Master Thesis

Henry Svedberg

August 12, 2024

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: a general increasing economic demand, increased firm expenses due to higher wages production cost or supply chocks, or due to the expectation about future inflation causing firms and people to adjust price and wages accordingly (Sveriges Riksbank 2022).

Monetary policy is the most important factor alongside fiscal policy makers 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; that is, that the central banks try to maintain the inflation rate at a constant value typically at or around 2 percent. This was first adopted in the 90's 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 reason behind having an inflation target is that it helps in creating favourable conditions for economic development by keeping a stable inflation rate (Sveriges Riksbank 2024).

The main tool the central banks rely on to steer the inflation is the policy rate, which are the daily interest rates that banks use when lending and borrowing to one another. This, in turn, affects other interest rates in the economy such as the banks and mortgage institutions. Thus, by adjusting the policy rates the central banks are able to influence the pace and direction of the overall economic activity affecting the aggregate output, employment and the inflation (Friedman 2000). Although not of interest for this study, there exists other complementary tools as

well which the central banks can use to affect the interest rate. For example quantitative easing which developed after the financial crisis 2008 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 exists a large amount of literature on this subject (Boivin et al. 2010). This is of course not without good reason; the transmission mechanism of the monetary policy has been thoroughly examined due to its importance for various reasons. Policymakers must have an adequate assessment regarding how the monetary policy affects the economy as well as the timing of the effects to determine the monetary policy at a specific point in time (Boivin et al. 2010).

However, when it comes to estimating the actual causal effects of the policy rate on inflation, there is very little available litterature. Much of the previous research based on the econometrics framework about the inflation rate and policy rates has typically revolved around making predictions. Usually by methods such as estimating Vector autoregression models and Granger causality for time series data, for example, revolving around analyzing monetary shocks. Although these types of analyses have their usage, these are in broad sense more oriented towards causal discovery, which is the task of identifying and understanding causal relationships in the data. However, such methods lack causal effect estimation, which is the estimation of the actual effect of a policy or treatment on a target variable. This is instead commonly done by other methods using metrics such as Average Treatment Effect (ATE), commonly associated with the causal inference framework as laid out by the Rubin Causal model (Moraffah et al. 2021).

In comparison to the traditional methods using the policy rate to forecast inflation, the literature based on the causal inference framework is almost non existent. This is possibly due to the fact that the causal inference framework has typically been developed around the context of a treatment variable being binary or polytomous discrete, thus there also exist much less literature around for continuous exposure variables even within the causal inference framework. However, the work with continuous exposure variables has seen recent development 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. More specifically, this study aims to estimate d $E\left[(Y(R_1|x) - Y(R_0|x)\right] = \eta(R_1, R_0, x)$ the causal dose-response curve. By using a new test devopled by Westling (2022) causal effect, this study will

answer the question if there is an actual causal effect from the policy rate on inflation. Furthermore using nonparametric regression, this study will also try to obtain the full density of the dose-response curve. By doing so, this study will provide an alternative way of seeing the actual causal effect of the policy rate on the inflation and lay out the foundation for others to develop.

I need to make sure this is consistent with the method section

The rest of this paper is structured as follows: A brief summary of the transmission mechanism is given in section 2.4.... ... ..

ction 2.1.

The Section 2.4 covers the history of the literature regarding the causality of interest rate and inflation. However, as a limitation, this is focused on the methodological econometrics and statistical modeling perspective of this area, rather than giving a full overview based more on a more traditional macroeconomics framework.

Will update this a bit more in the end to make the introduction mroe relevant to the actual methods and results

2 Previous research and theoretical framework

2.1 The monetary transmission mechanism

The traditional channels of the transmission mechanism stems 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. That is, it will not analyze 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 of the history of how the transmission mechanism has developed over time.

In accordance with the traditional Keynesian models, one of the channels of the transmission mechanism is interest rate channel; the more traditional way by adjusting the nominal money stock which affects the interest rate or more directly by through the policy rate. In any event though, in practise they fulfill same effects in that it impacts the aggregate output, employment and overall price levels (Mishkin 2016). Though it should be noted that it the real interest rates and not the nominal interest rates that affects 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 an 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 due to investors seeking to eliminate disparities of risk-adjusted expected return. This in turn will translate into changes in the real interest rates as well. Hence the user capital cost rises and the capital asset demand decreases due to increasing borrowing costs for both firms and individuals leading to lower spending on investment and in turn a decline in the aggregate spending and demand, which has a dampening effect on the inflation (Ireland 2010). Furthermore, monetary policy also has direct effect on household consumption. Lowering short-term interest rates, for instance, increases demand for assets like

common stocks and housing and 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 due to the return on domestic assets decreasing compared to that of foreign assets, causing domestic goods to be cheaper compared to foreign goods and an increase in foreign demand for domestic goods and thereby 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 from the higher aggregate demand raises the inflation. Increased interest rates has 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 exists several different mechanisms through which the monetary policy affects inflation and hence 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 nmot 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 individual effects. Nevertheless, the channels mentioned here are simply the main ones commonly mentioned within the literature. Furthermore it should also be noted that there 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 which affects the inflation rate, but it does nevertheless play an important role.

2.2 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 labour 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 develop 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 the account of the role of the expectations regarding inflation. The idea was that the unemployment rate does not affect the the inflation in levels, but rather the rate at which it changes. However, the amount of literature around this topic is enormous and this paper will not go into more details about this aspect. See for example Gordon (2011) and Mavroeidis et al. (2014) for an overview and more recent developments. However, due to the Philips Curve, the unemployment rate is commonly used in other studies inflation and interest rates, and hence why it is also included for this study. Omformulera

2.3 Causality debate

Though the effects of monetary policy on inflation is in widely accepted, there are some debate about its actual effectiveness; especially for developing countries (Islam & Ahmed 2023). A key aspect is the financial structure is different: 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 in relation to 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 distinct 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 generalize for other countries, it may be even more so for developing countries. For the remainder of the study though, it is assumed based on the macroeconomic theories that there already is an established causal relationship between interest rate and inflation.

