UPPSALA
UNIVERSITET

CAUSAL EFFECT ESTIMATION OF POLICY RATES ON INFLATION RATES: A STUDY BASED ON SWEDISH AND US DATA

Submitted by

Henry Svedberg

A thesis submitted to the Department of Statistics
in partial fulfillment of the requirements
for a two-year Master of Science degree in Statistics
in the Faculty of Social Sciences

Supervisor
Yukai Yang

Spring, 2024

ABSTRACT

Central banks can steer the inflation rate using monetary policy by its use of policy rates. The effect of monetary policy is one of the most researched areas within economics. Understanding the effects and timings of the causal transmission mechanism is of great importance to policymakers to determine the policy action at the correct point in time. However, much of the available literature analyzing the causal effects is typically done by prediction-based methods which are not able to estimate actual causal effects, such as Granger Causality. Instead, this study uses a method based on the causal inference framework to determine if there is a causal effect of police rates on inflation. Specifically, this study uses a newly developed nonparametric test based on the influence function for the primitive parameter for a nonparametric model to determine if the causal-dose response curve is flat. Furthermore, this study also uses nonparametric regression to estimate the relationship between policy rates and inflation. Using Swedish and US data, the results showed that there may be a causal effect of policy rates on inflation after about 2-4 months for Sweden and 2-5 months for the USA. However, the nonparametric regression did not capture any kind of relationship.

Keywords: Inflation, policy rates, interest rates, causal inference, continuous exposure, nonparametric regression, time series.

Contents

1 Introduction

Inflation is the increase of the general price levels, resulting in a decrease in the purchasing power of goods and services given the same amount of money. It can arise from various factors such as general increasing economic demand, increased firm expenses due to higher wages, production costs, or supply shocks. It can also be because of the expectation about future inflation causing firms and people to adjust prices and wages accordingly (Sveriges Riksbank 2022).

The most important factor alongside fiscal policy is monetary policy, which policymakers, typically the central banks, can use to control the price levels (Tran 2018). The modern monetary policy framework in most developed countries stems from the idea of inflation targeting, where central banks try to maintain the inflation rate at a constant value, typically at or around 2 percent. This was first adopted in the 1990s by some industrial countries, including Sweden, in the wake of the difficulties these countries faced in implementing monetary policy using an exchange rate peg or some monetary aggregate as an intermediate target (International Monetary Fund 2006). The Federal Reserve officially adopted it in 2012, although before that monetary policy in the USA was still being conducted without an explicit target (Board 2023). The reason for having an inflation target is that it helps create favorable conditions for economic development by maintaining a stable inflation rate (Sveriges Riksbank 2024).

The primary tool on which central banks rely to steer inflation is the policy rate, which is the daily interest rate that banks use when lending and borrowing to each other. This affects other interest rates in the economy, such as the banks and mortgage institutions. Thus, by adjusting policy rates, central banks can influence the pace and direction of the overall economic activity affecting aggregate output, employment, and inflation (Friedman 2000). Although not of interest for this study, there are other complementary tools that central banks can use as well to affect interest rates, an example being quantitative easing, which developed after the 2008 financial crisis and is the large-scale purchase of financial assets such as government bonds (Bernanke 2020).

The mechanism through which monetary policy affects the economy is called the monetary transmission mechanism. It is one of the most studied areas of monetary economics, and there is a large amount of literature on this subject (Boivin et al. 2010). This, of course, has a good reason behind it. Policymakers must have an adequate assessment regarding how the monetary policy affects the economy and the timing of the effects to determine the monetary policy at a specific point in time (Boivin et al. 2010).

However, for estimating the actual causal effects of the policy rate on inflation, there is very little available literature. Much of the previous research based on the econometrics framework about the inflation rate and policy rates has typically revolved around making predictions. Methods such as estimating vector autoregression models and Granger causality for time series data are commonly used, for example, revolving around analyzing monetary shocks. Although these types of analyses have their usage, these are in a broad sense more oriented toward causal discovery, which is the task of identifying and understanding causal relationships in the data. A problem with such methods is that they lack causal effect estimation, which is the estimation of the actual effect of a policy or treatment on a target variable. Instead, this is commonly done by other methods using metrics such as the average treatment effect (ATE) for binary variables to estimate the effect of a treatment variable on an exposure variable, which is the outcome, commonly associated with the causal inference framework as laid out by the Rubin Causal model (Moraffah et al. 2021). In contrast to traditional methods using the policy rate to forecast inflation, there is very little literature that bases itself on the causal inference framework, possibly because researchers have typically developed the causal inference framework around the context of a treatment variable being binary or polytomous discrete. Thus, within the causal inference framework there also exists much less literature on continuous exposure variables even within the causal inference framework. Nevertheless, the research with continuous exposure variables has seen recent developments opening up for more opportunities to test for causality. This paper then, will try to fill in this gap in the literature that exists for the causal effect estimation of the policy rate on inflation.

For continuous exposure variables, within the causal inference framework, the estimand of interest is the causal dose-response curve, which can be interpreted as the effect of a treatment variable on an exposure variable for any given level of treatment. More specifically, this study aims to estimate the causal dose-response curve where the policy rate is seen as the treatment variable and the inflation rate is seen as the exposure variable. Using a new nonparametric test developed by Westling (2022) causal effect, this study will answer the question if there is an actual causal effect from the policy rate on inflation for different lags of the interest rates. Furthermore, using nonparametric regression, this study will also try to plot the relationship between the inflation rate and policy rate to give an idea of how the causal dose-response curve may look like. Although, the nonparametric regression method itself is not based on the causal inference framework. To the best of my knowledge, this has not been done before and may provide results for others to continue the research on.

Is this an ok motivation for the use of the nonparametric regression? Also, would it be correct to say teh the nonparametric regression tries to estimate the density, although not in a causal setting?

The rest of this paper is structured as follows: Section 2 contains the previous research and the theoretical framework based on macroeconomics theory. The literature review in 2.1 presents presents review of the literature within the general area of interest rates and its effect on inflation. It covers the most commonly used models and their problems with regard to causality. However, it should be noted this this is a more general review of previous literature as there is very little literature regarding this subject based on the methods used in this paper. Furthermore, the section also presents a summary of the transmission mechanism as well as the effect of the unemployment rate as a potential confounder, and a small discussion regarding some of the causal effects of the policy rate.

The data used for this study are described in Section 3, which discusses how and where it was obtained, as well as some minor problems with the data.

The methods section consists of different subsections for the different statistical tests and methods used by this study such as the nonparametric test for estimating the effects of the causal dose-response curve as well as the nonparametric regression used and some other tests. The results are given in section 5 followed by a discussion and the conclusions in section 6

2 Previous research and theoretical framework

2.1 litterature review

There is an abundance of literature within the general area on the effects of interest rates on inflation, even before central banks officially adopted the idea of maintaining an inflation target by the policy rate. Though interest rates and policy rates are not strictly the same thing as policy rate refers to the one officially set by the central banks, they are of course very related, and the work of using policy rates stems from the early work of interest rates in a more general sense, and they also move in the same direction. Although the focus has not typically been so much about trying to estimate any direct causal effect, but rather using interest rates to predict the inflation rate. Much of the early work on this subject was based on regression analyses, often in some autoregressive setting. Some early notable works are for example Fama (1975) or Nelson & Schwert (1977), using short-term interest rates as predictors for inflation, and in which there are found to be significant results. Fama derived the relationship between interest rates and inflation as:

$$E\left(Y_t \mid I_{t-1}\right) = E\left(R_t - r_t \mid I_{t-1}\right) = R_t - E(r_t \mid I_{t-1}) \tag{1}$$

where R_t is the nominal return from one month treasury bills and Y_t is the inflation rate.

But as Xu et al. (2022) points it, there are limitations with this type of linear regression-based framework: most importantly, there is reason to suspect that the relationship between inflation and interest rates may not be constant over a longer period. Furthermore, the conditional mean represents the average return, but it doesn't provide much insight into the distribution of the return. Two distributions can have the same mean, yet exhibit very different shapes, and overlooking these differences can lead to incorrect conclusions. In their work, Xu et al. (2022) used locally stationary quantile regression instead to model inflation and interest rates and came to two conclusions: the first was that the inflation rate is positively correlated with a one-month lagged inflation rate for the quantiles considered. The second conclusion was that the relation between inflation and interest rates exhibits more complicated quantile-specific and time-varying features.

, researchers have traditionally dominated the methodological framework of the causality of interest rates on inflation using VAR models that began with the works by Sims (1980).

However, for a long time, the use of VAR models has traditionally dominated the methodological framework of the causality of interest rates on inflation, starting with the works by Sims (1980). In contrast to autoregression, which is a single-equation, single-variable linear model, where a current variable is regressed on on its past values, a VAR is an n-equation, n-variable linear model where each variable is regressed on its past values as well as the current and lagged values of the other n variables (Stock & Watson 2001). VAR models are commonly used with the Granger causality concept attributed to Granger (1969). A variable X is said to cause some other variable Y if past values of X help explain future values of Y, that is, it contains information about Y that is not available in Y's past values (Moraffah et al. 2021). With VAR models this is typically modeled as:

$$Y_t = \sum_{\tau=1}^{\tau_{\max}} \phi(\tau) Y_{t-\tau} + e_t, \tag{2}$$

where $Y_t = (Y_{1t}, \ldots, X_{nt})$ indicates time series Y at time step t, $\phi(\tau)$ is the $N \times N$ coefficient matrix at lag τ, $\tau_{\max}$ denotes the maximum time lag, and e represents an independent noise. If any coefficient in $\phi_{ji}(\tau)$ is nonzero, then $_i$ is said to cause Granger to cause Y_j with the lag τ.

The relationship can be shown by $Y^i_{t-\tau} \to Y^j_t$, which demonstrates the causal link between Y_i and Y_j at lag τ.

One notable study is Stock & Watson (2001), in which they use VAR and do not find evidence that the federal funds rate helps predict inflation, but instead that inflation helps predict the federal funds rate. Boivin et al. (2010) also using VAR models, does find in contrast to Stock & Watson that the Federal Reserve has played a role in reducing the volatility of inflation by its use of the Federal Funds rate. Tran (2018) uses a Vector error corrected model (VECM), which is a restricted VAR model, and finds that there is a short-run causal relationship from policy rate to inflation based on data from Vietnam,

Adolfson et al. (2007) makes the argument that although VAR models may have good statistical properties and can be useful as a way to make good forecasts, they come with little economic theory, and a drawback is that they typically are over-parameterized. Having many parameters causes imprecise estimates leading to large intervals for the pulse response functions. Hence, VAR models cannot give more in-depth answers about the monetary transmission mechanism, and the actual effects of monetary policy are not captured. One proposed solution is to instead incorporate Bayesian methods such as Bayesian Retrogressive (BVAR) models. These models can include some macroeconomic theory by the use of priors, for example, such as the concept of the steady state of the system (Villani 2006) and in which they found that BVARS can make inflation forecasts as good as more complicated models used at central banks. Another example is Berger & Österholm (2011), where they use BVAR to determine whether money growth Granger causes inflation based on US data, for which including interest rate consistently gave better results.

However, the literature within this area is well-developed. From a methodological perspective, they are not as much related to this paper. They are related in trying to establish whether there is an effect, as well as the magnitude, of policy rates on inflation. It should also be noted that, about casualty, the results from these models cannot be attributed to a direct causal interpretation, although they offer an in-depth analysis of varying statistical dependencies within a set of economic variables to better evaluate causality. (Doan et al. 1984). As mentioned briefly in the introduction, this paper aims to take an approach based on the causal inference framework, which is very scarce in comparison.

