ESTIMATING THE SWEDISH PHILLIPS RELATIONSHIP IN A MARKOV-SWITCHING VECTOR AUTOREGRESSION

COURSE 5350: THESIS IN ECONOMICS
Stockholm School of Economics

Svante Midander (21717)
Sebastjan Wassermeyer (40445)

Abstract

The Swedish Phillips relationship was recently examined by Svensson (2015), who found that the long-run trade-off is downward-sloping. Hence, there is an unemployment cost of inflation. He argues that this has occurred because inflation expectations are anchored to the inflation target, while average inflation has deviated from the target. This empirical finding has a large implication on the estimation of the Phillips curve. We wish to examine the Phillips relationship in light of the debate on whether surveyed expectations are anchored, by using an econometric method that is robust to whether this assumption holds or not. We study the Swedish Phillips curve in a regime shifting framework using the same data as Svensson (2015) for the period 1997Q4-2011Q4. We estimate a bivariate Markov-switching VAR with inflation and the unemployment rate, with regime-dependent dynamics and heteroskedasticity. We confirm the existence of a non-vertical long-run Phillips curve and conclude that linear estimation of the inflation–unemployment relation seems to be robust with regard to expectations over this period.

Keywords: Markov-switching, Phillips curve, vector autoregression, inflation expectations, non-linear time series models.

JEL classification: C32, C52, D84, E24, E31.

Supervisor: Rickard Sandberg and Johanna Wallenius
Date submitted: January 7, 2015
Date examined: January 14, 2015
Discussant: Julie Rondeau and Linnea Jahnson
Examiner: Kelly Ragan

CONTENTS

1 Introduction **1**

2 The Phillips Relationship **3**
- 2.1 Foundations of the New Keynesian Phillips Curve 5
- 2.2 An Econometric Formulation of the Phillips Curve 7
- 2.3 Recent Empirical Findings of the Swedish Phillips Relationship 8

3 A Regime Shifting Approach to the Phillips Curve **10**
- 3.1 State-Space Representation of the MS-VAR 13

4 Estimation **14**
- 4.1 Data 15
- 4.2 ARMA Representation Based Model Selection 16
- 4.3 Model Selection 17

5 Discussion **23**

6 Conclusion **28**

A Appendix **36**
- A.1 Derivations 36
  - A.1.1 The BLHK-Filter and Smoother 36
  - A.1.2 The EM Algorithm 37
  - A.1.3 Constrained Estimator for MSAH-VAR 38
- A.2 Tables 40
- A.3 Figures 40
- A.4 Matlab 47
  - A.4.1 Simulation 47
  - A.4.2 Loading and preparing data 50
  - A.4.3 Algorithm for starting values 52
  - A.4.4 EM-algorithm 53
  - A.4.5 BLHK 55

List of Figures

1 Average CPI inflation and inflation expectations at different horizons. . . 9
2 CPI inflation and unemployment 16
3 Smoothed probability for regime 1, $\Pr(s_t = 1)$. 24
4 Parametric encompassing of empirical moments 26
5 CPI inflation and Prospera inflation expectations. 40
6 CPI inflation and NIER inflation expectations. 41
7 CPI inflation and NIER and Prospera inflation expectations. 41
8 Empirical autocorrelations 42
9 ACF's of unemployment for different sample periods 42
10 Smoothed probability for regime 1 of all models 43
11 Frequency and log likelihood for different starting values 44
12 Conditional residuals of the MSAH-VAR(2) 45
13 Generalized and Rosenblatt-transformed residuals of the MSAH-VAR(2) . 45
14 Smoothed residuals of the MSAH-VAR(2) 46

Acknowledgement

We would like to thank Rikard Sandberg, Johanna Wallenius, Pär Stockhammar, Pär Österholm, Emma Hutchinson, Emelie Gustafson, Jonatan Samuelsson, Sören Thurell and Martin Rassl for comments and help with the work of this thesis.

1 INTRODUCTION

The economic relationship between inflation and unemployment, widely known as the Phillips curve, is probably one of the most researched areas in economic literature. Phillips (1958) found the first empirical evidence of a clear negative correlation between wage inflation and the unemployment rate in the United Kingdom over the period 1861-1957. This correlation was subsequently interpreted as causal, and suggested a trade-off for policy makers to be made between inflation and unemployment. Over the years, the Phillips curve has been rejected, reborn, and redeveloped in numerous different forms and specifications.

Today, the Phillips curve has a rather different interpretation than it had 50 years ago. The main critique of the original curve was formulated by Friedman (1968) and Phelps (1968), who argued that it was impossible for a policy maker to achieve a different long-run unemployment ratio other than the "natural rate"; which was consistent with the aggregate production potential of the economy. Lucas (1972, 1973) introduced rational expectations to the Phillips curve and argued that agents would incorporate the behavior of a policy maker, trying to exploit the relationship, in their formation of expectations of future inflation. The role of expectations is cardinal in the subsequent literature and a key feature in both the Neoclassical and New Keynesian Phillips curve (NKPC), where the latter has become the mainstream component of policy analysis and will provide the theoretical foundation for our analysis.

Naturally, expectations also have a critical role in the empirical investigations of the Phillips curve. As expectations are unobservable and hard to measure, the main econometric effort has been to develop methods for estimating the curve in the absence of expectations. Moreover, as the theoretical foundations of the Phillips curve suggest a long-run Phillips curve that would be vertical, and as a consequence also impossible to estimate, emphasis has been put on estimating the short-run Phillips curve.[1] In the short-run, the trade-off can still be exploited by the policy maker if the changes come as a surprise to the rational agent.

Recent empirical findings by Svensson (2015) have indicated that the long-run Phillips curve in Sweden and Canada has been sloping downward since 1997, and in the US since 2000. Svensson derives the long-run Phillips curve from his short-run estimates and finds that the long-run curve is non-vertical and has a negative slope of approximately 0.75, 0.42 and 0.23 for Sweden, Canada and the US respectively. Svensson's approach rests on the assumption of constant inflation expectations. In defense of this assumption, Svensson points to Sweden and the US In Sweden expectations seem to have been anchored

---

[1] One important exception is the literature following King and Watson (1994), who suggested that the long-run Phillips curve in fact could be tested when inflation has a unit root. That is, there are low-frequency, permanent changes in the inflation rate, such as in the U.S. inflation data of the 1970s. When inflation is a random walk, the data contain the relevant 'experiment' and the econometrician can estimate the response of the unemployment rate to a permanent change in inflation. This exercise typically fails to reject the hypothesis of a vertical long-run Phillips curve covering several countries and relevant time periods (see e.g. King and Watson (1994), Rudebusch and Svensson (1999), Benati (2012)).

to the inflation target,[2] and as noted by Fuhrer (2011), the US has maintained a stable inflation rate with an implicit inflation target of 2 percent—commonly perceived to have been followed by the Federal Reserve since the early 2000. In the Swedish case, Svensson claims that expectations are non-rational and stable at the inflation target, even though average inflation clearly has been undershooting the target over the period 1997-2012. This would imply a cost to the economy corresponding to the increase in unemployment caused by undershooting the target inflation. This cost has, in contrast to Sweden, not been realized in Canada or the US, since inflation on average has been on target.

The usefulness of the survey on which Svensson bases his argument of constant inflation expectations in Sweden is however subject to dispute.[3] The problem, as stated by Jonsson and Österholm (2012), is that it is impossible to assess whether expectations are irrational and constant at the target, or just measured incorrectly. It would therefore be attractive to reevaluate the existence of a non-vertical Phillips relationship in Sweden by relaxing this assumption.

One approach for doing this would be to use an econometric method that allows the relationship to change over the sample period due to different economic conditions, such as changing expectations. This could be done with the Markov-switching model of Hamilton (1989). The Markov-switching (MS) model assumes that an underlying, unobserved regime process affects the data-generating process. This allows the parameters of the econometric model to change according to the regimes such that the Phillips relationship can be estimated with time-varying coefficients.

The MS model has, since the seminal work of Hamilton (1989) largely been applied to the business cycle. Usage of the model outside of this area has been relatively rare. This has occurred despite it being an appealing framework for studying any time-varying economic relationship as the regimes in the MS model, in contrast to many other non-linear time series models, are treated as exogenous. The reason for using this method would hence be to see whether the model picks up shifts and changes in expectations that from a theoretical perspective clearly would determine the data generating process, but which are unobservable to the econometrician. If Svensson's findings of a non-vertical Phillips curve are robust using this approach, the assumption must also be eligible.

Studying inflation and the Phillips curve in a Markov-switching framework has previously been done by Nadal-De Simone (2000), Demers (2003), and Pagliacci and Barraez (2010) (the Phillips curve), and by Evans and Wachtel (1993), Ricketts and Rose (1995), Simon (1996), Blix (1999) Demers and Rodríguez (2001) (inflation dynamics). Of the former, Demers (2003) studied the Phillips curve in Canada using two different models; and found that an MS model with three regimes was more efficient than the methodology used by Bai and Perron (1998).[4] Pagliacci and Barraez (2010) estimated a NKPC on data from Venezuela and used it to distinguish between periods when expectations were "backward-looking" compared to when they were "rational".

---

[2] Expectations are anchored in the sense that average inflation expectations equals the target although average inflation falls short on the target.

[3] See e.g. Andersson et al. (2012), Flodén (2012), Andersson and Jonung (2014).

[4] In this approach, an unknown number of structural breaks are occurring at an unknown point in time.

Simon (1996) uses the model to study the inflation process in Australia, which has been characterized by high volatility since the early 1960s, and is therefore difficult to model in a linear framework. He interpreted the differences in the time series as shifts in expectations that were either due to changes in the shocks hitting the system (such as the oil-price shocks of the early 1970s) or because of changes in the monetary regime (as in the early 1990s). Blix (1999) uses an MS-VAR on Swedish data to account for the monetary policy regime shift in 1993, which allows the use of longer time series to improve forecasts and overall precision.[5]

Since our empirical approach captures non-linearities in the data, it relates to the empirical literature on non-linear Phillips curves. Laxton et al. (1995), Debelle and Vickery (1997), Debelle and Laxton (1997), Tambakis (1998), Schaling (1999), Laxton et al. (1999) and Zhang and Semmler (2004) , have studied convex Phillips curves in the US, UK, Australia, Canada and other G7 countries. Eisner (1996) finds evidence consistent with a concave Phillips curve and Filardo (1998) investigates a convex–concave Phillips curve using US data. Regarding Sweden, Eliasson (2001) estimated the short-run Phillips curve using a Smooth Transition Regression (STR) model for the period 1979Q2-1997Q4. She rejected linearity, but since the non-linear form of the model that she used and tested against the null does not have to be specified, the procedure does not propose any particular shape of the curve. This seems to be a general problem in the literature, as it is difficult to assess the exact form of non-linearity when comparing alternative shapes, nonetheless, the time-variation of the parameters are robust (Dupasquier and Ricketts, 1998). This also motivates the use of the MS model as it allows for any type of non-linearity and estimates the particular form of it without the need for *a priori* assumptions. Our study of the Phillips relationship also can be seen as a follow up on Eliasson's study as we estimate the Phillips curve for the period 1997Q4-2011Q4.

The rest of the text is structured as follows: Section 2 is a brief summary of the Phillips curve, its theoretical foundations and empirical findings of the relationship in Sweden. Section 3 then continues to explain the Markov-switching model. Section 4 describes our results followed by a discussion in Section 5 and Section 6 concludes our findings.

2 The Phillips Relationship

The form of the NKPC has evolved out of a continuous debate between the Keynesian and classical schools of economics over the past half-century. The original Phillips curve was first observed by Phillips (1958) as the inverse relationship between unemployment and money wage changes in the United Kingdom 1861-1957, and was later extended

---

[5] Even if a motivation in the literature is the model's ability to handle data issues such as structural breaks, the MS model has only shown small improvements to forecasts compared to linear models. Inflation forecasting has become very problematic (there is a vast literature concerning the difficulties of forecasting using U.S. data, see e.g. Stock and Watson (2007, 2009, 2010)), but an intrinsic deficiency of the model is its inability to predict future regimes (Bessec and Bouabdallah, 2005). We are primarily interested in the use of the model in the area of economic analysis rather than forecasting.

by Samuelson and Solow (1960) to the more common trade-off between inflation and unemployment. In an era that rejected the classical dichotomy and the neutrality of demand-side policies, the neoclassical synthesis combined Keynes' proposal of sticky prices in the short-run, but maintained the classical theory of neutrality in the long-run, thus making room for monetary policy affecting the real economy. Policy makers believed during this time that an optimal policy consisted of choosing the ideal level of inflation and unemployment along a non-shifting Phillips curve.

After the stagflation of the 1970s,[6] the relationship took a step back towards the classical view as the theoretical foundations of the curve were scrutinized by Friedman (1968) and Phelps (1968). They argued in their "Natural Rate Hypothesis" that only unexpected inflation can affect real variables and once policies become expected, the classical dichotomy is restored and aggregate variables return to their potential level. They modeled expectations about inflation in an adaptive manner, in which expectations are adjusted in accordance with recent forecast errors about inflation.

The backward-looking assumption was subject to heavy criticism by Lucas (1972, 1973) popularly known as the *Lucas critique*. He argued that by being entirely backward-looking, agents fail to take into account additional information available at the time, which would not be rational. Adaptive inflation would under a scenario of accelerating inflation, despite being announced in advance, systematically predict the inflation level inaccurately. In essence, the Lucas critique stated that the relation between any two macroeconomic aggregates, results from decisions by individuals considering the economic environment. When the economic environment changes, we can no longer expect the decisions of individuals to remain the same and hence the correlations between macroeconomic aggregates alter. The most important implication for the Phillips curve is that policy makers seeking to exploit the relationship between inflation and unemployment, can not systematically surprise the agents of the economy, which calls for a micro-founded approach when modeling the economy.

With the introduction of King et al. (1988)'s Real Business Cycle (RBC) theory, the sticky price assumption was abandoned and the classical dichotomy was restored, in which monetary policy can only influence the price-level and the Phillips curve becomes a vertical line.

Although RBC models performed surprisingly well, their assumptions about fully flexible prices stood in contrast with reality, Christiano et al. (1998). This sparked the emergence of the New Keynesian school of economics. The new approach combined the important developments of RBC theory, comprising the complexity of micro-founded dynamic stochastic general equilibrium (DSGE) models capturing rational expectations, but simultaneously departed from the assumption of fully flexible prices and the neutrality of monetary policy. In the New Keynesian school the scope of monetary policy is, however, restricted to restoring the economy to its natural flexible price level, done by managing forward looking expectations, featured in the Phillips curve.

In practice, and regardless of the theoretical foundations, the forward-looking com-

---

[6] The case of both high inflation and stagnating economic activity, a situation that would not be possible according to the early Phillips relationship.

ponent is in many econometric specifications proxied by past inflation. But more recent studies have to a larger extent incorporated surveyed inflation expectations as a measure of pure expectations of future inflation. There is, however, a debate over whether or not surveyed inflation expectations can be used with confidence when estimating the Phillips curve, a matter to which we will return later.

2.1 Foundations of the New Keynesian Phillips Curve

The micro-founded New Keynesian economy features identical monopolistic competitive firms facing constraints on their price adjustment ability, often modeled à la Calvo (1983) or Rotemberg (1982). Depending on the chosen price-adjustment mechanism and its inherent parameters, different degrees of price-rigidity can be achieved. Similarly, the degree of monopolistic competition is governed by the elasticity of substitution, ε.[7] Assuming perfectly flexible prices, monopolistically competitive firms reset their prices in each period as a markup over nominal marginal costs. By introducing nominal rigidities, firms maximize expected discounted future payoffs by setting the optimal current price, taking into account the possibility of not being able to reset the price in each period. Assuming Calvo-pricing in which $(1-\theta)$ represents the probability of a firm being able to reset its price, the optimal reset price can be described in a log-linearization as a markup over current and expected future weighted marginal costs

$$p^*_{i,t} = (1-\theta\beta)\sum_{j=0}^{\infty}(\theta\beta)^j E_{i,t}\left(mc^n_{i,t+j,t}\right), \tag{1}$$

where β is the discount factor, $E_{i,t}\left(mc^n_{i,t+j,t}\right)$ is the expected nominal marginal cost in period $t+j$ for firm i faced in period t, when optimally setting price $p^*_{i,t}$. This captures the forward-looking characteristics of the NKPC in determining inflation. If there is a continuum of firms $i \in [0,1]$ optimally setting their prices, using the law of large numbers, the aggregate price level, p_t, will be determined as a convex-combination of the previous period's prices and the average of the optimized prices in the current period:

$$p_t = \theta p_{t-1} + (1-\theta)\int_0^1 p^*_{i,t}di \tag{2}$$

To relate the firm-specific nominal marginal cost to the average marginal cost requires the specification of the production function. Assuming a Cobb-Douglas production function, it can be shown that

$$mc^n_{i,t+j,t} = mc^n_{t+j} + \frac{\alpha\varepsilon}{1-\alpha}\left(p^*_{i,t} - p_{t+j}\right), \tag{3}$$

---

[7] As ε tends to infinity, goods constitute perfect substitutes.

where $mc^n_{i,t+j,t}$ represents the firms specific nominal marginal cost and $mc^n_{t+j,t}$ represents the average nominal marginal cost. Inserting (3) into (1) gives

$$p^*_{i,t} - p_{t-1} = (1-\theta)\sum_{j=0}^{\infty}(\theta\beta)^j\left[\kappa E_{t,t}\left(mc_{t+j}\right) + E_{i,t}\left(\pi_{t+j}\right)\right], \tag{4}$$

with $\kappa = \frac{1-\alpha}{1-\alpha+\alpha\varepsilon} \leqslant 1$, $\pi = p_t - p_{t-1}$, and $mc_t = mc^n_t - p_t$. Substituting (4) into (2) gives the following equation

$$\pi_t = (1-\theta)(1-\theta\beta)\sum_{j=0}^{\infty}(\theta\beta)^j\left[\kappa\hat{E}_{t,t}\left(mc_{t+j}\right) + \hat{E}_{i,t}\left(\pi_{t+j}\right)\right], \tag{5}$$

where $\hat{E}_t = \int\limits_0^1 E_{i,t}di$ is the cross-sectional average expectation operator. Shifting the equation forward by one period and assuming that expectations across firms equal rational expectations $E_{i,t} \equiv E_t$ we arrive at

$$E_t\left(\pi_{t+1}\right) = (1-\theta)(1-\theta\beta)\sum_{j=0}^{\infty}(\theta\beta)^j\left[\kappa E_{t,t}\left(mc_{t+j+1}\right) + E_{i,t}\left(\pi_{t+j+1}\right).\right] \tag{6}$$

Combining (6) with (5) gives a difference equation expression for inflation:

$$\pi_t = \beta E_t\left(\pi_{t+1}\right) + \frac{(1-\theta)(1-\theta\beta)}{\theta}\kappa mc_t \tag{7}$$

By letting x_t represent a proxy for real marginal cost and u_t denote an unobservable disturbance term, the model can be rewritten as a forward-looking NKPC:

$$\pi_t = \beta E_t\left(\pi_{t+1}\right) + \lambda x_t + \varepsilon_t \tag{8}$$

The causal link between interest rates, marginal costs and inflation goes through consumption.[8] When consumers see nominal interest rates go down, the real interest rate also goes down and an inter-temporal substitution of consumption towards more consumption today, takes place through the Euler equation,

$$c_t = E_t\left\{c_{t+1}\right\} - \frac{1}{\sigma}\left(i_t - E_t\left\{\pi_{t+1}\right\} - \rho\right)$$

where σ and ρ are parameters from the consumers' utility function. This triggers an increase in demand and as firms see production quantities go up, marginal costs also go up due to decreasing marginal productivity. In response, firms increase prices, but in a step-wise fashion. Thus, the decrease in the real interest rate is maintained for a few periods, indicating that monetary policy has an impact on the real economy.[9]

---

[8] And possibly investment depending on the model, which is omitted in this example for simplicity.

[9] For a more thorough introduction to the benchmark New Keynesian model, see e.g. Gali (2008), Chapter 3.