Maybe just remove this part

2.4 literature overview

Im wondering if should make the it strictly more about the policy rate, rather than interest rates as a whole

There is an abundance of literature within the general area of 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 rate 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 from using policy rates stems from the early work of interest rates in a more general sense. Although the focus has not typically been so much about trying to estimate any direct causal effect, but rather using inflation rate to predict the inflation rate. Much of the early work on this subject were based on regression analyses, often in some autoregressive setting. Some early notable work are for example Fama (1975) or Nelson & Schwert (1977), using short-term interest rates as predictors for inflation, and in which there is 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 that there is reason to suspect that the relationship between inflation and interest rates may not be constant over a longer period of time. 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 inflation rate is positively correlated with one month lagged inflation rate for the quantiles considered. The second conclusion was the the relation between inflation and interest rates exhibits more complicated quantile-specific and time-varying features.

Nevertheless, for a long time the methodological framework of the causality of interest rates on inflation has traditionally been dominated by the use of VAR models, which began 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 its past values, a VAR is an n-equation, n-variable linear model where each variables is regressed on its own past values as well as the current and lagged values of the other n variables (Stock & Watson 2001). The VAR models are

commonly used with the concept of Granger causality attributed to Granger (1969). A variable X is said to Granger 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 (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 $\tau, \tau_{\max}$ denotes the maximum time lag, and e represents an independent noise. Using this equation, we say $_i$ Granger causes Y_j with lag τ if any of the coefficients in $\phi_{ji}(\tau)$ is nonzero. This relationship can be shown by $Y^i_{t-\tau} \rightarrow Y^j_t$ which demonstrates the causal link between Y_i and Y_j at lag τ.

own note: förbättra formulering

One notable study is Stock & Watson (2001), in which they uses VAR and does 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 Federal reserve though the federal funds rate has played an role in reducing the volatility of the inflation Tran (2018) uses an 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 a large number of 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 thus actual effect of monetary policy is not captured. One proposed solution is to instead incorporating Bayesian methods such as Bayesian Retrogressive (BVAR) models. These models are able to 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 are able to 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, whereas the litterature within this area is well developed. From a methodological

perspective they are not so much related to this paper. They are related in the context of trying to establish whether there is an effect, as well as the magnitude, of policy rates on inflation. It should also be noted that with regards to casualty noted that the results from these model cannot be attributed an 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 more based on the causal inference framework, which is almost non existent in comparison. Though there are some examples of it being used in a more broader context of analyzing moonetary policy as for example Angrist & Kuersteiner (2011). To the best of my knowledge there is no paper at all which has attempted estimate the causal dose-response curve of policy rate on inflation.

Depending on the final methods used,, i will try to add some more litterature related to that, as for example specific density estimation or such

3 Method

The method used for this paper can be split into two different parts. The first part is about testing the null hypothesis that the density is flat. This is done in section [to be filled in later]. Given that some results are found, then the next part consist of estimating the density itself, which is done in section4.

3.1 Non parametric test of the causal null

Perhaps a short introduction to the causal inference framework. Key concepts, Rubin causal model, counterfactual outcomes, average treatment effects, propensity score,

In traditional classical tests for association, causality can often be asserted due to randomization. In the case with observational data its 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 for example matching estimators (Rubin 1973), inverse probability weighted (IPW) estimators (Horvitz & Thompson 1952) or augmented inverse propensity weighted estimators (Robins et al. 1994). Then it its 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.

With a continuous exposure, the estimand of interest is usually the causal dose-response curve, which for each value of the exposure is the average outcome if all the units are assigned to that value of exposure. If there is a causal relationship between the exposure and outcome, then it corresponds to the dose-response curve being non-flat. However, this is typically much more difficult to estimate and the literature within this area is not as developed as in the discrete case. Hence, a common way to deal with this is to make the exposure variable discrete and then use other methods, such as the 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 its not always obvious which is the best choice, hence the result will vary depending on how the discretization is done. Converting an continuous exposure variable to discrete also causes information loss, leading to test typically having less power than 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 this study will use non-parametric methods, as they are more robust due to them making less assumptions.

Instead, this study will use a relatively new test as develop 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 which considers global inference for a continuous exposure.

3.2 The test

this is more of a placeholder subtitle, while thinking about if the test should be its own section

Im thinking of ways to change the notations in a consistent way

The test consist of multiple different steps. First the exposure of interest, which in this case represents the policy rate, is denoted as $A \in \mathcal{A} =$ with support $A_0 \subseteq \mathcal{R}$. Considering an outcome that would be observed if a unit could be assigned to a specific value of the exposure (potential outcomes), then for each possible values of the exposure, $a \in \mathcal{A}_0$, the potential outcomes are: $Y(a) \in \mathcal{R}$ by setting exposure to $A = a$. The counterfactual means are defined as $m(a) := E[y(a)]$, which is the average outcome under assignment of the entire population to exposure level $A = a$. The function $m : \mathcal{A} \rightarrow \mathcal{R}$ is the causal dose-response curve, also called the average dose-reponse function. Then the distribution from which the observed data is generated from is denoted P_0 such that: $O_1, \ldots, O_n \sim P_0$ for the data unit $O = (Y, A, W)$ where $Y := Y(A)$ and $W \in \mathcal{R}^d$. Throughout the rest of the paper, functions or parameters indexed by 0 will denote that it depends on the unknown distribution P_0

We want to test the null hypothesis: $m(a) = \gamma_0$ for all $a \longrightarrow \mathcal{A}_0$ and some $\gamma_0 \in \mathcal{R}$, that is, the dose-response curve is flat.

Thew null hypothesis is the formally defined as:

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

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

The null hypothesis corresponds to the causal dose-response curve being flat. 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, due to the fact that we cannot observe $Y(a)$ for all units, but instead the outcome $Y := Y(A)$, then m is not actually a mapping of the joint distribution of the pair (A, y). Hence this null hypothesis cannot actually be tested by the observed data. In order to make this testable, some assumptions has to be made:

1. The potential outcomes of each units are independent of the exposure of all other units (the stable unit treatment assumption).
2. The observed outcome Y almost surely equals $Y(A)$.