There are some examples of it being used in a broader context of analyzing monetary policy, for example, Angrist et al. (2018) which analyzes the effect of the policy rate on different macroeconomic variables using inverse probability weighting estimators. However, their analysis is restricted to a discretization of the policy rate. Vonnák (2021) extends the work of Angrist et al. by generalizing to the continuous case, but seems to rely on parametric assumptions about the specified propensity score. Also, even though they are estimating the causal dose-response curve, it does not test whether the causal dose-response curve is flat. Hence, to the best of my knowledge, this study is the first one to try and determine whether it is flat or not.

2.2 The transmission mechanism

The traditional channels of the transmission mechanism stem from the neoclassical framework of the models of investment, consumption, and international trade, which operate through different channels. For this paper, the focus is not on the actual transmission mechanism itself. Thus, no emphasis will be placed on the individual channels. Instead, the focus is on a more aggregate scale to answer the question if there can be determined to be any causal effect or not. For this reason, it is nevertheless important to have an understanding of the transmission mechanism. See for example Gertler & Karadi (2015) for more details and the history of how the transmission mechanism has developed over time.

According to the traditional Keynesian models, one of the channels of the transmission mechanism is the interest rate channel; the more traditional way by adjusting the nominal money affecting the interest rates interest rate or more directly by through the policy rate. In any event though, in practice they fulfill the same effects in that they impact the aggregate output, employment, and overall price levels (Mishkin 2016). However, it should be noted that it is the real interest rates and not the nominal interest rates that affect asset pricing and spending through the transmission channel. Furthermore, it is not only the current interest rates that matters, but also the expected interest rates. Nevertheless, setting the short-term nominal interest rates gives the central banks control over the contemporaneous real rates as well as the expected rates in the future for some horizon ahead (Gertler & Karadi 2015).

Businesses and households tend to look at the long horizon when factoring variation in interest rates into investment decisions. Although a contractionary monetary policy by increasing the policy rate increases the short-term nominal interest rates, it also increases the longer-term nominal interest rate, for example, because investors seek to eliminate disparities of risk-adjusted expected return. This will translate into changes in the real interest rates as well. Hence, the user capital cost rises and the capital asset demand decreases as a result of increasing borrowing costs for both firms and individuals leading to lower spending on investment and a decline in the aggregate spending and demand, which has a dampening effect on the inflation (Ireland 2010). Furthermore, monetary policy also has a direct effect on household consumption. Lowering short-term interest rates, for instance, increases demand for assets like common stocks and housing thus raising their cost. This results in an overall increase in total wealth stimulating household consumption and the aggregate demand (Gertler & Karadi 2015).

Another way that the transmission mechanism works is through the exchange rate channel for open economies; as a result of falling interest rates, the domestic currency depreciates vis-à-vis other currencies. This is because the return on domestic assets decreases compared to that of foreign assets, causing domestic goods to be cheaper than foreign goods and an increase in foreign demand for domestic goods, leading to expenditure switching - a switch between foreign and domestic goods - and hence a rise in net exports. Thus, the increase in economic activity as a result of the higher aggregate demand raises the inflation. Increased interest rates have the opposite effect. Hence, the exchange rate channel plays an important role in how monetary policy affects the economy (Taylor 2001).

There is also the asset price channel as described by Tobins q-theory (Tobin 1969), where tobins q is a ratio of the market value of a company's assets divided by the replacement cost of those assets. Low interest rates typically increase the q ratio by raising market values and lowering replacement costs, while high interest rates typically decrease the q ratio by lowering market values and raising replacement costs. For example, when q is lower, each firm has to issue more new shares of stock to finance any new investment project, making investment more expensive for the firm. Consequently, across all firms, investment projects that were only marginally profitable before monetary tightening now go unfunded due to the decline in q. This results in a decrease in output and employment (Ireland 2010).

Besides those channels, it should also be mentioned that there are several different mechanisms through which the monetary policy affects inflation and not only the ones being mentioned here. These act in parallel as well with different time frames. For example, the credit channel is another one. It will not be expanded upon here, but see for example Bernanke & Gertler (1995) or Kashyap & Stein (1994) for further details. However, the main point moving forward is that the way monetary policy works is an incredibly complex phenomenon and it is difficult to single out each effect. Nevertheless, the channels mentioned here are simply the main ones commonly mentioned within the literature. Furthermore, it is worth noting there are other factors affecting the interest rates in a country and not only its monetary policy. For example, the general level of global interest rates. Hence the policy rate is not solely responsible for being the factor that affects the inflation rate, but it does nevertheless play an important role.

2.3 Unemploymeyment rate and inflation

The relationship between inflation and unemployment stems from the concept of the Philips Curve originally developed by Phillips (1958), which, in short, was about the fact that there seemed to be an inverse relationship such that when the inflation rate was high, the unemployment rate was low and vice versa: this cause was attributed to that high demand on labor lead to increased wages, whereas during times of low demand it workers would be willing to accept lower wages. Philips did, however, note that the relationship between the unemployment rate and nominal wage change should be highly non-linear due to the rigidity stemming from the reluctance of workers to low rates even during times of high unemployment. Since then, the concept of the Philips curve has continuously been developing to incorporate new ideas. For example Friedman (1968) together with Phelps (1967, 1968) extended it to include the idea of the natural rate of unemployment, which revolves around the idea that there is one specific level of unemployment where the inflation rate will continue at an existing pace. If the unemployment rate falls below that threshold or gets above it, then the inflation rate will either increase or decrease. Their work also contributed to the idea of the augmented Philips curve, which took into account the role of the expectations regarding inflation. The idea was that the unemployment rate does not affect the inflation in levels, but rather the rate at which it changes. However, the amount of literature on this topic is vast and this paper will not go into more detail about this aspect. See for example Gordon (2011) and Mavroeidis et al. (2014) for an overview and more recent developments. However, because of its importance to the Philips Curve, the unemployment rate is commonly used in similar studies, and hence why it is also included in this study as it is an important confounder variable.

2.4 Causality debate

Though the effects of monetary policy on inflation are widely accepted, there is some debate about its actual effectiveness; especially for developing countries (Islam & Ahmed 2023). A key aspect is the different financial structure; for example the absence of effective and functional markets for things such as fixed-income securities, equities, and real estate. Furthermore, even though banks are the largest financial institutions, the financial system typically is small about the total economy (Mishra & Montiel 2013). Developing countries typically have limited and imperfect connections with private international capital markets, leading their central banks to heavily intervene in foreign exchange markets(Mishra et al. 2012). This institutional context indicates that the monetary transmission mechanism in low-income countries may significantly differ from that in advanced and emerging economies. However, as this study is limited to only US and Swedish data, this aspect is not further analyzed since these are considered advanced economies as per the IMF definition (Nielsen 2011), (IMF 2023). Although one should keep in mind moving forward that potential results may not be generalized for other countries, it may be even more so for developing countries.

3 Data

The data used in this study are monthly time series for the variables inflation rate, unemployment rate, and policy rate for Sweden and the United States. The data consists of 355 months covering the periods starting from June 1994, when the Swedish central bank Riksbanken officially started to track the policy rate, to December 2023. Thus, to make the results more comparable, the other time series have been limited to the same period although they have data available from an earlier time point. All the data is collected from various public governmental sources. The Swedish policy interest rate comes from Riksbanken, while the Swedish inflation rate and the unemployment rate are obtained from Statistics Sweden (Statistika Centralbyrån, SCB). The US policy rate (also known as the Federal Funds rate) was obtained from the US Federal Reserve Bank. The US data on unemployment and inflation is retrieved from the U.S. Bureau of Labor Statistics. Although in some literature, a larger collection of variables is used, such as in example Leeper et al. (1996) using 13 variables in VAR-based models, much of the literature considers these three variables in particular with a notable being Stock & Watson (2001). In this case, it is also preferable to try to limit the number of variables to keep the models as parsimonious as possible because of the somewhat limited available data.

A price index is used to operationalize the concept of inflation for both the US and Swedish data. Typically, the consumer price index (CPI) is used as a measure of inflation, which is a measurement of the change in prices of goods and services acquired for private domestic consumption, based on weighted averages for specific sets or baskets of products. However, this study will use the 12-month percentage difference of the consumer price index with fixed rates (CPIF) for Swedish inflation and the 12-month percentage difference of the personal consumer price index (PCE) for the US data since these are the measures used by the respective central banks as the inflation target. Both measures are closely related to the CPI; CPIF is measured in the same way as the CPI while discarding the effects of mortgages for consumers, whereas the PCE accounts for how much consumers are spending their money at a given time and more quickly adapts to changes in spending patterns. (Sveriges Riksbank n.d.), (Federal Reserve n.d.).

Regarding the unemployment rate, it should be noted that although they are roughly similarly defined in the Swedish and US cases, some differences exist. Generally, the unemployment rate for both countries refers to people who are part of the labor force, which means that they do not currently have a job but are actively seeking one. Hence, people without jobs but who are not actively seeking jobs are not considered unemployed. For the US data, these data are originally collected from surveys done by the Bureau of Labor Statistics, whereas the Swedish data are obtained from Arbetskraftundersökningarna (AKU), a survey done by Statistics Sweden. Thus, there are some differences in the exact method by which the surveys were conducted, such as the target population for the survey and the definitions used in the surveys. There have also been some changes over the years. For example, in 2021 AKU made changes due to a new framework law implemented by the EU which included changes in the target population of the surveys as well as some definitions used. Thus, comparisons between the periods before and after this date are not strictly straightforward (SCB 2023). However, SCB has worked on making the time series comparable through different means, such as imputations. However, in general, this is considered a minor problem for this study, and potential differences are not further analyzed.

4 Method

4.1 Nonparametric test of the causal dose-response function

This study will use a relatively new test as developed by Westling (2022). It is a doubly-robust nonparametric test relying on a cross-fitted estimator of an integral of the dose-response curve. To my knowledge, it is the first such nonparametric test that considers global inference for continuous exposure. Thus, the notations and explanations are largely based on the ones from the article. The focus of this section will be to explain the test while keeping the most relevant parts based on the results from Westling (2022). More details are provided in the original paper, and the proofs are available in the supplementary material. The test is also available in R with the package ctsCausal.

maybe I will remove this part in the end. But for now, I need to make sure that the notations are correctly used in the same way as in the article, then maybe change them

First, the exposure variable, or the treatment variable, which in this case represents the policy rate, is denoted as A with support $\mathcal{A} \in \mathbb{R}$. Using the potential outcomes from the causal inference framework, then for each possible value of the exposure, $a \in \mathcal{A}_0$, the potential outcomes are: $Y(a) \in \mathbb{R}$ under intervention by setting the exposure to $A = a$. Meaning, this is the outcome that would be observed if it was possible to assign a unit to that value of exposure.

The causal parameter of interest is the counterfactual means defined as $m(a) := E[y(a)]$, which is the average outcome under the assignment of the entire population to exposure level $A = a$. The function $m : \mathcal{A} \to \mathbb{R}$ is the causal dose-response curve, also called the average dose-response function.

The fundamental problem of causal inference is that it is impossible to observe all the potential outcomes; instead, it is only possible to observe one of the outcomes for each unit. The observed data is assumed to be generated from some unknown distribution P_0 such that: $O_1, \ldots, O_n \sim P_0$ for the data unit $O = (Y, A, W)$ where $Y := Y(A)$ is the observed outcome and $W \in \mathbb{R}^d$ are some pre-exposure covariates. Throughout the rest of the paper, functions or parameters indexed by 0 will denote that it depends on the unknown distribution P_0

The null hypothesis and alternative hypotheses that we would want to aim to test are:

$$H_0 : E\left[Y(a)\right] = c, \text{ for all } a \in \mathcal{A} \tag{3}$$

$$H_1 : E\left[Y(a_1)\right] \neq E\left[Y(a_2)\right] \text{ for some } a_1, a_2 \in \mathcal{A}. \tag{4}$$

The null hypothesis corresponds to the causal dose-response curve m being flat on $\mathcal{A}$. An alternative way to view it is that the average outcome under assigning all units to $A = a$ is the same for all $a \in \mathcal{A}$

However, because we cannot observe $Y(a)$ for all units, but instead the outcome $Y := Y(A)$, then m is not a mapping of the joint distribution of the pair (A, Y). Thus this null hypothesis cannot be tested by the observed data. To make this testable, some assumptions have to be made:

1. The potential outcomes of each unit are independent of the exposure of all other units (the stable unit treatment assumption).

