Seminar Forecasting - BSc3

Draft report

Back to the future: Noncausal regime-switching models for U.S. inflation forecasting

Team 4
Floris van den Doel: 595906
Max Lopes-Cardozo: 662957
Imme Schraag: 658072
Berend Wibbens: 665072

Abstract

1 Introduction

Inflation - a general increase in the prices of goods and services in an economy - is a well-known economic indicator. High inflation, for example, decreases purchasing power, makes it harder for businesses to plan investments, and impacts international trade. Furthermore, central banks - like the Federal Reserve or European Central Bank - use inflation (forecasts) to set interest rates. Accurate inflation numbers help prevent excessive inflation or deflation. In addition, companies use inflation forecasts to plan pricing strategies, wage adjustments, and investment decisions as inflation significantly influences the real return on assets (Boudoukh and Richardson, 1993). Governments need inflation predictions for tax revenue projections and social spending adjustments (e.g., pensions, subsidies). Therefore, accurate inflation forecasting is essential.

When describing the dynamics of economic time series such as inflation, univariate autoregressive (AR) models are commonly used. Meese and Geweke (1984) provide a comparison of the actual performance of several automated univariate autoregressive forecasting procedures. Further, several empirical examples in the literature exist such as Andrews and Chen (1994), which includes forecasting and the measurement of persistence. Also in theoretical macroeconomic models AR models are used (Stock and Watson, 1996). Besides, several financial models aim to use this univariate AR approach as well (Collins and Hopwood (1980) and Campbell et al. (1998)).

An assumption of many linear models is stationarity. However, some stationary time series can contain some non-linear characteristics. One example is locally explosive episodes — prolonged increases followed by abrupt crashes — often referred to as *bubbles* in the literature (Voisin, 2022). These bubbles, which are difficult to model with linear models, are common in commodities, stocks, and cryptocurrencies. Such dynamics often require nonlinear models or regime-switching frameworks, which involve estimating many parameters. Our research aims to address this problem.

Until approximately a decade ago, existing economic applications had been limited to (simple) causal autoregressive models, where the value of the variable of interest depends solely on present and past error terms. In contrast, Lanne and Saikkonen (2011) introduce a noncausal autoregressive model that incorporates both leads and lags, allowing for dependence on the future. This feature is particularly relevant in economic contexts where expectations play a central role, for example in the New Keynesian Phillips Curve (Lanne and Luoto, 2013). These noncausal AR models' errors can be predicted using past values of the series. This can be used to improve forecast accuracy if noncausality is detected. One possible interpretation of this predictability is that the errors capture the influence of omitted variables that are, in turn, predictable from the observed series.

Although some empirical studies have found a limited role for expectations in inflation dynamics (Rudd and Whelan, 2005), they play an important role in theoretical models like the New Keynesian Phillips curve (NKPC). While their formal incorporation into forecasting models is relatively recent, the idea that expectations influence economic behavior has long been recognized. Muth (1961) introduces the concept of rational expectations, asserting that economic agents utilize all available information to form predictions about the future. This implies that individuals anticipate policy responses and adjust their behavior accordingly. Consequently, if agents expect higher future inflation, they can adjust their current pricing, wage-setting, and spending decisions, thereby influencing present inflation rates.

To clarify the economic differences between the backward and forward-looking models, we further analyse some literature: Roeger and Herz (2012) compare the Traditional Phillips Curve (Phillips, 1958) with the New Keynesian Phillips Curve (Batini et al., 2005). The traditional framework assumes inflation is backward-looking, with firms setting prices based on past inflation. This creates a strict inter-temporal trade-off, where monetary policy shocks only have temporary effects, and output eventually returns to its baseline. In contrast, the forward-looking NKPC suggests that firms and consumers anticipate future inflation and output changes, allowing monetary shocks to have persistent effects. Roeger and Herz (2012) present empirical evidence that supports this view, showing that output does not fully revert to its baseline, implying long-term effects of monetary policy. Managing inflation expectations is therefore crucial, as forward-looking agents incorporate future policy decisions into pricing and wage setting. Further evidence from Korkmaz (2010) finds that, while the relative importance of past versus future expectations may vary by country, forward-looking behavior dominates in explaining inflation in Turkey, reinforcing the broader importance of inflation expectations in macroeconomic modelling. This evidence suggests that incorporating the noncausal part is highly relevant for the inflation forecasting framework.

As an addition to current literature on the topic of noncausal inflation forecasting, this paper will consider non-linear inflation dynamics in the form of regime-switching models. Regime switching models allow the process to switch between different states (regimes), each with its own parameters. This is useful for data showing different statistical properties across time (e.g. high vs. low inflation). In terms of monetary policy of the Fed, clear structural changes exist. For example, the Fed started targeting an annual inflation rate of 2% from 1996 onwards (Wells, 2024). Monetary policy is also commonly referred to as either being hawkish or dovish, where hawkish policy means tighter monetary policy to prevent excess inflation, whereas dovish means relaxing monetary policy to stimulate economic growth. The Fed usually tries to balance price stability with maximum sustainable employment, but during the times of the 2008 global financial crisis, the emphasis switched more towards employment (Van Dieijen and Lumsdaine, 2019). As Fed speeches play an important role in the formation of expectations, this can bring forth a fundamental change in the inflation dynamics. Regime switching within the inflation framework has also been researched in the literature. Amisano and Fagan (2013) investigate the relationship between money growth and inflation using a regime switching model, revealing that the strength and nature of this relationship vary significantly across different monetary regimes. Further, Nalewaik (2015) uses regime-switching models to improve inflation forecasting by estimating the likelihood of shifts to high-variance regimes, enhancing prediction intervals with more accurate and timely assessments of inflation uncertainty. So, it can be said that regime-switching inflation models have been examined in the literature before, such as in Dahl and Hansen (2001), but to our knowledge, never within a noncausal framework, despite expectations playing an important role in the justification for the use of these regime-switching models.

Now, a need for noncausal dynamics in inflation forecasting has been established. Our research builds upon Lanne and Saikkonen (2011) and Lanne et al. (2012) where a mixed causal noncausal autoregressive (MAR) model is presented and used for (inflation) forecasting. Our research question therefore is: to what extent can a mixed causal-noncausal (MAR) model with regime-switching improve inflation forecasts in the United States relative to a standard AR model with regime switching?

This research uses the FRED-MD dataset (Federal Reserve Bank of St. Louis, 2025). A large monthly macro-economic dataset which is commonly used for empirical applications. We use the non seasonally adjusted time series of the Consumer Price Index (CPI) obtained from U.S. Bureau of Labor Statistics (2025).

When performing our research, we first estimate a standard autoregressive (AR) model, which is followed by diagnostic tests for normality (Jarque-Bera and Kolmogorov-Smirnov) and independence of residuals. After establishing non-normality and non-i.i.d. behavior, MAR models are fit using approximate maximum likelihood estimation with scaled-t distributed errors. The number of lags and leads is selected using information criteria (AIC, BIC, HQIC), with a preference for BIC due to its stronger penalty on model complexity, an important consideration given the already larger number of parameters due to the threshold specification. Further, regime-switching is incorporated through both threshold (MART) and smooth transition (SMART) to capture non-linear inflation dynamics using different thresholds. Forecasts are made using an expanding window and are compared using the root mean squared forecast error and the Diebold-Mariano test to compare forecasting accuracy across different forecasting horizons up to two years. While these models are used in this paper to forecast inflation, the MART and SMART models could be considered for any process where MAR is deemed appropriate.

MAIN RESULTS AND CONCLUSION [1]

This report is structured as follows. Section 2 describes the data we aim to use for our research. Section 3 proposes the econometric techniques and MAR models in order to answer our research question. Further, in Section 4 we show our results. Finally, Section 5 covers our conclusions and discussion.

[1] still some discussions going on, will be included in final report

2 Data

For this research, we use the FRED-MD (Federal Reserve Bank of St. Louis, 2025) data set described in McCracken and Ng (2016). It is a large macroeconomic database that is designed for empirical applications. We consider monthly data in our research which is consistent with previous literature (Stock and Watson (1999), Moshiri and Cameron (2000) and Hecq et al. (2020)). The data is updated regularly and incorporates changes and revisions as well. Further, as the inflation numbers are reported on a monthly basis, we get more observations per year than a quarterly dataset. As in Lanne and Saikkonen (2011), we use the Consumer Price Index (CPIAUCSL) as a measure of inflation. The CPIAUCSL reported in the FRED-MD dataset is a seasonally adjusted series obtained from the US Bureau of Labor Statistics. However, for our research, we are looking for a non-seasonally adjusted time series. We want to investigate whether adding noncausal variables introduces better forecasting performance. The use of non adjusted data ensures that noncausal dynamics, if present, are not artifacts of seasonal adjustment filters as shown by Hecq et al. (2017c). Therefore, by using raw data we avoid this distortion. Further, using non-adjusted data can help avoid filtering out economically meaningful seasonal dynamics. Thus, we separately obtain the "raw" CPI for All Urban Consumers from the U.S. Bureau of Labor Statistics (2025). We consider the suggested transformation of the CPI by McCracken and Ng (2020), scaled to percentages $\pi_t = 100 * \log(CPI_t/CPI_{t-1})$ to obtain the monthly inflation rate (Faust and Wright, 2013). We use data from May 1959 up to December 2024. The first month of the dataset is omitted to avoid issues related to transformations resulting in 787 total observations. Further variables to consider are:

- UNRATE: the Civilian Unemployment Rate
- IPFINAL: Final Products (Market Group). This index includes final products, which are closer to a GDP-style measure since GDP measures the value of final goods and services.
- CUMFNS: Capacity Utilization: Manufacturing. While not a direct GDP measure, it indicates how much of the existing industrial capacity is being used, which can be a proxy for economic activity.
- RPI: Real Personal Income, an indicator of consumer purchasing power.
- RETAILx: Retail and Food Services Sales, proxy for consumer demand, which impacts inflation.
- VIXCLS [2]: Volatility Index, can be an indicator of market uncertainty affecting inflation expectations.

