 0.259^{***}$) suggests moderate persistence in differenced exchange rate movements. GDP's key drivers include the AR(1) term ($\phi_1 = 0.489^{**}$) and MA terms ($\theta_1 = -0.770^{***}$, $\theta_4 = -0.241^{***}$), with a large intercept ($c = 1832.987^{***}$) reflecting persistent upward drift. The models demonstrate heterogeneous dynamics: Trade balance exhibits complex error correction, exchange rate shows momentum effects, and GDP combines trend persistence with multi-period shock absorption.

	Trade Balance		Exchange Rate		GDP	
Component	Coeff.	SE	Coeff.	SE	Coeff.	SE
AR(1)	–	–	0.259^{***}	0.092	0.489^{**}	0.202
MA(1)	-0.775^{***}	0.057	–	–	-0.770^{***}	0.199
MA(2)	-0.100	0.074	–	–	0.093	0.093
MA(3)	-0.030	0.080	–	–	0.135^{*}	0.075
MA(4)	0.278^{***}	0.064	–	–	-0.241^{***}	0.061
Intercept	–	–	–	–	1832.987^{***}	232.754

Notes: Coefficient estimates (Coeff.) with standard errors (SE). – = Component not included in model. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. AR = Autoregressive term, MA = Moving Average term. GDP intercept in original units.
Source: Author's calculations using UK Statistical Office data.

Table 2: ARIMA Model Estimates

2.4 Forecasting

Table 3 reports forecast accuracy metrics for the first-differenced GDP, Trade Balance, and Exchange Rate series, calculated on the training set. The GDP series exhibits large absolute errors, with a Root Mean Squared Error (RMSE) of 9,021.2 and a Mean Absolute Error (MAE) of 3,505.4, while its Mean Absolute Scaled Error (MASE) of 0.77 suggests comparative performance relative to a benchmark. For the Trade Balance, the RMSE of 4,513.0 and MAE of 2,255.2 are accompanied by undefined percentage errors (MPE and MAPE), likely due to zero-crossings in the differenced series. The Exchange Rate demonstrates smaller-scale errors, with an RMSE of 0.035 and MAE of 0.027, alongside percentage errors exceeding 100%. Residual diagnostics show near-zero first-order autocorrelation (ACF1) across all series, ranging from -0.003 for GDP to 0.006 for the Trade Balance. The MASE values, all below 1, indicate that the models improve upon a naive forecast. These metrics reflect the distinct scaling and volatility profiles of the macroeconomic series under study.

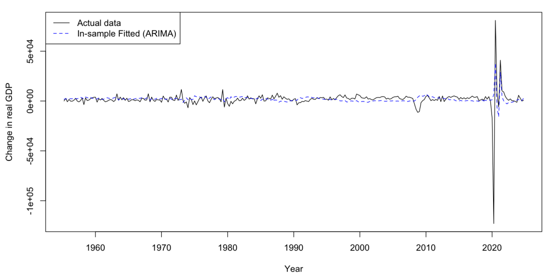
Series	**ME**	**RMSE**	**MAE**	**MPE**	**MAPE**	**MASE**	**ACF1**
GDP	-30.6	9021.2	3505.4	-106.5	452.6	0.770	-0.003
Trade Balance	-196.1	4513.0	2255.2	–	–	0.686	0.006
Exchange Rate	-0.001	0.035	0.027	137.8	152.9	0.682	-0.008

Notes: All metrics calculated on training set. Scaling factors: GDP RMSE/MAE in original units ($x10^{-3}$), Exchange Rate ME ($x10^{-3}$). "–" indicates infinite values due to zero-base percentage calculations. ME: Mean Error, RMSE: Root Mean Squared Error, MAE: Mean Absolute Error, MPE: Mean Percentage Error, MAPE: Mean Absolute Percentage Error, MASE: Mean Absolute Scaled Error, ACF1: Lag-1 Autocorrelation.
Sources: XX.