2. The observed outcome Y almost surely equals $Y(A)$.

Furthermore, if the following assumption were to hold:

3. $Y(a) \perp\!\!\!\perp A$ for all $a \in \mathcal{A}_0$,

Then m would be identifiable using univariate regression, and the null hypothesis would be testable using parametric approaches. However, in the presence of confounder variables affecting both A and $Y(a)$, it would cause nonparametric regression to have invalid type 1 error rates. Instead, the following assumptions are made;

4. $Y(a) \perp\!\!\!\perp A \quad W$ for all $a \in \mathcal{A}_0$

5. all $a \in \mathcal{A}_0$ are in the support of the conditional distribution of A given $W = w$ for almost every w. Or alternatively : $p(a \mid w) > 0$

Under assumption 4, then all confounding variables are accounted for in W. Under assumption 5, then all units can receive any treatment a on its support. Given that all the assumptions 1-2 and 4-5 hold, then:

$$E[Y(a) \mid W = w] = E_0(Y \mid A = a, W = w) := \mu_0(a, w), \text{ so}$$
$$E[Y(a)] = E_0\left[\mu_0(a, W)\right] := \theta_0(a),$$

which is known as the *backdoor formula*, or G-*computation* (Robins 1986). This is saying that if the conditions hold, then the G-formula, as defined by the expectation of the outcome regression, can be used to estimate the counterfactual mean parameter. $\mu_0(a, W)$ is the conditional mean of the observed Y of a by taking the expectation over the marginal distribution of the covariates W.

Under these conditions it is possible to redefine the null and alternative hypotheses in terms of the observed data distribution:

$$H_0 : \theta_0(a) = \gamma_0 \text{ for all } a \in \mathcal{A} \text{ and } \gamma_0 := E_0\left[\theta_0(A)\right]$$
$$H_1 : \theta_0(a) \neq \gamma_0, \text{ for some } a \in \mathcal{A}$$

In practical terms, it means that the null hypothesis holds if the g-computed regression function $\theta_0(a)$ is constant on the support of a, which only holds if it's equal to its average value. γ_0 is the average value of the causal dose-response function over the marginal distribution of the exposure A.

Next is to derive the actual test. Since it is difficult to estimate θ_0 itself, the proposed method by Westling (2022) is to estimate a primitive parameter of θ_0. Consider for example a density function g_0 which is flat on $[0, 1]$. Then the CDF G_0 of g_0 is the identity function on $[0, 1]$. Hence the density g_0 being flat corresponds to $G_0(X)$ being the identity function on $[0, 1]$, which in turn corresponds to $G_0(x) - x = 0$ for all $x \in [0, 1]$.

Then the null hypothesis can be expressed in terms of a primitive transformation of θ_0

With the following definitions:

$$\begin{aligned} F_0(a) &= P_0(A \leq a) \\ \gamma_0 &= \int_{-\infty}^{\infty} \theta_0(a) dF_0(a), \\ \Gamma_0(a) &= \int_{-\infty}^{a} \theta_0(u) dF_0(u), \\ \Omega_0(a) &= \Gamma_0(a) - \gamma_0 F_0(a). \end{aligned}$$

$F_0(a)$ is the the marginal cumulative distribution of the observed exposure a. γ_0 which was defined earlier is the average of the dose-response function over the marginal distribution of A. Then $\gamma_0(a)$ is the partial average of $gamma_0$ up to a.

This means that the causal

This leads to the following proposition:

> **Proposition 1** *if θ_0 is continuous on $\mathcal{A}$, then θ_0 is constant on $\mathcal{A}$ if and only if $\Omega_0(a) = 0$ for all $a \in \mathcal{A}$*

Unlike θ_0, $\Omega_0(a)$ is pathwise differentiable relative to the nonparametric model at each a with an estimable influence function. This can be done regardless of the form of the form of F_0. Thus it is possible to construct a uniformly asymptotically linear estimator of Ω_0. In the supplementary materials to Westling (2022), Westling proves that the efficient influence function is:

$$\begin{aligned}
D^*_{a_0,0}(y,a,w) :=& [I_{(-\infty,a_0]}(a) - F_0(a_0)] \\
& \times \left[\frac{y-\mu_0(a,w)}{g_0(a,w)} + \theta_0(a) - \gamma_0\right] \\
& + \int [I_{(-\infty,a_0]}(u) - F_0(a_0)]\,\mu_0(u,w)F_0(du) \\
& - 2\Omega_0(a_0).
\end{aligned} \tag{5}$$

If V denotes the number of folds, then a cross-fitted estimator $\Omega_n(a_0)$ of $\Omega_0(a_0)$ is given by:

$$\begin{aligned}
\frac{1}{V}\sum_{v=1}^{V} \Bigg\{ & \frac{1}{N_v}\sum_{i\in\mathcal{V}_{n,v}} [I_{(-\infty,a_0]}(A_i) - F_{n,v}(a_0)] \\
& \times \frac{Y_i - \mu_{n,v}(A_i,W_i)}{g_{n,v}(A_i,W_i)} \\
& + \frac{1}{N_v^2}\sum_{i,j\in\mathcal{V}_{n,v}} [I_{(-\infty,a_0]}(A_i) - F_{n,v}(a_0)]\,\mu_{n,v}\left(A_i,W_j\right)\Bigg\}.
\end{aligned} \tag{6}$$

This also requires the estimators for each fold $u_{n,v}$ of the outcome regression u_0 as well as $g_{n,v}$, which is the hybrid standardized propensity score defined as:

$$g_0(a,w) := \begin{cases} \frac{P_0(A=a|W=w)}{P_0(A=a)}, & \text{for } a \text{ such that } P_0(A=a)>0 \\ \frac{p_0(a|w)}{f_0(a)}, & \text{for } a \text{ where } F_0 \text{ is absolutely continuous.} \end{cases}$$

Both u_0 and g_0 are estimated using the Superlearner package consisting of flexible libraries of Generalized linear models, and generalized additive models, but it also has support for many other methods such as XGBOOST, Random Forest, LASSO, and neural networks to name a few. The methods used for this study are described more in detail in section 4.3. Using cross-validation to estimate the performance of the chosen models, it then creates an ensemble using a weighted average of the performance of the models combined.

Furthermore, (Westling 2022, p. 6) shows that $n^{1/2}\Omega_n - \Omega_0(a)$ under some standard robust doubly robust conditions will converge weakly to a Gaussian process Z_0. Thus by using the cross-fitted estimator Ω_n of Ω_0, it is possible to construct an approximation $T_{n,p,a}$ of the $1-\alpha$ quantile of the limiting distribution of $n^{1/2}\left\|\Omega_n - \Omega_0\right\|_{F_0,p}$.

The test procedure then consists of the following five steps:

Step 1: Split the sample into V sets $\mathcal{V}_{n,1}, \ldots, \mathcal{V}_{n,V}$ of approximately equal size.

Step 2: For each $v \in \{1, \ldots, V\}$, construct estimates $\mu_{n,v}$ and $g_{n,v}$ of the nuisance functions μ_0 and g_0 based on the training set $\mathcal{T}_{n,v}$ for fold v.

Step 3: For each a in the observed values of the exposure $\mathcal{A}_n := \{A_1, \ldots, A_n\}$, use $\mu_{n,v}$ and $g_{n,v}$ to construct $\Omega_n^\circ(a)$ as defined in (2).

Step 4: Let $T_{n,\alpha,p}$ be the $1-\alpha$ quantile of $\left(\frac{1}{n}\sum_{i=1}^n |Z_n(A_i)|^p\right)^{1/p}$ for $p < \infty$ or $\max_{a \in \mathcal{A}_n} |Z_n(A_i)|$ for $p = \infty$, where, conditional on $O_1, \ldots, O_n$, $(Z_n(A_1), \ldots, Z_n(A_n))$ is distributed according to a mean-zero multivariate normal distribution with covariances given by $\Sigma_n\left(A_i, A_j\right) := E\left[Z_n(A_i) Z_n\left(A_j\right) \mid O_1, \ldots, O_n\right] = \frac{1}{V}\sum_{v=1}^V \mathbb{P}_{n,v} D^*_{A_i,n,v} D^*_{A_j,n,v}$ for $D^*_{a_0,nv}(y,a,w)$ equal to

$$
\begin{aligned}
&[I_{(-\infty,a_0]}(a) - F_{n,v}(a_0)]\left[\frac{y-\mu_{n,v}(a,w)}{g_{n,v}(a,w)} + \theta_{n,v}(a) - \gamma_{n,v}\right] \\
&\quad + \int [I_{(-\infty,a_0]}(u) - F_{n,v}(a_0)]\, \mu_{n,v}(u,w) F_{n,v}(du) \\
&\quad - 2\Omega_{\mu_{n,v},F_{n,v},Q_{n,v}}(a_0)
\end{aligned}
$$

where $\theta_{n,v}(a) := \int \mu_{n,v}(a,w) dQ_{n,v}(w)$ and $\gamma_{n,v} := \iint \mu_{n,v}(a,w) dF_{n,v}(a) dQ_{n,v}(w)$

Step 5: Reject H_0 at level α if $n^{1/2}\left\|\Omega_n^\circ\right\|_{F_n,p} > T_{n,\alpha,p}$.

There are some notational stuff here I don't quite understand still

4.2 Alternative considerations for testing the causal effects

In traditional classical tests for association, causality can often be asserted due to randomization. In the case of observational data, it's more difficult due to potential confounding between the exposure and outcome. Much of the literature in the causal inference framework handles cases with a binary exposure, for which different methods have been developed such as matching estimators (Rubin 1973), inverse probability weighted (IPW) estimators (Horvitz & Thompson 1952) or augmented inverse propensity weighted estimators (Robins et al. 1994). Then, it is for example possible to estimate each of the counterfactual means under 0 and 1 and test the null hypothesis if they are equal or not. These methods can also be extended to handle polytomous discrete cases.

However, when the causal dose-response function is the estimand of interest as in this case with continuous variables. It is typically much more difficult to estimate and the literature within this area is not as developed as in the discrete case. One alternative way, which is also commonly used, is to make the exposure variable discrete and then use other methods, such as the methods above mentioned for example. Although It is possible, it is not ideal for various reasons. First, there are multiple ways to do the discretization and it is not always obvious which is the best choice, hence the result will vary depending on how the discretization is done. Converting a continuous exposure variable to discrete also causes information loss, leading to the test typically having less power than the test based on the original continuous exposure (Westling 2022). Furthermore, it makes the causal parameter have a more complicated interpretation (Young et al. 2019).

As Westling argues, one possibility of estimating the dose-response curve would be to assume that the regression for the outcome on the exposure and potential confounders follows a linear model. Given that the model is correctly specified, this would mean that the coefficient for the exposure corresponds to the slope of the dose-response curve. However, if the model is not correctly specified, the results would be inconsistent. Although IPS may be used, it would also cause inconsistent results if the model for the propensity score was wrong. Hence, for this reason, the nonparametric test as developed by Westling seems preferable.

4.3 Generalized linear models

Generalized linear models (GLMS) extend traditional linear regression to be able to model outcomes with non-modal response distributions as well as nonlinear functions of the mean. For this case, they are used specifically to estimate the nuisance estimators g_0 and μ_0. A GLM consists of three components: a random component a linear predictor and a link function.