Table 1: Summary Statistics FRED-MD

Variable	Mean	Median	SD	Skewness	Kurtosis
CPI [3]	139.59	139.10	83.47	0.22	-1.14
UNRATE	5.88	5.60	1.69	0.91	1.10
IPFINAL	71.00	69.91	27.28	-0.23	-1.44
CUMFNS	79.30	79.00	4.94	-0.12	0.41
RPI	9,616.38	8,457.39	5,050.94	0.44	-0.97
RETAILx	224,728.49	164,213.00	189,584.84	0.81	-0.27
VIXCLSx	19.30	17.62	7.07	1.89	6.15
Inflation	0.30	0.29	0.36	0.06	2.87

[2] Available from July 1962
[3] From U.S. Bureau of Labor Statistics (2025)

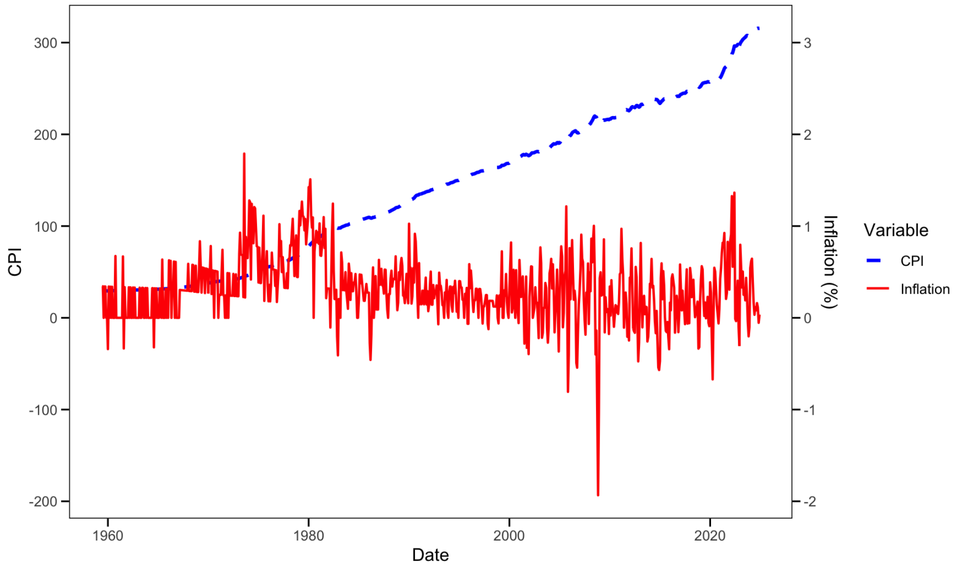

Figure 1: Monthly Inflation and CPI from 1959-6 to 2024-12.

Figure 1 shows the monthly inflation and CPI from June 1959 until December 2024. We observe some characteristics: the period up until the late 1970's is also referred to as the go/stop phase of monetary policy (Goodfriend, 2005). This period is characterized by an underestimation of the economic effects of inflation. As a result, prices had already begun to reflect high inflation expectations. As can be seen in Figure 1, this period was one of high inflation. This is followed by the Volcker disinflation period from 1979 to 1987. In order to restore the credibility of the Fed, sharp increases in interests rates were implemented in 1979. However, due to geopolitical situations, inflation continued to rise, which gave way to a massive 3 percentage point interest rate hike. After the Volcker disinflation period, Figure 1 shows a more stable trend. Some notable spikes can be seen in 2008, where a decrease in wealth - due to the 2008 financial crisis - caused a decrease in aggregate demand, leading to deflation (Spilimbergo et al., 2009). Recent years have also been quite inflationary, partly due to a tighter labour market during and post-Covid (Ball et al., 2022) as well as an increase in geopolitical risk Caldara et al. (2022).

We observe positive autocorrelation even at high lags which is shown in Figure 2. This is confirmed by the Ljung-Box test which indicates significant autocorrelation. If we look at the Figure 1 and after performing a unit root test (ADF), we conclude that the series can be considered as stationary.

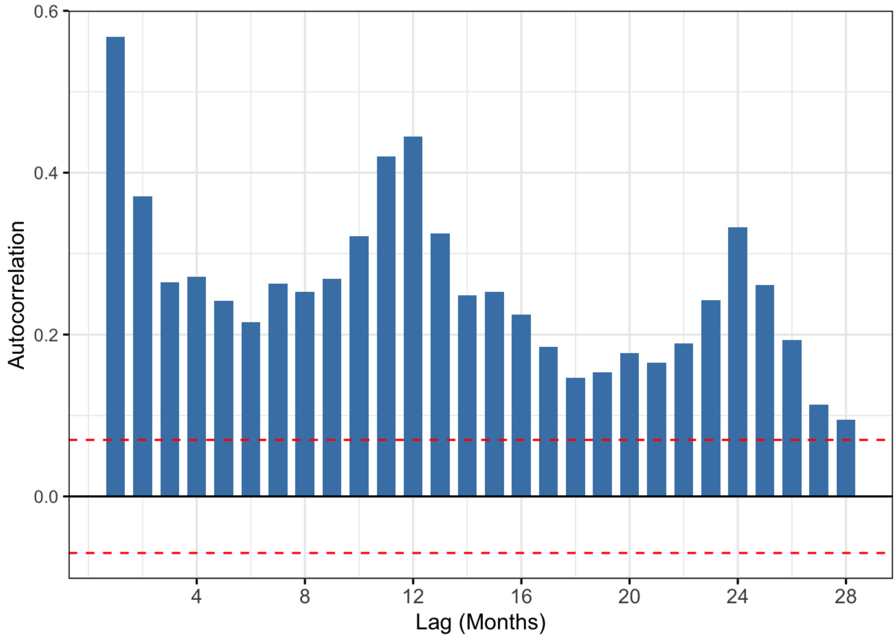

Figure 2: Autocorrelation of monthly inflation shows persistent dynamics up to lag 28

3 Methodology

3.1 Model

An autoregressive AR(p) process y_t depends on its last p values as follows,

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t. \tag{1}$$

The roots of the lag polynomial $(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p)$ all lie outside the unit circle for stationary processes, where L represents the lag operator $L^i y_t = y_{t-i}$ (Franses, 1998). The error term ε_t is assumed to be white noise. The one-step ahead forecast $\hat{y}_{t+1}$ can easily be calculated using the last p observed values of the process, and subsequent forecasts can be calculated recursively. However, if the underlying process were to not depend on the past values of y_t, but instead on the future values $y_{t+1}, y_{t+2}, \ldots$, then the AR(p) model is not necessarily the best way to model this process.

Models where a value y_t depends on its future values are called noncausal processes. Let y_t follow a purely noncausal AR(1) model (Telg, 2017):

$$y_t = \varphi_1 y_{t+1} + \varepsilon_t, \text{ with } |\varphi_1| < 1. \tag{2}$$

Noncausal processes have the same second-order dependence structures as causal processes, meaning that the auto-covariance $Cov(y_t, y_{t-k})$ is equal in a causal AR(p) model and a noncausal AR(p) model (Hecq and Velásquez-Gaviria, 2025). Importantly, this implies that if in this noncausal process the error term ε_t follows a Gaussian distribution, a noncausal AR(1) model contains the same information as a purely causal AR(1) model (Telg, 2017). This is the case because Gaussian distributions are fully characterized by their second-order structure. Hence, noncausal models are only of interest when the error terms are non-Gaussian.

Next to purely causal and purely noncausal processes, mixed causal noncausal autoregressive (MAR) models have been extensively studied in economic time series (Gouriéroux et al. (2013), Lanne and Saikkonen (2011) and Voisin (2022)). MAR models are autoregressive models not only using past values of the process but also future values. The model is constructed in a multiplicative structure of the backward- and forward-looking lag polynomials as follows:

$$\Phi(L)\Psi(L^{-1})y_t = \varepsilon_t, \tag{3}$$

where L^{-i} is the lead operator, $L^{-i}y_t = y_{t+i}$ and deterministic terms are left out for clarity. A MAR(r,s) process, with r lags and s leads, therefore has a lag polynomial $\Phi(L)$ of order r and a lead polynomial $\Psi(L^{-1})$ of order s. While having both lag and lead polynomials stationary, this model manages to capture non-linear features such as bubbles (Telg, 2017). A condition for the identification of MAR models is the non-Gaussianity of the error term, which appears to be quite common in practice and thus not restrictive.

A MAR process can be filtered into a purely causal and a purely noncausal component. This decomposition, or filtration, can be derived from equation (3) and proceeds as follows:

$$u_t = \Phi(L)y_t, \tag{4}$$

which defines u_t as a purely noncausal MAR(0,s) process. Substituting this into the original model yields:

$$\Psi(L^{-1})u_t = \varepsilon_t. \tag{5}$$

This purely noncausal component u_t is the process of interest in most of this paper since it is the forward-looking component of MAR models that makes forecasts more complicated compared to purely causal models. Lanne et al. (2012) show that a MAR process can be written as,

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_r y_{t-r} + u_t, \tag{6}$$

and equivalently,

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_r y_{t-r} + \sum_{j=1}^{\infty} \beta_j \varepsilon_{t+j}, \tag{7}$$

where β_j is the coefficient of z^j in the power series expansion of $\Psi(L^{-1})$.

This formulation explicitly shows how future error terms affect the process. It is important in the forecasting aspect of this research, as this formula consists of the familiar causal AR(r) process, leaving the noncausal part to be forecast. MAR models can be used as a powerful tool for modelling processes that tend to a future value, such as inflation in the NKPC. Though, in contrast to the New Keynesian Phillips Curve, not through expectations but through the actual values of future inflation.

Our research will consider a MAR model with scaled-t distributed error terms as deemed adequate for MAR models to model inflation in Lanne and Saikkonen (2011). To evaluate this choice of distribution for the error term, several misspecification tests will be performed, such as unit root testing using the Augmented Dickey–Fuller (ADF) test (Dickey and Fuller, 1979), as was also done in Hecq et al. (2016). We use it to test if macroeconomic data contains a stochastic trend, which is equivalent to testing for a unit root.

3.2 Estimation

The estimation of MAR models differs from that of purely causal AR(p) processes due to issues related to second-order dependence, as discussed in Section 3.1 (Hecq and Velásquez-Gaviria, 2025). For parameter estimation, we follow the approaches outlined by Lanne and Saikkonen (2011) and Telg (2017). We begin by estimating a purely autoregressive model and then test the residuals for both Gaussianity and independence. If both hypotheses are rejected, we consider this sufficient justification to explore a noncausal or mixed causal–noncausal specification (Telg, 2017). Lanne and Saikkonen (2011) find the scaled-t distribution to be well suited for modeling U.S. quarterly inflation data, while also satisfying the regularity conditions discussed in Andrews et al. (2006). Lanne and Saikkonen (2011) and Andrews et al. (2006) argue that symmetric, non-Gaussian density functions that are sufficiently smooth, such as the rescaled Student t-distribution, meet these requirements. As previously mentioned, we use monthly data instead of quarterly, which we consider appropriate for our analysis, particularly since Hecq et al. (2020) also employ monthly inflation data. Hecq and Velásquez-Gaviria (2025) propose a novel approach for estimating MAR models that does not require specifying a distribution for the error term. While this method lies beyond the scope of the present study, it represents a promising avenue for future research on inflation dynamics.