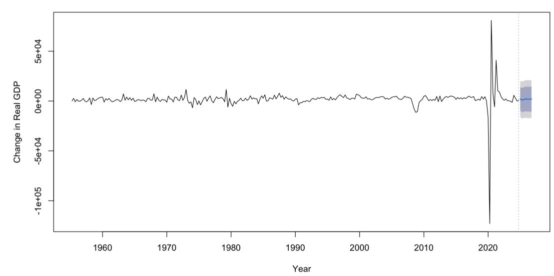
Table 3: Forecast Accuracy Metrics for First-Differenced Series

Figure 2 compares the true quarterly GDP change (black line) to the ARIMA model's fitted values (dashed blue line). The model seems to fit the real data generally quite well. Important deviations occur when huge spikes occur in the true data such as the COVID-19 spike around 2020. Here, the model predictions severely underestimate the real development. Additionally, small deviations occur around turning points but overall, the blue and black lines are quite close to each other, indicating a good in-sample fit. Note that although the MPE and MAPE errors are quite high, these percentage metrics blow up as the GDP growth is close to 0. Hence, the relatively high values of these specific accuracy metrics do not alarm us. Looking at the out-of-sample forecast, the model predicts GDP growth to continue hovering around a constant mean close to zero. However, we face rising uncertainty of our predictions as we go further into the future, showing the increasing variance of the Best Linear Forecast errors.

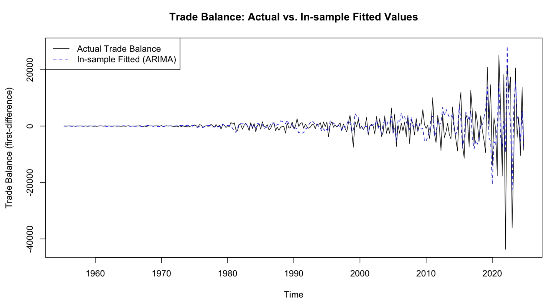
Figure 2: Forecast performance for UK GDP, Exchange Rate, and Trade Balance



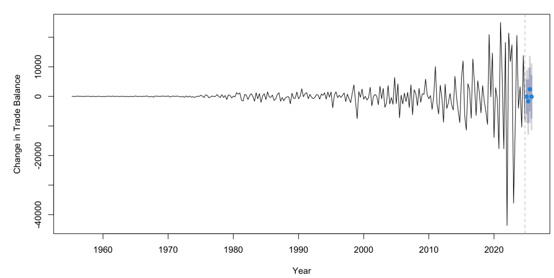
(a) GDP: In-sample forecast



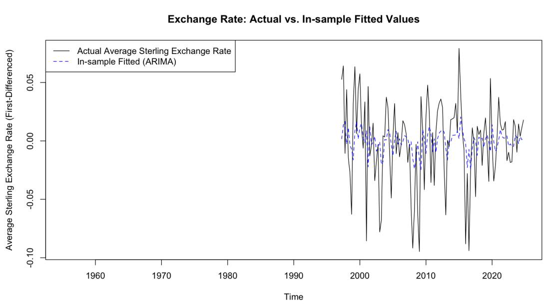
(b) GDP: Out-of-sample forecast



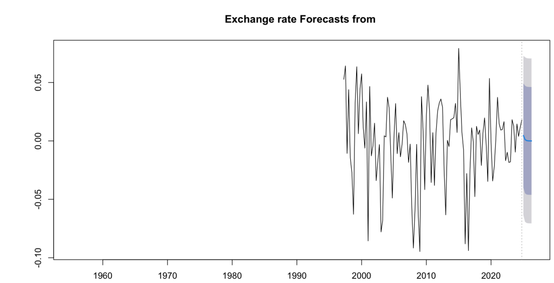
(c) Trade Balance: In-sample forecast



(d) Trade Balance: Out-of-sample forecast



(e) Exchange Rate: In-sample forecast



(f) Exchange Rate: Out-of-sample forecast

3 Multivariate Analysis

3.1 Theoretical framework

Present a simple model based on paper to guide our analysis especially the var ordering for cholesky decomp.