Futhermore, 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 is 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 it is possible for all units to receive any treatment a on its support. Given that all the assumptions 1-2 and 4-5 hold,

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

$$E[Y(a)] = E_0\left[\mu_0(a, W)\right] := \theta_0(a),$$

which is known as the *backdoor formula*, or G-*computation* (Robins 1986).

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

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

$$H_1 : \theta_0(a) \neq \gamma_0, \text{ for some } a \in \mathcal{A}$$

Next is to derive the actual test. However, it is difficult to estimate θ_0 itself. Instead the proposed method by Westling (2022) is to estimate an 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 primitve transformation of θ_0

With the following definitions:

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

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 $a \in \mathcal{A}$*

$\Omega_0(a)$ is pathwise differentiable relative to the nonparametric model at each a with an estimable influence function. 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:

$$\begin{aligned} D^*_{a_0,0}(y, a, w) :=& [I_{(-\infty,a_0]}(a) - F_0(a_0)] \\ & \times \left[\frac{y - \mu_0(a, w)}{g_0(a, w)} + \theta_0(a) - \gamma_0 \right] \\ & + \int [I_{(-\infty,a_0]}(u) - F_0(a_0)] \mu_0(u, w) F_0(du) \\ & - 2\Omega_0(a_0). \end{aligned} \tag{3}$$

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

$$\begin{aligned} \frac{1}{V} \sum_{v=1}^{V} \Bigg\{ & \frac{1}{N_v} \sum_{i \in \mathcal{V}_{n,v}} [I_{(-\infty,a_0]}(A_i) - F_{n,v}(a_0)] \\ & \times \frac{Y_i - \mu_{n,v}(A_i, W_i)}{g_{n,v}(A_i, W_i)} \\ & + \frac{1}{N_v^2} \sum_{i,j \in \mathcal{V}_{n,v}} [I_{(-\infty,a_0]}(A_i) - F_{n,v}(a_0)] \mu_{n,v}\left(A_i, W_j\right) \Bigg\}. \end{aligned} \tag{4}$$

Where $F_{n,v}$ and $Q_{n,v}$ are the marginal empirical distributions of A and W respectively within each fold.

I dont know whether i really need Q here, since its not explicitly in the formula. However, in the paper Westling talks about Q as well

This also requires the estimators $u_{n,v}$ 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, generalized additive models as well as multivariate splines , 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 details 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. This has proven to yield good results according to **??** and provides an easy way to use multiple different models.

maybe i will just exclude multivariate splines part

Furthermore, Westling (2022) shows that $n^{1/2}\Omega_n - \Omega_0(a)$ under some 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 appropximation $T_{n,p,a}$ of the $1-\alpha$ quantile of the limiting distribution of

$$n^{1/2} \left\| \Omega_n - \Omega_0 \right\|_{F_0,p} .$$

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).

Note, I don not intend to simply leave it like for all the steps mentioned below. However, this is the steps that Westling highlights in his article and which i don't fully understand.

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 $\Sigma_n\left(A_i, A_j\right) := E\left[Z_n(A_i) Z_n\left(A_j\right)\right.$

$\mid O_1, \ldots, O_n] = \frac{1}{V}\sum_{v=1}^{V} \mathbb{P}_{n,v} D^*_{A_i,n,v} D^*_{A_j,n,v}$ 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 $n^{1/2} \left\|\Omega_n^\circ\right\|_{F_n,p} > T_{n,\alpha,p}$.

4 Nonparametric regression for density estimation

Will fix this section a bit as well. Though my main concern here is if im including the relevant detaisl to describe the test or if i should add something more that i amy be missing

In order to estimate the density, we use non parametric methods.

Density estimation is the process of estimating some underlying probability density function by using the observed data. There are various methods; both parametric methods where the data generating process is from a known family of distribution, or nonparametric methods which are able to more flexible estimate unknown distributions

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.

Nonparametric estimation makes it possible to model relationships among variables while robust against functionfal form of misspecification, making it able to detect structure that may otherwise not would be detected Racine & Li (2004).

Smoothing refers to that predictions being weighted averages of the observed responses in the training data:

$$
\widehat{r}(x) = \sum_{i=1}^{n} y_i w(x, x_i, b) \tag{5}
$$

where h is the bandwidth that controls the degree of smoothing

Smoothing very little ($h \to 0$), means that the model is able to very small details of any potential "true" regression function- Less smoothing leads to less bias. However, it also causes each of the predictions to be an average of fewer observations, making the predictions more noisy. Hence smoothing increases the variance.

Since we have the relationship: Total error = noise + bias^2 + variance, and that changing the amount of smoothing affects both bias and variance, there is an optimal amount of smoothing which we want to find. This is 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.

More specifically in order to estimate the density by adapting a Cross-validated local linear estimator with the method developed by Racine & Li (2004) and Li & Racine (2004). Consider a non parametric regression model

$$y_j = g\left(x_j\right) + u_j, \quad j = 1, \ldots, n, \tag{6}$$

Here we have that x_j is a continuous random vector of dimension q. . Then the derivative of $g(x)$ is defined as $\beta(\boldsymbol{x}) \stackrel{\text{def}}{=} \nabla g(\boldsymbol{x}) \equiv \partial g(\boldsymbol{x})/\partial x (\nabla g(\cdot)$ is a $q \times 1$ vector).

Then

$$\delta(x) = (g(x), \beta(x)')'$$

Hence, $\delta(x)$ is a $(q + 1) \times 1$ vector-valued function whose first component is $g(x)$ and whose remaining q components are the first derivatives of $g(x)$.