2.2 AN ECONOMETRIC FORMULATION OF THE PHILLIPS CURVE

As mentioned, one main issue with the purely forward-looking NKPC (8) was its inability to mimic U.S. inflation data and more specifically, its inherent sluggishness. Different solutions were suggested, e.g. Gali and Gertler (1999)'s "hybrid" NKPC, with additional lag terms explaining inflation, assuming that some firms set their prices according to a backward-looking rule of thumb instead of the rational expectation of future inflation. Using this combination of the old and new Phillips curve, econometricians were often more successful in describing inflation persistence. A general formulation of these types of Phillips curves can be expressed as

$$\gamma(L)\pi_t = \gamma_f E_t\left(\pi_{t+1}\right) + \lambda x_t + \eta' w_t + \varepsilon_t \tag{9}$$

where $\gamma(L) = 1 - \gamma_1 L - \gamma_2 L^2 - \cdots - \gamma_l L^l$ is a lag operator and w_t denotes exogenous control variables. Setting $\eta = 0$, this formulation nests most of the different Phillips curves found in the literature, such as (8) with $\gamma(L) = 1$, the old Phillips curve with $\gamma_f = 0$, and the hybrid with $\gamma(L) = 1 - \gamma_b L$ and $\gamma_b + \gamma_f = 1$.

Before Lucas (1972), the standard practice when estimating the old Phillips curve was to check whether the lags of inflation summed to one, $\gamma(1) = \gamma_1 + \gamma_2 + \cdots + \gamma_l = 1$, inspired by Solow (1968) and Tobin (1968). A non-rejection of $\gamma(1)$ translates to a Phillips curve with an infinite slope and hence a non-existing long-run relationship. In a specification similar to (9), where $\gamma_f \neq 0$, such procedure is seldom rejected by data.

There are two approaches for estimating the Phillips curve in the literature. Either you estimate (9) as a single equation, called limited-information methods, or you estimate (9) as a part of a larger system, full-information methods, in which (9) is one out of several structural equations explaining the economy. This latter approach typically consists of more or less complex dynamic stochastic general equilibrium (DSGE) models and has the advantage of improving the precision of the estimates by imposing a theoretical structure on the variables. The downside is that potential misspecification in any of the other equations of the theoretical model also induces the risk of bias on the Phillips curve estimates.

The limited-information methods are more frequently used but are entailed with some serious econometric problems according to Mavroeidis et al. (2014). Except for the obvious one that x_t could be correlated with the error term ε_t, the crucial problem arise because of weak identification due to that $E_t\pi_{t+1}$ is unobserved and endogenous. Mavroeidis et al. (2014) focus on this latter point, and summarize three different limited-information approaches to estimate (9) when $E_t\pi_{t+1}$ is absent from the equation: (i) use realizations instead of $E_t\pi_{t+1}$ and estimate the equation using a generalized instrumental variable (GIV) approach, (ii) derive $E_t\pi_{t+1}$ from a reduced-form VAR, (iii) proxy for $E_t\pi_{t+1}$ with measured inflation expectations from surveys.

The simplest case of (i) replaces $E_t\pi_{t+1}$ with the realization π_{t+1}, but there are variants of exclusion restrictions on any of the lags $\gamma(L)$ in (9) that are then used as instruments instead. The VAR approach (ii) basically forms an estimate of $E_t\pi_{t+1}$ from the prediction of a reduced-form VAR of a vector consisting of, among other variables, π_t and x_t. This estimate can then be inserted into (9). The survey approach (iii) simply uses some measure of inflation expectations, such as inflation forecasts in (9). One important implication is that the microfoundations of (9) are constructed under the assumption that firms' expectations are rational, and when surveyed inflation expectations are used, e.g. forecasts, an additional error term $\epsilon_t = E_t\pi_{t+1} - \pi^s_{t+1|t}$, must be included. If this error term ϵ_t is correlated with either the forcing variable x_t or inflation π_t, the estimates will be biased.

All approaches (i)-(iii) then typically proceed to estimate (9) using Generalized Method of Moments (GMM). The identification of the parameters of (9) requires that the GMM moment conditions are satisfied, but useful estimators of γ and λ can also be obtained under weak identification when the moment conditions are nearly satisfied. This ends up often being the case when estimating the Phillips curve in (9). Mavroeidis et al. (2014) show that when identification is weak, the GMM estimates of (9) can be biased in different directions and extremely sensitive to different specifications. This may explain some of the conflicting results found in the literature and calls for a general need to develop new methods and collect new data sets.

2.3 Recent Empirical Findings of the Swedish Phillips Relationship

The Phillips relationship in Sweden, Canada and the US has recently been investigated by Svensson (2015), amongst others.[10] Svensson (2015) examines surveyed inflation expectations, collected for the Swedish Riksbank, and concludes that they seem to be anchored to the official inflation target of the Riksbank, at 2 percent, rather than following actual inflation. Expectations are anchored in the sense that average inflation expectations equals the target although average inflation has been lower over the period, see Figure 1.[11] Anchored inflation expectations are not uncommon, especially not in countries where the central bank has committed to an inflation target based on inflation forecasts and has transparently reported its policy decisions (Gürkaynak et al., 2007, 2010). Beechey et al. (2011) for example, find that long-run inflation expectations are less anchored in the US compared to the Euro area, the latter having had an explicit inflation target of 2 percent since 2003.[12] The point is that average inflation in Sweden over the period has fallen short compared to the target level,[13] and that expectations, if they were rational, should have adjusted down towards the actual inflation rate.

---

[10] Svensson (2015)'s findings has spurred a debate amongst economists (see e.g. Assarsson (2011), Andersson et al. (2012), Flodén (2012)), as well as among monetary policy makers and the national parliament.

[11] Svensson (2015) uses measured inflation expectations from TNS Sifo Prospera, that has been surveying a panel of Swedish labor market parties, purchasing managers and money market players since 1996, with the aim of mapping inflation expectations (Prospera, 2014).

[12] The Federal Reserve has had an official inflation target of 2 percent since 2012.

[13] A couple of studies have calculated average inflation over the recent period. Assarsson (2011) studies the period 1995-2010 and finds that revised CPI inflation has been on average 1.3 percent. Svensson (2015) calculate average real-time inflation to be 1.4 percent and average revised CPI inflation to be 1.3 percent over the period 1997-2011 respectively. Andersson et al. (2012) calculates average real-time CPI inflation to be 1.5 percent during the period 1995-2011.

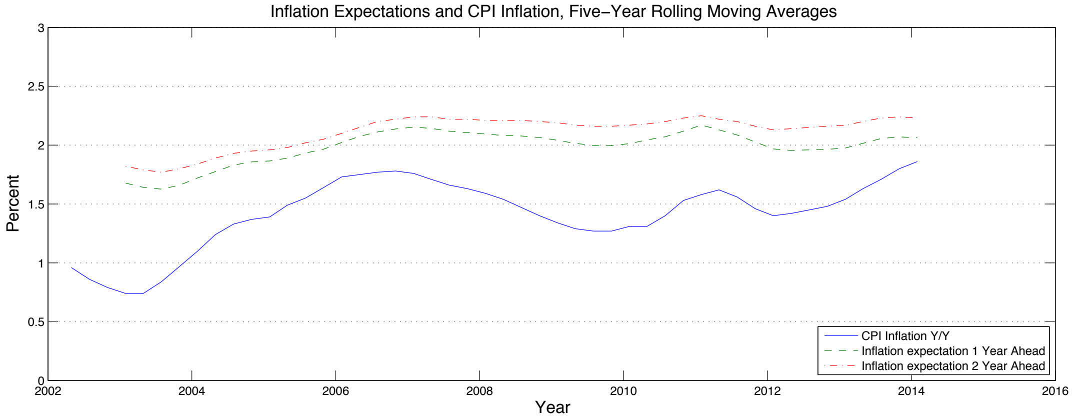


Figure 1: Average CPI inflation and inflation expectations at different horizons.

An explanation of the observed anchoring of expectations to the inflation target rather than actual inflation, could be that the expectation formation process is biased towards the communication of the inflation target, rather than actual inflation (Svensson, 2015). The expectations could be *near-rational*, as in Akerlof et al. (2000), where agents neglect deviations from an inflation target when deviations are small enough. The only difference in this case is that this applies to an inflation target of 2 percent instead of the zero-inflation application in Akerlof et al. (2000). Jonsson and Österholm (2012) also confirm that inflation expectations in the survey are irrational,[14] which they conclude can either be because the expectations' formation process being suboptimal, or because the recorded expectations are not the respondent's true inflation expectations. They point out that there is at the moment no method to establish whether expectations are irrational or simply mis-measured, if not both.

Svensson (2015) further argues that with respect to the wage setting process in Sweden, the inflation target has become more important than anything else for determining future wage levels. The wage formation process in Sweden is dominated by annual central wage negotiations between labor unions and employer organizations.[15] In these negotiations, the inflation target of 2 percent is used rather than actual or forecasted inflation (Morin, 2009). Hence, evidence suggests an inflation/unemployment relationship where inflation expectations have no or little importance in determining inflation.

Nonetheless, Svensson (2015) estimates the Phillips curve including surveyed expectations for Sweden's inflation rate over 1997Q4-2011Q4 using CPI inflation and the

---

[14] To be precise, the inflation expectations in the survey are biased and inefficient, and evaluating these expectations' forecast ability shows that they are worse predictors of future inflation compared to professional forecasting institutions and simple autoregressive models.

[15] Negotiations between the trade unions and employer organizations in the manufacturing industry have set the standard for the entire Swedish wage-setting process since the establishment of the Industrial Cooperation and Negotiation Agreement (ICNA) in 1997. (Konjunkturinstitutet, 2004). Thus, the wage formation in Sweden is based on highly centralized coordination, which lead to a very similar development in wages in the whole business sector over 1998-2011 (Konjunkturinstitutet, 2012).

unemployment rate as a measure of the forcing variable x_t. He uses two alternative specifications for inflation expectations, corresponding to a NKPC and a New Classical Phillips curve.[16] The model with the best fit includes the first difference and a one quarter lag of the unemployment rate. The coefficient for surveyed expectations is insignificant and since the rest of the estimates are robust to whether expectations are included or not, they are discarded from the equation. Svensson then derives the long-run Phillips curve from his short-run estimates by taking the unconditional expectation of the variables and rejects the Solow-Tobin test that $\gamma(1) = 1$. He concludes that the long-run Phillips curve seems to be non-vertical and downward-sloping, with a slope of about 0.75.

An alternate reason, pointed out by Svensson, for surveyed expectations to be insignificant, is a lack of variation in current inflation.[17] This is typically observed in situations where monetary policy is very effective at anchoring short-term expectations. In this regard, effective monetary policy is bad for the econometric environment (Mavroeidis, 2010, Cochrane, 2011). However, there have been other surveys on inflation expectations that show that they vary more. Flodén (2012) compares the survey of expectations commissioned by the Riksbank, with those of the National Institute of Economic Research (*Konjunkturinstitutet*),[18] and finds that the Riksbank's survey and the NIER's household survey show stable expectations close to the target, while the business survey does not, see Appendix Figure 7. Svensson (2015), however argues that the panel surveyed on the commission of the Riksbank is a better predictor of the expectations that are important for the wage formation in Sweden.

All in all, this debate do evoke some skepticism regarding the assumption of anchored expectations and defines the rational for a re-investigating of the long-run Phillips curve.

3 A Regime Shifting Approach to the Phillips Curve

At the moment, it seems like the empirical literature based on limited-information methods has reached an upper bound of how robust inference one can make on the Phillips curve. Nor do more complex full-information systems such as a DSGE model solve the problem of unobserved expectations. Therefore, we instead propose the use of an econometric approach that intuitively would handle unobservable elements that are important in economic relationships.

The strategy is simply to use a Markov-switching model that can manage unobserved

---

[16] Using 1-year-ahead inflation expectations, the NKPC is specified with expectations lagged one quarter back to avoid simultaneity problems, $E_{t-1}\pi_{t+4}$, while the New Classical Phillips curve has expectations lagged four quarters back, $E_{t-4}\pi_{t+4}$.

[17] Surveyed inflation expectations do actually vary rather much compared to the averages shown in Figure 1, see Appendix Figure 5.

[18] The NIER's *Economic Tendency Survey* consists of two parts, one business and one household survey, see Appendix Figure 6, and aims at compiling consumers and businesses views on the economy. The business tendency survey consists of a panel of 6,000 firms in the business sector, including all Swedish firms with more than 100 employees and a random sample of smaller sized firms. The consumer tendency survey takes a random sample of 1,500 households between 16 and 84 years old.

changes in the economic conditions implied by changes in expectations. This is not a novel idea: the MS model has been used in the purpose of inflation and the Phillips curve under periods of changes in the structure of the economy and monetary policy regimes, just because these have had large impacts on the underlying formation of expectations. However, the MS model has in comparison been very underutilized outside of its first obvious application on business cycles.

The MS model was first introduced by Goldfeld and Quandt (1973) and later Lindgren (1978) who built on Baum et al. (1970) and became popularized in the time series literature by Hamilton (1989). In the MS model, the data generating process consists of two components: (i) the Gaussian autoregressive model as the conditional data generating process and (ii) the regime shift function that is governed by a Markovian process of order one. The model uses an iterative algorithm for determining the regimes given an initial estimate of the parameters, and then estimates the parameters given these regimes. The model does no attempt to explain the periodization of the regimes or why they shift.

Since we are interested in the relationship between inflation and unemployment, we use a Markov-switching Vector Autoregressive (MS-VAR) model, which is a multivariate extension of Hamilton's original univariate model, and was formalized into a general econometric framework by Krolzig (1997). Following Krolzig (1997), the parameters θ of the VAR process will be time-invariant, conditional on an unobservable regime, denoted s_t, and letting $s_t \in \{1, 2, \ldots, M\}$. The conditional probability density function of a time series is then

$$p(y_t|Y_{t-1}, s_t) = \begin{cases} f(y_t|Y_{t-1}, \theta_1) & \text{if } s_t = 1 \\ \vdots & \\ f(y_t|Y_{t-1}, \theta_M) & \text{if } s_t = M \end{cases} \tag{10}$$

where $y_t = (y_{1t}, \ldots, y_{Kt})'$ is a K-dimensional observed time series vector, Y_{t-1} are the observations $\{y_{t-j}\}_{j=1}^{\infty}$, and θ_m are the parameters in regime m. Given regime s_t, the series is generated by a vector autoregressive process of order p, which in its general form can be written as

$$y_t = v(s_t) + \sum_{j=1}^{p} A_j(s_t) y_{t-j} + e_t, \quad e_t \sim NID(\mathbf{0}, \Sigma(s_t)), \tag{11}$$

where, again, y_t is a K-dimensional time series and e_t is assumed to be a Gaussian white noise process with mean zero and variance-covariance given by $\Sigma(s_t)$.

The intercept $v(s_t)$, the autoregressive parameters $A_1(s_t), \ldots, A_p(s_t)$, and the variance-covariance matrix $\Sigma(s_t)$ do all depend on s_t. The following example illustrates how the intercepts change according to the prevailing regime:

$$v(s_t) = \begin{cases} v_1 & \text{if } s_t = 1 \\ \vdots & \\ v_M & \text{if } s_t = M \end{cases}$$

It is possible to allow for different parameter-shifts and subsequently, the model can be specified according to the regime-dependent parameters; I (shifting intercept), A (shifting autoregressive paramters), and H (shifting heteroskedasticity).[19]

The regime s_t is generated by a discrete-regime, homogeneous first-order Markov process, with a finite number of regimes, defined by the transition probabilities

$$p_{ij} = Pr(s_{t+1} = j | s_t = i), \quad \sum_{j=1}^{M} p_{ij} \quad \forall i, j \in \{1, \ldots, M\}. \tag{12}$$

Each transition probability p_{ij}, describes the probability that the Markov chain moves from regime i at time $t-1$ to regime j at time t. Since the Markov chain is of order one, the current regime s_t only depends on the previous regime s_{t-1} (called the *Markov property*),

$$Pr(s_t | \{s_{t-j}\}_{j=1}^{\infty}, \{y_{t-j}\}_{j=1}^{\infty}) = Pr(s_t | s_{t-j}; \rho), \tag{13}$$

where ρ denotes the parameters of the regime generating process. If we consider only two regimes, $s_t \in \{1, 2\}$, the notation gets simpler and we can write the transition probabilities as:

$$\begin{gathered} Pr(s_t = 1 | s_{t-j} = 1) = p_{11} \\ Pr(s_t = 2 | s_{t-j} = 1) = p_{12} \\ Pr(s_t = 1 | s_{t-j} = 2) = p_{21} \\ Pr(s_t = 2 | s_{t-j} = 2) = p_{22} \end{gathered}$$

The probabilities can then be arranged into a matrix $\mathbf{P}$:

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \tag{14}$$

For the transition probabilities to be properly defined, they have to be non-negative and sum to one over each row of $\mathbf{P}$, i.e. for a two-regime Markov chain, $p_{11} + p_{12} = 1$, $p_{21} + p_{22} = 1$, which allows us to write $\mathbf{P}$ as

$$\mathbf{P} = \begin{bmatrix} p_{11} & 1 - p_{11} \\ 1 - p_{22} & p_{22} \end{bmatrix}. \tag{15}$$

In addition, it is assumed that the Markov chain is *ergodic* and *irreductible*,[20] which guarantees stationarity of the MS model.

---

[19] Hamilton (1989)'s model was in fact mean shifting, so that the mean $\mu = (\mathbf{I}_K - \sum_{j=1}^{p} A_j)^{-1} v$ would make a one-time jump after a change in the regime. This turns out to complicate the model unnecessarily, and will not be considered in this study.

[20] A Markov chain is *irreductible* when its state space contains a single communicating class. That is, all regimes are accessible and communicate with each other and there are no regimes outside the set of communicating regimes, i.e. the probabilities of moving from one regime to another are all non-zero. For a Markov process to be *ergodic*, exactly one of the eigenvalues of $\mathbf{P}$ has to be unity and all others must be inside the unit circle. When this holds, a stationary probability distribution of the regimes exists.

Further, information about the realization of the regimes are collected in the regime vector ξ_t,

$$\xi_t = \left[\begin{array}{c} I(s_t = 1) \\ I(s_t = 2) \end{array} \right] \tag{16}$$

where $I(\cdot)$ is an indicator variable,

$$I(s_t = m) = \begin{cases} 1 & \text{if } s_t = m \\ 0 & \text{otherwise.} \end{cases} \tag{17}$$

The regime vector ξ_t thus represents the unobserved regime of the model. The expectation of ξ_t, $\mathsf{E}(\xi_t)$, denotes the probability distribution of s_t and has the following properties:

$$\mathsf{E}[\xi_t] = \left[\begin{array}{c} \Pr(s_t = 1) \\ \Pr(s_t = 2) \end{array} \right] = \left[\begin{array}{c} \Pr(\xi_t = \iota_1) \\ \Pr(\xi_t = \iota_2) \end{array} \right] \tag{18}$$

where ι_m is the m-th column of a $(M \times M)$-dimensional identity matrix $\mathbf{I}_M$, where $M = 2$, i.e. the number of different regimes. The unobserved regimes can be derived from the data using a special filter and smoother algorithm.[21]

3.1 State-Space Representation of the MS-VAR

In order to work further with a MS-VAR model as in (11), it has proven useful to write in a *state-space* form, as is often done for studies of time series with unobserved states. [22] If we let the time series vector be $y_t = (\pi_t, u_t)'$, where π_t is inflation in period t, and u_t is the corresponding unemployment rate, the model in (11) can be written in state-space form as

$$\begin{aligned} y_t &= X_t \mathbf{B} \xi_t + e_t, \qquad e_t \sim NID(\mathbf{0}, \mathbf{\Sigma}) \\ \xi_{t+1} &= \mathbf{F} \xi_t + z_{t+1}, \quad\ z_{t+1} \sim MDS \end{aligned} \tag{19}$$

where X_t is the system inputs $X_t = \bar{x}_t' \otimes \mathbf{I}_K$, where $\bar{x}_t' = (1, \mathbf{y}_{t-1}') = (1, y_{t-1}', \ldots, y_{t-p}')$, $\mathbf{B}$ is the matrix of coefficients $\mathbf{B} = (\beta_1, \ldots, \beta_M)$, β_m is the coefficient vector in regime m, $\beta_m = (v_m', \alpha_{m,t-1}', \ldots, \alpha_{m,t-p}')'$, and $\mathbf{F}$ is the transposed matrix of transition probabilities, $\mathbf{F} = \mathbf{P}'$. The mean innovation process $\{z_t\}$ is a martingale difference series and e_t is Gaussian as in (11).[23]

Hence, the component

$$\pi_t = (\bar{x}_t' \otimes \iota_1) \mathbf{B} \xi_t + e_{t,\pi}$$

---

[21] The filter and smoothing algorithm was developed by Baum et al. (1970), Lindgren (1978), Hamilton (1988, 1989, 1994a) and Kim, hence called the *Baum–Lindgren–Hamilton–Kim (BLHK) filter* and *smoother*. The algorithm recursively reconstructs the path of the regime, $\{\xi_t\}_{t=1}^T$, and derives three different regime probabilities; the *predicted*, the *filtered*, and the *smoothed* regime probabilities, which build on three different information sets (see Appendix, Section A.1.1 for details).