3.2 Optimal lag selection

We use the VARselect command in R to estimate the optimal lags. The results in table 4 present the optimal lags as suggested by four different tests. These tests aim to balance out the goodness-of-fit against model complexit with lower values indicating better models. While AIC (Akaike Information Criterion) and FPE (Final Prediction Error) impose a constant penalty per additional coefficient (and hence, lag) added to the model, they opt for more complex dynamics if that means reducing the residual variance. By contrast, SC (Schwarz Information Criterion, also known as Bayesian Information Criterion) and HQ (Hannan-Quinn) have penalties that increase with the number of lags used, hence selecting only the most essential lags.

Criterion	Optimal Lag
Akaike Information Criterion (AIC)	7
Hannan-Quinn Information Criterion (HQ)	1
Schwarz Criterion (SC) / Bayesian Information Criterion (BIC)	1
Final Prediction Error (FPE)	7

Notes: Coefficient estimates (Coeff.) with standard errors (SE). – = Component not included in model. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. AR = Autoregressive term, MA = Moving Average term. GDP intercept in original units. *Source:* Author's calculations using UK Statistical Office data.

Table 4: Optimal Lag Selection Based on Information Criteria

This explains why AIC and FPE suggest 7 lags while SC and HQ suggest only 1 lag. In the following, we go with AIC's and FPE's results using 7 lags for our analysis. Our rationale is its widespread use in the econometric modeling literature and its robustness in capturing model fit while penalizing complexity less stringently than SC or HQ. This makes AIC particularly suitable for applications where predictive accuracy is prioritized, as it allows for a slightly more flexible model specification, which is often beneficial in a context with potential structural nuances.

3.3 VAR Estimation

Statistical model. In R, we estimate the model using the VAR command, which when using 7 lags implies an OLS regression of the following kind:

$$x_{i,t} = \alpha_{i,0} \sum_{i=1}^{3} \sum_{k=1}^{7} \alpha_{i,t-k} x_{i,t-k} + u_{i,t}$$

In this equation, $i \in \{1, 2, 3\}$ stands for GDP, Trade Balance, and Exchange Rate respectively.

3.4 Residual testing

Testing our residuals as shown in Table (??) shows that there is neither serial autocorrelation nor conditional heteroscedasticity (ARCH Effects) left in our model. However, the normality tests strongly rejects Gaussianity of our error terms suggesting that the residual distribution exhibits non-Gaussian features such as heavy tails or skewness, implying that standard t-tests and confidence intervals may be invalid unless we use sufficiently robust inference methods. This issue is common in VAR models and hence does not alarm us too much.

The test statistics for the behavior of our residuals are summarized in the following table:

Test	ChiSq	df	p.value
Serial Correlation	70.6377	45	0.0086591
ARCH	290.9468	288	0.4403334
Normality	7261.2574	6	0.0000000

Table 5: Residuals tests

3.5 VAR forecasts

Our out-of-sample forecasts for Δ GDP remain tightly centred on zero—consistent with the stationary behavior we established in our VAR—while the fan chart's gradually widening bands reflect the growing uncertainty as the forecast horizon extends. Importantly, even after the dramatic COVID-19 shock in early 2020, the model shows a rapid return of quarterly GDP growth to its long-run average of essentially zero, underscoring both the economy's resilience and the appropriateness of our stationary specification.