After rejecting normality for the purely causal AR(p) model, we will proceed to estimate the model using approximate MLE based on a scaled-t distribution following Telg (2017). We define the error process ε_t with a density $f_\varepsilon(\varepsilon_t; \boldsymbol{\lambda}|y_t)$ where $\boldsymbol{\lambda}$ is the vector of distributional parameters: scale and degrees of freedom.

The probability density function of ε_t:

$$f_{\varepsilon}\left(\varepsilon_t \mid \sigma, \nu\right) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}\sigma}\left(1+\frac{1}{\nu}\left(\frac{\varepsilon_t}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}}, \tag{8}$$

implies the following log-likelihood function:

$$\begin{aligned} l_y(\boldsymbol{\phi}, \boldsymbol{\varphi}, \alpha, \boldsymbol{\lambda} \mid \boldsymbol{y}) = (T-p)\left[\ln\left(\Gamma\left(\frac{\nu+1}{2}\right)\right) - \ln\left(\sqrt{\nu\pi\sigma^2}\right) - \ln\left(\Gamma\left(\frac{\nu}{2}\right)\right)\right] \\ - \frac{\nu+1}{2}\sum_{t=r+1}^{T-s}\ln\left(1+\frac{1}{\nu}\left(\frac{\phi(L)\varphi(L^{-1})y_t-\alpha}{\sigma}\right)^2\right). \end{aligned} \tag{9}$$

Where the error term ε_t is defined as in Equation (3), with the addition of an intercept. This function will be maximised using the BFGS algorithm to obtain the approximate MLE estimates (Broyden, 1970).

The estimation procedure for MAR models was originally implemented in R by Hecq et al. (2017a), although the package is no longer available on CRAN. Slight adjustments were made to the original functions to ensure compatibility with newer versions of R. The modified functions, along with additional functionality used in this paper, are available on GitHub[1].

3.3 Lag and lead selection

Lanne and Saikkonen (2011) propose to estimate a conventional causal autoregressive (AR) process by OLS or Gaussian maximum likelihood and determine the lag order p using information criteria. Several diagnostic tests can be performed to investigate whether the lag order p adequately captures the autocorrelation on the error process. Hecq et al. (2016) favour the use of the Bayesian Information Criterion (BIC). This criterion is likelihood-based and assigns different weights to the "parsimony" and "fit" of the model, with lower values indicating a better balance between the fit of the model and the complexity. We also use the following selection criteria: Akaike Information Criterion (AIC), Hannan-Quinn Information Criterion (HQ) and the Log-likelihood (LL).

Following the determination of the lag order, the null hypothesis of normality of the residuals of the AR(p) model is tested for. This is done using the Jarque-Bera (JB) and one-sample Kolmogorov-Smirnov (KS) tests. The JB test uses the Lagrange multiplier approach to jointly test for residual normality, homoskedasticity and serial independence (Jarque and Bera, 1980). The one-sample KS test is used to test the goodness of fit of a given set of data to a theoretical distribution (Berger and Zhou, 2014) and is in that way used to test the null hypothesis that a set of data comes from a normal distribution. Note that if the null hypothesis of normality is not rejected, there is no need to consider noncausal or mixed causal-noncausal models, as the causal and noncausal polynomials cannot be distinguished (Wong and Li, 2000).

In addition, we test for i.i.d.-ness of the AR(p) residuals. If the true data generating process (DGP) is i.i.d causal, then the errors from the AR(p) models must be i.i.d. too. The null hypothesis is then rejected if a causal model is using data generated by an i.i.d. noncausal DGP. Several tests exist, but for simplicity the regression of $\hat{\epsilon}_t$ on $\hat{\epsilon}^2_{t-1}$ up to $\hat{\epsilon}^2_{t-m}$ is used, where the residuals from the causal AR(p) are denoted as $\hat{\epsilon}_t$ (Hecq et al., 2016):

$$\hat{\epsilon}_t = \mu + \delta_1\hat{\epsilon}^2_{t-1} + \cdots + \delta_m\hat{\epsilon}^2_{t-m} + u_t. \tag{10}$$

To test for i.i.d.-ness the null hypothesis $H_0 : \delta_1 = \delta_2 = \cdots = \delta_m = 0$ is used. This test is asymptotically

[1]The GitHub repository is available at: https://github.com/Fdoel/Forecasting.

χ^2_m-distributed under normality. Because this normality assumption is not the case in the study at hand, its distribution can be tabulated. This is the natural extension to the procedure proposed by Gouriéroux et al. (2013) for $m = 1$ lag. In the different models we evaluate, we choose m to be equal to the number of lags p, where p is determined as outlined above.

Furthermore, Hecq et al. (2017b) state that it is advisable to perform additional tests on the residuals to check whether all serial correlation has been deleted. To do so we use the Ljung-Box test, which null hypothesis is that there is no autocorrelation between the residuals. The test uses a number of lags; Hassani and Yeganegi (2020) found that when the total number of observations is not to large, number of lags $= ln$(number of observations) (Tsay, 2005).

After detecting non-normality, finding that the i.i.d-ness of the causal residuals can be rejected and after checking whether all serial correlation has been deleted, the final step is to select the best lead and lag components in the MAR(r,s). In order to determine the individual numbers of lags (r) and leads (s) such that $r + s = p$, the approximate maximum likelihood (AML) approach that is explained in Section 3.2 is applied according to Lanne and Saikkonen (2011). In that study, after specifying the distribution of the error term, the AML based on Breidt et al. (1991) was applied. A similar approach using the sum p and the AML was also employed in Blasques et al. (2023) and Hecq et al. (2020).

To ensure the optimal selection of lag and lead lengths, we implement a grid search procedure that jointly determines the optimal number of lags (r) and leads (s) in the model. Grid search is a systematic optimization algorithm that requires the specification of an objective function to achieve non-linear optimization Gan (2014). Consistent with the goal of the pseudo-causal approach, we minimize the Bayesian Information Criterion (BIC) across the grid. The optimal solution corresponds to the point in the grid that yields the lowest BIC value.

3.4 Making forecasts

Forecasting using a MAR model is different from standard AR models. For a purely causal AR(p) model with $\varepsilon \sim D(0, 1)$, the expectation of the future error term is 0, so $\hat{y}_{t+1|t} = \phi_1 y_t + \cdots + \phi_p y_{t-p+1}$ (Franses, 1998). However, in a MAR(r,s) setting, since this error term is already embedded in y_t through the noncausal component, even under the i.i.d. assumption, the conditional expectation of the future error term is generally not equal to 0.

Taking the expectation on both sides of Equation (7) gives the following expectation of the value of y_{T+h}:

$$E(y_{T+h|T}) = \phi_1 E(y_{T+h-1|T}) + \cdots + \phi_r E(y_{T+h-r|T}) + \sum_{j=1}^{\infty} \beta_j E(\varepsilon_{T+h+j|T}). \tag{11}$$

Unlike traditional autoregressive models, where forecasts depend only on past values, MAR models require an estimation of future error terms. For a practical application, Lanne et al. (2012) propose to approximate this last noncausal term as $E(u_{T+h|T}) \approx \sum_{j=0}^{M-h} \beta_j \epsilon_{T+h+j}$. This sum can reasonably be approximated through simulations of the error terms, under the assumption that this distribution is correctly specified.

Our main focus will be a forecast horizon of one, three and twelve months ahead and possibly up to 24 months ahead (two years ahead) as this is consistent with previous literature (Stock and Watson (2002) and Clarida et al. (2000)). The forecasts are evaluated using root mean squared forecast error (RMSFE). As we use a large sample of delayed-mode quality controlled observations the root mean squared error (RMSE) might be preferred where the second moment sensitivity to large model errors is important (Brassington, 2017). Like in Lanne and Saikkonen (2011) we want to use an expanding window approach, where the first estimates are based on an initial set of observations. As the estimation discards $r+s$ observations, an expanding window approach has the advantage of using more observations compared to a rolling window approach. Furthermore, the Diebold-Mariano (DM) Test will be used to compare the predictive accuracy of two competing time series forecasting models. It checks whether the forecast errors between models are significantly different (Diebold and Mariano, 1995). The truncation parameter M is used to approximate the infinite moving average (MA) representation of the noncausal part of the model. Since this MA representation involves an infinite sum of future shocks, M limits the number of terms to a finite, computationally manageable amount. This truncation makes it feasible to simulate future values of the noncausal component while still capturing the essential dynamics of the model. Choosing an appropriate value for M is important to balance the trade-off between computational efficiency and forecast accuracy. An example of a common and well-chosen parameter can be found in Nyberg and Saikkonen (2014).

3.5 Regime-switching

3.5.1 MART(s,r) model

A glance at Figure 1 suggests that non-linear dynamics may exist in the inflation data. A possible reason for this could be a difference in inflation dynamics between higher and lower inflation periods. However, most formal statistical tests for detecting the presence of regimes rely either on comparing the in-sample performance of threshold models versus non-threshold models, or on assessing the differences between parameters within threshold models (Franses, 1998). Thus to test for the presence of regimes we will first need to develop a framework estimating these models.

The first regime switching model we consider is a threshold model, the mixed causal-noncausal autoregressive threshold (MART) model. Consider the MART(s,r) model, where the dynamics of y_t depend on both past ($\boldsymbol{x}_t$) and future ($\boldsymbol{z}_t$) values of y_t, with regime-specific coefficients.