[22] A model written in state-space form will typically consist of two sets of equations: The *measurement* equation, which describes the relationship between the observed time series and the unobserved state, and the *transition* equation which describes the process of state.

[23] A series is a martingale difference series if its expectation, conditional on past values, is null, hence the mean of $\{z_t\}$, $\mathsf{E}[z_t] = \mathsf{E}[z_t | \{\xi_{t-j}\}_{j=1}^{\infty}] = 0$.

of the first equation in (19), where ι_1 is the first row of a $(K \times K)$-dimensional identity matrix $\mathbf{I}_K$, corresponds to the Phillips curve in (9) with Markov-switching, where inflation expectations $E_t\pi_{t+1}$ and exogenous variables w_t are absent from the equation.

The first equation in (19) is the *measurement* equation and the second equation is the *transition* equation. Given the definitions above, the measurement equation can also be written as

$$y_t = (\xi_t' \otimes X_t)\text{vec}(\mathbf{B}) + e_t = \xi_{1t}X_t\beta_1 + \cdots + \xi_{Mt}X_t\beta_M + e_t, \tag{20}$$

where ξ_{mt} functions as a dummy variable for being in regime m.

The transition equation could be interpreted as a representation of the expectation formation process, and can be written as a VAR(1) process, as in Hamilton (1994b),

$$\xi_{t+1} = \mathbf{F}\xi_t + z_{t+1}, \quad z_{t+1} \equiv \xi_{t+1} - \mathsf{E}[\xi_{t+1}|\xi_t], \tag{21}$$

where the second equality follows as the innovation process is a martingale difference series. With this interpretation, expectations evolves according to the transition matrix $\mathbf{F}$. This is comparable to the interpretation of a time-varying Phillips curve by Nadal-De Simone (2000), where expectations were modelled to evolve according to a random walk. Similarly, (21) changes the economic conditions of the system in (20), as the parameters are allowed to change over time. The number of regimes can be thought of as a representation of the shapes that economy can take as a consequence of changes in expectations. However, the econometric model leaves the causal interpretation of the regime process open.

4 ESTIMATION

The MS-VAR model in (19) can be estimated by Maximum Likelihood (ML). However, since the regime generating process is unobserved, the ML estimation problem is highly nonstandard (Franses and van Dijk, 2000). Therefore, as suggested by Hamilton (1990), the estimation could be done using the *Expectation Maximization* (EM) algorithm developed by Dempster et al. (1977).[24]

As already mentioned, we estimate the Phillips curve as one of the equations in a reduced bivariate VAR system, similar to King and Watson (1994).[25] There is no

---

[24] The iterative algorithm will increase the value of the likelihood function and guarantees that a maximum is reached (see the Appendix Section A.1.2). There is however a risk of getting stuck in a local maxima of the likelihood function, and therefore the estimation should be started from *some* different initial values. The initial values are chosen from a grid over an appropriate parameter space. In our case, the grid typically consists of 10 steps. The possible combinations often turns out to be over 10^{12}, so we randomize a plausible number of these, often 10 000. Evaluation of the behavior of the likelihood function can be done by examining the distribution of the final log likelihood values that each initial value yields. This gives an indication of how complex the likelihood function of a specific model are. See A.4 for a brief overview of the implementation of the estimation procedure in MATLAB.

[25] In a previous version, a univariate Markov-switching autoregressive model with exogenous variables were instead used. This could be more easily compared to the limited-information Phillips curve estimation as the forcing variable is left unexplained. However, we choose to model unemployment endogenously in the system, as there is no clear theoretical reason for why the casual link actually is one-way.

imposed structure on the system and we allow for any number of autoregressive lags and let the data determine the most appropriate and parsimonious specification.

Further, as pointed out by Breunig et al. (2003), many studies that make use of the MS model simply specify, estimate and report a model without being concerned of testing whether the chosen model actually is correct.[26] Since it is easy to reject a linear over a non-linear model, the model selection procedure calls for a transparent and extensive method for choosing the appropriate model.

We have followed a bottom-up strategy proposed by Krolzig (1997) which works in conjunction with the usual Box-Jenkins, general-to-specific approach. The main steps are (i) analysis of the individual variables of the time series vector of interest to determine the lag order and number of regimes in the model and (ii) estimation of a preliminary MS-VAR model and its evaluation against different alternatives. The alternative models are characterized by different Markov-switching parameters (intercepts I, dynamics A, and heteroskedasticity H) that may increase the fit of the data, but since overfitting is a distinct possibility, we will favor the most parsimonious specification.

4.1 Data

We use the same Swedish quarterly data and sample period as Svensson (2015), ranging from 1997Q4–2011Q4, see Figure 2. The data comes from Statistics Sweden (SCB) and we use Svensson's own seasonal adjustment of the time series.[27] The inflation rate is the real-time CPI inflation, measured as the change in CPI over each quarter and annualized to correspond to a yearly change.[28] The unemployment rate is the number of unemployed divided by the labor force for the group of people between the age of 15 to 74. The sample consists of a total of 57 quarters and pre-sample periods to allow for the autoregressive lags.

There are other possible measures of the underlying inflation rate: CPIF that keep interest rate for households mortgage interest payment at a constant rate and CPIX that in addition controls for changes in subsidies and indirect taxes. These has been used as robustness checks in Svensson (2015) as well as CPI inflation at an annual rate, but since they do not change the estimates substantially, they will not be considered here.[29]

---

[26] Strangely, many studies even present the estimates without reporting standard deviations and p-values, and given how important this procedure is within the time-invariant, linear time series framework, we see this as a major deficit in the empirical literature.

[27] The seasonal component has been adjusted using ARIMA-X12 and Eviews additive. The data can be downloaded from http://larseosvensson.se/

[28] Svensson discusses whether it is appropriate to use real-time inflation data, or the revised data that are published later but are corrected for any errors or mismeasurement. The former is available for policy makers, firms, and households and will thus better reflect policy measures and economic outcomes at a given point in time, while the latter probably more accurately correspond to the actual level of inflation in the economy.

[29] CPI inflation at an annual rate introduces serial correlation because of overlapping data, and Svensson (2015) does not use this further his the analysis. However, this seems to be the main point of Andersson and Jonung (2014)'s critique as they rather uses CPI at an annual rate instead of quarterly annualized CPI, which they argue varies to much and therefore cannot be explained by surveyed inflation expectations.

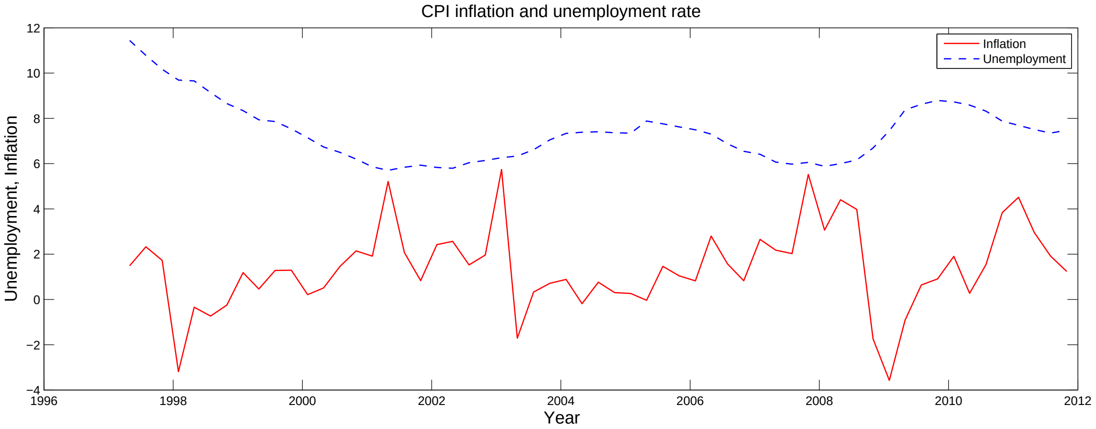


Figure 2: CPI inflation and unemployment

4.2 ARMA Representation Based Model Selection

An ARMA structure in the autocorrelation function of the individual time series components may reveal the characteristics of a data generating MS-AR process. The selection procedure of Krolzig (1997) builds on the theoretical ARMA representations theorems of MS models as a generalization of the results by Poskitt and Chung (1996) and Krolzig (1995). Hence, by studying the autocorrelation function (ACF) and the partial autocorrelation function (PACF), an ARMA(p^*,q^*) model can be identified, which then can be translated to predetermine the (maximum) lag order and number of regimes of an MSI(M)-AR(p) model.[30]

The inflation rate shows ARMA(1,1) characteristics, see the Appendix Figure 8.[31] This would according to the ARMA representation theorems correspond to an MSI(2)-AR(0) model ($M = q^* + 1 = 2$, $p = p^* - q^* = 0$), i.e. a *Hidden Markov-Chain* process.[32] The ACF of the unemployment rate shows a non-decaying pattern, especially when adding observations to the beginning of the sample, see the Appendix Figure 9. An augmented Dickey-Fuller test can not reject the null of a unit root (not reported). This can however be done by the first difference of the series and the ACF and PACF indicates an AR(1) process (see Figure 8).

Even though a unit root motivates the use of an MS model, the ARMA representation procedure is not very clear of how to interpret an integrated process. If we for the moment

---

[30] Under regularity conditions the ARMA representations will show the actual lag order and number of regimes, but the equality is not guaranteed to hold and then only the maximum p and M can be determined.

[31] Spikes in the ACF and PACF at the first lag indicate this, even though it is hard to tell whether the autocorrelation function cuts (an MA(1) process), or tails off, since the correlation coefficient in this case is rather small.

[32] The Hidden Markov process with no dynamics and only a shifting intercept is seldom used model in economics, but more usual in the engineering literature.

overlook the $I(1)$[33] and study the ACF, we can rule out both a pure MA(q) process as the ACF does not cut to zero, as well as a pure AR(p), since the decay is not geometric. The PACF with oscillating decay is in addition characteristic for a positive MA coefficient. An ARMA(2,1) process could be plausible for the unemployment time series. This would correspond to an MSI(2)-AR(1) model ($M = q^* + 1 = 2$, $p = p^* - q^* = 1$).

Given the tentative results of an MSI(2)-AR(0) and an MSI(2)-AR(1) of the two AR components in our VAR, we will consider a two-regime VAR.[34] As the unemployment series is integrated, the ARMA representation selection of one lag may not be appropriate, but we will begin with estimating an MSI(2)-VAR(1) (MSI-VAR(1) from now on), and check whether the model is consistent before considering additional lags.

4.3 MODEL SELECTION

The estimates of the MSI-VAR(1) and MSI-VAR(2) are shown in the first and the second column of Table 1. The regimes are rather persistent and equally distributed over the sample (see Appendix Figure 10). A plot of the various log likelihood found by the search algorithm shows a well-behaved likelihood function with relatively few local maxima (see Appendix Figure 11). Most of the coefficients of the MSI-VAR(1) are significant and have the expected sign. Examining the internal consistency of the MSI-VAR(1) however reveal autocorrelation in the residuals series of the unemployment rate. A Ljung-Box test of the generalized one-step ahead prediction errors and the Rosenblatt-transformed standard normal series (both reported) finds enough evidence to reject the null hypothesis of no autocorrelation.[35]

A likelihood ratio (LR) test of the null hypothesis of only one lag gives the likelihood ratio of 52.95 and can be rejected. Hence, we move over to the MSI-VAR(2) model. The regimes of the MSI-VAR(2) are quite different from the MSI-VAR(1) and much more volatile even though the expected duration of each regime is rather similar to the MSI-VAR(1). The model has slightly more insignificant parameters. The residuals of

---

[33] Eliasson (2001) similarly cannot reject an ADF test of a unit root, but chooses to neglect this as it seems to be due to asymmetries rather than non-stationarity.

[34] There is a general concern with parsimony with regard to the number of regimes in small sample sizes, as the parameters rapidly increase: an MSIAH(2)-VAR(1) with $K = 2$ has 20 parameters, an MSIAH(3)-VAR(1) has 29.

[35] The generalized, one-step prediction errors are defined as in Krolzig (1997) as

$$\hat{e}_{t|t-1} = y_t - \mathsf{E}[y_t|Y_{t-1}; \lambda = \tilde{\lambda}] = y_t - X_t\tilde{\mathbf{B}}\tilde{\mathbf{F}}\hat{\xi}_{t-1|t-1}$$

and has just been divided by the standard deviation. The Rosenblatt transformation of the residuals are calculated as

$$Z_t = \Phi^{-1}\left(\sum_{m=1}^{M} Pr\left(\xi_t = \iota_m|\mathbf{Y}_{t-1}\right)\int_{-\infty}^{e_t} f\left(\hat{e}_{mt}|\xi_t = \iota_m, \mathbf{Y}_{t-1}\right) de_t\right)$$

where Φ^{-1} is the inverse of the standard normal cumulative distribution function (see Smith (2008) for details).

TABLE 1: Estimation Results of the MSI, MSIH, MSIA and MSIAH

	MSI-VAR(1)		MSI-VAR(2)		MSIH-VAR(2)		MSIA-VAR(2)		MSIAH-VAR(2)	
	Regime 1	Regime 2	Regime 1	Regime 2	Regime 1	Regime 2	Regime 1	Regime 2	Regime 1	Regime 2
v_π	4.666***	6.364***	6.747**	5.340*	6.479***	7.179***	4.348**	14.847	1.555	24.067**
	[1.771]	[1.960]	[2.640]	[3.102]	[1.530]	[1.666]	[2.108]	[43.420]	[1.457]	[11.546]
v_u	1.266***	1.496***	0.562**	0.345	0.603***	0.857***	0.204	1.026	0.552*	0.036
	[0.371]	[0.432]	[0.278]	[0.292]	[0.224]	[0.243]	[0.296]	[1.082]	[0.334]	[1.263]
a_{11}	0.167		0.238		0.241**		0.494**	-0.398	0.384***	-0.531
	[0.108]		[0.159]		[0.120]		[0.230]	[2.611]	[0.116]	[1.843]
a_{21}	-0.089***		-0.044**		-0.056***		-0.024	-0.055	-0.053**	-0.007
	[0.021]		[0.019]		[0.018]		[0.023]	[0.066]	[0.024]	[0.043]
a_{12}	-0.559**		-2.002		-0.833**		-0.845	-8.054	0.039	-10.966
	[0.243]		[1.487]		[0.504]		[0.905]	[18.393]	[0.565]	[20.544]
a_{22}	0.826***		1.619***		1.430***		1.487***	1.530**	1.498***	1.630
	[0.053]		[0.115]		[0.118]		[0.091]	[0.768]	[0.165]	[1.262]
b_{11}			-0.182		-0.199**		-0.029	-0.570	0.112	-1.555
			[0.198]		[0.114]		[0.215]	[0.834]	[0.092]	[4.511]
b_{21}			0.046**		0.022		0.044*	0.018	0.028	0.146
			[0.021]		[0.026]		[0.025]	[0.068]	[0.031]	[0.180]
b_{12}			1.359		0.071		0.402	6.130	-0.150	8.327
			[1.196]		[0.463]		[0.754]	[13.067]	[0.529]	[19.974]
b_{22}			-0.680***		-0.527***		-0.530***	-0.631	-0.573***	-0.667
			[0.098]		[0.114]		[0.076]	[0.722]	[0.141]	[1.125]
Var(π)	2.119		1.991		0.214	4.665	1.169		0.505	1.942
	[0.580]		[0.745]		[0.101]	[1.893]	[0.345]		[0.172]	[8.901]
Cov(π,u)	-0.217		-0.130		-0.017	-0.184	0.029		-0.011	0.062
	[0.144]		[0.112]		[0.030]	[0.256]	[0.053]		[0.079]	[0.733]
Var(u)	0.054		0.022		0.011	0.026	0.015		0.034	0.004
	[0.016]		[0.005]		[0.006]	[0.008]	[0.005]		[0.012]	[0.008]
p_{11}	0.893		0.230		0.534		0.823		0.793	
	[0.084]		[0.154]		[0.121]		[0.117]		[0.100]	
p_{22}	0.755		0.548		0.592		0.279		0.269	
	[0.206]		[0.131]		[0.116]		[0.298]		[0.280]	
T	38.75	18.25	21.70	35.30	26.13	30.87	45.993	11.008	44.687	12.313
LL	-84.77		-58.30		-43.16		-45.32		-34.82	
AIC	195.549		150.597		126.314		140.647		125.644	
BIC	222.109		185.329		167.175		191.723		182.849	
MSC	253.278		199.466		168.707		181.129		157.699	
SSR$_\pi$	123.62		153.11		28.75		139.70		134.68	
R_π^2	0.358		0.205		0.851		0.275		0.301	
$\bar{R}_\pi^2$	0.201		0.000		0.791		0.055		0.021	
SSR$_u$	3.04		2.18		29.19		1.87		1.90	
R_u^2	0.957		0.969		0.583		0.973		0.973	
$\bar{R}_u^2$	0.946		0.959		0.416		0.965		0.962	
LB $\hat{e}_{\pi,t\|t-1}$	0.761		0.987		0.827		0.996		0.998	
LB $\hat{e}_{u,t\|t-1}$	0.000		0.997		0.792		0.998		0.991	
LB $Z_{\pi,t}$	0.759		0.987		0.912		0.996		0.975	
LB $Z_{u,t}$	0.000		0.995		0.810		0.998		0.991	
LB $\hat{e}_{\pi.t/t-1}^2$	0.544		0.531		0.888		0.583		0.752	
LB $\hat{e}_{u.t/t-1}^2$	0.139		0.203		0.821		0.127		0.262	
LB $Z_{\pi.t}^2$	0.616		0.525		0.009		0.583		0.975	
LB $Z_{u.t}^2$	0.121		0.206		0.601		0.127		0.303	

Notes: Estimated standard errors are shown in brackets. ***,**,* indicates significance on a 1 %, 5% and 10% level. The variance-covariance matrix has been calculated using the conditional score method in Hamilton (1993,1996). T is the expected duration in each regime. The AIC, BIC are calculated as usual (see e.g. Frühwirth-Schnatter (2006), Lütkepohl (2007)) while the MSC is calculated according to Smith et al (2006). p-values are reported on the Ljung-Box tests of autocorrelation and heteroskedasticity.

unemployment look like white noise, but we can reject the null of homoskedasticity in the Rosenblatt-transformed unemployment residual.

From this model we can test for alternative models in several directions, with regime dependent heteroskedasticity and dynamics as the main options. The LR test can reject the null hypothesis of homoskedasticity conditional on the same lag order.[36] The log likelihood of the MSIH-VAR(2) is –43.16 compared to –58.30 of the MSI. The LR test of regime dependent dynamics can also not reject the null of regime independent lags, as the log likelihood of the MSIA model is –45.32. Here a considerable problem however arises as the MSIA model seems to be hard to estimate. The regime duration is highly skewed with about 46 quarters in the first regime but only 11 quarters in the second. As a consequence, almost all parameters in the second regime are insignificant and we fear that the low duration in that regime results in poor estimates. A fundamental prerequisite of the estimation of the Markov-switching model is that the regimes are somewhat prolonged to be identifiable. If this requirement is met, the estimation algorithm has the ability of finding a potential maximum. We therefore argue that this model most likely is a bad alternative, given the small number of observations in our sample.

We also conduct the LR test for the joint hypothesis of homoskedasticity and regime independent lags and can reject the null. The MSIAH-VAR(2) has a log likelihood of –34.82, the lowest of all estimated models. However, the MSIAH model has the same problems of not fulfilling the prerequisites regarding the regimes as it has a similar skewness towards one regime. Compared to the MSI-VAR(2), there are 11 more parameters to estimate, adding up to a total of 26, which may be the upper bound with a sample size of 57 observations.

Given the estimation problems of the MSIA and MSIAH models and considering the increasing risk of local maxima as the likelihood function can be flat in more directions when more complex models are evaluated, we consider the MSIH models as our best alternative. The expected regime duration are evenly distributed with 26 quarters in the first and about 31 quarters in the second regime. The regimes are very volatile as there is about a 0.5 chance of switching out of each regime. This makes the model a bit hard to interpret. The intercept of the first equation in the system shifts from 6.48 in the first regime to 7.18 in the second and the intercept in the second equation is higher in regime 2 as well. Regime 2 seems to be the high-volatility regime as the variance of inflation is more than 20 times as high as it is in regime 1. All parameters are significant at acceptable levels except the second lag of inflation in the second equation and the second lag of unemployment in the first.