By Taylor expanding $g(x_j)$ at x_i, the result is:

$$g\left(x_j\right) = g\left(x_i\right) + \left(x_j - x_i\right)' \beta\left(x_i\right) + R_{ij} \tag{7}$$

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 6 can be rewritten as:

$$\begin{aligned} y_j &= g\left(x_i\right) + \left(x_j - x_i\right)' \nabla g\left(x_i\right) + R_{ij} + u_j \\ &= \left(1, \left(x_j - x_i\right)'\right) \delta\left(x_i\right) + R_{ij} + u_j. \end{aligned} \tag{8}$$

A leave-one-out local linear kernel estimator of $\delta\left(x_i\right)$ is obtained by a kernel weighted regres-

sion of y_j on $\left(1, \left(x_j - x_i\right)'\right)$ given by

$$\begin{aligned}\hat{\delta}_{-i}(x_i) &= \begin{pmatrix}\hat{g}_{-i}(x_i)\\ \hat{\beta}_{-i}(x_i)\end{pmatrix} \\ &= \left[\sum_{j\neq i} W_{h,ij}\begin{pmatrix} 1, & \left(x_j - x_i\right)' \\ x_j - x_i, & \left(x_j - x_i\right)\left(x_j - x_i\right)'\end{pmatrix}\right]^{-1} \sum_{j\neq i} W_{h,ij}\begin{pmatrix}1\\ x_j - x_i\end{pmatrix} y_j,\end{aligned} \tag{9}$$

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.

Define a $(q+1)\times 1$ vector e_1 whose first element is one with all remaining elements being zero. The leave-one-out kernel estimator of $g(x_i)$ is given by $\hat{g}_{-i}(x_i) = e_1'\hat{\delta}_{-i}(x_i)$, and we choose $h_1, \ldots, h_q$ to minimize the least-squares crossvalidation function given by

$$CV\left(h_1, \ldots, h_q\right) = \sum_{i=1}^{n}\left[y_i - \hat{g}_{-i}(x_i)\right]^2. \tag{10}$$

We use $\hat{h} = \left(\hat{h}_1, \ldots, \hat{h}_q\right)$ to denote the cross-validation choices of $h_1, \ldots, h_q$ that minimize equation 9. Having computed $\hat{h}$ we then estimate $\delta(x)$ 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 $W_{\hat{h},ix} = \prod_{s=1}^{q} \hat{h}_s^{-1} w\left(\left(x_{is} - x_s\right)/\hat{h}_s\right)$, and we estimate $g(x)$ by $\hat{g}(x) = e_1'\hat{\delta}(x)$.

4.1 Dickey-Fuller test for stationarity

The concept of stationary is an important 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 to 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:

$$\begin{aligned}\gamma_{jt} &= \mathrm{Cov}\left(Y_t, Y_{t-j}\right) = \mathrm{E}\left(Y_t - \mu_t\right)\left(Y_{t-j} - \mu_{t-j}\right) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(y-\mu_t)\left(x-\mu_{t-j}\right) f_{Y_t,Y_{t-j}}(y,x)\mathrm{d}y\ \mathrm{d}x\end{aligned} \tag{11}$$

Then 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:

$$\begin{aligned}\mathrm{E}\left(Y_t\right) &= \mu, && \text{for all } t; \\ \mathrm{E}\left(Y_t - \mu\right)\left(Y_{t-j} - \mu\right) &= \gamma_j < \infty, && \text{for all } t \text{ and any } j\end{aligned} \tag{12}$$

Time series need to be stationary to ensure that the properties remain the same over time and to make valid inference. Consequences of using nonstationary time series in a regression setting are for example 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 $\mathrm{E}\left(e_t \mid y_{t-1}, y_{t-2}, \ldots y_0\right) = 0$. Then $\{y_t\}$ has a unit root if and only if $\rho = 1$

should this be written with parenthesis as $\{e_t\}$?

The 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:

$$\Delta y_t = \alpha + \theta y_{t-1} + \delta t + \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \cdots + \zeta_k \Delta y_{t-k} + \epsilon_t \tag{13}$$

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

The test statistic does not have a standard t-distribution, but Hamilton (1994) derives the limiting distribution.

$$\frac{T \cdot (\hat{\rho}_T - 1)}{1 - \hat{\zeta}_{1,T} - \hat{\zeta}_{2,T} - \cdots - \hat{\zeta}_{p-1.T}} \xrightarrow{L} \frac{\frac{1}{2}\{[W(1)]^2 - 1\} \cdot \int W(r)\, dr}{\int [W(r)]^2\ dr - \left[\int W(r)\, dr\right]^2} \tag{14}$$

Is this the correct one?? Or is it only for without drift?? From page 553 in Hamilton

4.2 Ljung-Box

$$Q_{LB}(k) = T(T+2)\sum_{j=1}^{k}\frac{\hat{\rho}_j^2}{T-j} \overset{a}{\sim} \chi^2(k-p) \tag{15}$$

To estimate the outcome regression u_0 and the propensity score g_0 we used SuperLearner with flexible libraries consisting of genreralized linear models with the identity link function for u_0 and the logit link funciton for g_0, generalized additive models and multivariate regression splines

4.3 GLM

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. 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:

$$f\left(y_i\right)=\exp\left\{\frac{y_i\theta_i-b\left(\theta_i\right)}{a_i(\phi)}-c\left(y_i,\phi\right)\right\}, \tag{16}$$

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

$$\eta_i=g(u_i)\sum_{j=1}^{p}\beta_j x_{ij}, \quad i=1,\ldots,n. \tag{17}$$

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 he 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 because 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:

$$\text{logit}(\pi_i) = \log(\frac{\pi_i}{1-\pi_i}) = \sum_{j=1}^{p} \beta_j x_{ij} \tag{18}$$

The generalized additive model can can be considered as a general extension of the generalized linear model by replacing the linear predictor with additive smooth functions of the of the covariates. For example, instead of the previous GLM structure wit the logit link as described by equation 18, 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:

$$\text{logit}(\pi_i) = \log(\frac{\pi_i}{1-\pi_i}) = \sum_{j=1}^{p} s_j(x_{ij}) \tag{19}$$

GAM models provide more flexibility of 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 is able to discover patterns which 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 inference compared to GLMS as for example 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 inference, but rather estimate the nusiscance estimators for the Westling test.