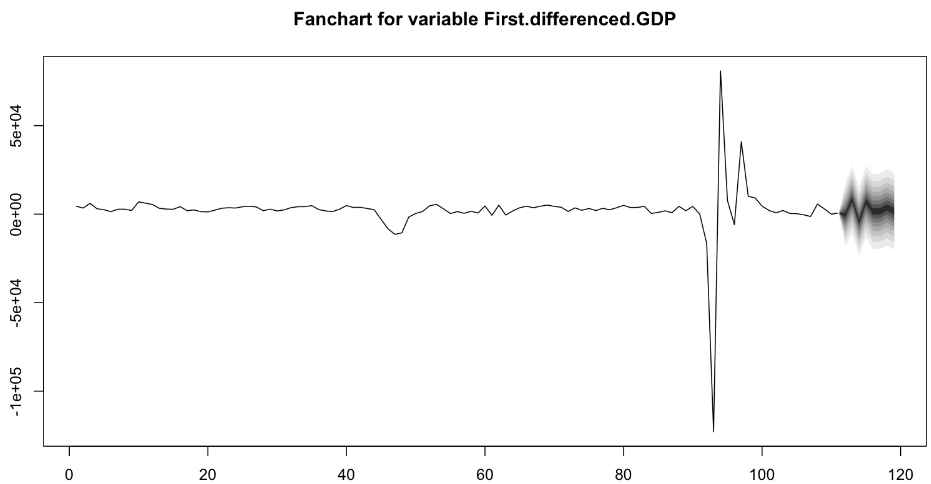


Figure 3: GDP - VAR Forecast

3.6 Cholesky decomposition

The Cholesky decomposition factorizes the residual covariance matrix of our VAR into a lower-triangular matrix, which yields a set of orthogonal (i.e. uncorrelated) structural shocks. By imposing a recursive identification scheme—where the first variable is assumed to be contemporaneously exogenous, the second may respond immediately to the first, the third to the first two, and so on—we obtain a uniquely defined causal ordering. This ordering lets us interpret each impulse-response function as the dynamic effect of a one-unit shock in one variable on all others in the system. In our analysis, we apply the Cholesky decomposition to the VAR residuals and adopt the ordering: Δ GDP →Exchange Rate →Trade Balance, so that shocks to Δ GDP are treated as exogenous and shocks to Trade Balance may reflect contemporaneous feedback from both Δ GDP and Exchange Rate.

3.7 Impulse response functions

A natural way to think about the very short-run causal ordering is that GDP shocks—say a surprise shift in domestic demand or productivity—are the "slowest moving" of the three: they show up in quarterly national-accounts data and take a full quarter to be measured, and they reflect broad real-economy fundamentals. By contrast, the exchange rate is set in continuous, high-frequency financial markets and will react almost instantaneously to any news about output or monetary-policy surprises. Finally, the trade balance is connected to contracts, value chains, etc. that make short-term adjustment difficult: export and import contracts, shipping lags and invoicing delays mean that net-export volumes likely respond with substantial delay to movements in both output and the exchange rate.

The impulse response functions for GDP are depicted as:

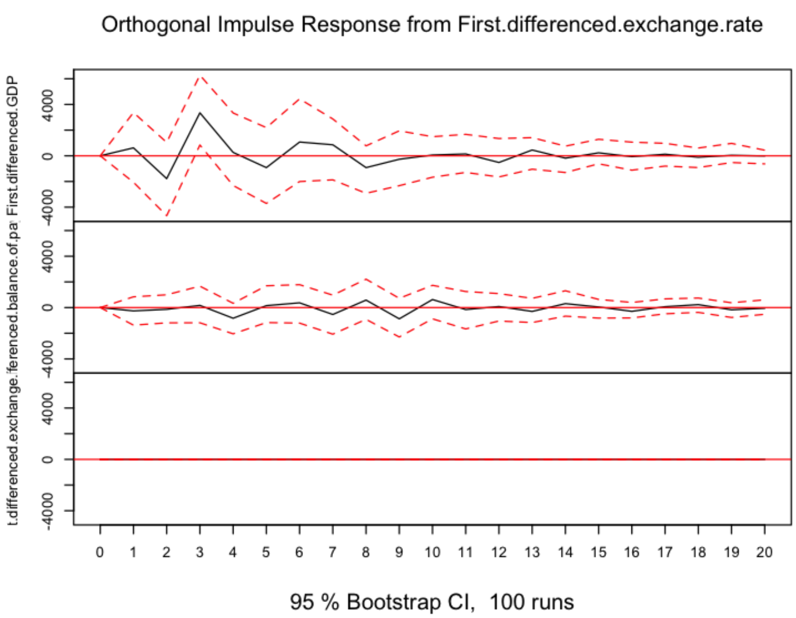


Figure 4: GDP - Shock reactions

The impulse response functions for Trade Balance are depicted as:

The impulse response functions for Exchange Rate are depicted as:

3.8 Robustness analysis: modify the ordering of the variables

After modifying the ordering, the IRFs look as follows:

The impulse response functions for GDP are depicted as:

4 Concluding Remarks

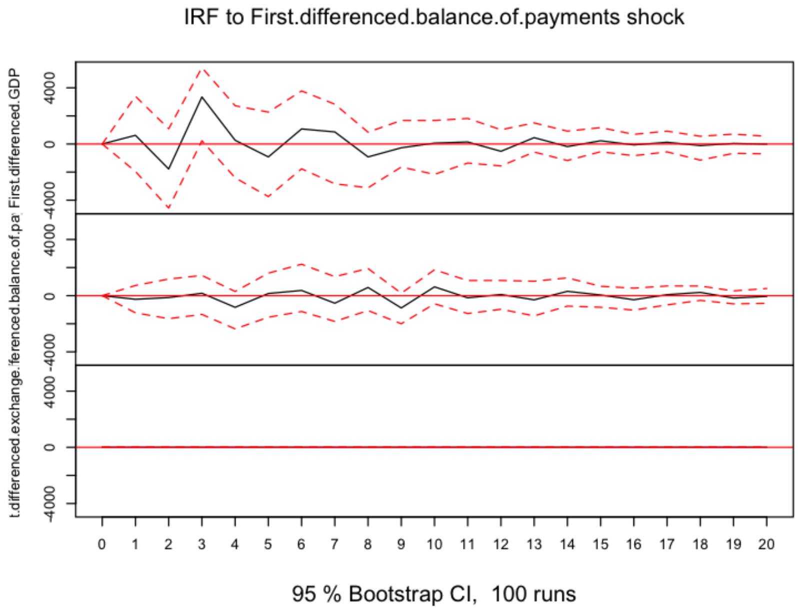


Figure 5: Trade Balance - Shock reactions

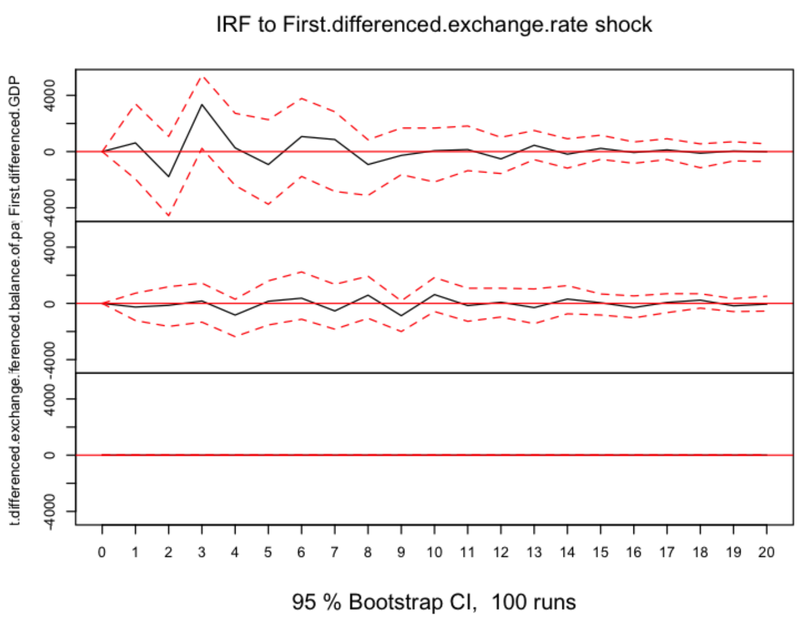


Figure 6: Exchange Rate - Shock reactions