As an alternative, an estimated MSA-VAR(1) model is shown in Table 2.[37] The

---

[36] We consider testing of models with different lag order as well, but as with the MSI-VAR(1), the MSIA and MSIAH with only one lag points at similar problems of remaining autocorrelation in the unemployment residuals (not reported) and we therefore will not consider them as alternatives.

[37] Assuming that the main reason for the econometrician to use the MS model is because of changes in the mean of the time series of interest, Krolzig (1997) proposes that estimating an MSI model often is an appropriate start for evaluating different Markov-switching models. However, if we suspect the dynamics to be an important feature of the time series, estimating an MSA model could also be appropriate as we

TABLE 2: Estimation Results of the MSA, MSAH and MSH

	MSA-VAR(1)		MSA-VAR(2)		MSAH-VAR(2)		MSH-VAR(2)	
	Regime 1	Regime 2	Regime 1	Regime 2	Regime 1	Regime 2	Regime 1	Regime 2
v_π	5.552***		9.749***		8.521***		2.948**	
	[1.524]		[2.338]		[1.926]		[1.291]	
v_u	1.142***		0.501		0.481*		0.405	
	[0.337]		[0.322]		[0.250]		[0.300]	
a_{11}	0.623***	-0.180	0.556**	-0.192	0.173	0.120	0.279**	
	[0.172]	[0.164]	[0.273]	[0.168]	[0.154]	[0.195]	[0.125]	
a_{21}	-0.118***	-0.039	-0.059	-0.013	-0.052	-0.041*	-0.034	
	[0.025]	[0.028]	[0.043]	[0.030]	[0.042]	[0.022]	[0.021]	
a_{12}	-0.600***	-0.609***	-4.686***	-2.078*	-0.855	-5.387***	-0.559	
	[0.227]	[0.204]	[1.302]	[1.263]	[0.616]	[1.726]	[0.565]	
a_{22}	0.879***	0.830***	1.651***	1.446***	1.392***	1.718***	1.589***	
	[0.043]	[0.045]	[0.260]	[0.147]	[0.157]	[0.179]	[0.156]	
b_{11}			-0.439***	-0.374	-0.292**	-0.309	-0.028	
			[0.163]	[0.274]	[0.121]	[0.272]	[0.105]	
b_{21}			0.022	0.036	0.031	0.030	0.054**	
			[0.036]	[0.024]	[0.032]	[0.031]	[0.025]	
b_{12}			3.647***	0.933	-0.172	4.462***	0.270	
			[1.167]	[1.200]	[0.511]	[1.671]	[0.546]	
b_{22}			-0.701***	-0.530***	-0.474***	-0.769***	-0.648***	
			[0.240]	[0.125]	[0.150]	[0.164]	[0.136]	
$\mathrm{Var}(\pi)$	1.869		1.069		0.234	3.046	0.398	7.077
	[0.522]		[0.321]		[0.119]	[1.378]	[0.206]	[5.917]
$\mathrm{Cov}(\pi,u)$	-0.150		-0.073		-0.023	-0.130	-0.037	-0.105
	[0.125]		[0.067]		[0.038]	[0.227]	[0.075]	[0.367]
$\mathrm{Var}(u)$	0.042		0.023		0.012	0.026	0.044	0.014
	[0.010]		[0.007]		[0.008]	[0.008]	[0.018]	[0.010]
p_{11}	0.730		0.453		0.490		0.777	
	[0.192]		[0.167]		[0.146]		[0.120]	
p_{22}	0.835		0.674		0.554		0.527	
	[0.127]		[0.136]		[0.122]		[0.253]	
T	22.34	34.66	21.94	35.06	26.11	30.89	37.649	19.351
LL	-79.39		-49.01		-38.90		-60.66	
AIC	188.78		144.02		117.81		167.33	
BIC	219.42		191.01		158.67		214.32	
MSC	241.54		180.85		160.20		204.72	
SSR_π	112.90		132.28		39.06		151.91	
$\mathrm{R}_\pi{}^2$	0.414		0.313		0.797		0.211	
$\bar{\mathrm{R}}_\pi{}^2$	0.270		0.105		0.716		-0.105	
SSR_u	2.52		1.77		40.70		1.92	
$\mathrm{R}_\mathrm{u}{}^2$	0.964		0.975		0.418		0.973	
$\bar{\mathrm{R}}_\mathrm{u}{}^2$	0.955		0.967		0.185		0.962	
LB $\hat{e}_{\pi,t/t-1}$	0.736		0.983		0.797		0.955	
LB $\hat{e}_{u,t/t-1}$	0.000		0.994		0.256		0.994	
LB $Z_{\pi,t}$	0.602		0.951		0.719		0.956	
LB $Z_{u,t}$	0.000		0.972		0.208		0.996	
LB $\hat{e}_{\pi.t/t-1}{}^2$	0.651		0.465		0.960		0.572	
LB $\hat{e}_{u.t/t-1}{}^2$	0.152		0.214		0.953		0.156	
LB $Z_{\pi.t}{}^2$	0.837		0.825		0.170		0.862	
LB $Z_{u.t}{}^2$	0.483		0.408		0.977		0.073	

Notes: Estimated standard errors are shown in brackets. ***,**,* indicates significance on a 1 %, 5% and 10% level. The variance-covariance matrix has been calculated using the conditional score method in Hamilton (1993,1996). T is the expected duration in each regime. The AIC, BIC are calculated as usual (see e.g. Frühwirth-Schnatter (2006), Lütkepohl (2007)) while the MSC is calculated according to Smith et al (2006). p-values are reported on the Ljung-Box tests of autocorrelation and heteroskedasticity.

MSA-VAR(1) is well estimated by the algorithm and has persistent regimes. The log likelihood is –79.39 compared to –84.77 of the MSI-VAR(1) and information criteria prefers the MSA model.[38] However, the unemployment residuals are autocorrelated. With an additional lag, this problem can be solved and the MSA-VAR(2) has well behaved residuals in both of the variables. It has a higher log likelihood, –49.01, compared to the MSI-VAR(2). The regimes are more volatile but the expected regime duration is similar to the model with only one lag. The LR test with null hypothesis of homoskedasticity can be rejected at all acceptable significance levels but the likelihood ratio of regime dependent intercepts is 7.37 and close to the critical value, even though we reject the null. A joint test of homoskedasticity and regime dependent intercepts can be rejected.[39] The MSAH-VAR(2) has a higher log likelihood than the MSIA-VAR(2) and the highest of all other models except MSIAH-VAR(2). It is also preferred by the AIC and BIC when comparing all models, also the unnested MSIH-VAR(2). The MSC, however chooses the MSIAH-VAR(2).[40]

If we consider the information criteria where the AIC and BIC selected the MSAH model over the MSIH, we can conclude that it seems that the data is best explained by changing dynamics rather than changes in the intercepts. Regime-dependent heteroskedasticity is however an important feature in both models. The sum of squared residuals SSR_π is significantly reduced in both the MSIH and MSAH compared to homoskedastic models and the amount of sample variation that can be explained by the model is high, even though it is worse in explaining the relatively low variation in the unemployment data.

We proceed according to the Box-Jenkins procedure and check whether we can make the MSAH-VAR(2) more parsimonious by restricting insignificant parameters, see Ta-

---

may find different, non-nested models that we cannot evaluate by extensions to the MSI.

[38] The Akaike (AIC) and Schwarz Baysian (BIC) information criteria of Kullback-Leibler divergence for the MS model are constructed as is common for multivariate models, see Lütkepohl (2007), Frühwirth-Schnatter (2006). In addition the Markov Switching criteria by Smith et al. (2006) is reported. This criteria is primarily for comparison of models with different number of regimes and variables and has not been tested for consistency for comparison of different Markov characteristics. We do however report this as the criteria penalizes the log likelihood with respect to the expected duration of the regimes. The MSC is constructed as

$$\text{MSC} = -2\ln L(\tilde{\lambda}|Y_T) + \sum_{s=1}^{M} \frac{\hat{T}_s(\hat{T}_s + A_s K)}{B_s \hat{T}_s - A_s K - 2} \tag{22}$$

where $\hat{T}_s = tr(\hat{W}_s)$, $\hat{W}_s = diag(\hat{\xi}_{1i|T}, \ldots, \hat{\xi}_{Ti|T})$. A_s and B_s are set as in Smith et al (2006) to $A_s = M$ $B_s = 1$.

[39] The tests containing the unrestricted models of MSIA and MSIAH have however the same estimation problems as discussed above.

[40] As there are only two more parameters in an MSIAH-VAR(2) compared to the MSAH-VAR(2) with a total of 24, we could suspect the same problem with estimating this model. However, the regimes are more equal in duration even though both regimes shift with a high probability. Hence we do not end up with all parameters in one regime being obviously badly estimated but the precision of the estimates are reduced compared to less complex models. This reveals the problem of evaluation and selecting different MS models when the sample size is small. This is since the log likelihood almost always increases when allowing more parameters to shift even though the precision of the estimates falls significantly.

TABLE 3: Estimation Results of Different Constrained MSAH-VAR(2) Models

	$b_{11}^{(2)} = 0$		$b_{21}^{(2)} = 0$		$a_{11}^{(2)} = b_{11}^{(2)} = 0$		$b_{11}^{(2)} = b_{21}^{(2)} = 0$		$a_{11}^{(2)} = b_{11}^{(2)} = b_{21}^{(2)} = 0$		$b_{12}^{(1)} = a_{11}^{(2)} = b_{11}^{(2)} = 0$	
	Regime 1	Regime 2	Regime 1	Regime 2	Regime 1	Regime 2	Regime 1	Regime 2	Regime 1	Regime 2	Regime 1	Regime 2
v_π	8.368***		8.362***		8.360***		8.320***		8.315***		8.539***	
	[1.772]		[1.665]		[1.646]		[1.684]		[1.581]		[1.584]	
v_u	0.494		0.549*		0.495		0.537*		0.538*		0.489	
	[0.325]		[0.284]		[0.325]		[0.292]		[0.288]		[0.318]	
a_{11}	0.180*	-0.005	0.181	0.073	0.180*		0.181*	-0.003	0.181		0.165	
	[0.108]	[0.242]	[0.112]	[0.268]	[0.109]		[0.110]	[0.246]	[0.112]		[0.103]	
a_{21}	-0.052**	-0.036	-0.053**	-0.031	-0.052**	-0.036	-0.053**	-0.030	-0.053**	-0.030	-0.051**	-0.036
	[0.022]	[0.027]	[0.023]	[0.025]	[0.022]	[0.026]	[0.022]	[0.026]	[0.022]	[0.024]	[0.022]	[0.025]
a_{12}	-0.811	-5.311***	-0.837	-5.354***	-0.809	-5.301***	-0.826	-5.331***	-0.824	-5.326***	-1.030***	-5.374***
	[0.782]	[1.848]	[0.739]	[1.806]	[0.764]	[1.688]	[0.756]	[1.845]	[0.760]	[1.647]	[0.208]	[1.645]
a_{22}	1.391***	1.717***	1.379***	1.696***	1.390***	1.716***	1.382***	1.703***	1.382***	1.702***	1.409***	1.723***
	[0.145]	[0.182]	[0.140]	[0.169]	[0.145]	[0.176]	[0.139]	[0.172]	[0.140]	[0.169]	[0.132]	[0.174]
b_{11}	-0.285**		-0.286**	-0.180	-0.284**		-0.283**		-0.283**		-0.297***	
	[0.112]		[0.115]	[0.251]	[0.112]		[0.114]		[0.110]		[0.107]	
b_{21}	0.030	0.019	0.028		0.0303	0.0187	0.028		0.028		0.031	0.018
	[0.022]	[0.025]	[0.023]		[0.022]	[0.025]	[0.021]		[0.021]		[0.022]	[0.025]
b_{12}	-0.196	4.367**	-0.170	4.434**	-0.1974	4.358***	-0.177	4.394**	-0.177	4.389***		4.407***
	[0.730]	[1.827]	[0.694]	[1.802]	[0.718]	[1.684]	[0.711]	[1.839]	[0.721]	[1.644]		[1.640]
b_{22}	-0.474***	-0.768***	-0.470***	-0.751***	-0.474***	-0.767***	-0.471***	-0.757***	-0.471***	-0.756***	-0.492***	-0.773***
	[0.133]	[0.171]	[0.132]	[0.162]	[0.134]	[0.166]	[0.129]	[0.164]	[0.130]	[0.161]	[0.120]	[0.163]
Var(π)	0.234		0.240		0.235	3.272	0.240	3.305	0.240	3.305	0.241	3.315
	[0.294]		[0.254]		[0.296]	[0.941]	[0.280]	[0.944]	[0.278]	[0.978]	[0.322]	[0.973]
Cov(π,u)	-0.024		-0.025		-0.024	-0.137	-0.025	-0.153	-0.025	-0.152	-0.024	-0.141
	[0.029]		[0.029]		[0.029]	[0.071]	[0.029]	[0.070]	[0.030]	[0.072]	[0.032]	[0.071]
Var(u)	0.012		0.013		0.012	0.026	0.013	0.027	0.013	0.027	0.012	0.026
	[0.007]		[0.007]		[0.007]	[0.008]	[0.007]	[0.008]	[0.007]	[0.008]	[0.007]	[0.008]
p_{11}	0.485		0.510		0.485		0.505		0.505		0.498	
	[0.155]		[0.153]		[0.155]		[0.151]		[0.150]		[0.153]	
p_{22}	0.549		0.559		0.549		0.551		0.551		0.549	
	[0.153]		[0.149]		[0.153]		[0.153]		[0.151]		[0.152]	
T	26.136	30.864	26.52	30.48	26.13	30.87	26.62	30.38	26.507	30.493	26.507	30.493
LL	-40.01		-39.87		-40.03		-40.05		-40.06		-39.88	
AIC	118.03		117.74		116.06		116.10		114.12		113.75	
BIC	156.847		156.554		152.831		152.876		148.848		148.486	
MSC	162.421		162.111		162.449		162.472		162.487		162.129	
SSR_π	39.70		38.17		39.67		39.36		39.35		39.54	
R_π^2	0.79		0.80		0.79		0.80		0.80		0.79	
$\bar{\mathrm{R}}_\pi^2$	0.71		0.72		0.71		0.71		0.71		0.71	
SSR_u	42.60		42.43		42.52		44.40		44.35		44.38	
R_u^2	0.391		0.393		0.392		0.365		0.366		0.365	
$\bar{\mathrm{R}}_\mathrm{u}^2$	0.147		0.151		0.149		0.111		0.112		0.112	
LB $\hat{e}_{\pi,t\|t-1}$	0.800		0.806		0.800		0.807		0.807		0.828	
LB $\hat{e}_{u,t\|t-1}$	0.526		0.313		0.536		0.483		0.489		0.527	
LB $Z_{\pi,t}$	0.748		0.747		0.748		0.756		0.756		0.770	
LB $Z_{u,t}$	0.116		0.111		0.118		0.081		0.082		0.116	
LB $\hat{e}_{\pi.t\|t-1}^2$	0.958		0.960		0.958		0.958		0.958		0.951	
LB $\hat{e}_{u.t\|t-1}^2$	0.948		0.949		0.947		0.957		0.956		0.941	
LB $Z_{\pi.t}^2$	0.739		0.727		0.740		0.735		0.735		0.680	
LB $Z_{u.t}^2$	0.736		0.944		0.737		0.682		0.682		0.760	

Notes: Estimated standard errors are shown in brackets. ***,**,* indicates significance on a 1 %, 5% and 10% level. The standard errors and p-values have been calculated by bootstrapings 10 000 simulated time series with size T+100 from the estimated model. The 100 first observations are discarded to get independence from starting values. T is the expected duration in each regime. The AIC, BIC are calculated as usual (see e.g. Frühwirth-Schnatter (2006), Lütkepohl (2007)) while the MSC is calculated according to Smith et al (2006). p-values are reported on the Ljung-Box tests of autocorrelation and heteroskedasticity.

ble 3.[41] This could be motivated both in terms of forecasting and economic analysis, as it increases the overall fit and precision of the estimates. We find that imposing restrictions on three of the parameters that are being close to zero does not significantly alter the log likelihood. We test for several different restrictions, but see evidence of autocorrelation in the unemployment residual when imposing any restriction on that equation in the system.

The best alternative is achieved when we impose the effect of the two quarter lag of unemployment on inflation to be equal to zero in regime 1 ($b_{12}^{(1)} = 0$) and the persistence of inflation being zero in regime 2 ($a_{11}^{(2)} = b_{11}^{(2)} = 0$). We can reject the null hypothesis when performing a LR test of the restricted model versus the unrestricted. All residuals are white noise, and there is no evidence of remaining heteroskedasticity (the residual series are plotted in figure 12, 13 and 14) The restricted model has 10 significant parameters with high precision (significant at about a 1% significance level), while most of the insignificant parameters are small in magnitude (the exception is v_u and a_{11} that have a standard deviation of about 3/5 of the estimate). The log likelihood is –39.88 and the model is selected by the AIC and BIC compared to all other alternative specifications.[42]

5 DISCUSSION

Applying the bottom-up strategy for selecting the Markov characteristics and lag order of our MS model lead us to a model with regime-dependent dynamics and heteroskedasticity while the intercepts were regime-independent. This is interesting since significant regime-independent intercepts points towards a rather stable Phillips relationship, which do not shift up and down as expectations are changing. Even though our sample size is small in relation to the complexity of the model, we were able to yield a model with relative high precision and fit. The estimated reduced-form system can be written in matrix form where $\hat{A}_m$, $\hat{B}_m$ are the shifting autoregressive parameter matrices of the first and second lag and $\hat{\Sigma}_m$ the regime-dependent variance-covariance matrix. The system can be written as

$$\begin{bmatrix} \pi_t \\ u_t \end{bmatrix} = \begin{bmatrix} \hat{v}_\pi \\ \hat{v}_u \end{bmatrix} + \begin{bmatrix} \hat{a}_{11} & \hat{a}_{12} \\ \hat{a}_{21} & \hat{a}_{22} \end{bmatrix}_m \begin{bmatrix} \pi_{t-1} \\ u_{t-1} \end{bmatrix} + \begin{bmatrix} \hat{b}_{11} & \hat{b}_{12} \\ \hat{b}_{21} & \hat{b}_{22} \end{bmatrix}_m \begin{bmatrix} \pi_{t-2} \\ u_{t-2} \end{bmatrix} + \begin{bmatrix} e_{\pi,t} \\ e_{u,t} \end{bmatrix}$$

$$\hat{V} = \begin{bmatrix} 8.539 \\ 0.489 \end{bmatrix} \quad \hat{A}_1 = \begin{bmatrix} 0.165 & -1.030 \\ -0.051 & 1.409 \end{bmatrix}_1 \quad \hat{A}_2 = \begin{bmatrix} 0 & -5.374 \\ -0.036 & 1.723 \end{bmatrix}_2$$

$$\hat{B}_1 = \begin{bmatrix} -0.297 & 0 \\ 0.031 & -0.492 \end{bmatrix}_1 \quad \hat{B}_2 = \begin{bmatrix} 0 & 4.407 \\ 0.018 & -0.773 \end{bmatrix}_2$$

$$\hat{\Sigma}_1 = \begin{bmatrix} 0.241 & -0.024 \\ -0.024 & 0.012 \end{bmatrix}_1 \quad \hat{\Sigma}_2 = \begin{bmatrix} 3.314 & -0.141 \\ -0.141 & 0.026 \end{bmatrix}_2$$

---

[41] Since Krolzig (1997) does not provide any estimators featuring parameter constraints we have for this purpose derived them ourselves, see Appendix, Section A.1.3

[42] The MSC selects a model with the only restriction that $b_{21}(s_t = 2) = 0$, but we choose the more parsimonious model with three restrictions.

$$\hat{\mathbf{P}} = \begin{bmatrix} 0.498 & 0.502 \\ 0.451 & 0.549 \end{bmatrix} \quad \hat{T} = \begin{bmatrix} 26.507 \\ 30.493 \end{bmatrix}$$

The processes in the two regimes are primarily characterized by the changing $\tilde{\Sigma}_m$. The variance of the forecast errors of inflation is significantly higher in the second regime compared to the first. The correlation between the two forecasting errors as well as the variance of the unemployment error are also larger in magnitude in the second regime. The changes in the dynamics indicates that inflation has much more persistence in first regime while apparently the opposite is true for the unemployment rate.