The random component is the response variable $y_1, \cdots y_n$ with a PDF or PMF from the exponential family in which case a general expression can be written as of:

$$f(y_i) = \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{a_i(\phi)} - c(y_i, \phi)\right\}, \tag{7}$$

maybe I will just remove this expression of the exponential family. Not sure if it's really necessary to write it out

Then a linear predictor relates parameters η_i concerning $E(y_i)$ to the independent variables $x_1, \cdots, x_p$ by a linear combination:

$$\eta_i = g(u_i)\sum_{j=1}^{p}\beta_j x_{ij}, \quad i = 1, \ldots, n. \tag{8}$$

The link $g(\cdot)$ function connects the random components with the linear predictor to model the expected outcome. That is, if $u_i = E(y_i)$, then η_i is linked to u_i by

The natural parameter for a normal distribution is the mean and the link function $g(u_i) = u_i$ is the identity link function. This has the form:

$$u_i = \sum_{J=1}^{P} \beta_j x_{ij}, \quad i = 1, \ldots, n$$

which is the ordinary linear model and corresponds to the ordinary least squares regression.

One advantage of expressing it in terms of a generalized linear model is that the logit model as well as the GAM model can easily be expressed in terms of a GLM.

For example, consider a Bernoulli trial for some observation i and with the binomial parameter π_i, so that $p(y_i = 1) = \pi_i$ and $p(y_i = 0) = 1 - \pi_i$. The natural parameter is the log odds $\log[u_i/1 - u_i)]$. Then logistic regression can be expressed in terms of a GLM by using the canonical logit link function:

$$\text{logit}(\pi_i) = \log(\frac{\pi_i}{1 - \pi_i}) = \sum_{j=1}^{p} \beta_j x_{ij} \tag{9}$$

The generalized additive model can be considered as a general extension of the generalized linear model by replacing the linear predictor with additive smooth functions of the covariates. For example, instead of the previous GLM structure with the logit link as described by equation 9, it is instead generalized to $g(\mu_i) = \sum_{j=1}^{p} s_j(x_{ij})$ where $s_j(\cdot)$ is an unspecified smooth function of the predictor j. Using the same logit link, the GAM model would then be expressed as:

$$\text{logit}(\pi_i) = \log(\frac{\pi_i}{1 - \pi_i}) = \sum_{j=1}^{p} s_j(x_{ij}) \tag{10}$$

GAM models provide more flexibility than the regular GLM models, as it does not require the relationship between the logit transformation of Y and $\boldsymbol{X}$ to be linear. Hence it can discover patterns that may otherwise be missed using GLM leading to better predictions. One drawback though is its loss of interpretability of the effects of covariates with a smooth term. It is also more difficult to make inferences compared to GLMS as it is unclear how to apply confidence intervals. GAMs also typically require a larger sample as smoothing methods can have a large number of parameters. Nevertheless. In this case, the GAM model, as well as the GLM models, are not used directly to make inferences, but rather estimate the nuisance estimators for the Westling test.

4.4 Nonparametric regression

Nonparametric smoothing regression is a type of regression analysis used to estimate the relationship between a dependent variable and one or more independent variables without assuming a predefined form (like a linear or polynomial form) for the relationship. This flexibility allows the model to adapt more closely to the actual data, making it particularly useful in situations where the relationship between variables is complex or unknown. Hence it can model relationships among variables while also being robust against functional form of misspecification and may detect structures that may otherwise be missed (Racine & Li 2004). As has been mentioned previously, the relationship between interest rates and inflation is assumed to be nonlinear, which is the reason for using nonparametric regression in this case. By using nonparametric regression, the goal is to hopefully be able to capture the relationship between the policy rates and inflation to give an idea of how the causal dose-response may look. However, it must be stated that this is not based on the causal inference framework and any result cannot be attributed to a direct causal interpretation.

Is this enough to mention regarding this??

The smoothing part of nonparametric smoothing regression refers to the predictions being weighted averages of the observed responses in the training data., In so-called kernel regression, the degree of smoothing is controlled by some bandwidth denoted as h.

Smoothing very little ($h \to 0$), means that the model can have very small details of any potential "true" regression function. However, it also causes each of the predictions to be an average of fewer observations, making the predictions more noisy.

Something to keep in mind is the relationship between bias and variance through the total error: total error = noise + bias^2 + variance. Changing the amount of smoothing affects both bias and variance and there is an optimal amount of smoothing which can be found. This is usually done by cross-validation, and the bandwidth with the lowest error under cross-validation is then used to fit the regression curve for all the data.

For this paper, the nonparametric regression used adapts a cross-validated local linear estimator with the method developed by Racine & Li (2004) and Li & Racine (2004). It is a nonparametric regression that can use both continuous and categorical data using kernels and where the bandwidth is chosen by a data-driven approach. In R it is computed by the `npreg` package. This works for both continuous and categorical data in a natural manner. However, since the data used for this study are only continuous, then the method applied is the one specifically from Li & Racine (2004) with the case of continuous regressors. It is estimated as follows:

First, the nonparametric regression model can be expressed as:

$$y_j = g\left(x_j\right) + u_j, \quad j = 1, \ldots, n, \tag{11}$$

and where $x_j \in \mathbb{R}^q$. Then the derivative of $g(x)$ is defined as $\beta(\boldsymbol{x}) \overset{\text{def}}{=} \nabla g(x) \equiv \partial g(x)/\partial x$ and the gradient $\nabla g(\cdot)$ is a $q \times 1$ vector. Then $\delta(x) = (g(x), \beta(x)')'$ so that $\delta(x)$ is a $(q+1) \times 1$ vector-valued function. The first component of $\delta(x)$ is $g(x)$ and the other q components are the first derivatives of $g(x)$.

is this the correct way to write out that x_j is a continuous random vector of dimension q?

By Taylor expanding $g(x_j)$ at x_i, it results in:

$$g\left(x_j\right) = g\left(x_i\right) + \left(x_j - x_i\right)' \beta\left(x_i\right) + R_{ij} \tag{12}$$

where $R_{ij} = g\left(x_j\right) - g\left(x_i\right) - \left(x_j - x_i\right)' \beta\left(x_i\right)$.

Then equation 11 can be rewritten as:

$$\begin{aligned} y_j &= g\left(x_i\right) + \left(x_j - x_i\right)' \nabla g\left(x_i\right) + R_{ij} + u_j \\ &= \left(1, \left(x_j - x_i\right)'\right) \delta\left(x_i\right) + R_{ij} + u_j. \end{aligned} \tag{13}$$

By using kernel weighted regression of y_j on $\left(1, \left(x_j - x_i\right)'\right)$, a leave-one-out local linear kernel estimator of $\delta\left(x_i\right)$ can be expressed as:

$$\begin{aligned} \hat{\delta}_{-i}\left(x_i\right) &= \begin{pmatrix} \hat{g}_{-i}\left(x_i\right) \\ \hat{\beta}_{-i}\left(x_i\right) \end{pmatrix} \\ &= \left[\sum_{j \neq i} W_{h,ij} \begin{pmatrix} 1, & \left(x_j - x_i\right)' \\ x_j - x_i, \left(x_j - x_i\right)\left(x_j - x_i\right)' & \end{pmatrix}\right]^{-1} \sum_{j \neq i} W_{h,ij} \begin{pmatrix} 1 \\ x_j - x_i \end{pmatrix} y_j, \end{aligned} \tag{14}$$

where $W_{h,ij} = \prod_{s=1}^{q} h_s^{-1} w\left(\left(x_{js} - x_{is}\right)/h_s\right)$ is the product kernel function and $h_s = h_s(n)$ is the smoothing parameter associated with the s th component of x.

Next e_1 is defined as a a $q+1 \times 1$ vector, with the first element being 1 and all the rest being 0. The leave-one-out kernel estimator of $g\left(x_i\right)$ is given by $\hat{g}_{-i}\left(x_i\right) = e_1' \hat{\delta}_{-i}\left(x_i\right)$, and $h_1, \ldots, h_q$ are chosen to minimize the least-squares cross-validation function given by

$$CV\left(h_1, \ldots, h_q\right) = \sum_{i=1}^{n} \left[y_i - \hat{g}_{-i}\left(x_i\right)\right]^2 . \tag{15}$$

$\hat{h} = \left(\hat{h}_1, \ldots, \hat{h}_q\right)$ denotes the cross-validation choices of $h_1, \ldots, h_q$ which minimize equation 14. After $\hat{h}$ is computed, then $\delta(x)$ is estimated by:

$$\begin{aligned}
\hat{\delta}(x) &= \begin{pmatrix} \hat{g}(x) \\ \hat{\beta}(x) \end{pmatrix} \\
&= \left[\sum_{i=1}^{n} W_{\hat{h},ix} \begin{pmatrix} 1, & (x_i - x)' \\ x_i - x, (x_i - x)(x_i - x)' & \end{pmatrix} \right]^{-1} \sum_{i=1}^{n} W_{\hat{h},ix} \begin{pmatrix} 1 \\ x_i - x \end{pmatrix} y_i
\end{aligned}$$

where $W_{\hat{h},ix} = \prod_{s=1}^{q} \hat{h}_s^{-1} w\left((x_{is} - x_s) / \hat{h}_s\right)$, and $g(x)$ is estimated by $\hat{g}(x) = e_1' \hat{\delta}(x)$.

4.5 Dickey-Fuller test for stationarity

The concept of stationary is an important concept for time series. A process is strictly stationary if its properties are not affected by a change in the time origin. That is, a process for the random variable Y_t is strictly stationary if for any integer values $j_1, j_2, \ldots, j_m$, the joint distribution of $\left(Y_t, Y_{t+j_1}, Y_{t+j_2}, \cdots, Y_{t+j_n}\right)$ depends not on time t, but only on $j_1, j_2, \ldots, j_n$.

However, the concept of strict stationarity is often too strong to verify in most practical cases and it is usually satisfactory for a process to be weakly stationary.

The jth autocovariance is defined as:

$$\begin{aligned}
\gamma_{jt} &= \operatorname{Cov}\left(Y_t, Y_{t-j}\right) = \mathrm{E}\left(Y_t - \mu_t\right)\left(Y_{t-j} - \mu_{t-j}\right) \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y - \mu_t)\left(x - \mu_{t-j}\right) f_{Y_t, Y_{t-j}}(y, x) \mathrm{d}y \, \mathrm{d}x
\end{aligned} \tag{16}$$

If neither the expectation μ_t nor the autocovariances γ_{jt} depend on the time t, then the process for Y_t is said to be covariance-stationary or weakly stationary:

$$\begin{aligned}
\mathrm{E}\left(Y_t\right) &= \mu, && \text{for all } t; \\
\mathrm{E}\left(Y_t - \mu\right)\left(Y_{t-j} - \mu\right) &= \gamma_j < \infty, && \text{for all } t \text{ and any } j
\end{aligned} \tag{17}$$

Time series typically need to be stationary to ensure that the properties remain the same over time and to make valid inferences. There are potential consequences of using nonstationary time series in regression settings, such as inefficient estimators, sub-optimal forecasts, and invalid significance tests (Granger & Newbold 1974). This is tested with an augmented dickey fuller test as follows:

Consider for example an AR(1) process as: $y_t = \alpha + y_{t-1} + \delta_t + e_t$ where e_t is a martingale difference sequence such that $\mathrm{E}(e_t \mid y_{t-1}, y_{t-2}, \ldots y_0) = 0$. Then y_t has a unit root if and only if $\rho = 1$

The(non-augmented) Dickey-Fuller test involves fitting the model by OLS and testing the null hypothesis of $\rho = 1$ against the alternative of $\rho < 1$. However, to control for serial correlation the augmented Dickey-Fuller test fits a model on the form:

$$\Delta y_t = \alpha + \lambda y_{t-1} + \delta t + \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \cdots + \zeta_k \Delta y_{t-k} + \epsilon_t \tag{18}$$

for some specified number of k lags. Testing $\mathrm{H}_0 : \lambda = 0$ against $\mathrm{H}_1 : \lambda < 0$ is then equivalent to test if $\rho = 1$ in the former model.