The model is given by:

$$y_t = \boldsymbol{\phi}_t \boldsymbol{x}_t + \boldsymbol{\varphi}_t \boldsymbol{z}_t + \varepsilon_t, \tag{12}$$

where

$$\boldsymbol{x}_t = (y_{t-1}, \ldots, y_{t-s})', \quad \boldsymbol{z}_t = (y_{t+1}, \ldots, y_{t+r})'. \tag{13}$$

The regime-specific coefficient structures are determined by the threshold variable q_{t-i}, such that:

$$\boldsymbol{\phi}_t = \begin{cases} \boldsymbol{\phi}_\nu, & \text{if } q_{t-d} \geq c, \\ \boldsymbol{\phi}_\mu, & \text{if } q_{t-d} < c. \end{cases} \tag{14}$$

Similarly, the noncausal coefficients follow:

$$\boldsymbol{\varphi}_t = \begin{cases} \boldsymbol{\varphi}_\nu, & \text{if } q_{t+d} \geq c, \\ \boldsymbol{\varphi}_\mu, & \text{if } q_{t+d} < c. \end{cases} \tag{15}$$

In this definition the regime specific parameters are defined as follows:

- $\boldsymbol{\phi}_\nu$ and $\boldsymbol{\phi}_\mu$ as the vectors of causal coefficients in regimes 1 and 2, respectively.
- $\boldsymbol{\varphi}_\nu$ and $\boldsymbol{\varphi}_\mu$ as the vectors of noncausal coefficients in regimes 1 and 2, respectively.

This definition implies that there are four possible states for the process to be in, as the causal and noncausal part can be in different regimes. This definition keeps the mixed model consistent with both purely causal as well as with purely noncausal processes. q_{t-d} and q_{t+d} could be any variable, but with the hypothesis that inflation dynamics differ between low and higher inflation periods, a logical choice is to let $q_{t-d} = y_{t-d}$ and $q_{t+d} = y_{t+d}$. This means that the state of the model is dependent on past inflation for the causal part of the process, and dependent on future inflation for the noncausal part of the process.

For the threshold in our regime-switching model, we initially use the median to ensure a sufficient number of observations in both regimes, before extending the approach with a grid search to jointly optimize the number of lags, leads, and the threshold value.

If there are no noncausal components in a MART model we have a causal autoregressive threshold (ART) model. This model is a standard AR model with a threshold, which is typically estimated via maximum likelihood using a scaled-t distribution.

3.5.2 SMART(s,r) model

The smooth mixed causal-noncausal autoregressive threshold model (SMART) is a logical extension of the MART model. In this model, the distinct threshold between regime one and two is replaced by a smoothing function $G(q_t, c, \gamma)$ as is commonly done in AR models (Franses, 1998). This model can be formulated very similarly to the MART model:

$$y_t = \boldsymbol{\phi}_t \boldsymbol{x}_t + \boldsymbol{\varphi}_t \boldsymbol{z}_t + \varepsilon_t, \tag{16}$$

where

$$\boldsymbol{x}_t = (y_{t-1}, \ldots, y_{t-s})', \quad \boldsymbol{z}_t = (y_{t+1}, \ldots, y_{t+r})', \tag{17}$$

$$\boldsymbol{\phi}_t = G(q_{t-d}, c, \gamma)\boldsymbol{\phi}_\nu + (1 - G(q_{t-d}, c, \gamma))\boldsymbol{\phi}_\mu, \tag{18}$$

$$\boldsymbol{\varphi}_t = G(q_{t+d}, c, \gamma)\boldsymbol{\varphi}_\nu + (1 - G(q_{t+d}, c, \gamma))\boldsymbol{\varphi}_\mu. \tag{19}$$

A common choice for $G(q_{td}, c, \gamma)$ is the logistic function:

$$G(q_{t\pm d}, c, \gamma) = \frac{1}{1 + e^{-\gamma(q_{t\pm d} - c)}}, \tag{20}$$

where γ is a smoothness parameter controlling the transition speed between regimes.

3.5.3 Parameter estimation in the (S)MART model

The approximate MLE estimator can easily be derived from Equation (8). Using the definition of the error term in Equations (12) & (23), we derive the following log likelihood function for the (S)MART model. Assuming i.i.d t-distributed error terms, and the distribution being equal for both regimes:

$$f_\varepsilon(\varepsilon_t \mid \sigma, \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}\,\sigma} \left(1 + \frac{1}{\nu}\left(\frac{\varepsilon_t}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}}, \tag{21}$$

$$\begin{aligned} \ell_y(\boldsymbol{\phi}, \boldsymbol{\varphi}, \alpha, \boldsymbol{\lambda} \mid \boldsymbol{y}) = {} & (T - \max(r,d) - \max(s,d)) \left[\ln \Gamma\left(\frac{\nu+1}{2}\right) - \ln \Gamma\left(\frac{\nu}{2}\right) - \ln\left(\sqrt{\nu\pi\sigma^2}\right)\right] \\ & - \frac{\nu+1}{2} \sum_{t=\max(r,d)+1}^{T-\max(s,d)} \ln\left(1 + \frac{1}{\nu}\left(\frac{\varepsilon_t}{\sigma}\right)^2\right), \end{aligned} \tag{22}$$

$$\varepsilon_t = y_t - [G(y_{t-d}, c, \gamma)\,\boldsymbol{\phi}_\nu + (1 - G(y_{t-d}, c, \gamma))\,\boldsymbol{\phi}_\mu]\,\boldsymbol{x}_t + [G(y_{t+d}, c, \gamma)\,\boldsymbol{\varphi}_\nu + (1 - G(y_{t+d}, c, \gamma))\,\boldsymbol{\varphi}_\mu]\,\boldsymbol{z}_t. \tag{23}$$

For the MART model, let $G(y_{t-d}, c, \gamma) = \mathbb{1}\{y_{t-d} \geq c\}$, and for the SMART model let G be the logistic function as defined in Equation (20). Like the regular MAR model, this function can be maximized using the BFGS algorithm to obtain approximate MLE estimates. These threshold models also introduce the parameters c, γ, and d. These parameters are be determined using a grid search approach (Franses, 1998). We use a grid-search approach for these variables, looking to maximize the log-likelihood. Future research could consider different estimation techniques for these variables.

If there are no noncausal components in a SMART model we have a smooth causal autoregressive threshold (SART) model. This model is a standard AR model expanded with a smooth function around the threshold, the model is, such as an ART model estimated via maximum likelihood using a scaled-t distribution.

3.6 Forecasting in the MART model

The threshold in the MART model add a layer of complexity to the forecasts. Consider the optimal h-step ahead forecast, then we can split forecasts into two categories: 1. $h \leq d$ 2. $h > d$. In the first case the causal regime is observed, and thus known. From Equation (11) & (12) we can derive the optimal h-step ahead forecast $E(y_{T+h|T})$. Assume without loss of generality that the causal process is regime #1, then the MART process can be defined using a piecewise function:

$$y_{T+h} = \begin{cases} \phi_{\nu,1}y_{T+h-1} + \cdots + \phi_{\nu,r}y_{T+h-r} + \sum_{j=1}^{\infty} \beta_{\nu,j}\varepsilon_{T+h-j}, & \text{if } y_{T+h+d} \geq c, \\ \phi_{\nu,1}y_{T+h-1} + \cdots + \phi_{\nu,r}y_{T+h-r} + \sum_{j=1}^{\infty} \beta_{\mu,j}\varepsilon_{T+h+j}, & \text{if } y_{T+h+d} < c. \end{cases} \tag{24}$$

Combining these into a singular function yields:

$$y_{T+h} = \phi_{\nu,1}y_{T+h-1} + \cdots + \phi_{\nu,r}y_{T+h-r} + \mathbb{1}(y_{T+h+d} \geq c) \sum_{j=1}^{\infty} \beta_{\nu,j}\varepsilon_{T+h-j} + \mathbb{1}(y_{T+h+d} < c) \sum_{j=1}^{\infty} \beta_{\mu,j}\varepsilon_{T+h+j}, \tag{25}$$

where $\beta_{\mu,j}, \beta_{\nu,j}$ are the power series coefficients of $\Psi_\nu(L^{-1})$ and $\Psi_\mu(L^{-1})$ respectively. Taking the conditional expectation on both sides gives the optimal h-step ahead forecast:

$$\begin{aligned} E[y_{T+h|T}] = \phi_{\nu,1}E[y_{T+h-1|T}] + \cdots + \phi_{\nu,r}E[y_{T+h-r|T}] \\ + \mathbb{P}(y_{T+h+d} \geq c \mid \mathcal{Y}_T) \sum_{j=1}^{\infty} \beta_{\nu,j}E[\varepsilon_{T+h+j|T}] \\ + \mathbb{P}(y_{T+h+d} < c \mid \mathcal{Y}_T) \sum_{j=1}^{\infty} \beta_{\mu,j}E[\varepsilon_{T+h+j|T}], \end{aligned} \tag{26}$$

where $\mathcal{Y}_t = (y_T, \ldots, y_0)$ is the information set of y at time T. Interestingly the non-causal part of the forecasted becomes a probability weighted average of forecasts of the two possible states. It thus becomes somewhat like the SMART model in its forecast. The same approximation for the sum of future error terms can be used as explained in section 3.4.

The difficult part here is the need to determine the probability of being in a regime. First, we implement a standard unconditional probability. We aim to also investigate using the conditional probability to provide for better forecasts.[4]

3.7 Forecasting in the SMART model

Forecasting in the SMART model is similar to that of the MART. Once again we consider two cases, 1. $h \leq d$ 2. $h > d$. For the first case the causal part of the forecast is known, and thus as in the MART

[4] We were unable to do this thus far, our goal is to complete this for the final report.

model, only the noncausal part is left to be determined. Combining Equations 23 & 18 & 19, y_{T+h} can be represented by the following equation,

$$\begin{aligned} y_{T+h} = & \left[G(y_{T+h-d}, c, \gamma)\, \boldsymbol{\phi}_\nu + (1 - G(y_{T+h-d}, c, \gamma))\, \boldsymbol{\phi}_\mu\right] \boldsymbol{x}_{T+h} \\ & + G(y_{T+h+d}, c, \gamma) \sum_{j=1}^{\infty} \beta_{\nu,j} \varepsilon_{T+h+j} \\ & + (1 - G(y_{T+h+d}, c, \gamma)) \sum_{j=1}^{\infty} \beta_{\mu,j} \varepsilon_{T+h+j}. \end{aligned} \tag{27}$$

Taking the conditional expectation yields,

$$\begin{aligned} y_{T+h} = & \left[G\left(E(y_{T+h-d|T}), c, \gamma\right) \boldsymbol{\phi}_\nu + \left(1 - G\left(E(y_{T+h-d|T}), c, \gamma\right)\right) \boldsymbol{\phi}_\mu\right] \boldsymbol{x}_{T+h} \\ & + G\left(E(y_{T+h+d|T}), c, \gamma\right) \sum_{j=1}^{\infty} \beta_{\nu,j} E(\varepsilon_{T+h+j}) \\ & + \left(1 - G\left(E(y_{T+h+d|T}), c, \gamma\right)\right) \sum_{j=1}^{\infty} \beta_{\mu,j} E(\varepsilon_{T+h+j}). \end{aligned} \tag{28}$$

The same approximation for the future error terms can be used here as in section 3.4. The same question arises as in the forecast for the MART, how to calculate the conditional expectation of y_{T+h+d}.