The regimes have been similar to other more complex models and show a rather volatile path between the two regimes (see Figure 3). The unconditional probability of being in a given regime, p_1 and p_2 are 0.473 and 0.527 respectively.[43] The regime probabilities for staying in each regime are around 0.5, and on the threshold for when the regimes are said to be persistent or not. This also makes the regimes rather hard to interpret from a structural perspective: They do not follow a clear pattern such that prolonged periods under one regime being followed by the other, a characteristic of the regime one often is interested to find with the MS model (such as e.g. when studying business cycle fluctuations). Comparing the time path of the inflation rate with the regime periodicity it is evident that it does not follow any clear pattern. If we return to the above discussion of inflation expectations, we could for example have suspected to find some co-movement with the inflation gap, contemporaneous or intertemporally correlated to some degree.

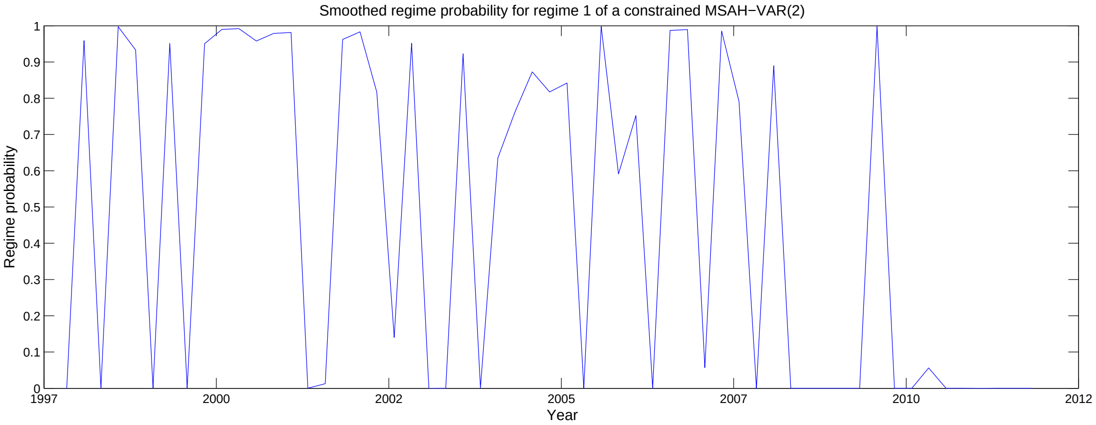


Figure 3: Smoothed probability for regime 1, $\Pr(s_t = 1)$.

The probably most important test we could conduct in this study is to test for linearity. Unfortunately, an intrinsic problem with the MS model is that it is not evident how to perform a test against a linear model. A time-invariant Gaussian VAR(p) model can

---

[43] The unconditional regime probabilities are calculated as $p_1 = [1 - p_{22}]/[2 - p_{11} - p_{22}]$ and $p_2 = [1 - p_{11}]/[2 - p_{11} - p_{22}]$.

be written as a MS(1)-VAR(p) which is nested in the MS(M)-VAR(p) models. However, equivalence of the VAR parameters $\theta_1 = \theta_2 = \cdots = \theta_M$ makes the transition probabilities p_{ij} unidentifiable, and these nuisance parameters cause a bias of the LR against the null as the probabilistic distribution is non-standard (Hansen, 1992).

A solution proposed by Krolzig (1997) would be to test a model with some regime-dependent parameters against a MSH model (with only regime-dependent heteroskedasticity) given that the alternative model also assumes heteroskedasticity. The LR ratio when testing an unconstrained MSAH-VAR(2) against a MSH-VAR(2) is 43.53 (critical value 15.51) and we can reject the null hypothesis that all parameter except Σ_m are regime-independent, see Table 2. Given that heteroskedasticity is an important feature of the data, we consider this as a plausible test against a linear alternative.[44]

An additional test would be to see whether we parametrically can encompass real data characteristics from simulations on our estimated model. Failure to encompass the sample mean and variance within plausible levels should be seen as a sign of misspecification and erroneous usage of the MS model, when the true data generating process is better explained in another way. In Figure 4 we display the sample means, variances and covariance (indicated by straight line) of our time series, and we encompass all of the important features seen in the data rather well.

Before continuing the discussion of the estimated model, it is in place to remember that the estimated VAR is in a reduced form. The forecasting errors $e_{\pi,t}$ and $e_{u,t}$ could be composed by a combination of the underlying structural shocks $\varepsilon_{u,t}$ and $\varepsilon_{\pi,t}$, and if we do not impose any identifying restriction on the specific form of how they relate to each other, the system remains unidentified. Since it is hard to find credible identification restrictions, especially for non-linear models, the large bulk of the literature have focused on either generalized impulse response analysis (GIRF) and forecast error variance decomposition (GFEVD)[45] (Krolzig, 2006, Karamé, 2012, Do et al., 2013, Lanne and Nyberg, 2014) or on recursive identification through Choleski decomposition (Weise, 1999, Ehrmann et al., 2003, Droumaguet, 2012).

However, since the use of identifying restriction imposes rather much ambiguity depending on the theoretical stance, see discussion in e.g. King and Watson (1994), Kilian (2011), we consider only the reduced form estimates as they provide enough information to characterize the Phillips relation.[46] Foremost, we wish to establish comparability with

---

[44] A possible future extension would be to follow the procedure suggested by Lenčuchová (2011), using the White (1987) test for serial correlation and Hamilton (1996)'s approach for dynamic specification test or, as is more common, by generating the probabilistic distribution with Monte Carlo methods to get appropriate critical values for the test statistics.

[45] The generalized versions of the IRF and FEVD basically consider shocks to each equation of the system, i.e. in e_t, instead of the uncorrelated structural shocks ε_t.

[46] Imposing restrictions can have rather large implication on the empirical conclusions and should be implemented with care. King and Watson (1994) use a similar bivariate system to study the long-run Phillips relationship under three different identifications, corresponding to a traditional Keynesian (TK), a rational expectation monetarist (REM), and a real business cycle (RBC) interpretation of the model economy. The TK and the RBC corresponds to different orderings of a recursive system and yield the most extreme results while the REM identification ends up being some form of 'mainstream' approach

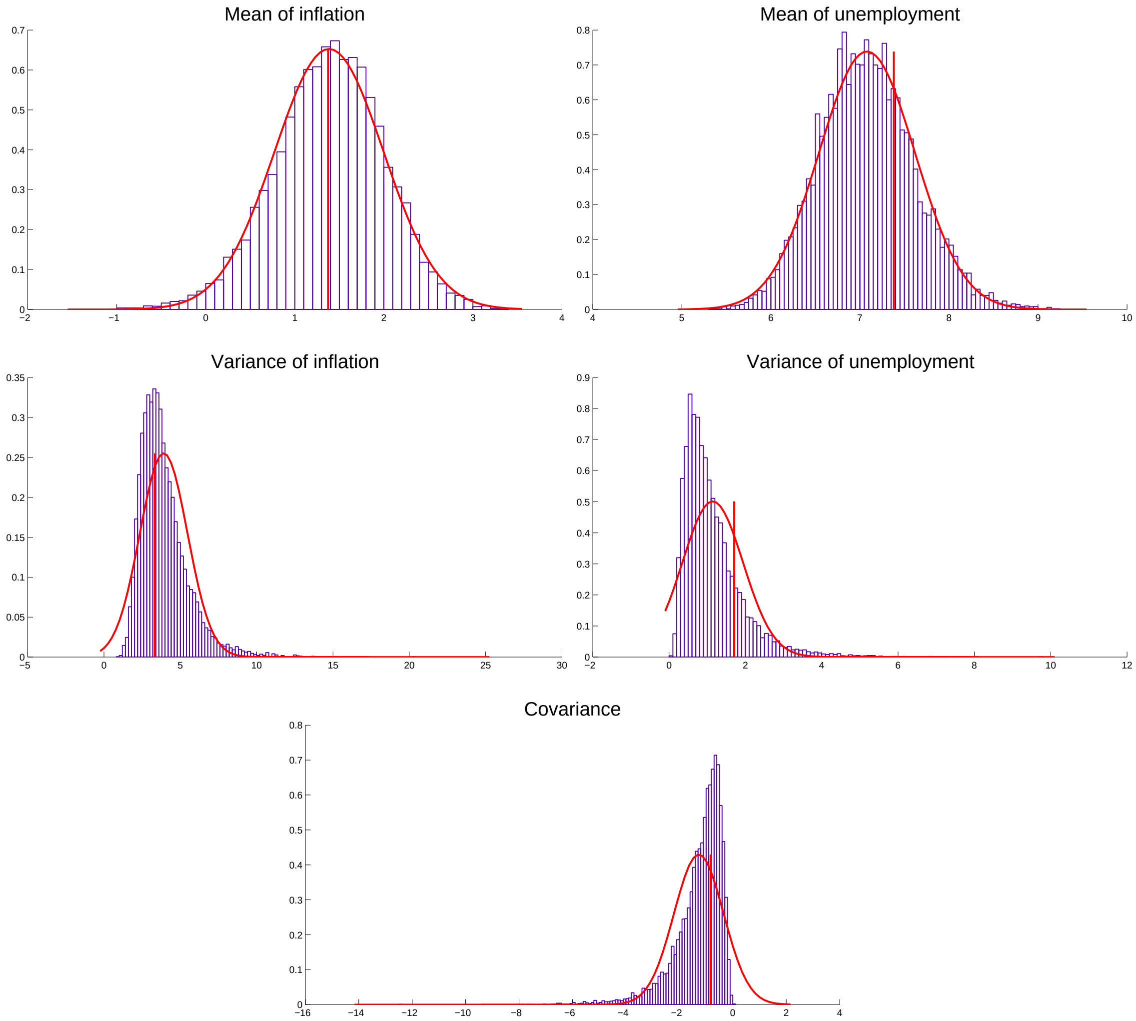


Figure 4: Parametric encompassing of empirical moments

Svensson (2015), and since he uses a single-equation estimation, we would be less able to compare our estimates if we imposed identifying restrictions. Alternatively, the results could be interpreted as a recursive ordering with π_t and u_t, and hence that structural shocks in the unemployment equation has no contemporaneous effect on inflation.[47] As pointed out in Staiger et al. (1997), it is rather plausible that inflationary effects of tight demand occur with a lag.

Concentrating on the component of the VAR system (19) corresponding to the short-run Phillips curve equation (9), we yielded two rather similar inflation processes in both regimes, that differ most with regard to the volatility in the error terms. In the first regime, inflation shows persistence where the two lags sum up to –0.132, keeping in mind that the first lag is insignificant, while inflation in the second regime has no persistence at all.

$$\begin{aligned}\text{Regime 1:}\quad & \pi_t = 8.539 + 0.165\pi_{t-1} - 0.297\pi_{t-2} - 1.030u_{t-1} + e_{\pi,t}\\ \text{Regime 2:}\quad & \pi_t = 8.539 - 5.374u_{t-1} + 4.407u_{t-2} + e_{\pi,t},\end{aligned}$$

It is rather clear that the Solow-Tobin test of $\gamma(1)$ can be rejected and it is therefore not far-fetched to establish a non-vertical long-run Phillips curve, as in Svensson (2015). Eliasson (2001) do not test for a non-vertical Phillips curve in her study of the period 1979Q2-1997Q4, but the sum of the lags of her short-run estimates were significantly larger. In addition, she also found that the short-run trade-off varied with time, as well as the intercept, which was excluded from the linear benchmark Phillips curve for the same period.

We look further at the long-run component of the inflation–unemployment trade-off. We can do the same exercises as in Svensson (2015), and back out the long-run coefficient from the short-run estimates by taking the unconditional expectation of each variable, $E\{\pi_t\} = \pi$, $E\{u_t\} = u$, and $E\{e_{\pi,t}\} = 0$.[48] This yields a Phillips curve trade-off of –0.910 in the first regime and –0.967 in the second where the slope in the first regime is possibly overestimated due to the lack of precision in the lag of inflation. Since the trade-off is rather similar in both regimes, it is not very far-fetched to talk about a long-run slope even though the regimes are not that persistent.

This implies a slightly steeper Phillips curve compared to Svensson's, average benchmark of –0.75. If we take the average with regard to the expected duration in each regime, we get a slope of –0.94. Calculating the unemployment cost that would be entailed with an average inflation 0.6 percent lower compared to the target, yield a 0.64 percent higher unemployment. As pointed out by Svensson, the slope of the curve is sensitive to starting date, but our estimate could be viewed as rather conservative as the curve flatten out including additional and fever observation at the beginning of the sample.[49]

---

that gives room for both demand and supply disturbances in the business cycle.

[47] In this case, the reduced-form estimates of the Phillips curve are in fact identical to the structural model.

[48] Basically we find it as $\partial\pi/\partial u = [\hat{a}_{11} + \hat{b}_{11}]/[1 - \hat{a}_{12} + \hat{b}_{12}]$.

[49] A flatter Phillips curve implies a higher unemployment cost.

To conclude, it seems like the estimates of our non-linear Phillips curve are rather similar compared to linear specifications even though we have considerable regime-dependent heteroskedasticity included in our model. Most notable is that the testing procedure found a model with regime-independent intercepts, which points towards a stable Phillips relationship which do not change because of varying expectations. As the regimes of the model were highly non-persistent, it is not clear how to interpret the non-linearity from an economic perspective, other than that some observations contributed to much higher residual variance.

6 Conclusion

We have in this study investigated the Phillips relationship in Sweden by estimating a Markov-switching VAR that corresponded to a bivariate system with inflation and the unemployment rate. We were able to fit an internally consistent and parsimonious model to the data that was modelled with time-varying and regime-dependent dynamics and heteroskedasticity. The model encompassed the empirical moments of the data well and was tested against several different MS model specifications.

The resulting processes in our two regime MSAH-VAR(2) model had a more persistent inflation process the first regime, while the other was dominated by higher variance and volatility in the residuals. The regime periodicity did not comove with the inflation time series path in such a way that it could be clearly interpreted as changes in expectations. The resulting long-run estimates of the Phillips trade-off had a slope of –0.910 in the first regime and –0.967 in the second. Even though the interpretation of a long-run relationship in a Markov framework are conceptually rather vague, we argue that the processes are so similar so that we can confirm the existence of a non-vertical long-run Phillips curve for the period 1997Q4-2011Q4. This would imply an unemployment cost of about 0.64 percent, deduced from our reduced form estimates, which of course should be interpreted with care in accordance with the structural VAR time-series literature.

A limitation in this study is that we have not considered different time-spans of our data, mainly because of the computational burden estimating the model. In addition, we have not used alternative measures or definitions of inflation or the forcing variable as robustness, other than those of the benchmark, Svensson (2015), model.[50]

Direction of future investigations of the Phillips curve in the Markow-switching framework can easily be pinned down as there are many aspects that we leave out. We have not considered the forecasting performance of the model and neither tested it more closely against a linear specification, other than a heteroskedastic MSH model. In addition, we have not looked further into the GIRF and GFEVD as is common practice in the non-linear time series literature, even though we rather would see an analysis of a structural model, perhaps using heteroskedasticity for identification which would enabling statistical testing of imposed restrictions, see e.g. Lanne and Lütkepohl (2008), Netšunajev (2013).

---

[50] Svensson consider e.g. the unemployment gap by subtracting a time-varying NAIRU, which however, as it varies very little over the period, do not change the results.

References

Akerlof, G. A., Dickens, W. T., and Perry, G. L. (2000). Near-Rational Wage and Price Setting and the Long-Run Phillips Curve. *Brookings Papers on Economic Activity*, 31(1):1–60.

Andersson, B., Palmqvist, S., and Österholm, P. (2012). The Riksbank's Attainment of its Inflation Target over a Longer Period of Time. Economic Commentaries 4, The Swedish Riksbank, Stockholm.

Andersson, F. N. G. and Jonung, L. (2014). The Return of the Original Phillips curve? An Assessment of Lars E. O. Svensson's Critique of the Riksbank's Inflation Targeting, 1997-2012. Working Paper 28, Department of Economics, Lund University.

Assarsson, B. (2011). Penningpolitiken i Sverige. *Ekonomisk Debatt*, 39(3):46–59.

Bai, J. and Perron, P. (1998). Estimating and Testing Linear Models with Multiple Structural Changes. *Econometrica*, 66(1):47–78.

Baum, L. E., Petrie, T., Soules, G., and Weiss, N. (1970). A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains. *The Annals of Mathematical Statistics*, 41(1):164–171.

Beechey, M. J., Johannsen, B. K., and Levin, A. T. (2011). Are Long-Run Inflation Expectations Anchored More Firmly in the Euro Area than in the United States? *American Economic Journal: Macroeconomics*, 4(2):104–129.

Benati, L. (2012). The long-Run Phillips Curve: A Structural VAR Investigation. Working paper, Department of Economics, University of Bern.

Bessec, M. and Bouabdallah, O. (2005). What Causes the Forecasting Failure of Markov-Switching Models? A Monte Carlo Study. *Studies in Nonlinear Dynamics & Econometrics*, 9(2):1–24.

Blix, M. (1999). Forecasting Swedish Inflation With a Markov Switching VAR. Working Paper Series 76, Sveriges Riksbank (Central Bank of Sweden).

Breunig, R., Najarian, S., and Pagan, A. (2003). Specification Testing of Markov Switching Models. *Oxford Bulletin of Economics and Statistics*, 65(s1):703–725.

Calvo, G. A. (1983). Staggered Prices in a Utility-Maximizing Framework. *Journal of Monetary Economics*, 12(3):383–398.

Christiano, L. J., Eichenbaum, M., and Evans, C. L. (1998). Monetary Policy Shocks: What Have We Learned and to What End? Working Paper 6400, National Bureau of Economic Research.

Cochrane, J. H. (2011). Determinacy and Identification with Taylor Rules. *Journal of Political Economy*, 119(3):565–615.

Debelle, G. and Laxton, D. (1997). Is the Phillips Curve Really a Curve? Some Evidence for Canada, the United Kingdom, and the United States. *IMF Staff Papers*, 44(2):249–282.

Debelle, G. and Vickery, J. (1997). Is the Phillips Curve a Curve? Some Evidence and Implications for Australia. RBA Research Discussion Papers 9706, Reserve Bank of Australia.

Demers, F. (2003). The Canadian Phillips Curve and Regime Shifting. Working Papers 03-32, Bank of Canada.

Demers, F. and Rodríguez, G. (2001). Estimation of the Taylor Rule for Canada Under Multiple Structural Changes. Working paper, Department of Economics, University of Ottawa.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38.

Do, H. X., Brooks, R. D., and Treepongkaruna, S. (2013). Generalized Impulse Response Analysis in a Fractionally Integrated Vector Autoregressive Model. *Economics Letters*, 118(3):462–465.

Droumaguet, M. (2012). *Markov-Switching Vector Autoregressive Models: Monte Carlo Experiment, Impulse Response Analysis, and Granger-Causal Analysis.* PhD thesis, European University Institute.

Dupasquier, C. and Ricketts, N. (1998). Non-Linearities in the Output-Inflation Relationship: Some Empirical Results for Canada. Working Paper 98-14, Bank of Canada, Ottawa.

Ehrmann, M., Ellison, M., and Valla, N. (2003). Regime-dependent Impulse Response Functions in a Markov-switching Vector Autoregression Model. *Economics Letters*, 78(3):295–299.

Eisner, R. (1996). A New View of the NAIRU. In Davidson, P. and Kregel, J. A., editors, *Improving the Global Economy: Keynesianism and the Growth in Output and Employment.* Edward Elgar Publishing, Cheltenham.

Eliasson, A.-C. (2001). Is the Short-run Phillips Curve Nonlinear? Empirical Evidence for Australia, Sweden and the United States. Sveriges Riksbank Working Paper Series 124.

Evans, M. and Wachtel, P. (1993). Inflation Regimes and the Sources of Inflation Uncertainty. *Proceedings*, 25(3):475–520.

Filardo, A. J. (1998). New Evidence on the Output Cost of Fighting Inflation. *Economic Review*, (Q III):33–61.

Flodén, M. (2012). A Note on Swedish Inflation and Inflation Expectations. Working paper, Stockholm University and CEPR.

Franses, P. H. and van Dijk, D. (2000). *Non-linear Time Series Models in Empirical Finance*. Cambridge University Press, Cambridge, UK.

Friedman, M. (1968). The Role of Monetary Policy. *American Economic Review*, 58(1):1–17.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models: Modeling and Applications to Random Processes*. Springer Series in Statistics. Springer, Berlin.