The test statistic does not have a standard t-distribution and there also exists several different versions of it. See for example Hamilton (1994, Chapter 17) where the different test statistics are derived. In R this is performed with the function `adf.test()` from the aTSA package.

4.6 Ljung-Box test for autocorrelation

To test for autocorrelation in the time series, this study will use the Ljung-Box test (Ljung & Box 1978). The empirical autocovariance function is given by: $\hat{\gamma}_j = 1/T \sum_{t=i+1}^{T} \left(Y_t - \bar{Y}\right)\left(Y_{t-j} - \bar{Y}\right)$, from which it follows that the empirical autocorrelation function is given by: $\hat{\rho}_j = \hat{\gamma}_j / \hat{\gamma}_0$

The null and alternative hypotheses are

$$H_0 : \rho_1 = \rho_2 = \ldots = \rho_k = 0$$
$$H_1 : \rho_j \neq 0 \text{ for some } j \in \{1, 2, \ldots, k\}$$

The test statistic is given by the Ljung-Box statistic as:

$$Q_{LB}(k) = T(T+2) \sum_{j=1}^{k} \frac{\hat{\rho}_j^2}{T-j} \overset{a}{\sim} \chi^2(k-p) \tag{19}$$

5 Results

5.1 Summary statistics

Table 1: Summary statistics for the different variables

	n	mean	sd	median	min	max	skew	kurtosis
us_interest	355	2.45	2.24	1.75	0.05	6.54	0.37	-1.52
swe_interest	355	2.26	2.30	2.00	-0.50	8.91	0.84	0.39
us_unemployment	355	5.62	1.82	5.20	3.40	14.80	1.43	2.27
swe_unemployment	355	7.60	1.16	7.60	4.90	10.50	0.15	-0.66
swe_CPIF	355	1.94	1.63	1.60	-0.40	10.20	2.89	9.80
us_PCE	355	2.76	2.54	2.60	-14.70	25.00	1.06	23.85

Table 1 shows some summary statistics for the different variables. Comparing the Swedish and US data it can be seen that the US inflation rate on average is higher than the Swedish inflation rate at 2.76 compared to 1.94 as defined by their respective consumer price indices (CPIF and PCE). These values seem reasonable with the inflation target rate of 2 percent in mind, and it shows that the central banks have been fairly successful in keeping the target. The US exhibits more variation as indicated by the higher standard deviation at 2.25 compared to 1.63, as well as having a large difference between the min and max values. Furthermore, the US inflation also has a considerably higher kurtosis at 23.85 compared to 9.80 for the Swedish inflation, meaning the distribution of the US values has a heavier tail. The other variables interest and unemployment rate are more similar concerning the different statistics. There are minor differences in interest rates. For the unemployment rate, it can be said that the Swedish unemployment on average is higher, but also shows slightly less variation and skewness.

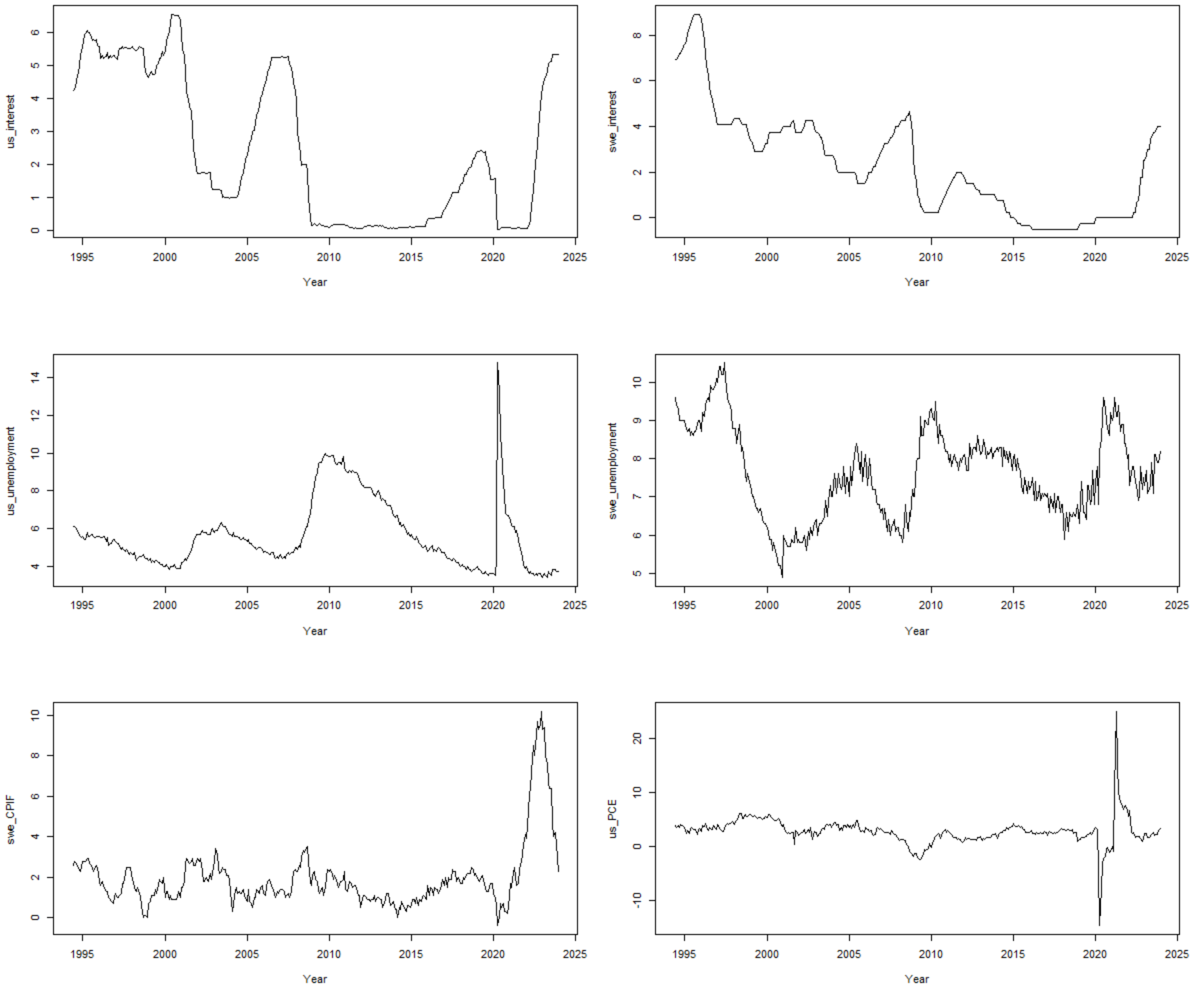

Figure 1: Variables over time

Figure 1 shows the different variables plotted over time. The visualized plots seem consistent with the results in table 1. The interest rates and unemployment rates for both countries seem to roughly follow a similar pattern to each other concerning the overall trend and their different peaks during the different years. This is likely because both Sweden and the USA are following the same economic cycle. One noticeable difference is that the us policy rates for a long time were very close to 0 for a long time around the years 2010-2015. As for the inflation rate of both countries, it seems as if they are fairly stable over time, until the year 2020 in which there can be seen to be a large positive spike for both countries, as well as a large negative spike for the USA. This can also be seen for the other variables, although the spikes do not stand out as much for them, in comparison to the previous values as there can be seen to be previous spikes as well.

5.2 Results

The first test performed was the adjusted Dickey-Fuller test for stationarity, with the results shown in table 2.

Table 2: Adjusted Dickey-Fuller test results

	statistic	p.value
us_interest	-2.31	0.45
swe_interest	-2.51	0.36
us_unemployment	-2.22	0.48
swe_unemployment	-3.40	0.05
swe_CPIF	-2.46	0.38
us_PCE	-4.01	0.01

Since all the p-values are not statistically significant except for the US PCE, it means that these time series for the variables are not stationary and the null hypothesis of a unit root cannot be rejected. This could potentially be problematic. However, based on the visual look at the time series in figure 1, it was determined that they are fairly stable over time concerning both mean and variance and there is no apparent long-running trend. ADF tests are sensitive to large spikes, such as the one apparent around the year 2020 which may affect the performance of the tests. Furthermore, the variables should all be relatively bounded based on economic theory. For example. the interest rates and inflation cannot be a negative for longer period and is typically only seen empirically during deep recessions, in which case large effort is put in to make them positive. Likewise, the unemployment t rate cannot be below 0 percent. Likewise, the unemployment rate cannot be below 0 percent while it has never been higher than 14.80 percent for the US and around 14 percent for Sweden. Hence, these processes cannot for example be said to follow a random walk or any other pattern of a long persistent trend. Thus, even though it is nonstationary, it should still be possible to use the time series without further adjustments. Although their first-order difference proved to be stationary, it would not make more sense to use it since it is still based on the difference between two consecutive values, in which case the overall pattern is still relatively similar.

I don't know how else I should motivate using nonstationary data. Is this OK?

Another concern is the autocorrelation of the data. The ljung box-test showed there being autocorrelation (Table 4 in the appendix). Also, a visual inspection of the Autocorrelation functions and partial autocorrelation functions showed the behavior of a typical AR process in a slow decay of the ACF, and with a sharp cutoff for the PACFs. The PACFs were significant for the first couple of lags, but a large spike for the 12 lag, which seems reasonable considering it is monthly data. Based on these results and the visual plots, positive values are typically followed by other positive values, whereas negative values are typically followed by negative values. Nevertheless, it is not feasible to, for example, consider using an AR(12) model to take into account the 12:th lag, and it is more desirable to keep the model parsimonious. Hence for the causal null test as well as the nonparametric regression used, the approach was to keep only one autoregressive part of the inflation rate. It is not clear if using more autoregressive parts would change the results in any major way. But by keeping one autoregressive part, the models are still fairly simple while also keeping it based on economic theory that the inflation at some period t for one time period can be modeled by the information given by the previous period $t - 1$, as done in Fama (1975) mentioned in section 2.1.

Table 3: Causal null tests for Swedish and US data

	Swedish Data	**US Data**
Lag	p.val	p.val
1	0.18	0.26
2	0.02	0.00
3	0.00	0.00
4	0.03	0.00
5	0.09	0.00
6	0.31	0.37

The results from the causal null test are shown in table 3. It was tested for six different models by using different lags of the unemployment rate and the policy rates, as well as keeping one autoregressive part of the inflation. For example, for the results relating to Lag 1, the results were based on using the policy rate as the treatment variable for inflation, with the one-month lag of interest rate and unemployment, as well as an autoregressive term of the inflation for the previous month, as covariates. For the results relating to lag 2, the lag length of interest rate and unemployment rates was increased to use the two-month lagged data instead, and so on up to lag 6.

The p-values are significant for lag 2,3 and 4 for the Swedish data. Likewise, for the US data, the results have significant p-values for lag 2,3, 4, and 5. Hence, the null hypothesis of a flat causal causal dose-response curve is rejected. Based on the significant lags, several nonparametric models were fitted by the method described in section 4.4 (figures 2-9 in the appendix). However, according to these figures, there does not seem to be any special relationship between the inflation rates and the policy rates while taking into account the unemployment rates as well as the lagged values of the inflation rate. The curves for both the Swedish data and US data seem to be fairly flat, although there are some minor curves around them. Furthermore, the automatically chosen bandwidths (tables 5, 6 in the appendix) for the Swedish variables were typically around 0-3 for all the variables. However, for some of the US unemployment and policy rates, the bandwidths were unreasonably high above several thousand. Hence, the results did not provide any insight into the relationship between the policy rate and inflation.