Interestingly, when $\mathbb{P}(y_{T+h+d} \geq c \mid \mathcal{Y}_T) = G(E(y_{T+h+d|T}), c, \gamma)$ the noncausal part of the MART and SMART forecast, when the parameters of per regime are equal in both models, becomes identical for the two forecasts. For y_{T+h-d} sufficiently far from the threshold such that $G(E(y_{T+h+d|T}), c, \gamma) \approx 0$ or $G(y_{T+h-d|T}, c, \gamma) \approx 1$, the two forecast become approximately identical.

3.8 Simulation Study

To study the finite-sample properties of the estimators and the tests of the (regime switching) MAR model, a small Monte Carlo simulation study is conducted, as was also done in Lanne and Saikkonen (2011). The results are based on 10,000 realisations and each realization is generated in two steps for the MAR model without regime switching:

1. A series from the causal AR(r) model $\Phi(L)u_t = \epsilon_t$ for $(t = r+1, \ldots, T)$ is generated.
2. Then y_t is computed recursively from $\Psi(L^{-1})y_t = u_t$ for $(t = T-s, \ldots, 1)$.

$\Phi(L)$ is the lag polynomial of order r and $\Psi(L^{-1})$ the lead polynomial of order s. We follow the procedure used in Lanne and Saikkonen (2011): a MAR(1,1) is considered. This follows from Breidt et al. (1991): a second-order process as the data-generating process (DGP) is used as this is the simplest model that still permits an analysis of different aspects of estimation and testing. The r and s initial observations, respectively, are set to zero, and to eliminate the initialization effect, 100 observations at the beginning and end of each realization are discarded. In all experiments, the error term ϵ_t is assumed to follow the t-distribution with 3 degrees of freedom and σ is set to 0.1. In our research, three different combinations of parameter estimates are considered, $(\phi_1, \psi_1) = \{(0.9, 0.9), (0.9, 0.1), (0.1, 0.9)\}$. In the first case, the roots of the lag polynomials are equal and close to the unit circle. In the two other cases, the roots of the "causal" and "noncausal" polynomials are clearly different. Three sample sizes in this simulation study are considered, namely 300, 500 and 800.

When regime switching is included in the simulation study, the two steps change trivially to incorporate the parameters from the MART and SMART models explained above.

3.9 Possible extension: external regressors

A potential extension of our MART/SMART model involves the incorporation of external regressors.[5] To prevent overfitting and ensure methodological rigor, we aim to follow a step-by-step approach rather than introducing an excessive number of variables simultaneously. Some relevant variables have already been discussed in Section 2 above.

In the existing literature, the simultaneous integration of external regressors and a mixed causal-noncausal autoregressive framework remains relatively novel. Hecq et al. (2020) incorporated commodity prices to enhance predictive performance and to distinguish between various MARX models using second-order properties. However, their approach relies on non-Gaussian maximum likelihood estimation, which poses challenges in our context due to the non-continuously differentiable nature of functions resulting from regime switching. Similarly, Stock and Watson (1999) examined an extensive set of external variables but ultimately found that the most effective models were relatively simple, often grounded in the well-established Phillips Curve.

This consideration leads us to an external regressor of particular interest: the unemployment rate. An insightful study, albeit in the context of Nigeria rather than the United States, is Adeleye et al. (2020). This research employs several external regressors that could be valuable for our analysis, provided that equivalent U.S. data are available. Alongside a few additional intuitive variables, these regressors form a promising set for further investigation. Nevertheless, a careful assessment is required to determine data availability and evaluate the computational feasibility and empirical relevance of incorporating these variables into our model:

- Interest rate
- Government expenditure
- Exchange rate
- Export/food export
- Inflation rate of trade of trading partners (import)
- Unemployment rate
- The other variables mentioned in Section 2

There are multiple ways to select which exogenous variables to include in the MARX model. Hecq et al. (2020) use the information criteria to select, however, other methods exist in the literature as well (general-to-specific and specific-to-general for example). As the inclusion of external regressors is an extension, we leave the exact details for the "research" part of our seminar.

[5]Not yet implemented in results, aim is to complete this for final report

4 Results

This section covers the results of our research. First, we implement a standard MAR model and compare its characteristics to a standard AR model. Further we extend it to a MAR with a threshold: MART. This model is even further extended to incorporate smooth transitions: SMART. Finally, we consider some exogenous regressors.

All simulations and results were performed on a MacBook Air with an Apple M2 chip and 8GB of RAM, using R version 4.4.2 within RStudio for all computations. As previously mentioned, the MARX package from Hecq et al. (2017b) served as the baseline for our own modifications and extensions needed for our models.

4.1 Standard MAR and AR

This part covers the basic MAR models. These are compared against their "simpler" AR counterparts. First a simulation study is conducted in order to examine finite-sample properties. Further, the models are selected. Finally the out-of-sample forecasting performance is analyzed.

4.1.1 Simulation study

A simulation study is done in order to examine the finite-sample properties of the estimators. We consider a MAR(1,1) model. Monte Carlo simulation is conducted based on 10,000 realizations. The 100 observations at the beginning and end of each realization are discarded. The error term ϵ_t is assumed to follow the t-distribution with 3 degrees of freedom and σ is set to 0.1. Sample sizes 300, 500 and 800 are used in this simulation study. Table 2 shows the estimates with standard deviation within parentheses.

Table 2: Simulation estimates MAR(1,1) model

		$\phi = 0.9,\ \psi = 0.9$		$\phi = 0.9,\ \psi = 0.1$		$\phi = 0.1,\ \psi = 0.9$	
Sample size	Param	Estimate	Scale / df	Estimate	Scale / df	Estimate	Scale / df
300	ϕ_1	0.882 (0.045)	0.089 /	0.696 (0.338)	0.037 /	0.274 (0.329)	0.031 /
	ψ_1	0.879 (0.045)	4.52	0.286 (0.352)	4.50	0.718 (0.316)	4.24
500	ϕ_1	0.895 (0.023)	0.094 /	0.872 (0.126)	0.098 /	0.150 (0.188)	0.095 /
	ψ_1	0.893 (0.024)	3.05	0.119 (0.125)	3.20	0.846 (0.190)	3.17
800	ϕ_1	0.897 (0.015)	–0.038 /	0.899 (0.014)	–0.068 /	0.110 (0.093)	–0.053 /
	ψ_1	0.896 (0.015)	3.15	0.097 (0.036)	3.08	0.887 (0.092)	3.05

In our first case, where $(\phi, \psi) = (0.9, 0.9)$: true parameters are large and near a unit root, the estimates for both ϕ and ψ are close to 0.9 across all samples. Further, as the sample size increases the standard deviations shrink, showing the estimators are consistent. For the second case $(\phi, \psi) = (0.9, 0.1)$ we expect a strongly causal process and weak noncausal component. ϕ is still estimated relatively well and converges nicely to 0.9 for large sample sizes. However, ψ estimate centers closer to 0.1, but with larger standard deviation — especially in small samples. For the final case $(\phi, \psi) = (0.1, 0.9)$ we observe the same pattern.

We see that the degrees of freedom (dof) decreases as the sample size increases and that suggests that the model is accurately capturing the true distribution of the data. As the sample size grows, the model becomes better at estimating the variance and reduces the need for excessively heavy tails, which aligns with the true dof of 3. This indicates that the model is stabilizing and converging to the correct distribution.

The scale parameter in table 2 reflects the dispersion of the residuals in the model, with smaller values indicating less variation in the errors as sample size increases. While negative scale values appear, they don't affect the parameter estimation because in AMLE (Approximate Maximum Likelihood Estimation), the scale parameter is squared in the likelihood function, making it effectively positive and ensuring that the estimation remains valid.

4.1.2 Model selection

We estimate a MAR model as described in Section 3: first a Pseudo-Causal AR(p) model is estimated for all $p = 1, ..., p_{max}$. The maximum possible value of p is set at 18 (1 year and 6 months) instead of 12 in this instance, this demonstrates that the number of lags selected is not equal to the maximum number of lags. For instance, when $p_{max} = 12$, the number of lags is equal to the maximum value. The results are presented in Table 3: the model selection criteria for each lag are estimated.

Table 3: Pseudo-Causal AR(p) model selection criteria by lag order

Lag (p)	**BIC**	**AIC**	**HQIC**
0	0.779	0.773	0.775
1	0.398	0.386	0.391
2	0.403	0.385	0.392
3	0.409	0.385	0.394
4	0.400	0.370	0.382
5	0.408	0.373	0.386
6	0.415	0.373	0.390
7	0.404	0.357	0.375
8	0.410	0.356	0.376
9	0.410	0.350	0.373
10	0.395	0.329	0.355
11	0.360	0.288	0.315
12	**0.349**	**0.271**	**0.301**
13	0.355	0.271	0.303
14	0.364	0.274	0.308
15	0.369	0.273	0.310
16	0.378	0.276	0.315
17	0.381	0.272	0.314
18	0.386	0.271	0.316

We observe that the minimum is obtained at lag $p = 12$ for all selection criteria. Therefore, we continue with $p = s + r = 12$ for further model selection. The KS-test rejects the null hypothesis of normally distributed residuals of the purely causal AR(12) model with a $p < 0.000$. This is again confirmed by the JB-test, which also rejects the same null hypothesis (p-value < 0.000). Thus, the causal and noncausal polynomial can be distinguished. Additionally, we consider the test for i.i.d.-ness of the AR(12) residuals as described in Section 3.3. The regression of $\hat{\epsilon}_t$ on $\hat{\epsilon}^2_{t-1}$ up to $\hat{\epsilon}^2_{t-m}$ with $m = 12$ leads to the rejection of the null hypothesis of all coefficients being equal to zero (χ^2-statistic $= 49.5$ corresponding to a p-value 0.000). We choose $m = 12$ as this is equal to the optimal number of lags determined in the AR model. Furthermore we test for no serial correlation between the residuals. Using the Ljung-Box test with number of lags equal 20. The p-value of the test equals 0.997, which is very large, meaning that we do not reject the null hypothesis and all serial correlation between the residuals has been deleted.

Now, the final step is to select the best lead and lag components in the MAR(r,s). This is done using the AML approach as described in Section 3.3. These results can be seen in Table 4. We find a MAR(1,11) as our optimal model.