5 Data

The data set used for this is Panel data obtained by combining several different individual monthly time series data for the inflation rate, policy interest rate as well as the unemployment for USA and Sweden. Originally the individual time series for the different variables have different lengths since the starting years for each of them vary, but they all end in either December 2023 or January 2024. However, the data for the Swedish policy rate is the shortest of the individual time series due to it starting in June 1994 since that is when the data consistently started being tracked monthly. Hence, for the combined panel data used in this study, the other time series have been limited to also start from June 1994 to make it more comparable.

All the data is collected from various public governmental sources. The Swedish policy interest rate is from the Swedish Central Bank Riksbanken, and the inflation rate as well as the unemployment rate is retrieved from Statistics Sweden (Statistika Centralbyrån, SCB). The US Fed Funds rate is retrieved from the US Federal Reserve Bank, and the data for the unemployment and consumer price index is retrieved from the U.S. Bureau of Labor Statistics.

For both the US and Swedish data, the concept of inflation is operationalized through the consumer price index CPI, 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 (U.S. Bureau of Labor Statistics), (Statistics Sweden). Hence this is used as a proxy for the overall rate of inflation. There are different measurements one can use for the CPI such as the actual CPI value, or for example monthly differences. For this study, the focus will be on the 12 month percentage change. That is, the difference percentage difference for one month compared to the same month one year ago, since this is the measurement commonly used when referring to the inflation rate Sweden (2024). Annother reason to use the annual percentage change rather than the direct CPI is the fact that the data will liekly be more stationary as the annual change is less likely to be dependent upon previous values.

Regarding the unemployment rate, it should be noted that whereas they are roughly similarly defined in the Swedish and US cases, there are some differences in some of the details. In general, the unemployment rate for both countries refers to the people who are part of the labor force, meaning that they do not currently have a job but are actively seeking a job. Hence, people without jobs but who are not actively seeking jobs are not considered unemployed. For the US data, this data is originally collected from surveys done by Bureau of Labor Statis-

omformulera lite sen

tics, whereas the Swedish data is obtained from Arbetskraftundersökningarna (AKU), a survey done by Statistics Sweden. Thus there are some differences in the exact method by which the surveys were done such as the target population for the survey as well as some of the definitions used in the surveys. There have also been some changes over the years. For example, in 2021 AKU made some changes due to a new framework law implemented by the EU; this included changes in the target population of the surveys as well as some definitions used, and hence 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. Nevertheless, overall this is considered a minor problem for this study, and potential differences are not further analyzed for this study.

Hence this study will consider three variables, inflation rate, unemployment rate and the policy rate. Although in some literature, a larger collection of variables are used, such as for example in Leeper et al. (1996), much of the literature considers these variables in particular such as for example one notable being Stock & Watson (2001). In this case it is also preferably to try to limit the number of variables to keep the models as parsimonous as possible due somewhat limited available data.

5.1 Summary statistics

Table 1: Summary statistics for the different variables

	n	mean	sd	median	min	max	skew	kurtosis
swe_CPI	355	1.72	2.25	1.40	-1.90	12.30	2.46	7.47
us_CPI	355	2.53	1.64	2.30	-2.00	9.00	1.22	3.26
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

Table 2 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 1.72 compared to 0.21 as defined by their respective consumer price indices (CPI). These values seem reasonable with the inflation target rate of 2 percent in mind. Furthermore,

the Swedish inflation rate exhibits more variation as indicated by the higher standard deviation at 2.25 compared to 1.64, as well as having a higher degree of asymmetry as indicated by the higher skewness. The Swedish CPI shows relatively high kurtosis at 7.47 and compared to 3.26 for the US, indicating a relatively fat tail for its distribution.

Comparing the interest rates, it can be seen that they are considerably closer to each other with regard to almost all aspects.

note: I was thinking of using PCE and KPIF instead of CPI as that is wahts officially Riksbanken and the FED is targeting, rather than the CPI

Table 2: 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_KPIF	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