6 Discussion and conclusions

The main goal of this study was to analyze whether the causal dose-response curves of policy rate on inflation are flat for the US and Swedish data and to use nonparametric regression to try and capture what that relationship may look like. The most important result was that the nonparametric test for the causal-dose response curve showed significant results for lags 2,3 and 4 for the Swedish data and lags 2-5 for the US data, indicating that the causal-dose response curve is not flat. Based on these results, it could mean that the causal effects of policy rates are seen within 2-4 months for Sweden and 2-5 months for the USA, which would also align with the economic theories in there being a lag between the effects. Thus these results provide new evidence of the causality of policy rates on inflation based on the causal inference framework which has not seen much use in this area.

The nonparametric regression curves did not capture any special relationship. Although this was not formally analyzed in this study, one potential problem could be the high spikes in inflation rates following the COVID-19 epidemic. It is not known whether the inclusion of these values may have affected the results in any way. Another concern regarding the US data is the fact that the policy rates were close to 0 for a long time during the years 2010-2015 and it is not clear whether this may have affected the results, but together with the large spikes, it may explain why the bandwidths for the US data specifically seemed unreasonably high as the models may have have had trouble smoothing the data trying to capture a relationship.

There is also room for improvement for future research. For example, the model specifications may have been wrong. In this case, the models using lagged versions of the unemployment rate and the interest rates used the same lag for both of them. it is possible that the real lagged effects if they can be singled out to a specific lag, are different for interest rates and unemployment rates. it is also possible that the inclusion of one autoregressive part of the inflation rate may have been misspecified in that it should have included more. Furthermore, the models used to specify and estimate the nuisance functions μ_0 and g_0 for the causal null test were based on the GLM framework. Another type of model may do better work at this that can better capture the dependencies of the time series data such as the autocorrelation. It is also unknown whether this may have affected the asymptotic of the test statistics in any way. For example, the cross-validation estimators may not be unbiased if the autocorrelation is not taken into account. Other types of cross-estimation techniques may be better suited for this type of time series data.

I don't know what more I should write about this since I

Nevertheless, to the best of my knowledge, these results are the first of their kind in that it has not been done before, at least about the nonparametric test of the causal dose-response curve. Thus, although the results are interesting from that perspective, they should still be interpreted with some caution due to the limitations of the study. However, future research based on the potential improvements is encouraged.

Acknowledgements

First and foremost, I would like to express my appreciation to my supervisor Dr. Yang for his guidance and support throughout the thesis. His insights have provided much help for the outcome of this study. I would also like to thank my friends and classmates at Uppsala University for supporting and helping me throughout the degree. Lastly, I want to express my gratitude to my family for their support and encouragement during all this time.

Here i am just wondering if your title is formally Dr. or if its Professor?

Bibliography

Adolfson, M., Andersson, M., Lindé, J., Villani, M. & Vredin, A. (2007), 'Modern forecasting models in action: Improving macroeconomic analyses at central banks', *Riksbank Research Paper Series* (23).

Angrist, J. D., Jordà, Ò. & Kuersteiner, G. M. (2018), 'Semiparametric estimates of monetary policy effects: string theory revisited', *Journal of Business & Economic Statistics* **36**(3), 371–387.

Berger, H. & Österholm, P. (2011), 'Does money matter for us inflation? evidence from bayesian vars', *CESifo Economic Studies* **57**(3), 531–550.

Bernanke, B. S. (2020), 'The new tools of monetary policy', *American Economic Review* **110**(4), 943–983.

Bernanke, B. S. & Gertler, M. (1995), 'Inside the black box: the credit channel of monetary policy transmission', *Journal of Economic perspectives* **9**(4), 27–48.

Board, F. R. (2023), 'Monetary policy report - march 2023'. Accessed: 2024-08-12.
URL: *https://www.federalreserve.gov/monetarypolicy/2023-03-mpr-summary.htm*

Boivin, J., Kiley, M. T. & Mishkin, F. S. (2010), How has the monetary transmission mechanism evolved over time?, *in* 'Handbook of monetary economics', Vol. 3, Elsevier, pp. 369–422.

Doan, T., Litterman, R. & Sims, C. (1984), 'Forecasting and conditional projection using realistic prior distributions', *Econometric reviews* **3**(1), 1–100.

Fama, E. F. (1975), 'Short-term interest rates as predictors of inflation', *American Economic Review* **65**(3), 269–282.

Federal Reserve (n.d.), 'Economy at a glance: Inflation'. Accessed: 2024-02-26.
URL: *https://www.federalreserve.gov/economy-at-a-glance-inflation-pce.htm*

Friedman, B. M. (2000), 'Monetary policy'.

Friedman, M. (1968), 'The role of monetary policy', *American Economic Review* **58**(1), 1–17.

Gertler, M. & Karadi, P. (2015), 'Monetary policy surprises, credit costs, and economic activity', *American Economic Journal: Macroeconomics* **7**(1), 44–76.

Gordon, R. J. (2011), ‘The history of the phillips curve: Consensus and bifurcation’, *Economica* **78**(309), 10–50.

Granger, C. W. (1969), ‘Investigating causal relations by econometric models and cross-spectral methods’, *Econometrica: journal of the Econometric Society* pp. 424–438.

Granger, C. W. & Newbold, P. (1974), ‘Spurious regressions in econometrics’, *Journal of econometrics* **2**(2), 111–120.

Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press.
URL: *https://www.worldcat.org/title/time-series-analysis/oclc/1194970663referer=brief$_r$esults*

Horvitz, D. G. & Thompson, D. J. (1952), ‘A generalization of sampling without replacement from a finite universe’, *Journal of the American statistical Association* **47**(260), 663–685.

IMF, I. M. F. (2023), ‘World economic outlook database, april 2023: Groups and aggregates’, `https://www.imf.org/en/Publications/WEO/weo-database/2023/April/groups-and-aggregates`. Accessed: 2024-07-31.

International Monetary Fund, I. (2006), ‘Inflation targeting and the imf’, *Policy Papers* **2006**(006), A001.
URL: *https://www.elibrary.imf.org/view/journals/007/2006/006/article-A001-en.xml*

Ireland, P. N. (2010), Monetary transmission mechanism, *in* ‘Monetary economics’, Springer, pp. 216–223.

Islam, T. U. & Ahmed, D. (2023), ‘Inflation targeting: A time-frequency causal investigation’, *Plos one* **18**(12), e0295453.

Kashyap, A. K. & Stein, J. C. (1994), Monetary policy and bank lending, *in* ‘Monetary policy’, The University of Chicago Press, pp. 221–261.

Leeper, E. M., Sims, C. A., Zha, T., Hall, R. E. & Bernanke, B. S. (1996), ‘What does monetary policy do?’, *Brookings papers on economic activity* **1996**(2), 1–78.

Li, Q. & Racine, J. (2004), ‘Cross-validated local linear nonparametric regression’, *Statistica Sinica* pp. 485–512.

Ljung, G. M. & Box, G. E. (1978), ‘On a measure of lack of fit in time series models’, *Biometrika* **65**(2), 297–303.

Mavroeidis, S., Plagborg-Møller, M. & Stock, J. H. (2014), 'Empirical evidence on inflation expectations in the new keynesian phillips curve', *American Economic Journal: Journal of Economic Literature* **52**(1), 124–188.

Mishkin, F. S. (2016), 'The economics of money, banking, and financial markets'.

Mishra, P. & Montiel, P. (2013), 'How effective is monetary transmission in low-income countries? a survey of the empirical evidence', *Economic Systems* **37**(2), 187–216.

Mishra, P., Montiel, P. J. & Spilimbergo, A. (2012), 'Monetary transmission in low-income countries: effectiveness and policy implications', *IMF Economic Review* **60**(2), 270–302.

Moraffah, R., Sheth, P., Karami, M., Bhattacharya, A., Wang, Q., Tahir, A., Raglin, A. & Liu, H. (2021), 'Causal inference for time series analysis: Problems, methods and evaluation', *Knowledge and Information Systems* **63**, 3041–3085.

Nelson, C. R. & Schwert, G. W. (1977), 'Short-term interest rates as predictors of inflation: On testing the hypothesis that the real rate of interest is constant', *The American Economic Review* **67**(3), 478–486.

Nielsen, L. (2011), *Classifications of Countries Basedon their Level of Development: How it is Done and How it Could Be Done*, International Monetary Fund.

Phelps, E. S. (1967), 'Phillips curves, expectations of inflation and optimal unemployment over time', *Economica* pp. 254–281.

Phelps, E. S. (1968), 'Money-wage dynamics and labor-market equilibrium', *Journal of political economy* **76**(4, Part 2), 678–711.

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.

Racine, J. & Li, Q. (2004), 'Nonparametric estimation of regression functions with both categorical and continuous data', *Journal of Econometrics* **119**(1), 99–130.

Robins, J. (1986), 'A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect', *Mathematical modelling* **7**(9-12), 1393–1512.

Robins, J. M., Rotnitzky, A. & Zhao, L. P. (1994), 'Estimation of regression coefficients when some regressors are not always observed', *Journal of the American statistical Association* **89**(427), 846–866.

Rubin, D. B. (1973), 'Matching to remove bias in observational studies', *Biometrics* pp. 159–183.

Sims, C. A. (1980), 'Macroeconomics and reality', *Econometrica: journal of the Econometric Society* pp. 1–48.

Stock, J. H. & Watson, M. W. (2001), 'Vector autoregressions', *Journal of Economic perspectives* **15**(4), 101–115.

Sveriges Riksbank (2022), 'What is inflation?'. Accessed: 2024-08-08.
URL: *https://www.riksbank.se/en-gb/monetary-policy/the-inflation-target/what-is-inflation/*

Sveriges Riksbank (2024), 'The inflation target'. Accessed: 2024-08-08.
URL: *https://www.riksbank.se/en-gb/monetary-policy/the-inflation-target/*

Sveriges Riksbank (n.d.), 'How is inflation measured?'. Accessed: 2024-02-26.
URL: *https://www.riksbank.se/en-gb/monetary-policy/the-inflation-target/how-is-inflation-measured/*

Taylor, J. B. (2001), 'The role of the exchange rate in monetary-policy rules', *American economic review* **91**(2), 263–267.

Tobin, J. (1969), 'A general equilibrium approach to monetary theory', *Journal of money, credit and banking* **1**(1), 15–29.

Tran, N. (2018), 'The long-run analysis of monetary policy transmission channels on inflation: a vecm approach', *Journal of the Asia Pacific Economy* **23**(1), 17–30.

Villani, M. (2006), 'Inference in vector autoregressive models with an informative prior on the steady state', *Riksbank Research Paper* (19).

Vonnák, B. (2021), NEW WAYS TO ESTIMATE THE DYNAMIC EFFECTS OF MACROECONOMIC POLICIES, PhD thesis, Central European University.

Westling, T. (2022), 'Nonparametric tests of the causal null with nondiscrete exposures', *Journal of the American Statistical Association* **117**(539), 1551–1562.

Xu, Z., Kim, S. & Zhao, Z. (2022), 'Locally stationary quantile regression for inflation and interest rates', *Journal of Business & Economic Statistics* **40**(2), 838–851.

Young, J. G., Logan, R. W., Robins, J. M. & Hernán, M. A. (2019), 'Inverse probability weighted estimation of risk under representative interventions in observational studies', *Journal of the American Statistical Association* .