Table 4: Comparison of AR(12) and MAR(1,11) models

Model	**AR(12)**	**MAR(1,11)**
Scale	0.215	-0.196
Degrees of Freedom	5.438	3.816
AIC	0.195	**0.168**
BIC	0.297	**0.264**
HQIC	0.234	**0.205**
Log-Likelihood	-58.41	**-49.23**

Table 4 shows that the mixed causal-noncausal autoregressive with 1 lag and 11 leads outperforms the standard AR(12) model. The MAR(1,11) model achieves lower values across all information criteria (AIC, BIC, HQIC), indicating a better balance between model fit and complexity. Additionally, the MAR model yields a notably higher log-likelihood value (-49.23 compared to -58.41), suggesting a superior in-sample fit.

4.1.3 Forecasting

We use the two models above to forecast inflation. An expanding window approach is implemented with 100 initial observations. The truncation parameter M, which plays a critical role in the moving average (MA) representation of the noncausal component of the model, is set to 50 (which is the literature standard). The number of simulations to forecast noncausal component N is set to 1,000 for computational feasibility. The results of the out-of-sample forecasting performance are shown in Table 5.

Table 5: RMSFE and Diebold-Mariano test comparison of MAR(1,11) and AR(12) models

Forecast horizon	**MAR(1,11) RMSFE**	**AR(12) RMSFE**	**DM p-value**
$h = 1$	0.471	0.306	< 0.001
$h = 6$	0.390	0.336	< 0.001
$h = 12$	0.375	0.373	0.883
$h = 24$	0.377	0.403	0.019

Table 5 presents the Root Mean Squared Forecast Error (RMSFE) values for the mixed causal-noncausal model and the purely causal model at forecast horizons $h = 1$, 6, 12, and 24. For the shortest horizon ($h = 1$), the AR(12) model outperforms the MAR(1,11) model considerably, with a RMSFE of 0.306 compared to 0.471. This is consistent with the economic idea that short-term inflation is mostly driven by past information. As our forecast horizon increases, the RMSFE gap decreases. For $h = 6$ the MAR(1,11) RMSFE drops down to 0.390 while the causal AR(12) model remains somewhat more accurate with a RMSFE equal to 0.336. For the one-year-ahead forecast horizon ($h = 12$) two models seem to perform equivalently (RMSFEs of 0.375 and 0.373, respectively), suggesting that the inclusion of noncausal variables may begin to capture relevant forward-looking influences in the inflation framework. Interestingly, at the longest horizon ($h = 24$), namely 2 years ahead, the MAR(1,11) performs better than the AR(12) model (RMSE of 0.377 vs. 0.403). This could be an indication that for longer horizons, structural dynamics and anticipatory characteristics become more important, and noncausal information may improve predictive performance. On the other side, the marginal differences also show that the forecast uncertainty increases as the forecast horizon expands. Overall, the causal model dominates in the short run, while the MAR(1,11) model becomes more competitive in the medium to long run.

The Diebold-Mariano test results as shown in Table 5 prove that the relative forecasting performance of the MAR(1,11) and AR(12) models differs across various forecast horizons. For horizons $h = 1$ and

$h = 6$, we observe extremely small p-values, indicating that the AR(12) model produces significantly more accurate forecasts than the MAR at short horizons. For the one-year-ahead forecasts ($h = 12$), the p-value becomes 0.883: no statistically significant difference in forecast accuracy. Finally, for the two-year-ahead forecast horizon ($h = 24$), the p-value becomes 0.019. This implies that the MAR(1,11) model significantly outperforms the AR(12) at longer horizons. These results showcase that the relative effectiveness of causal and noncausal models differs across forecast horizons, with a MAR model becoming more valuable in the long run.

4.2 MAR with threshold: MART

In this section the results obtain from the MART model are explained. We focus on model selection, coefficient estimates and forecasting.

4.2.1 Model selection

Before we start estimating a MART model, we first have to determine the threshold used in our model. For the threshold we use the median of the inflation data, which is approximately 0.29. This corresponds with 3.5% annual inflation. The grid search performed on the threshold, lags and leads, found that there was no clear pattern with negligible differences in BIC performance between the different models when looking at the threshold and lags and leads. The first 20 best results with the lowest BIC of a total of 1.859 results are shown in Table 13 in the Appendix A. Furthermore by choosing the median we ensure that sufficient amount of data is allocated to both regimes, allowing for more reliable and stable parameter estimates and reducing the risk of overfitting or instability in either regime due to data sparseness. Therefore we choose to estimate the model as described in Section 3, this method is also been done by Hecq et al. (2016) and Lanne and Saikkonen (2011).

To determine the parameter d, we conduct a separate grid search that considers the model's threshold but not the number of lags and leads, as these are not predefined. Upon reviewing the results of the grid search (the first 20 entries in Table 13), we note that, given the threshold must be set such that at least 15% of the data falls into one regime and the rest into the other (implying the threshold should be 0.6 or smaller), the number of lags and leads remains relatively small. Consequently, and to reduce computation time, we decide that, when determining d, the number of lags and leads can vary independently between 0 and 4. In Table 14 in Appendix A, the first best 10 results of the grid search are shown, selected on the best BIC. We see that $d = 3$ has the best BIC value, and therefore we choose to use this value for d.

The first step after determining the threshold and d is to determine the number of lags in a Pseudo-Causal AR(p) model, this model is estimated for all $p = 1, \ldots, pmax$. The maximum number of lags is chosen to be 12, which is equal to a year. The results of the selection criteria are shown in Table 6. As mentioned earlier in Section 3.3, Hecq et al. (2016) found favor for the use of the BIC by chosing the number of lags in the AR model. Therefore we observe that the minimum value is obtained at $p = 2$.

Table 6: Pseudo-Causal AR(p) model selection criteria by lag order for MART model

Lag (p)	BIC	AIC	HQIC
0	0.781	0.775	0.777
1	0.386	0.368	0.375
2	**0.378**	0.349	0.360
3	0.389	0.347	0.363
4	0.395	0.341	0.362
5	0.412	0.347	0.372
6	0.421	0.344	0.374
7	0.421	0.331	0.366
8	0.434	0.333	0.372
9	0.443	0.329	0.373
10	0.438	0.312	0.361
11	0.418	0.280	0.333
12	0.412	**0.270**	**0.327**

After determination of the lag order, we must test for the normality of the residuals of the AR(2) model. The KS-test and the JB-test both reject the null hypothesis of normally distributed residuals of the purely causal AR model with a p-value of 0 for both tests. This means that the causal and noncausal polynomial can be distinguished. Furthermore wes test for the i.i.d-ness of the AR(2) residuals. The regression of $\hat{\epsilon}_t$ on $\hat{\epsilon}^2_{t-1}$ and $\hat{\epsilon}^2_{t-2}$. We chose $m = 2$ to put in the general function in Section 3.3 as this is equal to the optimal number of lags determined in the AR model. The regression leads to the rejection of the null hypothesis $H_0 : \delta_1 = \delta_2 = 0$ of all coefficients being equal to zero (χ^2-statistic $=$ 26.528 corresponding to a p-value 0.00). In addition, we check if there is serial correlation between the residuals using the Ljung-Box test with number of lags equal to 6 as $ln(787) = 6,67$. The p-value of the test equals 0.090, which is a bit larger than 0.05, meaning that we do not reject the null hypothesis and all serial correlation between the residuals has been deleted.

Now we select the best lead (r) and lag (s) component for the MART(r,s) model. This is done using the AML approach also described in Section 3.3. We found MART(1,1) to be our optimal model ,with the median as threshold and d equal to 3. This model with the specified parameters is also found by the grid search in Table 13 and belongs to the best 20 models found based on the BIC. And this model is also found to be the best in Table 14.

Table 7 shows that the MART model with 1 lag and 1 lead outperforms the ART(2) model, which is the standard AR model with an added threshold. We choose the threshold and d to be the same value as for the MART(1,1) model. The MART(1,1) model achieves lower values across all information criteria (AIC, BIC, HQIC), indicating a better balance between model fit and complexity. Furthermore, the MART(1,1) model yielded a higher log-likelihood value (-104,369 compared to -129,331), suggesting a better in-sample fit.

Table 7: MART(1,1) and ART(2) model

Model	MART(1,1)	ART(2)
Threshold	0.29 (Median)	0.29 (Median)
Scale	-0.220	0.243
Degrees of Freedom	4.482	6.892
d	3	3
AIC	**0.288**	0.357
BIC	**0.335**	0.416
HQIC	**0.306**	0.380
Log-Likelihood	**-104.369**	-129.331

In Table 8 the coefficient estimates and their corresponding p-values are shown. All the coefficients are statistically significant, as they all have a p-value of zero.

Table 8: Coefficients estimates and p-values from the MART model and difference test between regime estimates.

Coefficients	**Estimate**	**p-value**
Causal Regime 1	-0.067	0.00
Causal Regime 2	-0.166	0.00
Non-Causal Regime 1	0.597	0.00
Non-Causal Regime 2	0.676	0.00

4.2.2 Forecasting with a MART model

There is probably a bug in our code that produces the following results:

Table 9: RMSE and Diebold-Mariano test comparison for MART(1,1) and ART(2)

horizon	**RMSE_mart**	**RMSE_art**	**DM_mart_vs_causal**
h1	671.0018	0.3255669	0.05210577
h2	138.6832	0.3535888	0.10240551
h3	388.4514	0.3486271	0.08175155

These results are obviously absurd. What is happening is that for some windows, $|\varphi| > 1$ for either one of the regimes, this causes the term $\sum_{j=0}^{M-h} \beta_j \epsilon_{T+h+j}$ to explode, leading to extremely inflated forecasts. Why these estimates can become larger than 1 when the series is clearly non-explosive is currently our highest priority question to solve.

4.3 SMART

In this section the results obtained from the SMART model are explained. We focus on model selection, coefficient estimates and forecasting.

4.3.1 Model Selection

Before we start estimating a SMART model, we first have to determine the threshold, d and gamma used in our model. All parameters are mentioned in 3.5. For the threshold we use the median of the inflation data, which is approximately 0.29. This corresponds with 3.5% annual inflation. By the same arguments as stated in Section 4.2, we use the median as the threshold in this model. The results of the grid search performed on the threshold, lags and leads are shown in Appendix A, where the best 20 results based on the BIC of the different models are displayed. Here you can see that the BIC values are very close to one and other, and therefore we found them again negligible.