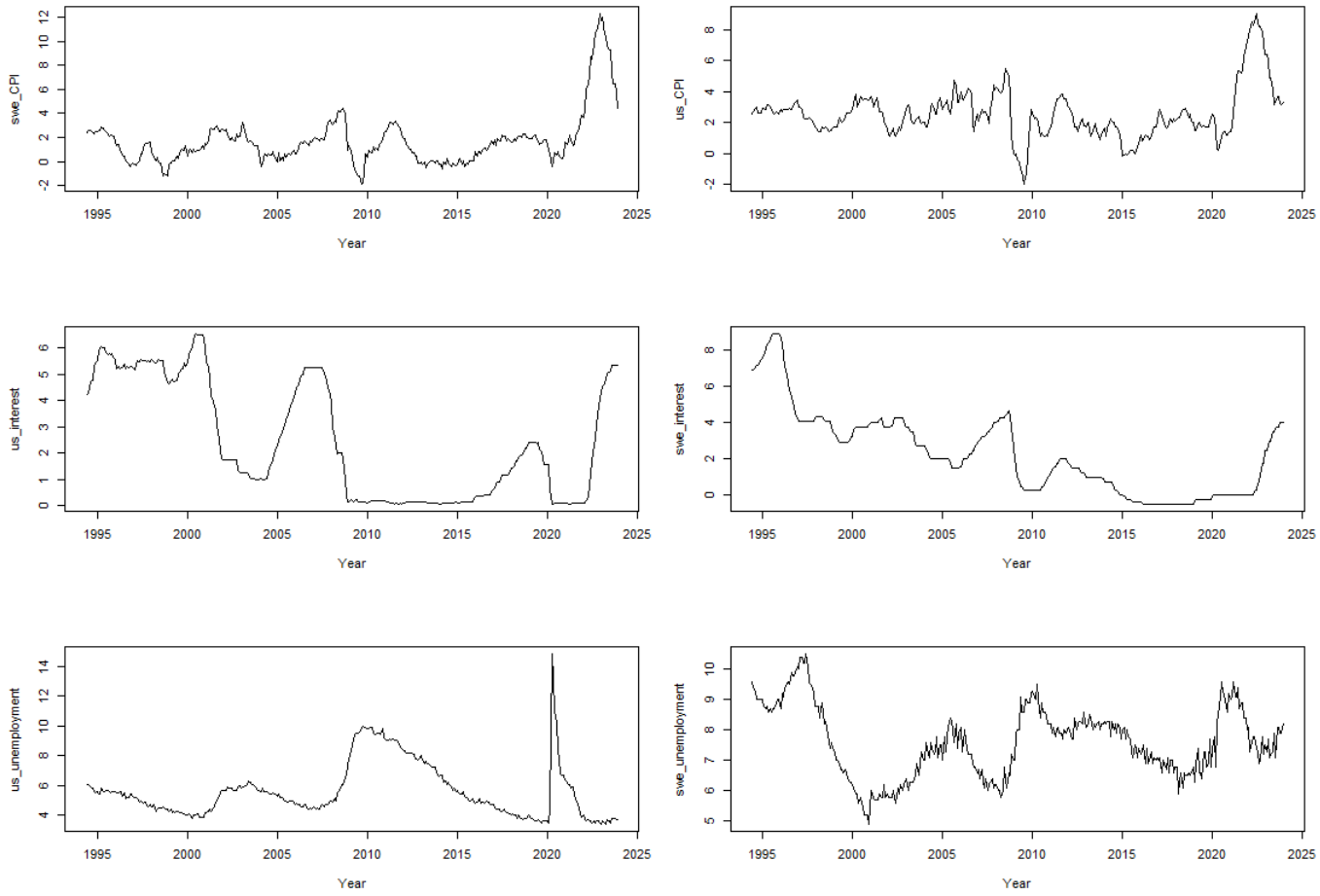

Figure 1: Variables over time

Figure 1 shows the different variables plotted over time. The visualized plots seem consistent with the results in table 2; the inflation rate for both countries seems fairly stable over time, though for the Swedish CPI there is one noticeable outlier for December 2022 at 12.3 percent. The interest rates are 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 due to both countries following the same economic cycle and thus having both of the respective central banks taking similar action.

6 Results

Table 3: ADF tests)

	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

The ADF tests showed that all the variables were nonstaionary except US PCE. However, the first order difference for all variables were stationary. Although us PCE was already stationary, for this was also differentiated to be able to compare the results.

Table 4: ADF tests for the differenced data

	statistic	p.value
us_interest	-4.33	0.01
swe_interest	-5.78	0.01
us_unemployment	-5.58	0.01
swe_unemployment	-4.05	0.01
swe_KPIF	-6.50	0.01
us_PCE	-7.20	0.01

Table 5: Causal null tests for the swedish differenced data

	p	obs.stat	p.val	ci.ll	ci.ul
lag 0	Inf	0.01	0.58	0.00	0.04
lag 1	Inf	0.01	0.30	0.00	0.04
lag 2	Inf	0.02	0.21	0.00	0.04
lag 3	Inf	0.02	0.01	0.00	0.05

Table 6: Causal null tests for the US differenced data

	p	obs.stat	p.val	ci.ll	ci.ul
lag 0	Inf	0.05	0.46	0.00	0.15
lag 1	Inf	0.05	0.51	0.00	0.14
lag 2	Inf	0.04	0.64	0.00	0.15
lag 3	Inf	0.06	0.30	0.00	0.16