Fuhrer, J. C. (2011). Inflation Expectations and the Evolution of U. S. Inflation. Public Policy Brief 11-4, Federal Reserve Bank of Boston.

Gali, J. (2008). *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework*. Princeton University Press, Princeton, NJ.

Gali, J. and Gertler, M. (1999). Inflation Dynamics: A Structural Econometric Analysis. *Journal of Monetary Economics*, 44(2):195–222.

Goldfeld, S. M. and Quandt, R. E. (1973). A Markov Model for Switching Regressions. *Journal of Econometrics*, 1(1):3–15.

Gürkaynak, R. S., Levin, A. T., Marder, A. N., and Swanson, E. T. (2007). Inflation Targeting and the Anchoring of Inflation Expectations in the Western Hemisphere. *Economic Review*, 2007:25.

Gürkaynak, R. S., Levin, A. T., and Swanson, E. (2010). Does Inflation Targeting Anchor Long-Run Inflation Expectations? Evidence From the US, UK, and Sweden. *Journal of the European Economic Association*, 8(6):1208–1242.

Hamilton, J. D. (1988). Rational-Expectations Econometric Analysis Of Changes In Regime: An Investigation Of The Term Structure Of Interest Rates. *Journal of Economic Dynamics and Control*, 12(2-3):385–423.

Hamilton, J. D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, 57(2):357–384.

Hamilton, J. D. (1990). Analysis of Time Series Subject to Changes in Regime. *Journal of Econometrics*, 45(1-2):39–70.

Hamilton, J. D. (1994a). State-Space Models. In Engle, R. F. and McFadden, D. L., editors, *Handbook of Econometrics*, volume 4, chapter 50. Elsevier.

Hamilton, J. D. (1994b). *Time Series Analysis*. Princeton University Press, Princeton, NJ.

Hamilton, J. D. (1996). Specification Testing in Markov-Switching Time-Series Models. *Journal of Econometrics*, 70(1):127–157.

Hansen, B. E. (1992). The Likelihood Ratio Test under Nonstandard Conditions: Testing the Markov Switching Model of GNP. *Journal of Applied Econometrics*, 7(S1):61–82.

Jonsson, T. and Österholm, P. (2012). The Properties of Survey-Based Inflation Expectations in Sweden. *Empirical Economics*, 42(1):79–94.

Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering, Transactions of the ASME*, 82, Series D:35–45.

Karamé, F. (2012). An algorithm for generalized impulse-response functions in markov-switching structural var. *Economics Letters*, 117(1):230–234.

Kilian, L. (2011). Structural Vector Autoregressions. CEPR Discussion Papers 8515, Centre for Economic Policy Research.

Kim, C.-L. Dynamic Linear Models with Markov-Switching. *Journal of Econometrics*, 60(1-2):1–22.

King, R. G., Plosser, C. I., and Rebelo, S. T. (1988). Production, Growth and Business Cycles: I. The Basic Neoclassical Model. *Journal of Monetary Economics*, 21(2-3):195–232.

King, R. G. and Watson, M. W. (1994). The Post-War U.S. Phillips Curve: a Revisionist Econometric History. In *Carnegie-Rochester Conference Series on Public Policy*, volume 41, pages 157–219. Elsevier.

Konjunkturinstitutet (2004). Wage Formation – Economic Conditions in Sweden 2004. Special studies, The National Institute of Economic Research.

Konjunkturinstitutet (2012). Wage Formation in Sweden – Summary 2012. Special studies, The National Institute of Economic Research.

Krolzig, H.-M. (1995). Specification of Autoregressive Processes with Markov Switching Regimes Based on ARMA Representations. Working papers, Humboldt Universität zu Berlin.

Krolzig, H.-M. (1997). *Markov Switching Vector Autoregressions. Modelling, Statistical Inference and Application to Business Cycle Analysis*, volume 454 of *Lecture Notes in Economics and Mathematical Systems.* Springer, Berlin.

Krolzig, H.-M. (2006). Impulse-response Analysis in Markov Switching Vector Autoregressive Models. Working paper, Economics Department, University of Kent. Keynes Collage.

Lanne, M. and Lütkepohl, H. (2008). Identifying Monetary Policy Shocks via Changes in Volatility. *Journal of Money, Credit and Banking*, 40(6):1131–1149.

Lanne, M. and Nyberg, H. (2014). Generalized Forecast Error Variance Decomposition for Linear and Nonlinear Multivariate Models. Technical report, School of Economics and Management, University of Aarhus.

Laxton, D., Clark, P. B., and Rose, D. (1995). Asymmetry in the U.S. Output-Inflation Nexus - Issues and Evidence. IMF Working Papers 95/76, International Monetary Fund.

Laxton, D., Rose, D., and Tambakis, D. (1999). The U.S. Phillips Curve: The Case for Asymmetry. *Journal of Economic Dynamics & Control*, 23(9-10):1459–1485.

Lenčuchová, J. (2011). Testing of Markov Assumptions Based on the Dynamic Specification Test. *Acta Polytechnica Hungarica*, 8(3):27–36.

Lindgren, G. (1978). Markov Regime Models for Mixed Distributions and Switching Regressions. *Scandinavian Journal of Statistics*, 5(2):81–91.

Lucas, R. J. (1972). Expectations and the Neutrality of Money. *Journal of Economic Theory*, 4(2):103–124.

Lucas, R. J. (1973). Some International Evidence on Output-Inflation Tradeoffs. *American Economic Review*, 63(3):326–334.

Lütkepohl, H. (2007). *New Introduction to Multiple Time Series Analysis*. Springer.

Mavroeidis, S. (2010). Monetary Policy Rules and Macroeconomic Stability: Some New Evidence. *American Economic Review*, 100(1):491–503.

Mavroeidis, S., Plagborg-Møller, M., and Stock, J. H. (2014). Empirical Evidence on Inflation Expectations in the New Keynesian Phillips Curve. *Journal of Economic Literature*, 52(1):124–88.

Morin, M. (2009). Löneutrymmet i Djup Lågkonjunktur. Technical report, Landsorganisationen i Sverige, LO.

Nadal-De Simone, F. (2000). Forecasting Inflation in Chile Using State-Space and Regime-Switching Models. IMF Working Papers 00/162, International Monetary Fund.

Netšunajev, A. (2013). *Structural Vector Autoregressions with Markov Switching: Identification via Heteroskedasticity*. PhD thesis, European University Institute.

Pagliacci, C. and Barraez, D. (2010). A Markov-Switching Model of Inflation: Looking at the Future During Uncertain Times. *Análisis Económico*, 59(25):25–46.

Phelps, E. S. (1968). Money-Wage Dynamics and Labor Market Equilibrium. *Journal of Political Economy*, 76(4):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.

Poskitt, D. and Chung, S.-H. (1996). Markov Chain Models, Time Series Analysis and Extreme Value Theory. *Advances in Applied Probability*, 28(2):405–425.

Prospera (2014). Inflation Expectations. Press release, TNS Sifo Prospera.

Ricketts, N. and Rose, D. (1995). Inflation, Learning and Monetary Policy Regimes in the G-7 Economies. Working Paper 95-6, Bank of Canada.

Rotemberg, J. J. (1982). Monopolistic Price Adjustment and Aggregate Output. *The Review of Economic Studies*, 49(4):517–531.

Rudebusch, G. and Svensson, L. E. O. (1999). Policy Rules for Inflation Targeting. In *Monetary Policy Rules*, pages 203–262. University of Chicago Press.

Samuelson, P. A. and Solow, R. M. (1960). Analytical Aspects of Anti-Inflation Policy. *American Economic Review*, 50(2):177–194.

Schaling, E. (1999). The Non-Linear Phillips Curve and Inflation Forecast Targeting. Working Paper 98, Bank of England.

Simon, J. (1996). A Markov-Switching Model of Inflation in Australia. RBA Research Discussion Papers 9611, Reserve Bank of Australia.

Smith, A., Naik, P. A., and Tsai, C.-L. (2006). Markov-Switching Model Selection Using Kullback-Leibler Divergence. *Journal of Econometrics*, 134(2):553–577.

Smith, D. R. (2008). Evaluating Specification Tests for Markov-Switching Time-Series Models. *Journal of Time Series Analysis*, 29(4):629–652.

Solow, R. M. (1968). Recent Controversy on the Theory of Inflation: An Eclectic View. In Rousseas, S. W., editor, *Inflation: Its Causes, Consequences, and Control*. New York University.

Staiger, D., Stock, J. H., and Watson, M. W. (1997). The NAIRU, Unemployment and Monetary Policy. *Journal of Economic Perspectives*, 11(1):33–49.

Stock, J. H. and Watson, M. W. (2007). Why Has U.S. Inflation Become Harder to Forecast? *Journal of Money, Credit and Banking*, 39(1):3–33.

Stock, J. H. and Watson, M. W. (2009). Phillips Curve Inflation Forecasts. In Fuhrer, J., Kodrzycki, Y., Little, J., and Olivei, G., editors, *Understanding Inflation and the Implications for Monetary Policy*, chapter 3. MIT Press, Cambridge.

Stock, J. H. and Watson, M. W. (2010). Modeling Inflation after the Crisis. Working paper, National Bureau of Economic Research.

Svensson, L. E. O. (2015). The Possible Unemployment Cost of Average Inflation below a Credible Target. *American Economic Journal: Macroeconomics*, 7(1):258–96.

Tambakis, D. N. (1998). Monetary Policy with a Convex Phillips Curve and Asymmetric Loss. IMF Working Papers 98/21, International Monetary Fund.

Tobin, J. (1968). Discussion. In Rousseas, S. W., editor, *Inflation: Its Causes, Consequences, and Control*. New York University.

Weise, C. L. (1999). The Asymmetric Effects of Monetary Policy: A Nonlinear Vector Autoregression Approach. *Journal of Money, Credit and Banking*, 31(1):85–108.

White, H. (1987). Specification Testing in Dynamic Models. In *Advances in Econometrics, Fifth World Congress*, volume 1, pages 1–58.

Zhang, W. and Semmler, W. (2004). Monetary Policy with Endogenous NAIRU. Computing in Economics and Finance 2004 64, Society for Computational Economics.

A APPENDIX

A.1 DERIVATIONS

A.1.1 THE BLHK-FILTER AND SMOOTHER

How do we make inference on the regime probabilities ξ_t? This, it turns out, can be done using a relatively simple and straight-forward filter and smoothing algorithm. The algorithm could be seen as a discrete version of the Kalman (1960) filter, but because of its discrete support, allows for full derivation of the conditional distribution of the state variable, instead of just the two first moments as in the Kalman filter. The filtered and smoothed regime probabilities are used further for setting up the likelihood function and are important components in the estimation algorithm.

The filter is an iterative algorithm for predicting ξ_{t+1} on the basis of the observations $Y_t = (y'_t, y'_{t-1}, \ldots, y'_{t-p})'$. For every period $t = 1, \ldots, T$, the filter makes inference on the present regime ξ_t and a one-period-ahead forecast of the next coming regime ξ_{t+1} using the information up to point t. Hence, for each iteration, the information set used will increase by that period's observation, until the whole sequence of regime probabilities is determined.

Assuming that the parameters θ are known, the filter collects the conditional probability distribution of ξ_{t+1} in a vector $\hat{\xi}_{t+1|t}$, called the *predicted regime probabilities*,

$$\hat{\xi}_{t+1|t} = \mathsf{E}[\xi_{t+1|Y_t}] = \begin{bmatrix} Pr(\xi_{t+1} = \iota_1|Y_t) \\ \vdots \\ Pr(\xi_{t+1} = \iota_M|Y_t) \end{bmatrix} \tag{23}$$

Because of the binary nature of the components of $\hat{\xi}_{t+1|t}$, this is both the conditional mean and the probability distribution of ξ_{t+1} conditional on Y_t. In the same manner, the inference on the current regime vector ξ_t, the *filtered regime probabilities*, are

$$\hat{\xi}_{t|t} = \mathsf{E}[\xi_{t|Y_t}] = \begin{bmatrix} Pr(\xi_t = \iota_1|Y_t) \\ \vdots \\ Pr(\xi_t = \iota_M|Y_t) \end{bmatrix}. \tag{24}$$

Using matrix notation, the filtered probabilities are computed as

$$\hat{\xi}_{t|t} = \frac{\eta_t \odot \hat{\xi}_{t|t-1}}{\mathbf{1}'_M(\eta_t \odot \hat{\xi}_{t|t-1})} = \frac{\eta_t \odot \mathbf{F}(\hat{\xi}_{t-1|t-1})}{\mathbf{1}'_M(\eta_t \odot \mathbf{F}(\hat{\xi}_{t-1|t-1})}, \tag{25}$$

where $\mathbf{1}_M$ is a $(M \times 1)$ vector of ones, $\odot$ denotes element-wise matrix multiplication, and the second equality follows from the formulation of $\mathbf{F} = \mathbf{P}'$, which is defined exactly as the transition matrix between two adjacent regime vectors $\xi_{t+1} = \mathbf{F}\xi_t$. Hence, we can forecast the next period's regime and get the predicted probabilities $\hat{\xi}_{t+1|t}$:

$$\hat{\xi}_{t+1|t} = \mathbf{F}\hat{\xi}_{t|t} = \frac{\mathbf{F}(\eta_t \odot \hat{\xi}_{t|t-1})}{\mathbf{1}'_N(\eta_t \odot \hat{\xi}_{t|t-1})}, \tag{26}$$

If we have an initial regime vector ξ_0, we can recursively derive the filtered and predicted regime probabilities by iterating (25) and (26) for $t = 1, \dots, T$, to get the whole sequence of regime probabilities $\{\hat{\xi}_{t|t}\}_{t=1}^{T}$ and $\{\hat{\xi}_{t|t-1}\}_{t=1}^{T}$.

The smoother algorithm updates the filtered probabilities by using the whole sample information up until time T to get the *smoothed regime probabilities*, $\hat{\xi}_{t|T}$. Hence, the filtered probabilities are updated with the unused information $Y_{t+1.T} = (y'_{t+1}, \dots, y'_T)'$. While Hamilton (1988, 1989) derived $\hat{\xi}_{t|T}$ using the whole past history of the regime, Kim proposed a much less computationally demanding version of the smoother using the Markov property (13) to condense the necessary information set to condition on. However, as Hamilton (1994b) points out, this algorithm is only applicable when the Markov process is first-order, and the conditional density function (10) depends on $s_t, s_{t-1}, \dots$ only through s_t.

The smoother algorithm computes $\hat{\xi}_{t|T}$ by iterating backwards, $j = 1, \dots, T-1$, using the last filtered probability $\hat{\xi}_{T|T}$ and can be written as

$$\hat{\xi}_{T-j|T} = \left[\mathbf{F}'(\hat{\xi}_{T-j+1|T} \oslash \hat{\xi}_{T-j+1|T-j})\right] \odot \hat{\xi}_{T-j|T-j} \tag{27}$$

where $\oslash$ denotes element-wise matrix division.

A.1.2 The EM Algorithm

The EM algorithm iterates over two separate steps: (i) The *Expectations step* (E), where the unobserved regimes ξ_t are estimated, using the smoothed probabilities $\hat{\xi}_{t|T}$. In addition, the conditional probabilities $Pr(\xi|Y, \lambda^{(j-1)})$ are computed from the BLHK-filter and smoother using the parameter vector $\lambda^{(j-1)}$ from the previous iteration. (ii) The *Maximization step* (M), where an estimate of λ is derived from the solution of the first order conditions of the maximization of the likelihood function using the smoothed probabilities $\hat{\xi}_{t|T}(\lambda^{(j-1)})$ from the last E-step.

Hence, each iteration of the EM algorithm involves a pass through the BLHK filter and smoother at the E-step, and the following estimation of the parameters at the M-step, that are subsequently used in the next iteration's E-step, and so on. The likelihood function can be derived as a by-product of the BLHK filter and smoother,

$$\begin{aligned} L(\lambda|Y) &\equiv p(Y_T|Y_0;\lambda) \\ &= \prod_{t=1}^{T} p(Y_t|Y_{t-1},\lambda) \\ &= \prod_{t=1}^{T} \sum_{\xi_t} p(y_t|\xi_t, Y_{t-1}, \theta) Pr(\xi_t|Y_{t-1},\lambda) \\ &= \prod_{t=1}^{T} \eta_t' \hat{\xi}_{t|t-1} = \prod_{t=1}^{T} \eta_t' \mathbf{F} \hat{\xi}_{t-1|t-1} \end{aligned} \tag{28}$$

where η_t is a vector containing the conditional densities $p(y_t|\xi_t, Y_{t-1})$ for each state. The maximum likelihood (ML) estimates are then obtained by maximizing the likelihood function $L(\lambda|Y)$ given the adding-up restrictions,

$$\begin{aligned}\mathbf{P1}_M &= 1\\ \mathbf{1}'_M\xi_0 &= 1,\end{aligned}$$

and the non-negativity restrictions,

$$\rho \geq \mathbf{0}, \sigma \geq \mathbf{0}, \xi_0 \geq \mathbf{0}.$$

The ML estimate $\tilde{\lambda}$ is the solution to the first-order conditions of the maximization of the constrained log-likelihood function,

$$\ln L(\lambda) \;\equiv\; \ln L(\lambda|Y_T) - \kappa'_1(\mathbf{P1}_M - \mathbf{1}_M) - \kappa_2(\mathbf{1}'_M\xi_0 - 1).$$

where κ_1 and κ_2 are the Lagrange multipliers for the adding-up respectively non-negativity restrictions. The FOCs are hence given by

$$\begin{aligned}\frac{\partial \ln L(\lambda|Y_T)}{\partial \theta'} &= \mathbf{0}\\ \frac{\partial \ln L(\lambda|Y_T)}{\partial \rho'} - \kappa'_1(\mathbf{1}'_M - \mathbf{I}_M) &= \mathbf{0}\\ \frac{\partial \ln L(\lambda|Y_T)}{\partial \xi'_0} - \kappa_2\mathbf{1}'_M &= \mathbf{0}.\end{aligned}$$

The iteration of the E and M-steps are guaranteed to increase the value of the likelihood function, so that it finally reaches a fixed-point where $\tilde{\lambda}^{(j)} = \tilde{\lambda}^{(j-1)}$ coincides with the maximum of the likelihood function $L(\lambda|Y)$.