Appendix

Table 4: Ljung-Box test for white noise

	statistic	parameter	p.value
us_interest	3444	12	0.00
swe_interest	3410	12	0.00
us_unemployment	2386	12	0.00
swe_unemployment	2904	12	0.00
swe_KPIF	2300	12	0.00
us_PCE	722	12	0.00

	lagged_swe_KPIF)	swe_unemployment	swe_interest
lag = 1	2.96	1.22	1.34
lag = 2	2.67	0.76	1.17
lag = 3	2.19	0.74	1.38
lag = 4	1.75	0.30	1.91
lag = 5	2.06	0.51	1.81
lag = 6	1.70	0.79	0.26

Table 5: Bandwidths for Swedish data

	lagged_us_PCE	us_unemployment	us_interest
lag = 1	5.74	3.86	4795118.89
lag = 2	5.94	1789395	0.31
lag = 3	5.59	1564231	3285292
lag = 4	5.58	6451167	5093350
lag = 5	5.56	2939304	2911466
lag = 6	1.16	0.28	5.67

Table 6: Bandwidths for US data

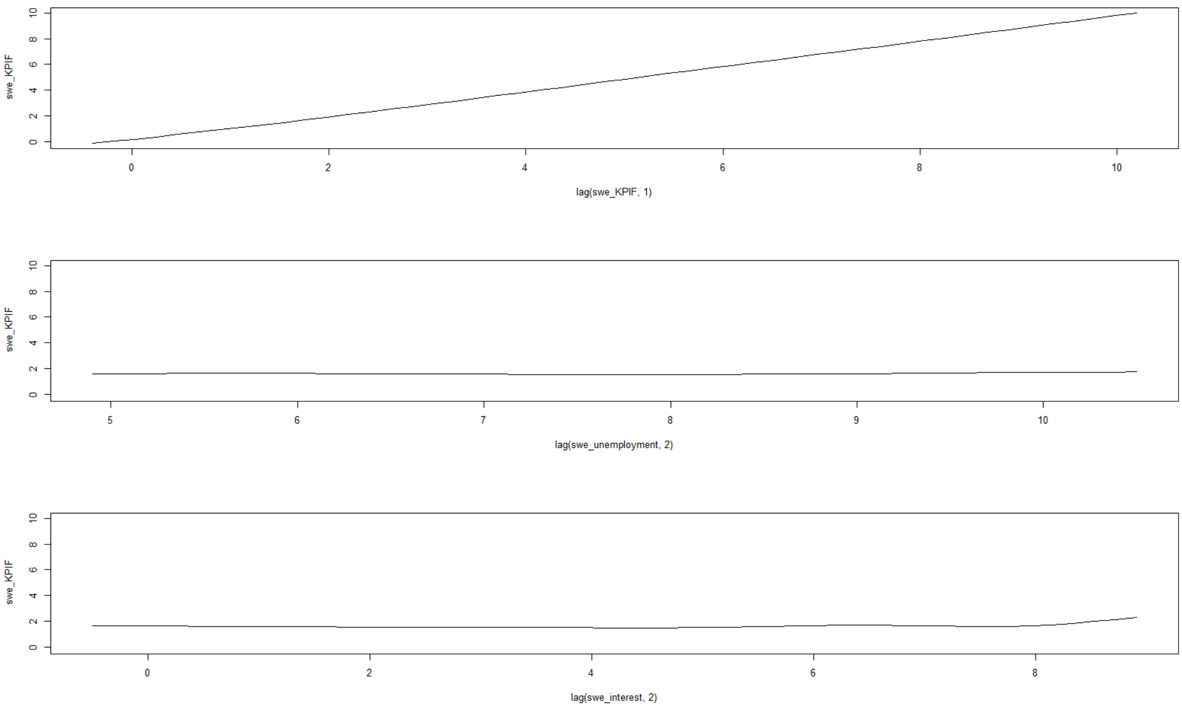

Figure 2: Nonparametric regression curves for Swedish CPIF using 2 lag