Example results of density: Swe data

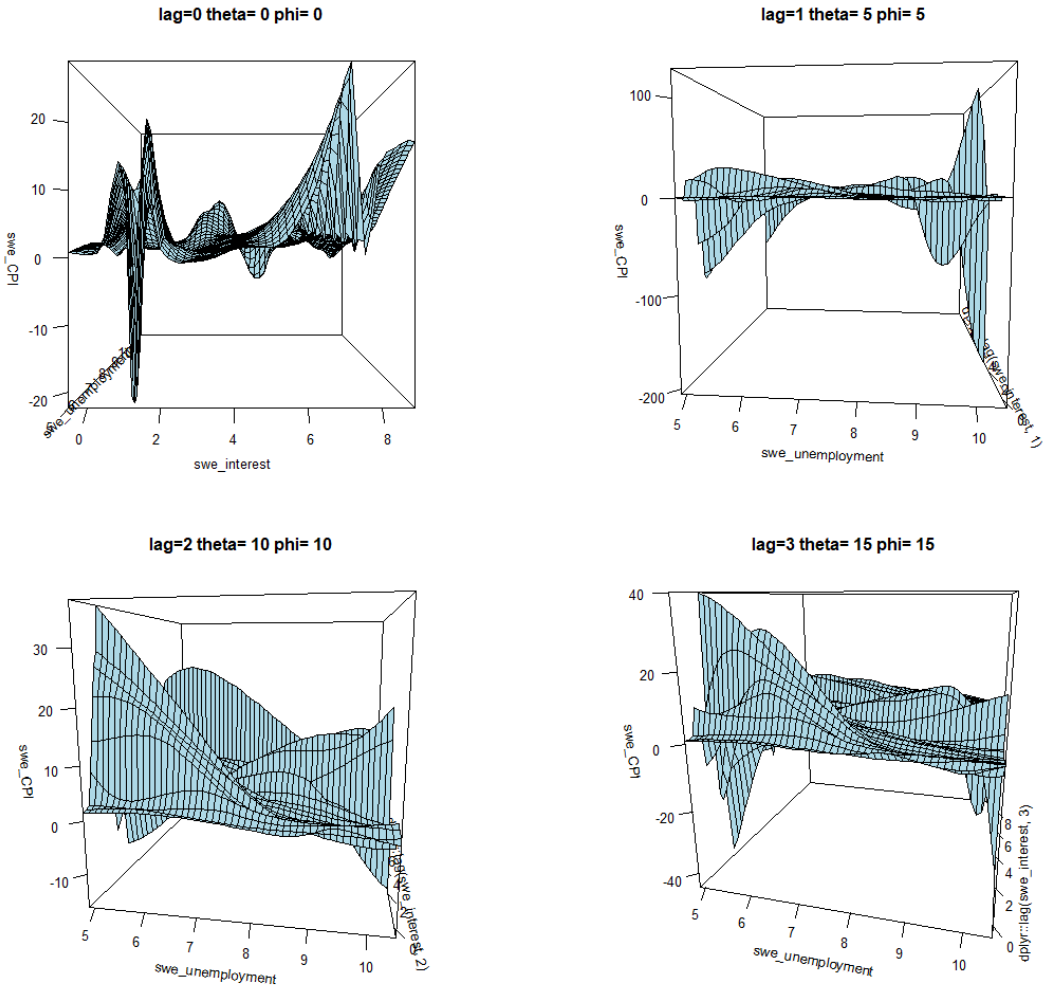

Figure 2: swe1

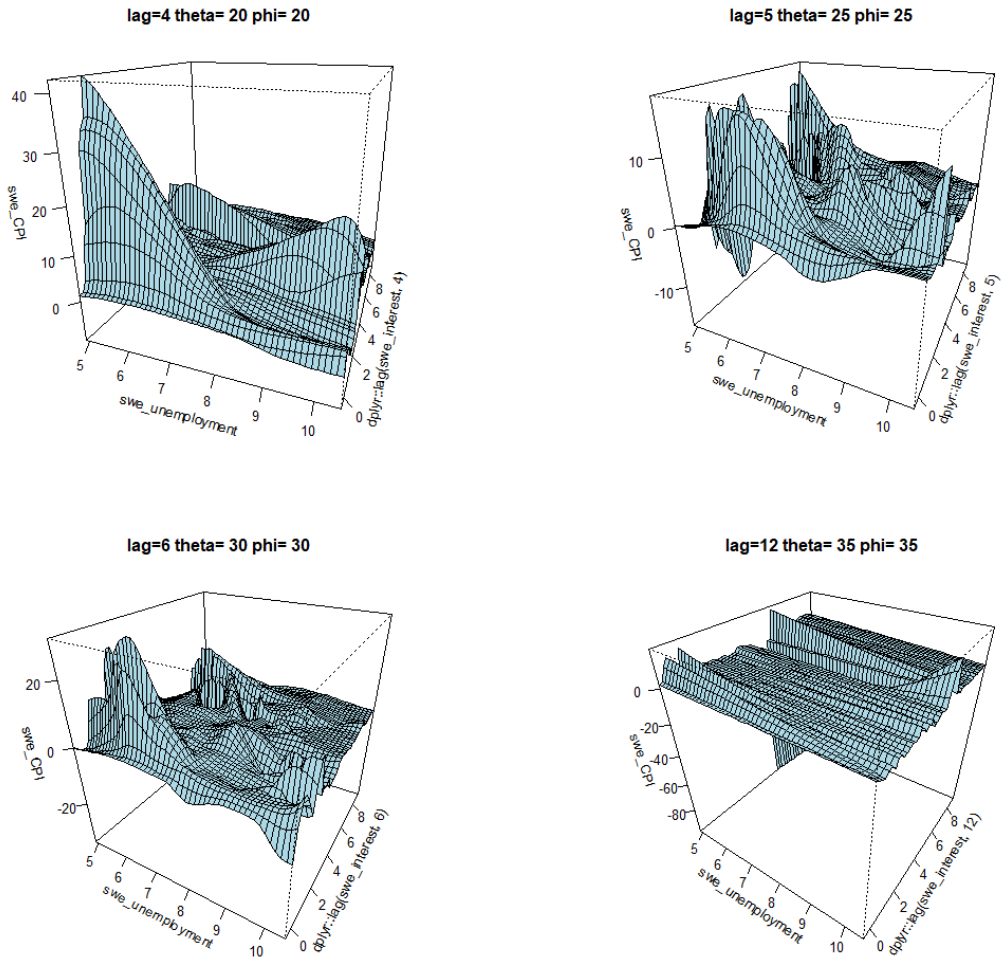

Figure 3: swe2

US results:

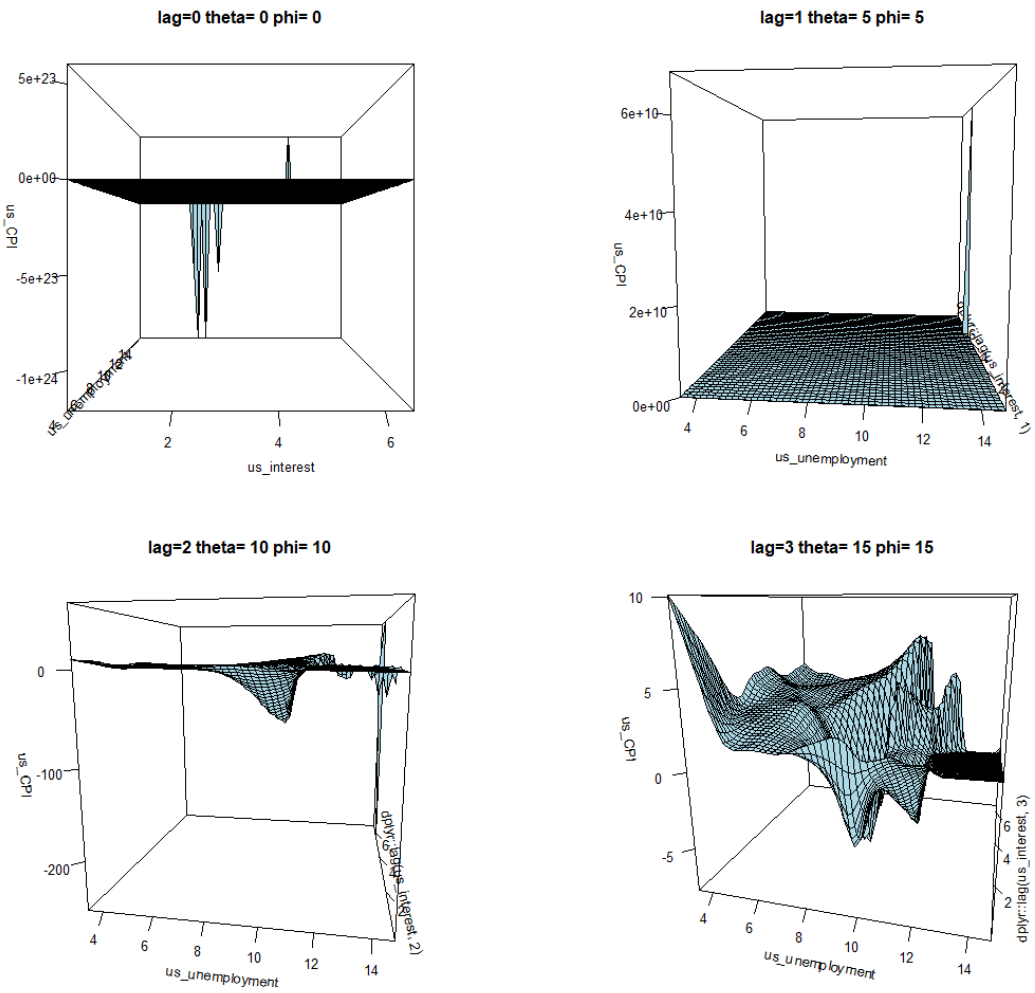
lag=0 theta= 0 phi= 0
5e+23
0e+00
-5e+23
-1e+24
us_CPI
us_unemployment
2
4
6
us_interest
lag=1 theta= 5 phi= 5
6e+10
4e+10
2e+10
0e+00
us_CPI
4
6
8
10
12
14
us_unemployment
dplyr::lag(us_interest, 1)
lag=2 theta= 10 phi= 10
0
-100
-200
us_CPI
4
6
8
10
12
14
us_unemployment
dplyr::lag(us_interest, 2)
lag=3 theta= 15 phi= 15
10
5
0
-5
us_CPI
4
6
8
10
12
14
us_unemployment
dplyr::lag(us_interest, 3)