However, we did use a grid search to determine the values for d and gamma. Both parameters have not a clear pattern and are not easily observable as multiple values have the same BIC. But we found a small favour for d equal to 6 and gamma equal to 15.

The first step after determining the threshold, d and gamma is to determine the number of lags in a Pseudo-Causal AR(p) model, this model is estimated for all $p = 1, \dots, pmax$. The maximum number of lags is chosen to be 12, which is equal to a year. The results of the selection criteria are shown in Table 10. As mentioned before, we use the BIC to select the number of lags, in this case we see that BIC has the lowest value at $p = 2$.

Table 10: Pseudo-Causal AR(p) model selection criteria by lag order for SMART model

Lag (p)	**BIC**	**AIC**	**HQIC**
0	0.783	0.777	0.779
1	0.392	0.375	0.381
2	**0.380**	0.350	0.361
3	0.394	0.352	0.368
4	0.396	0.342	0.362
5	0.408	0.343	0.368
6	0.426	0.348	0.378
7	0.420	0.331	0.365
8	0.434	0.332	0.371
9	0.446	0.332	0.376
10	0.435	0.309	0.357
11	0.410	0.272	0.325
12	0.399	**0.249**	**0.306**

Following the determination of the lag order, it is important to test the normality of the residuals of the AR(2) model. Utilising the KS-test and the JB-test, it is observed that the null hypothesis of normally distributed residuals of the purely causal AR model are both rejected, with a p-value of 0 observed for both tests. This outcome indicates that it is possible to distinguish between causal and non-causal polynomial models. Furthermore, an investigation is conducted to ascertain the independence of the AR(2) residuals. The regression of $\hat{\epsilon}_t$ on $\hat{\epsilon}^2_{t-1}$ and $\hat{\epsilon}^2_{t-2}$. The value of m was set to 2 in the general function presented in Section 3.3, as this corresponds to the optimal number of lags identified within the AR model. The regression leads to the rejection of the null hypothesis $H_0 : \delta_1 = \delta_2 = 0$ of all coefficients being equal to zero (χ^2-statistic = 26.528 corresponding to a p-value 0.0). Furthermore, the presence of serial correlation between the residuals is investigated using the Ljung-Box test, with the number of lags set at 6 as stated above. The p-value of the test is 0.09, which is just a bit larger than the significance level of 0.05. This indicates that the null hypothesis is not rejected and that all serial correlation between the residuals has been eliminated.

The next step is to select the best lead (r) and lag (s) components for the SMART(r,s) model. This is done using AML. We found SMART(0,2) to be our best model, with the median as threshold, gamma equal to 15 and d equal to 7.

As demonstrated in Table 11, the SMART model with 0 lags and 2 leads exhibits superior performance in comparison to the SART(2) model. The threshold, d and gamma in the SART model are set to match the values employed in the SMART(0,2) model. The SMART(0,2) model achieves lower values across all information criteria (AIC, BIC, HQIC), indicating a superior balance between model fit and complexity. In addition, the SMART(0,2) model yielded a higher log-likelihood value, suggesting a superior in-sample fit.

Table 11: SMART(0,2) and SART(2) model

Model	**SMART(0,2)**	**SART(2)**
Threshold	0.29 (Median)	0.29 (Median)
Scale	0.225	-0.242
Degrees of Freedom	4.998	6.925
d	6	6
Gamma	15	15
AIC	**0.321**	0.328
BIC	**0.381**	0.388
HQIC	**0.344**	0.351
Log-Likelihood	**-114.370**	-117.226

Table 12: Coefficients estimates and p-values from the SMART model.

Coefficients	**Estimate**	**p-value**
Causal Regime 1	0.00	-
Causal Regime 2	0.00	-
noncausal Regime 1.1	0.416	0.000
noncausal Regime 1.2	0.217	0.000
noncausal Regime 2.1	0.561	0.000
noncausal Regime 2.2	-0.088	0.000

4.3.2 Forecasting with a SMART model

The SMART forecasts currently suffer from the same issue as MART forecasts. We are working on solving this issue.

4.4 Exogenous regressors

These will be included in our final report.

5 Conclusion

Inflation is a widely recognized economic indicator, the predictions of which are utilized by governments, central banks and the private sector. Accurate inflation forecasts are therefore important from a theoretical as well as a practical standpoint. In this seminar, we have built upon the mixed causal-noncausal autoregressive (MAR) model introduced by Lanne and Saikkonen (2011), and extended this framework by incorporating regime-switching models. The main goal of this research was to investigate to what extent the inflation forecasting performance of these models can be improved by using MAR models with regime-switching - which is done by comparing these new models against standard autoregressive (AR) models with similar regime-switching definitions.

We used non-seasonally adjusted monthly CPI data from U.S. Bureau of Labor Statistics (2025) to measure inflation, and incorporated macroeconomic factors from the FRED-MD dataset (Federal Reserve Bank of St. Louis, 2025) for further exploration. After finding that a standard AR model does not show normally and i.i.d. distributed error terms, a MAR(1,11) model is implemented. This model's forecasting performance, especially for longer forecasting horizons (up to 24 months), gives some promising results. For this longer forecast horizon, the MAR model significantly outperformed the simple benchmark AR(12) model with regard to the Diebold-Mariano test and the root mean squared forecast error (RMSFE).

The MAR model was further extended by including regime-switching using threshold and a smooth-transition techniques, resulting in the MART and SMART models. FURTHER RESULTS MART AND SMART

These results have multiple important implications. Our research suggests that using noncausal components can benefit inflation forecasts for the medium and long-term horizon. This confirms the theoretical idea that expectations and forward-looking dynamics influence inflation numbers. This also justifies the use of MAR models - specifically in regimes characterized by structural changes for inflation dynamics. Theoretically, this research adds to the literature by introducing threshold specifications for an existing MAR model. This further supports the deviation from traditional backward-looking AR models, also giving attention to the importance of expectations and nonlinear characteristics of macroeconomic time series.

Besides these contributions, we find some limitations in our research. First, due to our focus on univariate inflation forecasts, we have not spent much attention looking at exogenous regressors in our modelling framework, which could result in more accurate forecasting performance. Second, due to the focus on implementing the threshold model, the estimation process for the selection of thresholds and the smoothing parameters was not thoroughly investigated. Other estimation techniques than grid search can possibly improve estimation by removing some initializations and computational constraints.

Therefore, future research could incorporate some of the following propositions. INCORPORATING EXTERNAL REGRESSORS Hecq et al. (2020).[6] Another possibly value addition could be to explore whether Bayesian or machine learning methods could be used to improve threshold, leads, and lags estimation in high-dimensional models. Lastly, the MAR and SMART models could be extended to a multivariate setting in order to better capture inflation and its key drivers, generating even more robust forecasts.

[6] Will be incorporated in final report

References

Adeleye, N., Ogundipe, A. A., Ogundipe, O., Ogunrinola, I., and Adediran, O. (2020). Internal and external drivers of inflation in nigeria.

Amisano, G. and Fagan, G. (2013). Money growth and inflation: A regime switching approach. *Journal of International Money and Finance*, 33:118–145.

Andrews, B., Davis, R. A., and Breidt, F. J. (2006). Maximum likelihood estimation for all-pass time series models. *Journal of Multivariate Analysis*, 97(7):1638–1659.

Andrews, D. W. and Chen, H.-Y. (1994). Approximately median-unbiased estimation of autoregressive models. *Journal of Business & Economic Statistics*, 12(2):187–204.

Ball, L., Leigh, D., and Mishra, P. (2022). Understanding us inflation during the covid-19 era. *Brookings Papers on Economic Activity*, 2022(2):1–80.

Batini, N., Jackson, B., and Nickell, S. (2005). An open-economy new keynesian phillips curve for the u.k. *Journal of Monetary Economics*, 52(6):1061–1071. The econometrics of the new Keynesian price equation.

Berger, V. W. and Zhou, Y. (2014). Kolmogorov–smirnov test: Overview. In *Wiley StatsRef: Statistics Reference Online*. John Wiley & Sons.

Blasques, F., Koopman, S. J., and Mingoli, G. (2023). Observation-driven filters for time-series with stochastic trends and mixed causal non-causal dynamics. Technical report, Tinbergen Institute Discussion Paper.

Boudoukh, J. and Richardson, M. (1993). Stock returns and inflation: A long-horizon perspective. *The American economic review*, 83(5):1346–1355.

Brassington, G. (2017). Mean absolute error and root mean square error: which is the better metric for assessing model performance? In *EGU general assembly conference abstracts*, page 3574.

Breidt, F. J., Davis, R. A., Lh, K.-S., and Rosenblatt, M. (1991). Maximum likelihood estimation for noncausal autoregressive processes. *Journal of Multivariate Analysis*, 36(2):175–198.

Broyden, C. G. (1970). The convergence of a class of double-rank minimization algorithms 1. general considerations. *IMA Journal of Applied Mathematics*, 6(1):76–90.

Caldara, D., Conlisk, S., Iacoviello, M., and Penn, M. (2022). The effect of the war in ukraine on global activity and inflation. *FEDS Notes*.

Campbell, J. Y., Lo, A. W., MacKinlay, A. C., and Whitelaw, R. F. (1998). The econometrics of financial markets. *Macroeconomic Dynamics*, 2(4):559–562.

Clarida, R., Gali, J., and Gertler, M. (2000). Monetary policy rules and macroeconomic stability: evidence and some theory. *The Quarterly journal of economics*, 115(1):147–180.

Collins, W. A. and Hopwood, W. S. (1980). A multivariate analysis of annual earnings forecasts generated from quarterly forecasts of financial analysts and univariate time-series models. *Journal of Accounting Research*, 18(2):390–406.

Dahl, C. M. and Hansen, N. L. (2001). The formation of inflation expectations under changing inflation regimes. *Studies in Nonlinear Dynamics & Econometrics*, 4(4).

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a):427–431.

Diebold, F. X. and Mariano, R. S. (1995). Com paring predictive accu racy. *Journal of Business and Economic Statistics*, 13(3):253–263.

Faust, J. and Wright, J. H. (2013). Forecasting inflation. In Elliott, G. and Timmermann, A., editors, *Handbook of economic forecasting*, volume 2, Part A, pages 2–56. Elsevier.

Federal Reserve Bank of St. Louis (2025). FRED-MD: A monthly database for macroeconomic research. Accessed: 2025-03-19.

Franses, P. H. (1998). *Time series models for business and economic forecasting*. Cambridge university press.

Gan, P. (2014). The optimal economic uncertainty index: A grid search application. *Computational Economics*, 43:159–182.