A.1.3 Constrained Estimator for MSAH-VAR

To rewrite the estimator into a form allowing for constraints on any number of arbitrary parameters, the order of parameters must be rearranged such that the ones constrained are positioned at the end of the vector. This implies that the matrix $\bar{X}_m$ also must be rearranged as to comply with the positions of the parameters. Since $\bar{X}_m$ contains the regressors for all K equations, changing the position of a given column will effect the corresponding parameter in all K equations and hence the possibility to apply restrictions for an individual equation is ruled out. To overcome this, we define $\bar{X}^*_m \equiv \bar{X}_m \otimes \Sigma^{-1}_m$ which allows us to rearrange columns corresponding to individual parameter constraints. Obviously, the shift of columns for a parameter in one regime, implies that we must shift the matching columns for all regimes.

$$\tilde{\beta} = \left(\sum_{m=1}^{M}\left(\bar{X}'_m\hat{\Xi}_m\bar{X}_m\right)\otimes\tilde{\Sigma}^{-1}_m\right)^{-1}\left(\sum_{m=1}^{M}\left(\bar{X}'_m\hat{\Xi}_m\right)\otimes\tilde{\Sigma}^{-1}_m\right)\mathbf{y} \tag{29}$$

The next step is to rearrange the estimator such that it can be rewritten in terms of $\bar{X}^*_m$. For the first term in the estimator we can rewrite it in the following manner.

$$\left(\bar{X}'_m\hat{\Xi}_m\bar{X}_m\right)\otimes\tilde{\Sigma}^{-1}_m = \left(\left(\hat{\Xi}_m\bar{X}_m\right)'\bar{X}_m\right)\otimes\left(\left(\tilde{\Sigma}_m\times\tilde{\Sigma}^{-1}_m\right)\tilde{\Sigma}^{-1}_m\right)$$

$$= \left(\left(\hat{\Xi}_m \bar{X}_m\right)' \otimes \left(\tilde{\Sigma}_m \times \tilde{\Sigma}_m^{-1}\right)\right) \left(\bar{X}_m \otimes \tilde{\Sigma}_m^{-1}\right)$$

$$= \left(\left(\hat{\Xi}_m \bar{X}_m\right) \otimes \left(\tilde{\Sigma}_m \times \tilde{\Sigma}_m^{-1}\right)\right)' \left(\bar{X}_m \otimes \tilde{\Sigma}_m^{-1}\right)$$

$$= \left(\left(\hat{\Xi}_m \otimes \tilde{\Sigma}_m\right) \left(\bar{X}_m \otimes \tilde{\Sigma}_m^{-1}\right)\right)' \left(\bar{X}_m \otimes \tilde{\Sigma}_m^{-1}\right)$$

$$\left(\bar{X}_m' \hat{\Xi}_m \bar{X}_m\right) \otimes \tilde{\Sigma}_m^{-1} = \left(\bar{X}_m \otimes \tilde{\Sigma}_m^{-1}\right)' \left(\hat{\Xi}_m \otimes \tilde{\Sigma}_m\right)' \left(\bar{X}_m \otimes \tilde{\Sigma}_m^{-1}\right) \tag{30}$$

For the second term, the preceding equivalence hold.

$$\left(\bar{X}_m' \hat{\Xi}_m\right) \otimes \tilde{\Sigma}_m^{-1} = \left(\hat{\Xi}_m \bar{X}_m\right)' \otimes \tilde{\Sigma}_m^{-1}$$

$$= \left(\left(\hat{\Xi}_m \bar{X}_m\right) \otimes \left(I \times \tilde{\Sigma}_m^{-1}\right)\right)'$$

$$= \left(\left(\hat{\Xi}_m \otimes I\right) \left(\bar{X}_m \otimes \tilde{\Sigma}_m^{-1}\right)\right)'$$

$$\left(\bar{X}_m' \hat{\Xi}_m\right) \otimes \tilde{\Sigma}_m^{-1} = \left(\bar{X}_m \otimes \tilde{\Sigma}_m^{-1}\right)' \left(\hat{\Xi}_m \otimes I\right)' \tag{31}$$

Substituting in $\bar{X}_m^*$ into (31) and (30), (29) can be rewritten as.

$$\tilde{\beta} = \left(\sum_{m=1}^{M} \bar{X}_m^{*'} \left(\hat{\Xi}_m \otimes \tilde{\Sigma}_m\right) \bar{X}_m^*\right)^{-1} \left(\sum_{m=1}^{M} \bar{X}_m^{*'} \left(\hat{\Xi}_m \otimes I\right)'\right) \mathbf{y}$$

Now we can impose restrictions using standard time-series procedures[51] and rewrite the estimator into its ultimate form.

$$\tilde{\gamma} = \left(\sum_{m=1}^{M} R' \bar{X}_m^{*'} \left(\hat{\Xi}_m \otimes \tilde{\Sigma}_m\right) \bar{X}_m^* R\right)^{-1} \left(\sum_{m=1}^{M} R' \bar{X}_m^{*'} \left(\hat{\Xi}_m \otimes I\right)' \left(\mathbf{y} - \bar{X}_m^{*'} r\right)\right)$$

Where R is a pick-matrix and r contains the constraints such that the following holds.

$$\tilde{\beta} = R\tilde{\gamma} + r$$

---

[51] See e.g. Lütkepohl (2007), section on linear constraints.

A.2 TABLES

TABLE 4: Likelihood Ratio Tests

	MSI-VAR(1)	MSI-VAR(2)	MSIH-VAR(1)	MSIH-VAR(2)	MSA-VAR(1)	MSA-VAR(2)	MSAH-VAR(1)	MSAH-VAR(2)	MSIA-VAR(1)	MSIA-VAR(2)	MSIAH-VAR(1)	MSH-VAR(2)
LR (H_0: vech(Σ_i) = vech(Σ_m))	39.625	30.283			23.729	20.210			18.765	21.003		
Critical value	7.815	7.815			7.815	7.815			7.815	7.815		
p-value	0.000	0.000			0.000	0.000			0.000	0.000		
LR (H_0: $a_i = a_m$)	25.592	25.950	4.731	16.670								
Critical value	9.488	15.507	9.488	15.507								
p-value	0.000	0.001	0.316	0.034								
LR (H_0: $v_i = v_m$)					14.819	7.370	9.855	8.164				
Critical value					5.992	5.992	5.992	5.992				
p-value					0.001	0.025	0.007	0.017				
LR (H_0: vech(Σ_i) = vech(Σ_m) $\cup$ $a_i = a_m$)	44.356	46.953										43.517
Critical value	14.067	19.675										15.507
p-value	0.000	0.000										0.000
LR (H_0: vech(Σ_i) = vech(Σ_m) $\cup$ $v_i = v_m$)					33.584	28.373						
Critical value					11.071	11.071						
p-value					0.000	0.000						
LR (H_0: $p = 1$)	52.952		43.610		60.759		57.240		53.310		55.549	
Critical value	9.488		9.488		9.488		9.488		9.488		9.488	
p-value	0.000		0.000		0.000		0.000		0.000		0.000	

Notes: LR statistic, critical values, and p-values. The null hypothesis in parenthesis. The test under the null has a χ^2-distribution, with r degrees of freedom.

A.3 FIGURES

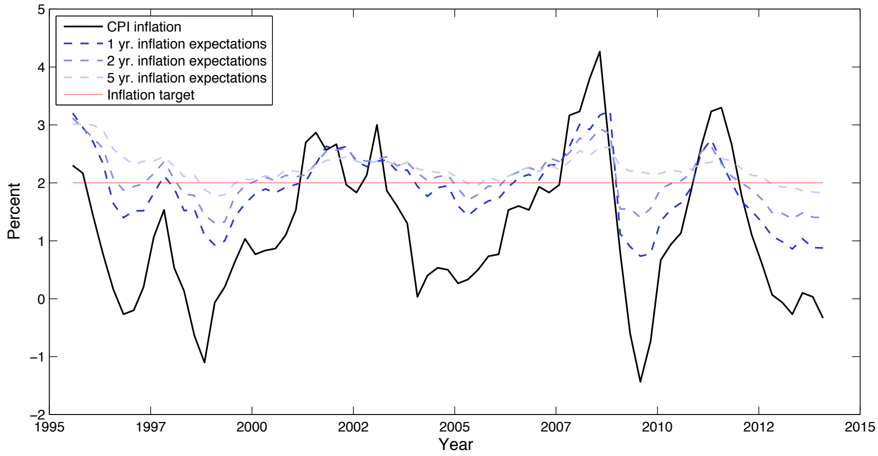


Figure 5: CPI inflation and Prospera inflation expectations.

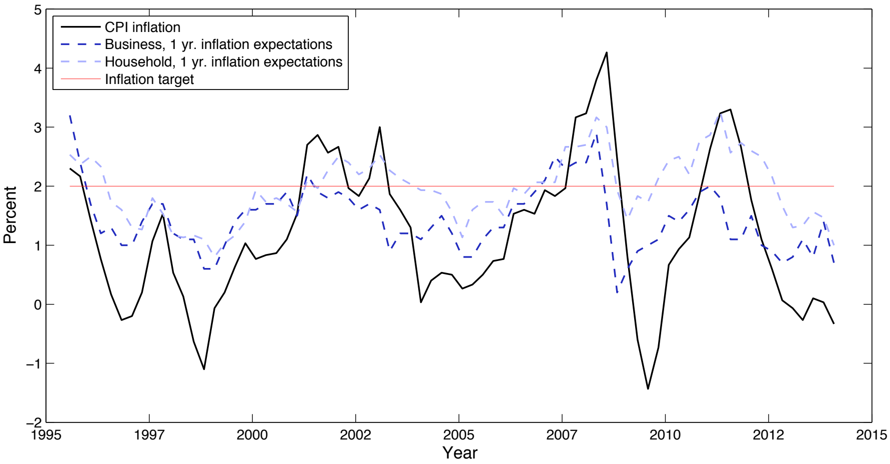


Figure 6: CPI inflation and NIER inflation expectations.

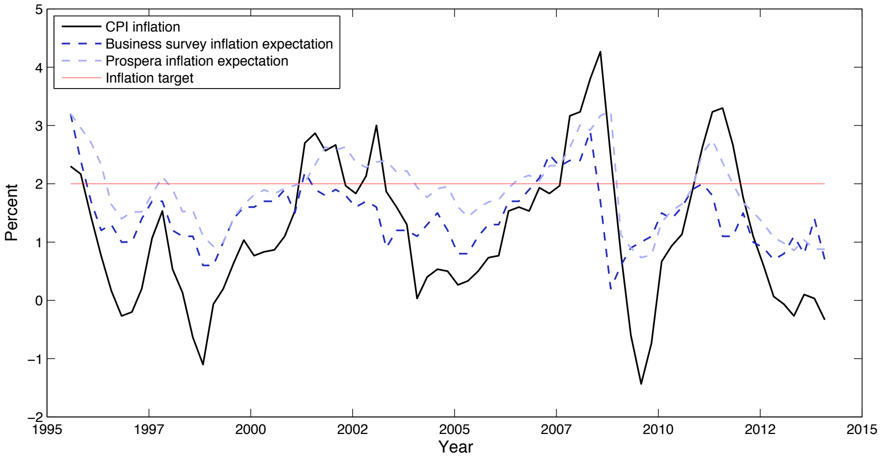


Figure 7: CPI inflation and NIER and Prospera inflation expectations.

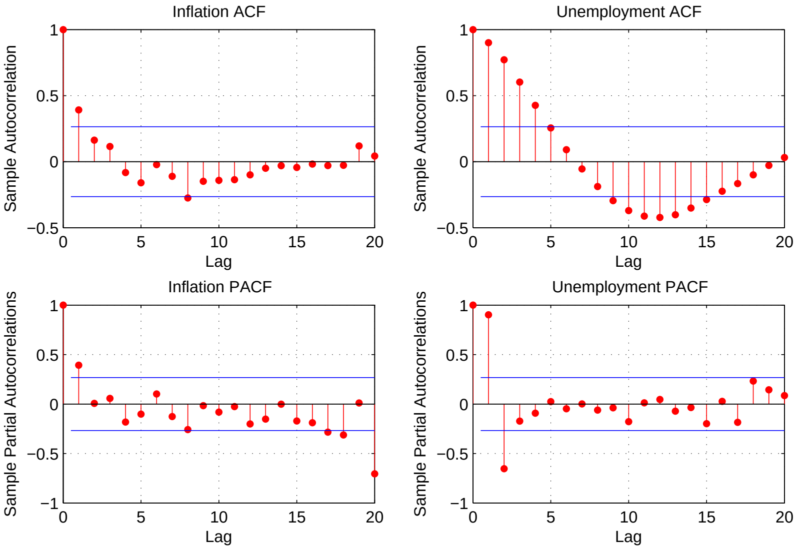


Figure 8: Empirical autocorrelations

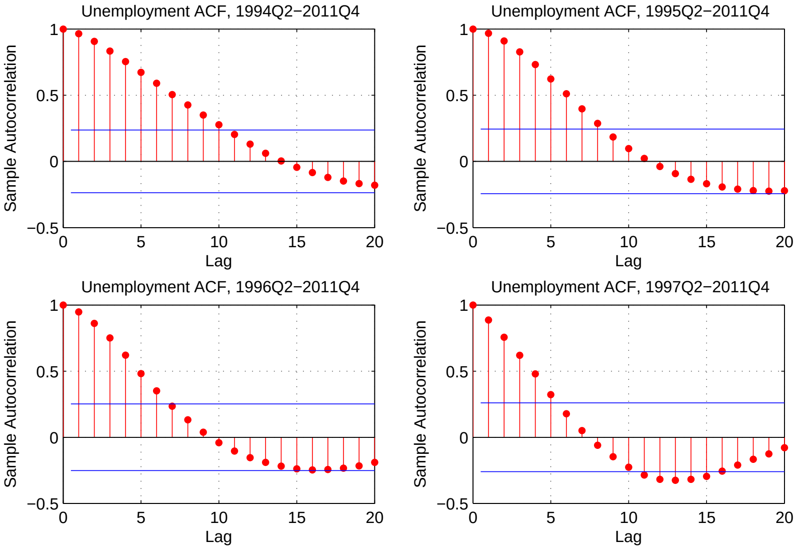
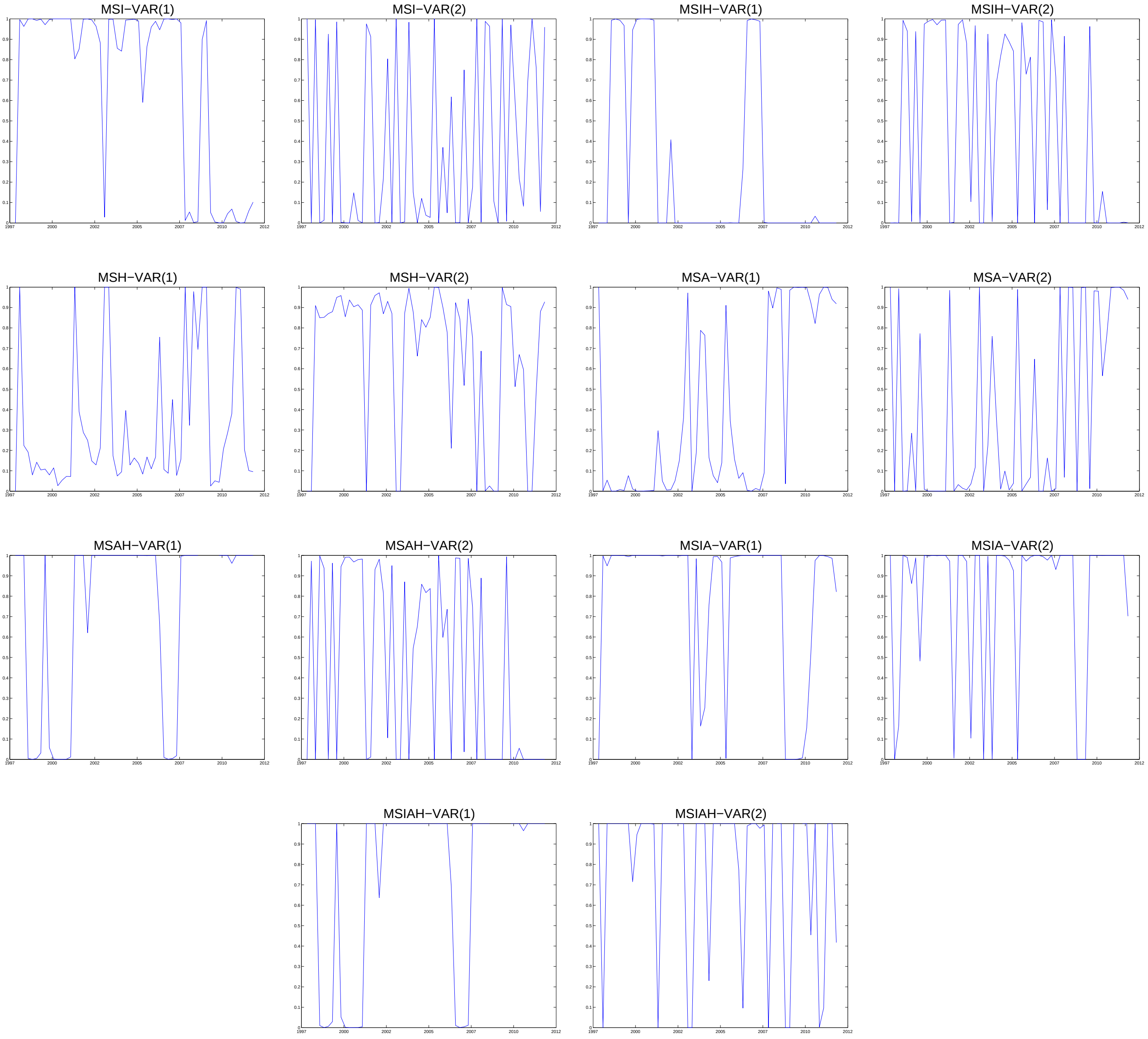


Figure 9: ACF’s of unemployment for different sample periods



MSI−VAR(1)
MSI−VAR(2)
MSIH−VAR(1)
MSIH−VAR(2)
MSH−VAR(1)
MSH−VAR(2)
MSA−VAR(1)
MSA−VAR(2)
MSAH−VAR(1)
MSAH−VAR(2)
MSIA−VAR(1)
MSIA−VAR(2)
MSIAH−VAR(1)
MSIAH−VAR(2)

Figure 10: Smoothed probability for regime 1 of all models

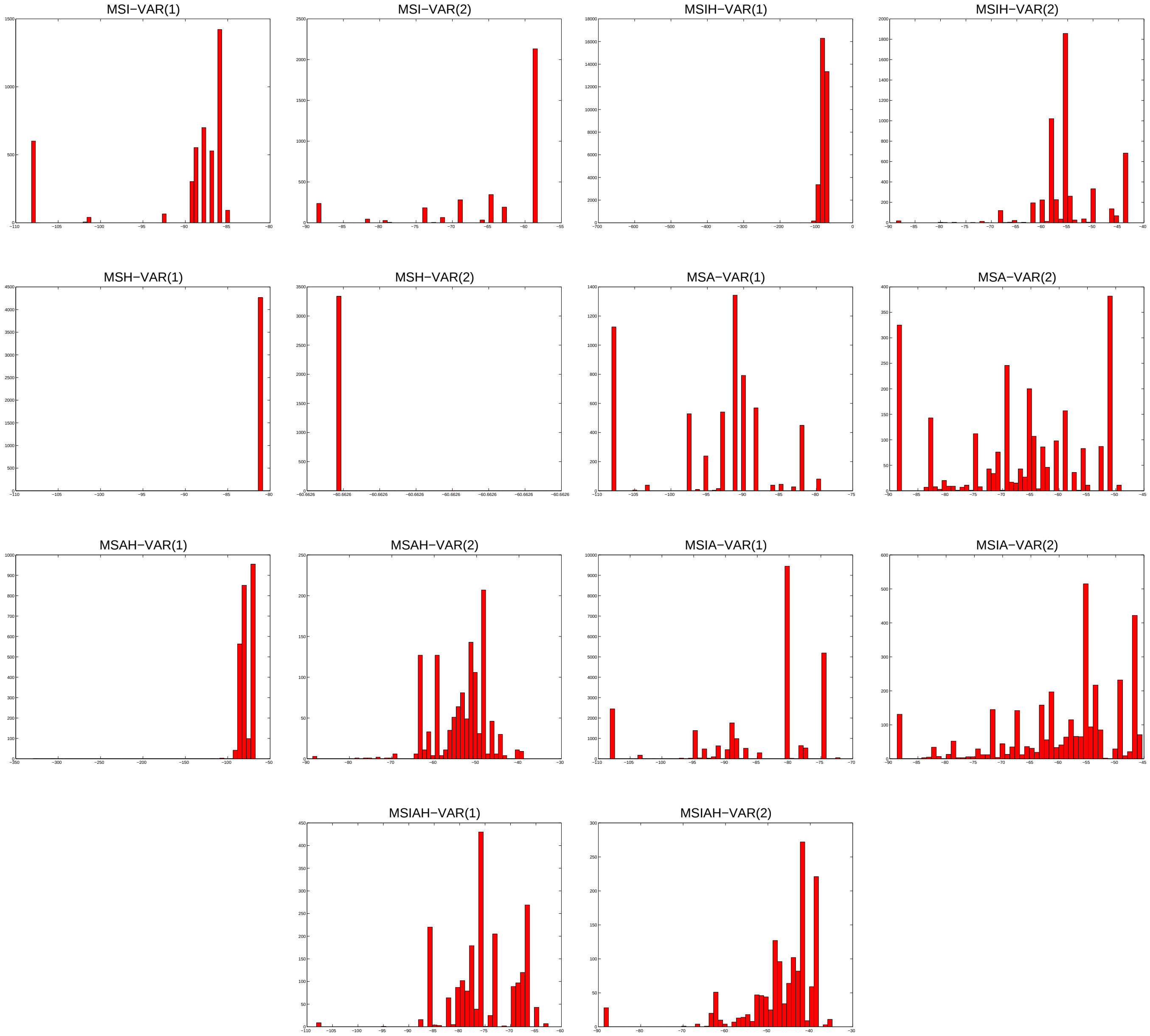


Figure 11: Frequency and log likelihood for different starting values

Figure 12: Conditional residuals of the MSAH-VAR(2)

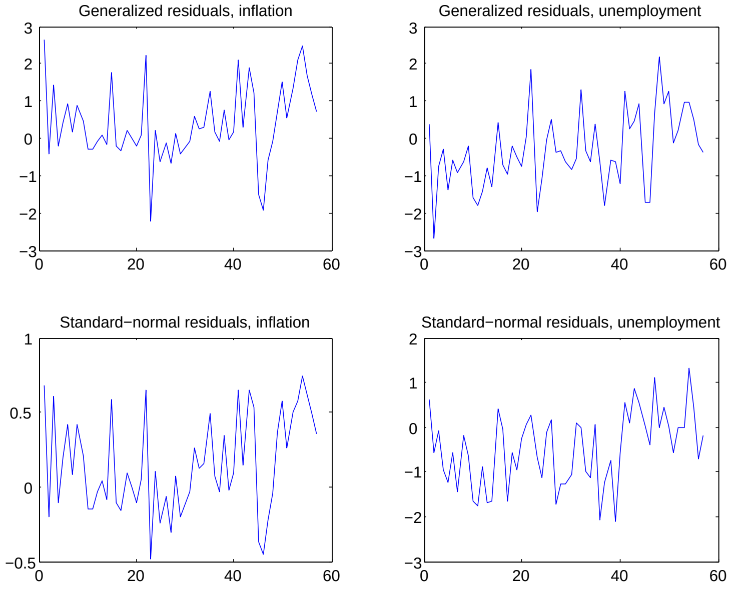


Figure 13: Generalized and Rosenblatt-transformed residuals of the MSAH-VAR(2)

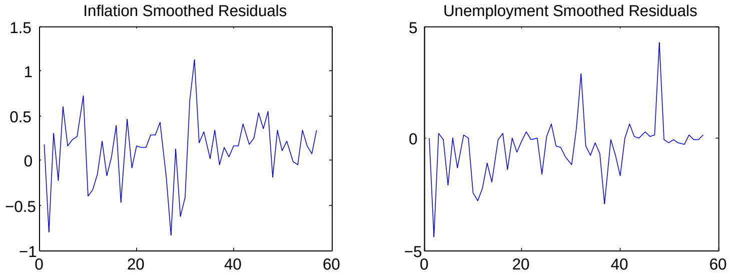


Figure 14: Smoothed residuals of the MSAH-VAR(2)

A.4 Matlab

In this section follows a short description of our approach when building the estimation models in matlab, followed by an example of the code for the most flexible model, MSIAH-VAR. To ensure consistency and avoid errors in the code, we have created a simulation section (see: A.4.1), in which we can generate data from some given parameters. The advantage of this approach is that we know the true parameters of the data and thus can evaluate the performance of our estimators regarding both the estimation of parameters from the observable data and the hidden Markov-chain. The next script A.4.2 prepares the data and sorts the order of the regressors in the event of constraints, such that it complies with the estimator. The algorithm for starting values A.4.3 makes a linear regression and places a grid around it. From the combinations of the grid around the linear regression, we draw a random sample, which serves as the starting values for our estimation. The starting values are entered into the EM-script A.4.4, which starts the BLHK-filter A.4.5 and generates state probabilities for the given estimates. The state probabilities are returned into the EM-script where they serve as weights for the ML-estimation. The procedure continues iteratively until the parameters have converged and the estimates and state probabilities are presented.