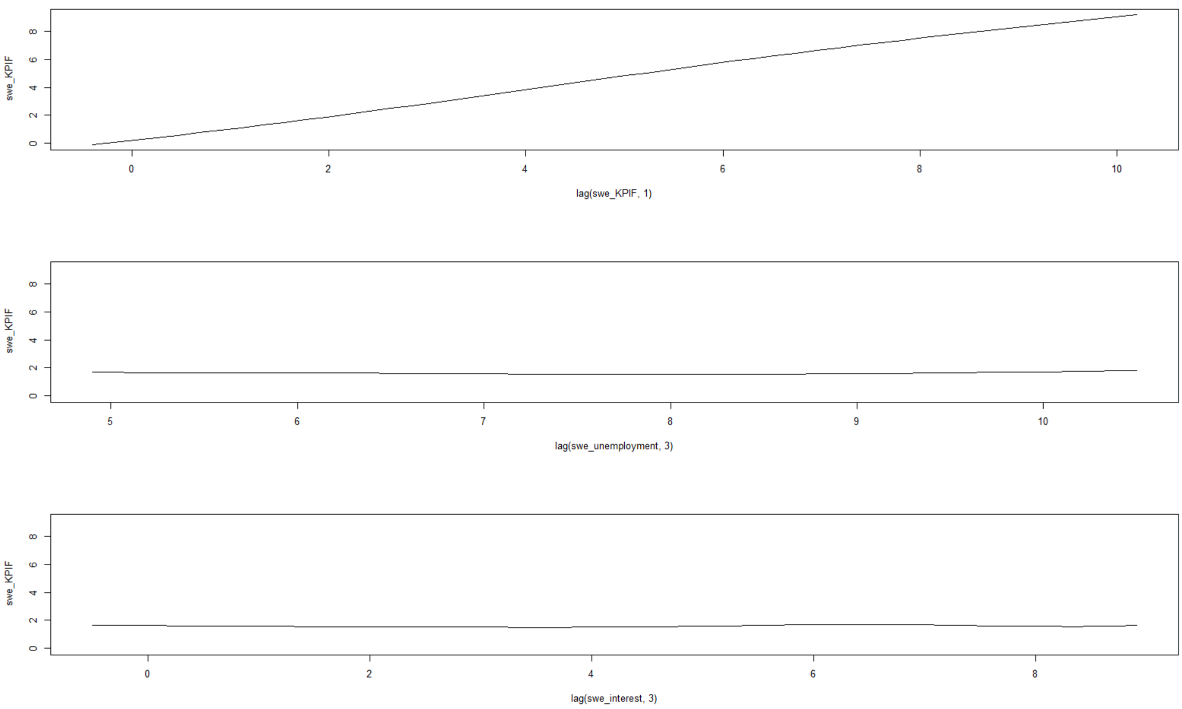

Figure 3: Nonparametric regression curves for Swedish CPIF using 2 lags

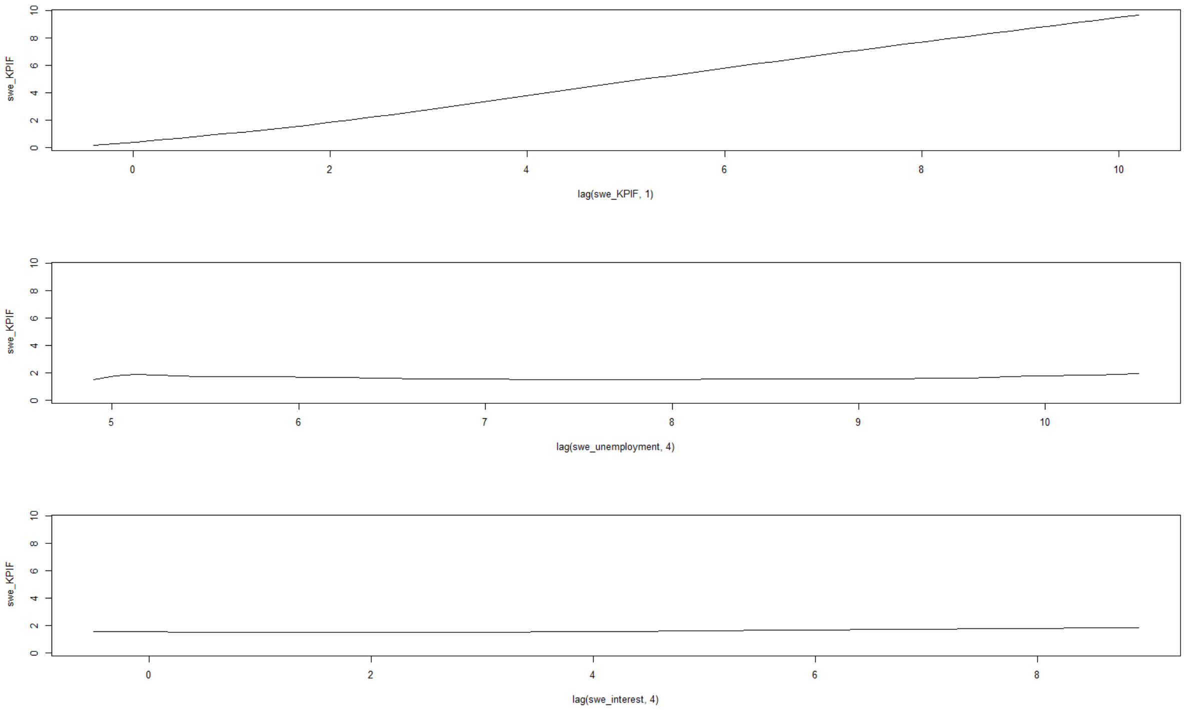

Figure 4: Nonparametric regression curves for Swedish CPIF using 4lags

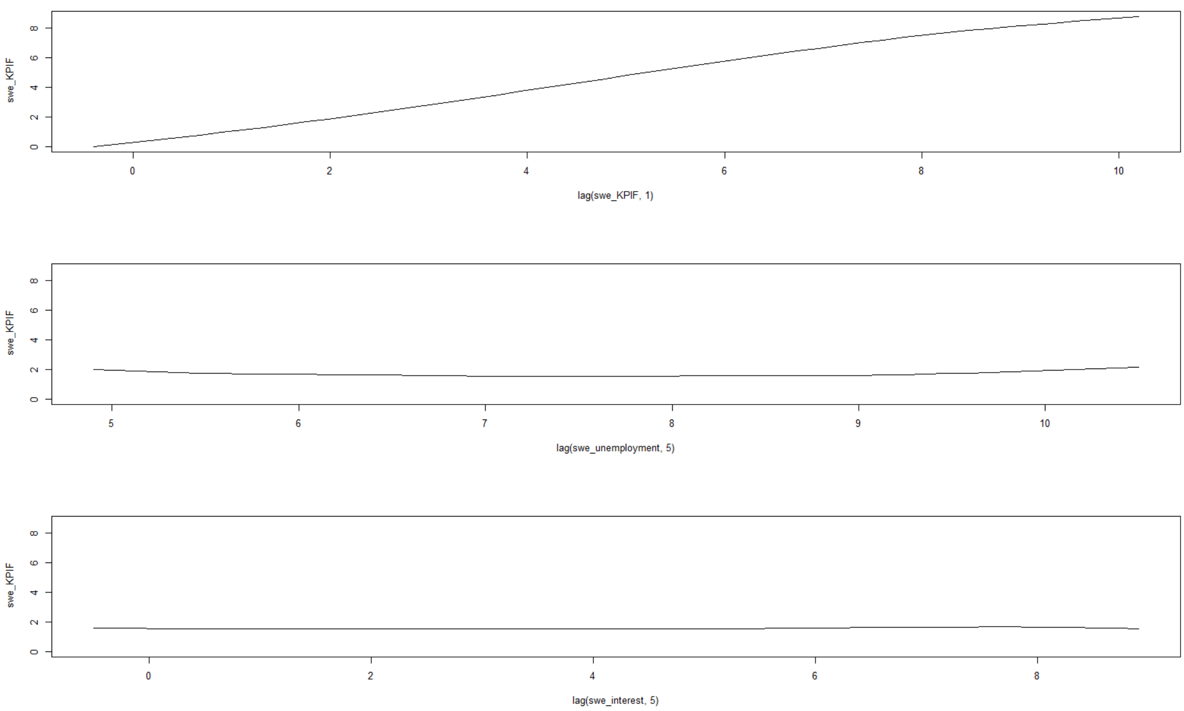

Figure 5: Nonparametric regression curves for Swedish CPIF using 5 lags

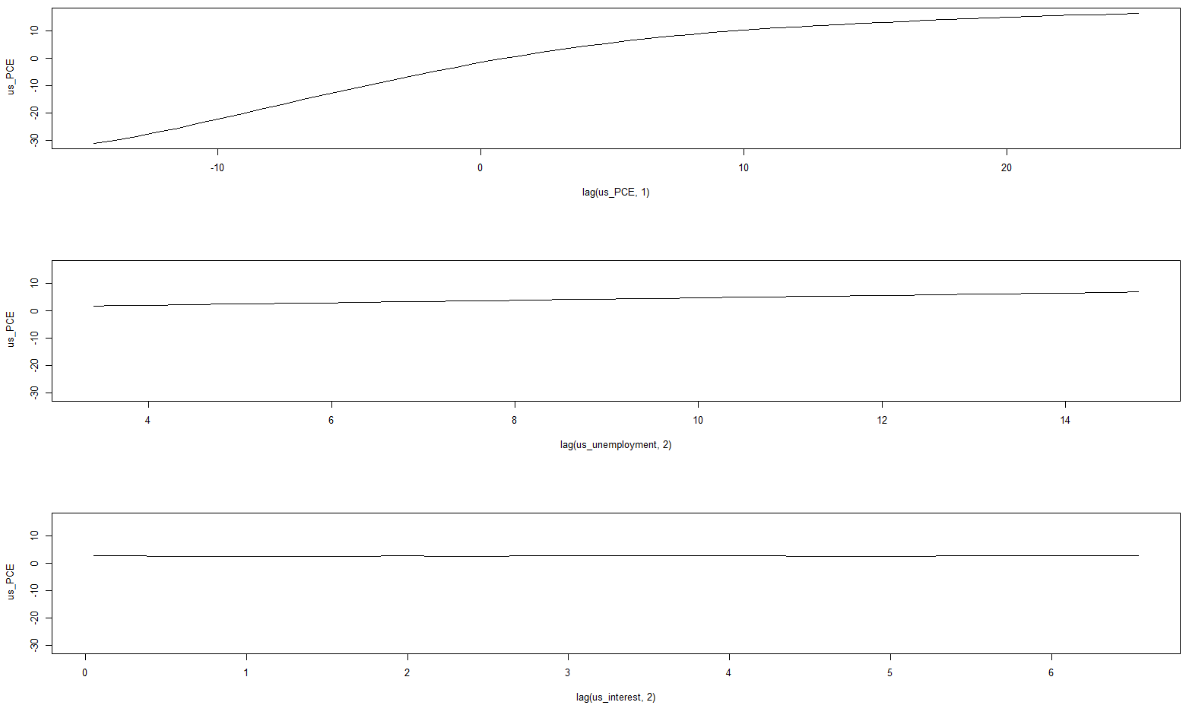

Figure 6: Nonparametric regression curves for US PCE using 2 lags

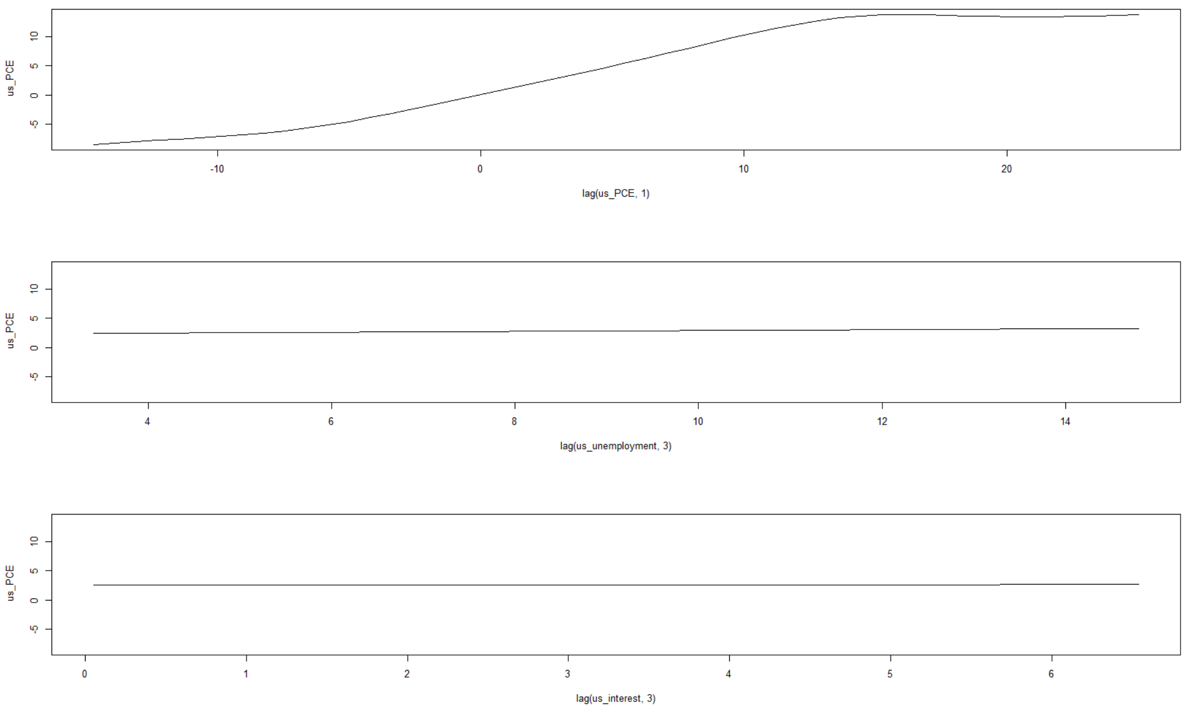

Figure 7: Nonparametric regression curves for US PCE using 3 lags

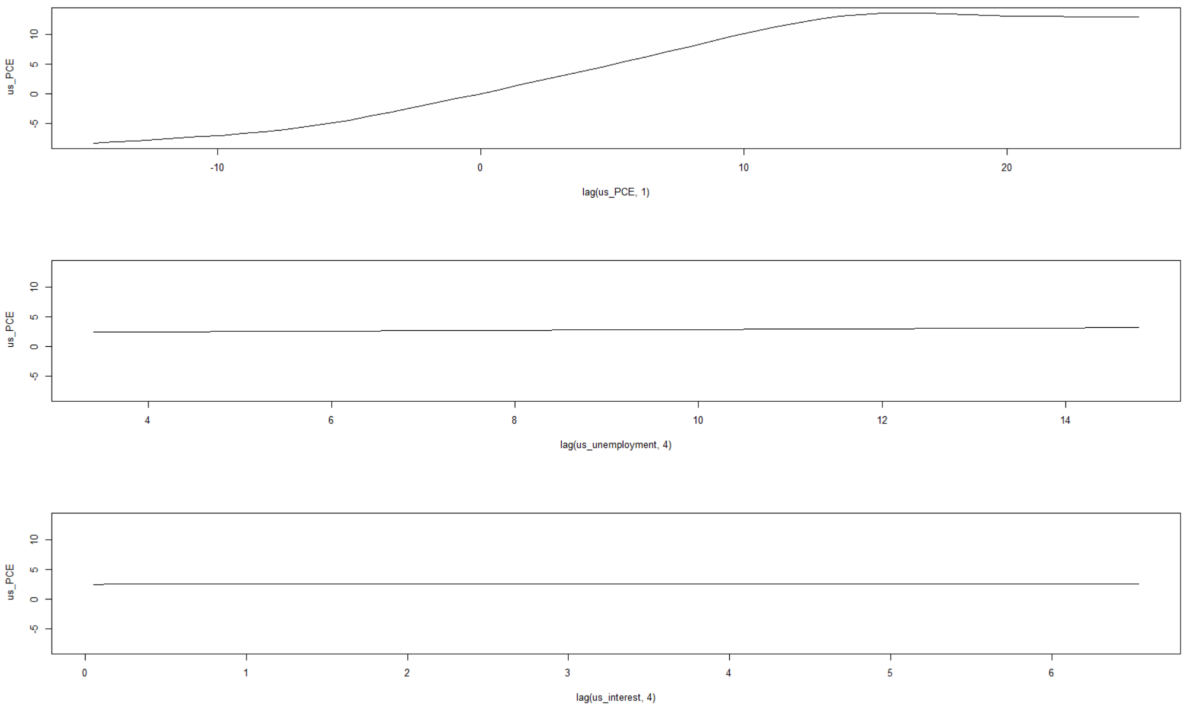

Figure 8: Nonparametric regression curves for US PCE using 4 lags

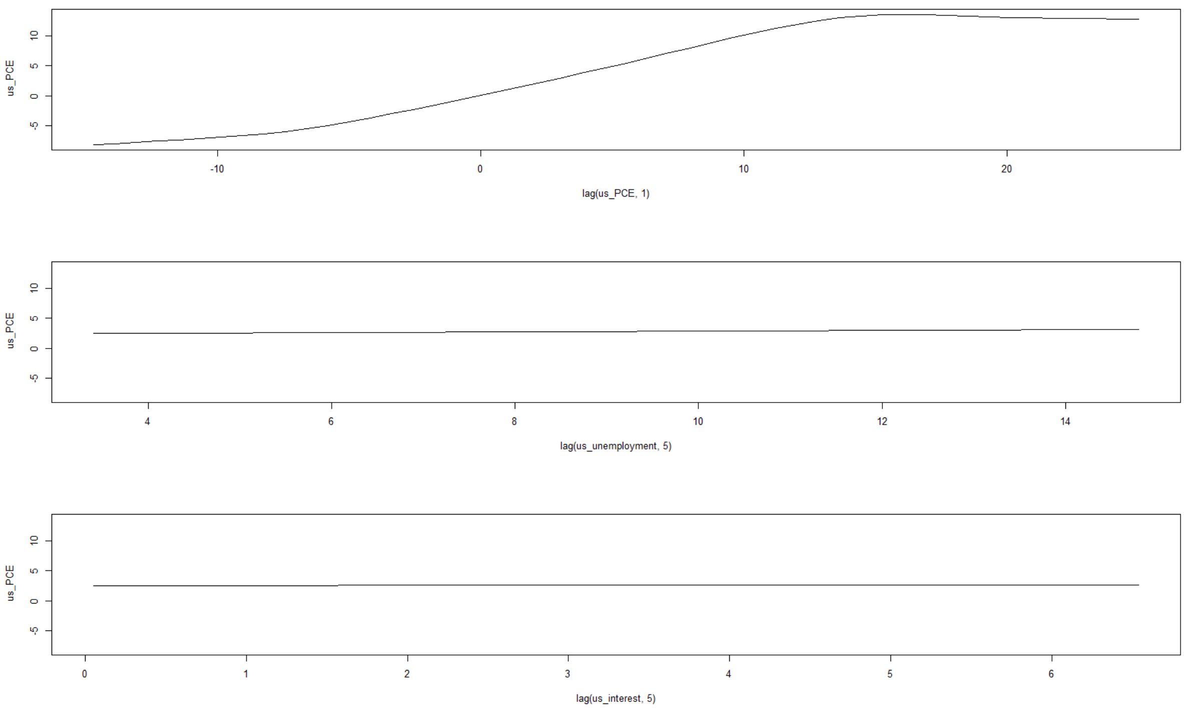

Figure 9: Nonparametric regression curves for US PCE using 5 lag