Goodfriend, M. (2005). Inflation targeting in the united states? In Bernanke, B. S. and Woodford, M., editors, *The Inflation-Targeting Debate*, pages 311–337. University of Chicago Press, Chicago.

Gouriéroux, C., Zakoian, J.-M., et al. (2013). *Explosive bubble modelling by noncausal process*. CREST.

Hassani, H. and Yeganegi, M. R. (2020). Selecting optimal lag order in ljung–box test. *Physica A: Statistical Mechanics and its Applications*, 541:123700.

Hecq, A., Issler, J. V., and Telg, S. (2020). Mixed causal–noncausal autoregressions with exogenous regressors. *Journal of Applied Econometrics*, 35(3):328–343.

Hecq, A., Lieb, L., and Telg, S. (2016). Identification of mixed causal-noncausal models in finite samples. *Annals of Economics and Statistics*, (123/124):307–331.

Hecq, A., Lieb, L., and Telg, S. (2017a). Simulation, estimation and selection of mixed causal-noncausal autoregressive models: The marx package. Workingpaper, SSRN e-library.

Hecq, A., Lieb, L., and Telg, S. (2017b). Simulation, estimation and selection of mixed causal-noncausal autoregressive models: The marx package. Working Paper 3015797, SSRN Electronic Journal. Available at SSRN: https://ssrn.com/abstract=3015797 or http://dx.doi.org/10.2139/ssrn.3015797.

Hecq, A., Telg, S., and Lieb, L. (2017c). Do seasonal adjustments induce noncausal dynamics in inflation rates? *Econometrics*, 5(4):48.

Hecq, A. and Velásquez-Gaviria, D. (2025). Spectral estimation for mixed causal-noncausal autoregressive models. *Econometric Reviews*, pages 1–24.

Jarque, C. M. and Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics letters*, 6(3):255–259.

Korkmaz, S. (2010). Backward-looking phillips curve, forward-looking phillips curve and application on turkey. *International Research Journal of Finance and Economics*, 55:170–176.

Lanne, M. and Luoto, J. (2013). Autoregression-based estimation of the new keynesian phillips curve. *Journal of Economic Dynamics and Control*, 37(3):561–570.

Lanne, M., Luoto, J., and Saikkonen, P. (2012). Optimal forecasting of noncausal autoregressive time series. *International Journal of Forecasting*, 28(3):623–631.

Lanne, M. and Saikkonen, P. (2011). Noncausal autoregressions for economic time series. *Journal of Time Series Econometrics*, 3(3).

McCracken, M. W. and Ng, S. (2016). Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4):574–589.

McCracken, M. W. and Ng, S. (2020). Fred-qd: A quarterly database for macroeconomic research. Working Paper 28283, National Bureau of Economic Research.

Meese, R. and Geweke, J. (1984). A comparison of autoregressive univariate forecasting procedures for macroeconomic time series. *Journal of Business & Economic Statistics*, 2(3):191–200.

Moshiri, S. and Cameron, N. (2000). Neural network versus econometric models in forecasting inflation. *Journal of forecasting*, 19(3):201–217.

Muth, J. F. (1961). Rational expectations and the theory of price movements. *Econometrica*, 29(3):315–335.

Nalewaik, J. (2015). *Regime-switching models for estimating inflation uncertainty*. Federal Reserve Board.

Nyberg, H. and Saikkonen, P. (2014). Forecasting with a noncausal var model. *Computational statistics & data analysis*, 76:536–555.

Phillips, A. W. (1958). The relation between unemployment and the rate of change of money wage rates in the united kingdom, 1861-1957. *Economica*, 25(100):283–299.

Roeger, W. and Herz, B. (2012). Traditional versus new keynesian phillips curves: Evidence from output effects. *International Journal of Central Banking*, 8:87–109.

Rudd, J. and Whelan, K. (2005). New tests of the new-keynesian phillips curve. *Journal of Monetary Economics*, 52(6):1167–1181.

Spilimbergo, A., Symansky, S., Blanchard, O., and Cottarelli, C. (2009). Fiscal policy for the crisis. *CESifo Forum*, 10(2):26–32.

Stock, J. H. and Watson, M. W. (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business Economic Statistics*, 14(1):11–30.

Stock, J. H. and Watson, M. W. (1999). Forecasting inflation. *Journal of monetary economics*, 44(2):293–335.

Stock, J. H. and Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–162.

Telg, S. (2017). *Mixed causal-noncausal models: identification, estimation and inference*. PhD thesis, Maastricht University, Netherlands.

Tsay, R. S. (2005). *Analysis of Financial Time Series*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc. First published: 19 August 2005. Online ISBN: 9780471746195. 605 pages.

U.S. Bureau of Labor Statistics (2025). Consumer price index. Accessed: 2025-04-10.

Van Dieijen, M. and Lumsdaine, R. L. (2019). What say they about their mandate? a textual assessment of federal reserve speeches. *A Textual Assessment of Federal Reserve Speeches (August 14, 2019)*.

Voisin, E. (2022). *Modelling and forecasting economic time series with mixed causal-noncausal models.* PhD thesis, Maastricht University, Netherlands.

Wells, M. (2024). Federal reserve: The origins of the 2 percent inflation target. *Econ Focus*, 24(1Q/2Q):10–13.

Wong, C. S. and Li, W. K. (2000). On a mixture autoregressive model. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 62(1):95–115.

A Tables

Table 13: Top 20 Grid Search Results on MART

Threshold	Lags	Leads	AIC	BIC	HQ	Loglikelihood	k	N	df	scale
0.90	1	3	0.25	0.32	0.28	-86.25	12	783	4.05	-0.21
0.70	1	11	0.16	0.33	0.22	-33.51	28	775	3.79	-0.19
0.80	1	11	0.16	0.33	0.22	-33.81	28	775	3.82	-0.19
0.80	1	1	0.28	0.33	0.30	-101.95	8	785	4.28	-0.22
0.60	1	2	0.27	0.33	0.29	-95.49	10	784	4.34	-0.22
0.90	1	11	0.16	0.33	0.23	-34.57	28	775	3.68	0.19
0.80	2	1	0.27	0.33	0.29	-96.22	10	784	4.60	0.22
0.60	0	2	0.27	0.33	0.30	-97.06	10	785	4.81	-0.22
0.90	1	4	0.25	0.33	0.28	-83.94	14	782	4.13	-0.21
0.00	1	1	0.29	0.33	0.31	-104.77	8	785	4.45	0.22
1.00	12	1	0.15	0.33	0.22	-29.85	30	774	5.20	-0.21
0.60	2	1	0.28	0.34	0.30	-98.31	10	784	4.53	-0.22
0.90	2	2	0.26	0.34	0.29	-91.72	12	783	4.25	0.21
0.80	2	10	0.17	0.34	0.23	-37.78	28	775	4.79	-0.20
0.10	1	1	0.29	0.34	0.31	-106.44	8	785	4.43	0.22
0.80	3	1	0.27	0.34	0.30	-92.96	12	783	4.37	-0.21
0.00	3	1	0.27	0.34	0.30	-93.05	12	783	4.65	-0.22
0.60	0	1	0.29	0.34	0.31	-107.12	8	786	4.63	0.22
0.70	1	3	0.27	0.34	0.30	-93.31	12	783	4.05	-0.21
0.30	1	1	0.29	0.34	0.31	-107.37	8	785	4.50	0.22

Table 14: Top 10 Grid Search Results for d in MART Model

Threshold	d	Lags	Leads	AIC	BIC	HQ	Loglikelihood	k	N	df	scale
0.29	3	1	1	0.29	0.34	0.31	-104.37	8	781	4.48	-0.22
0.29	5	1	1	0.29	0.34	0.31	-105.03	8	777	4.42	0.22
0.29	2	1	1	0.29	0.34	0.31	-106.86	8	783	4.37	-0.22
0.29	1	1	4	0.26	0.34	0.29	-87.29	14	782	4.02	-0.21
0.29	1	1	1	0.30	0.34	0.31	-107.97	8	785	4.47	-0.22
0.29	9	1	1	0.30	0.34	0.31	-105.54	8	769	4.42	-0.22
0.29	1	1	2	0.29	0.35	0.31	-102.21	10	784	4.55	-0.22
0.29	7	1	1	0.30	0.35	0.32	-107.32	8	773	4.46	-0.22
0.29	11	1	4	0.26	0.35	0.30	-86.54	14	765	3.86	-0.21
0.29	1	1	3	0.28	0.35	0.31	-97.51	12	783	4.21	-0.21

Table 15: Top 20 Grid Search Results on SMART

Threshold	Lags	Leads	AIC	BIC	HQ	Loglikelihood	k	n	df	Scale
1.00	1	3	0.25	0.32	0.27	-83.98	12	783	4.14	-0.21
0.90	1	2	0.26	0.32	0.28	-90.83	10	784	4.34	-0.21
0.90	1	3	0.25	0.32	0.27	-84.11	12	783	4.16	-0.21
0.80	1	2	0.26	0.32	0.28	-91.00	10	784	4.36	-0.21
1.00	1	2	0.26	0.32	0.28	-91.31	10	784	4.31	-0.21
0.90	1	1	0.27	0.32	0.29	-98.13	8	785	4.45	-0.22
0.80	1	1	0.27	0.32	0.29	-98.20	8	785	4.48	-0.22
0.80	1	3	0.25	0.32	0.27	-84.72	12	783	4.17	-0.21
1.00	1	4	0.24	0.32	0.27	-77.91	14	782	3.87	-0.20
0.70	1	2	0.26	0.32	0.28	-91.78	10	784	4.39	-0.21
0.90	1	4	0.24	0.32	0.27	-78.27	14	782	3.89	-0.20
1.00	1	1	0.27	0.32	0.29	-98.75	8	785	4.42	-0.22
0.70	1	1	0.27	0.32	0.29	-99.00	8	785	4.50	0.22
1.00	1	2	0.26	0.32	0.28	-92.45	10	784	4.53	-0.22
0.80	1	4	0.24	0.32	0.27	-78.86	14	782	3.89	-0.20
0.90	1	2	0.26	0.32	0.28	-92.57	10	784	4.54	-0.22
0.40	6	6	0.15	0.32	0.22	-31.38	28	775	5.01	0.20
0.70	1	3	0.25	0.32	0.28	-85.83	12	783	4.19	-0.21
0.30	6	6	0.15	0.32	0.22	-31.39	28	775	5.02	0.20
0.50	6	6	0.15	0.32	0.22	-31.39	28	775	5.01	0.20