A.4.1 Simulation

```
function [ data, beta_gen,P, A,S,eps,sigma_gen ] = ...
data_simulation_MSIAH(K,T,p,M,constraints,r)

if M<1|M>3|K<1|K>3 |p<1|p>4
    error('The following intervals are accepted as inputs M [2,3],K [1,3], p [1,4]')
end
%% I Initialization

                %% TRANSITION PROBABILITIES
                % Transitions probabilities from state 1
p11 = 0.8;
p12 = 0.1;
p13 = 1-p11-p12;

                % Transitions probabilities from state 2
p21 = 0.1;
p22 = 0.8;
p23 = 1-p21-p22;
                % Transitions probabilities from state 3
p31 = 0.03;
p32 = 0.05;
p33 = 1-p31-p32;

P = [p11 p12 p13; p21 p22 p23; p31 p32 p33];

if M==2
```

```
    P = P(1:M,1:M);
    P(1,:) = P(1,:)./sum(P(1,:),2);
    P(2,:) = P(2,:)./sum(P(2,:),2);
end

if min(min(P))==0
    error('All transition probabilities must be positive')
end
                %% DATA GENERATING PARAMETERS
                % Starting values for y
start_data(:,1) = [0.3 0.4 0.2 0.12 0.4 0.32 0.08 0.24 0.1 0.64 0.43 0.31]';
                % Standard errors for the data generating process
                % State 1
std_dev_f1 = .3;
std_dev_x1= .5;
std_dev_z1 = .5;

                % State 2
std_dev_f2 = .2;
std_dev_x2 = .5;
std_dev_z2 = .7;

                % State 3
std_dev_f3 = 0.3;
std_dev_x3 = 0.5;
std_dev_z3 = 1.2;


std_dev(:,:,1)  = [std_dev_f1; std_dev_x1; std_dev_z1] ;
std_dev(:,:,2)  = [std_dev_f2; std_dev_x2; std_dev_z2] ;
std_dev(:,:,3)  = [std_dev_f3; std_dev_x3; std_dev_z3] ;

std_dev = std_dev(:,:,1:M);

for s=1:M
sigma_gen(:,:,s) = diag(std_dev(1:K,:,s).^2);
end

varz = char('f','x','z');

% STATE 1
% Intercepts
f_0_1 = -2;
x_0_1 = -0.3;
z_0_1 = -0.2;

% AR lag 1 state 1
f_f_1_1 = 0.1;
x_f_1_1 = 0.7;
z_f_1_1 = 0.1;
.
.
.
```

We have excluded the rest of the parameters since they make up a very extensive list

```
for i = 1:M         % This loop runs over the states
for j = 1:K         % This loop runs over the variables
for s = 1:p         % This loop runs over the lags

                    % This if-condition is in order to write out the
                    % intercepts and the parameters of the first lag
    if s==1

    eval(['A(j,1,i)=' sprintf('%s',varz(j)) '_0_' num2str(i) ';']);

    end

                    % This loop writes out the the variables we are
                    % regressing on for all the lags
    for q = 1:K

       eval(['A(j,(K*(s-1)+1)+q,i)=' sprintf('%s',varz(j)) '_' ...
        sprintf('%s',varz(q)) '_' num2str(s) '_' num2str(i) ';']);
    end

end
end
end

A(find(constraints==1))=r(find(constraints==1));
for i = 1:M
beta_gen(:,i) = reshape(A(:,:,i),K^2*p+K,1)';
beta_gen(:,i) = beta_gen([find(constraints(:,i)==0); find(constraints(:,i)==1)],i);
    %eval(['beta' num2str(i) '_gen = reshape(A(:,:,i),K^2*p+K,1);'])
end

rand_st = randn(T,1);
S=zeros(T,1);
S(1,1) = 1;     % State that the first states is equal to 1

for t = 1:T-1
                % This sums upp P along the 2nd dimension for the state
                % row
    temp_P = cumsum(P(S(t),1:M),2);
    i =1;
    while cdf('Normal',rand_st(t),0,1)-temp_P(i)>0

      i = i+1;
    end
    S(t+1) =  i;
end

%% This part generates the data by using the generated states in the previous section
```

```
data = zeros(K*T+K*p,1);
data(1:K*p,1) = start_data(1:K*p,1);
% Plugging in the starting values as first data point.

for t=1:T
eps(:,t)=(diag(sigma_gen(:,:,S(t)))).^(1/2).*randn(K,1);
end

for t=1:T;
    data(t*K+p*K-K+1:t*K+p*K,1) = flipud(A(:,1,S(t))+A(:,2:end,S(t))...
        *flipud(data(K*(t-1)+1:t*K+p*K-K,1))+eps(:,t));
end

data=reshape(data,K,T+p);
data = flipud(data);
data = data';

end
```

A.4.2 LOADING AND PREPARING DATA

```
den_var1 = 1;
den_var2 = 1;
den_var3 = 0;
den_use = [den_var1 den_var2 den_var3];
K = 2;
p = 1;
M = 2;

%% LOADING DATA

load_inf_s = 1;                 % Svensson inflation
load_unem_s = 1;                % Svensson unemployment
load_inf_exp_1_pro = 0;         % Loads inflation expectations from prospera 1 yr
load_inf_exp_2_pro = 0;         % Loads inflation expectations from prospera 2 yr
load_inf_exp_5_pro = 0;         % Loads inflation expectations from prospera 5 yr
load_inf_exp_hh = 0;            % Loads inflation expectations from hh, KI
load_inf_exp_co = 0;            % Loads inflation expectations from companies, KI
load_inter = 0;                 % Loads repo-rate
load_gdp_gap = 0;               % Loads gdp gap

load = [0 0 0 load_inf_s load_unem_s load_inf_exp_1_pro...
load_inf_exp_2_pro load_inf_exp_5_pro...
load_inf_exp_hh load_inf_exp_co load_inter load_gdp_gap];

cutoff_end = 9;                 % How many periods are cut off at the end
cutoff_beg = 13;                % How many observation are cut off at the
                                % beginning.

data_temp = importdata('Data_SMSW.xlsx');
data = data_temp.data;
date = datenum(data(1:end,1),data(1:end,2),0,0,0,0);
```

```
data = data(1+cutoff_beg:end-cutoff_end,load==1);
date = date(1+cutoff_beg:end-cutoff_end);

%% CONSTRAINTS

con_ind = []';                    % Indices of the parameters to be constrained
constraints = zeros(K^2*p+K,M);
constraints(con_ind)=1;
con_val = []';                    % Set the value for the constrained parameter
r = zeros(K^2*p+K,M);
if sum(con_ind)>0
    r(con_ind) = con_val;
end

T = size(data,1)-p;
y = reshape(data(p+1:end,:)',K*T,1);

% Matrix of regressors
Xbar0 = ones(T,K*p+1);
for t = 1:T
    for k = 1:p
        Xbar0(t,2+k*K-K:1+k*K)=data(t+p-k,:);
    end
end

Xbar = repmat(kron(Xbar0,eye(K)),[1 1 M]);
for s = 1:M
    Xbar(:,:,s) = Xbar(:,[find(constraints(:,s)==0); find(constraints(:,s)==1)],s);
end
if (size(Xbar,1)~=T*K) | (size(Xbar,2)~=K*(K*p+1))
    error('Xbar dimensions not correct')
end

R = cell(M,1);
temp = eye(K^2*p+K);
for s = 1:M
    temp2 = temp;
    temp2(end-sum(constraints(:,s),1)+1:end,:)=[];
    R{s} = temp2';
    r_est(:,s)= r([find(constraints(:,s)==0); find(constraints(:,s)==1)],s);
end

initial_algo_max_MSIAH          % Algorithm for generating starting values

%% Start iteration
iterate = zeros(size(initial_beta,3),1);

parfor w=1:size(initial_beta,3)
beta = initial_beta(:,:,w);
[lambda,xit,xiT,xiT2,xit1,eta,Utilde,i] = EM_MSIAH(sigma,beta,P,xi0,data,...
Xbar,K,y,p,T,den_use,M,R,r_est,w);
iterate(w,1) = i;
final_estimate(:,:,w) = lambda;
```

```
final_estimate_prob(:,:,w) = catpad(1,xiT,xit);
likelihood(w,1) = sum(diag(log(eta)'*xiT));
end
```

A.4.3 ALGORITHM FOR STARTING VALUES

```
if M~=2
error('This script works only for M=2')
end

% INITIAL VALUE ALGORITHM IS USED FOR MAX PARAMETERS

min_initial = -3;           % Min. value that we run the grid over
max_initial = 3;            % Max. value that we run the grid over
grid_steps = 5;             % Number of steps on our grid

for i=1:grid_steps
    grid(i) = min_initial + (i-1)*((max_initial-min_initial)/(grid_steps-1));
end


initial0=kron(inv((Xbar0'*diag(ones(1,T))*Xbar0))*Xbar0'*diag(ones(1,T)),eye(K))*y;

P = [0.5 0.5; 0.5 0.5];     % Initial guess for the Markov parameters
xi0 = [0.4 0.6]';           % Initial guess for the initial state

% Initial guess for covarince matrix

B_1=reshape(initial0,K,(K*p+1));
Y(1:T,:)=data(p+1:T+p,:);
One_M=ones(2,1);            % Vector of ones with M as length.

Utilde=kron(One_M,Y)-kron(eye(2),Xbar0)*([B_1 B_1]');
Xi=diag([ones(1,T) zeros(1,T)]');
sigma=T^(-1)*Utilde'*Xi*Utilde;
sigma = repmat(sigma,[1 1 M]);

var_weights = reshape(repmat(diag(sigma(:,:,1)),1,K*p+1),K^2*p+K,1);


if sum(den_use,2)<K         % We use a limited number of equations to determine the density

error('The search algorithm is not checked for sum(den_use,2)<K')

initial1 = reshape(getfield(reshape(initial0,K,K*p+1),...
{find(den_use==1),1:K+1}),sum(den_use)*(K+1),1);

elements = sum(den_use)*(K*p+1);
% This an auxilary vector for the loop
beta_intermed = zeros(elements,1);
n=0;
k=0;
```

```
beta = zeros(sum(den_use)*(K+1),1,size(grid,2)^elements);
[k,beta] = loop(k,elements,grid,n,beta,beta_intermed,initial1,var_weights);

temp = repmat(reshape(initial0,K,K*p+1),[1,1,size(grid,2)^elements]);

temp(find(den_use==1),:,:) = reshape(beta,[sum(den_use),(K+1),size(grid,2)^elements]);

temp = reshape(temp,[K+K^2,size(grid,2)^elements,1]);

beta1(:,1,1:size(grid,2)^(2*elements))=KronProd({temp,ones(1,size(grid,2)^elements)});
beta2(:,1,1:size(grid,2)^(2*elements))=kron(ones(1,size(grid,2)^elements),temp);

clear beta temp initial1

elseif sum(den_use,2)==K

    elements=size(initial0,1);
    % This an auxilary vector for the loop
    beta_intermed=zeros(elements,1);
    n=0;
    k=0;
    beta=zeros(K+p*K^2,1,size(grid,2)^elements);
    [k,beta,n] = loop(k,elements,grid,n,beta,beta_intermed,initial0,var_weights);
end
                            % Number of chosen values from the initial
                            % values algorithm

if nr_init>size(grid,2)^(M*(elements-K^2*p)+K^2*p)
    nr_init=size(grid,2)^(M*(elements-K^2*p)+K^2*p);
end
                            % Indices for the nr_init selected
                            % startingvalues
                            if size(grid,2)^(M*elements)< 10^10
initial_beta_ind = [datasample(linspace(1,size(grid,2)^(M*elements),...
size(grid,2)^(M*elements)),nr_init,'Replace',false)];
% Reducing the number of initial values by drawing a random sample
                            else
initial_beta_ind = randi([0 size(grid,2)^(M*elements)],1,nr_init);
                            end
                            % The set of starting values for all states are
                            % created
initial_beta = [beta(:,:,ceil(initial_beta_ind/size(beta,3))) ...
 beta(:,:,initial_beta_ind-size(beta,3)...
    *(ceil(initial_beta_ind/size(beta,3))-1))];
```

A.4.4 EM-ALGORITHM

```
function [lambda,xit,xiT,xiT2,xit1,eta,Utilde,i] = ...
EM_MSIAH(sigma,beta,P,xi0,data,Xbar,K,y,p,T,den_use,M,R,r_est,w)

it = w
```

```
% Check for positive definite Sigma
for s=1:M
[~,pos] = chol(sigma(:,:,s));
if pos~=0
    error('Sigma is not positive definite')
end
clear pos
end

tol = 10^−5;                            % Tolerance of convergence critera
i = 1;
lambda(:,:,i) = catpad(1,beta,reshape(sigma,K,K*M),P,xi0');


while convergence(lambda,i)>tol
        if i>1000
        break
end
%% II Expectation Step

[xit, xiT, xiT2,xit1,eta] = BLHK_filter_MSIAH(data, Xbar, lambda, i,K,T,p,den_use,M);

% Obtaining the filtered and smoothed regime probabilities from the
% BLHK filter.

%% III Maximization Step

joint = sum(xiT2(:,1:end−1),2);
marg = sum(xiT(:,1:end−1),2);

for s = 1:M
    for q = 1:M
        P(s,q) = joint(q*M−M+s)/marg(s);
    end
end

% 2. Regression Step: Normal Equations for the parameters

% Estimation of the beta coefficients
Y(1:T,:) = data(p+1:T+p,:);
Ttilde = sum(xiT,2);
for s = 1:M
    Xi1(:,:,s) = diag(xiT(s,:));
    beta(:,s) = R{s}*(inv(R{s}'*(Xbar(:,:,s)'*KronProd({Xi1(:,:,s),...
    eye(K)},[2 1])*Xbar(:,:,s)*R{s}))*R{s}'*((KronProd({Xi1(:,:,s),eye(K)},...
    [2 1]))*Xbar(:,:,s))'*(y−(Xbar(:,:,s)*r_est(:,s))))+r_est(:,s);
    B(:,:,s) = reshape(beta(:,s),K,(K*p+1));
end

% Estimation of the sigma matrix (heteroskedasticity)
Utilde = kron(ones(M,1),Y)−...
reshape(multiprod(Xbar,permute(beta(:,1:M),[1 3 2]),[1 2]),K,T*M)';
```

```
for s = 1:M
    sigma(:,:,s) = (Ttilde(s))^(−1)*Utilde(1+(s−1)*T:s*T,:)'*...
    Xi1(:,:,s)*Utilde(1+(s−1)*T:s*T,:);
end

xi0 = xiT(:,1);

%% IV Final steps

% Updating lambda

i = i+1;
lambda(:,:,i) = catpad(1,beta,reshape(sigma,K,K*M),P,xi0');

end
lambda = lambda(:,:,end);

end
```

A.4.5 BLHK

```
function [xit,xiT,xiT2,xit1,eta] = BLHK_filter_MSIAH(data,Xbar,lambda,i,K,T,p,den_use,M)
% BLHK_filter
K2 = sum(den_use,2);        % Number of variables that are used in the BLHK filter
beta = lambda(1:K^2*p+K,1:M,i);
sigma = reshape(lambda(K^2*p+K+1:K^2*p+2*K,1:M*K,i),K,K,M);
P = lambda(K+K^2*p+K+1:K+K^2*p+K+M,1:M,i);
xi0 = lambda(K^2*p+2*K+3,1:M,i)';
F = P';             % The transition matrix (transformed P matrix)
xit = zeros(M,T);   % Matrix containing the filtered regim probabilities
xit1 = zeros(M,T);  % Matrix containing the predicted regim probabilities
xiT = zeros(M,T);   % Matrix containing the smoothed regim probabilities
xiT2 = zeros(M^2,T);% Matrix containing the joint regim probabilities
eta = zeros(M,T);   % Vector containing the probability densities

%% Start the filter

for t=1:T           % Forward recursion t = 1, ... , T

for s =1:M
    temp(:,s) = Xbar((t−1)*K+1:(t−1)*K+K,:,s)*beta(:,s);
    ybar(:,s) = temp(find(den_use==1),s);
end

sig_index = [kron(find(den_use==1)',ones(sum(den_use,2),1)) ...
kron(ones(sum(den_use,2),1),find(den_use==1)')];

for s = 1:M
    sigma2(:,:,s) = (reshape(sigma(sub2ind(size(sigma(:,:,s)), ...
        sig_index(:,1), sig_index(:,2))+(s−1)*(K*K))...
        ,sqrt(size(sig_index,1)),sqrt(size(sig_index,1))))';
    eta(s,t) = (2*pi)^(−K2/2)*det(sigma2(:,:,s))^(−1/2)*...
```

```
    exp(-(1/2)*(data(t+p,find(den_use==1))'-ybar(:,s))'...
        *(sigma2(:,:,s)^-1)*(data(t+p,find(den_use==1))'-ybar(:,s)));
end

% Solving for the posterior inference (filtered regime probabilities),
% using the prior inferences and the conditional densities.

if t==1
    xit(:,t) = (eta(:,t).*xi0)/(ones(M,1)'*(eta(:,t).*xi0));
else
    xit(:,t) = (eta(:,t).*xit1(:,t-1))/(ones(M,1)'*(eta(:,t).*xit1(:,t-1)));
end
    xit1(:,t) = F*xit(:,t);
end

%% Start the smoother

for s=T-1:-1:1      % Backward recursion j = 1, ... , T-1

if s==T-1
    xiT(:,s) = xit(:,s).*(F'*(xit(:,s+1)./xit1(:,s)));
else
    xiT(:,s) = xit(:,s).*(F'*(xiT(:,s+1)./xit1(:,s)));
end
end

xiT(:,T) = xit(:,T);


for s=1:T-1
    xiT2(:,s) = P(:).*kron((xiT(:,s+1)./xit1(:,s)), xit(:,s));
end
end
```