Figure 4: us1

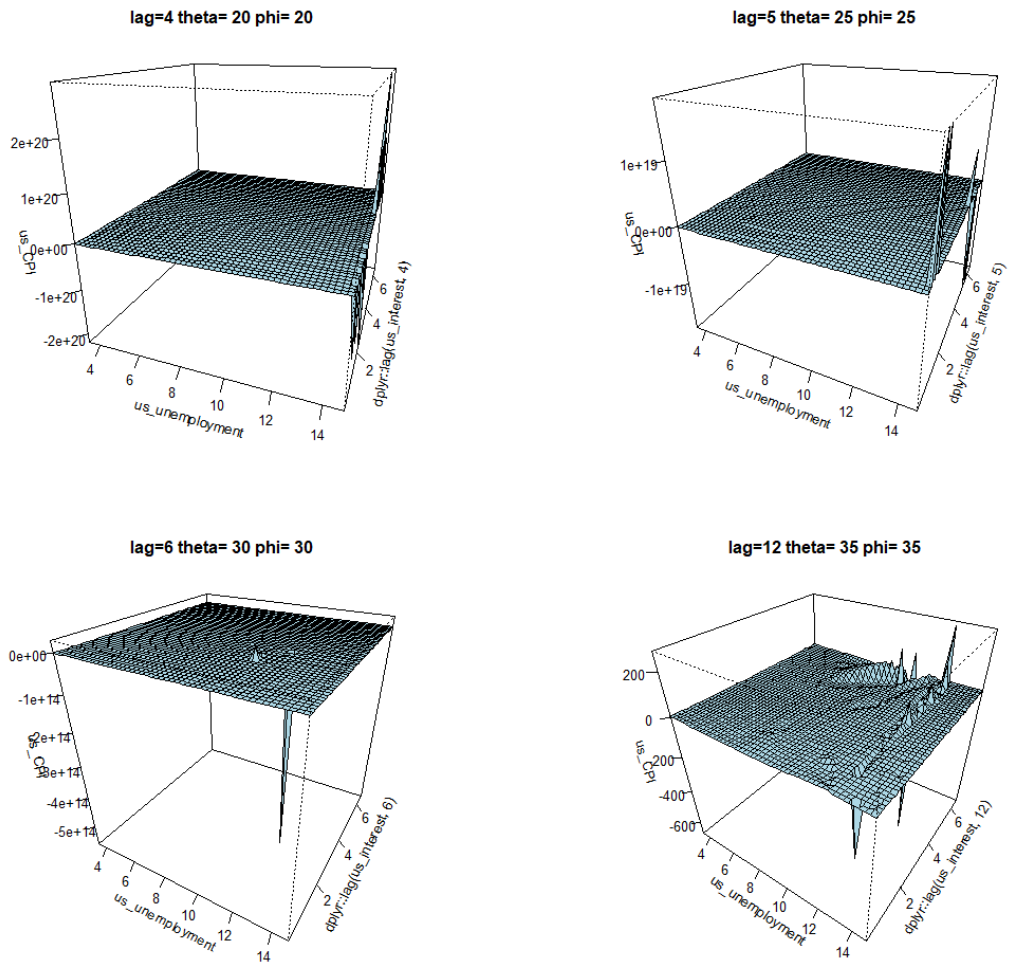
lag=4 theta= 20 phi= 20
lag=5 theta= 25 phi= 25
lag=6 theta= 30 phi= 30
lag=12 theta= 35 phi= 35
us_CPI
us_unemployment
dplyr::lag(us_interest, 4)
dplyr::lag(us_interest, 5)
dplyr::lag(us_interest, 6)
dplyr::lag(us_interest, 12)

Figure 5: us2

References

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. & Kuersteiner, G. M. (2011), 'Causal effects of monetary shocks: Semiparametric conditional independence tests with a multinomial propensity score', *Review of Economics and Statistics* **93**(3), 725–747.

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.

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.

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,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.

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/*

Sweden, S. (2024), 'Consumer price index (kpi)', `https://www.scb.se/hitta-statistik/statistik-efter-amne/priser-och-konsumtion/konsumentprisindex/konsumentprisindex-kpi/`. Accessed: 2024-07-31.

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).

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* .