

CAMP Working Paper Series
No 4/2015

Sigma Point Filters For Dynamic Nonlinear Regime Switching Models

Andrew Binning and Junior Maih



© Authors 2015

This paper can be downloaded without charge from the CAMP website <http://www.bi.no/camp>

Sigma Point Filters For Dynamic Nonlinear Regime Switching Models¹

This Version:
May 18, 2015

Andrew Binning^a, Junior Maih^{a,b}

^a*Monetary Policy Department, Norges Bank*
^b*BI Norwegian Business School*

Abstract

In this paper we take three well known Sigma Point Filters, namely the Unscented Kalman Filter, the Divided Difference Filter, and the Cubature Kalman Filter, and extend them to allow for a very general class of dynamic nonlinear regime switching models. Using both a Monte Carlo study and real data, we investigate the properties of our proposed filters by using a regime switching DSGE model solved using nonlinear methods. We find that the proposed filters perform well. They are both fast and reasonably accurate, and as a result they will provide practitioners with a convenient alternative to Sequential Monte Carlo methods. We also investigate the concept of *observability* and its implications in the context of the nonlinear filters developed and propose some heuristics. Finally, we provide in the RISE toolbox, the codes implementing these three novel filters.

Keywords: Regime Switching, Higher-order Perturbation, Sigma Point Filters, Nonlinear DSGE estimation, Observability

1. Introduction

Many important problems in modern macroeconomics require the estimation of parameters and unobserved variables in dynamic nonlinear (DSGE) models with switching regimes. Examples include [Aruoba & Schorfheide \(2013\)](#) and [Bi & Traum \(2014\)](#). Unfortunately the existing procedures for the filtering of such models are cumbersome, compelling researchers to take shortcuts such as linearization, constant parameters and Kalman Filtering for example. In this paper we propose alternative filtering procedures that are fast and reasonably accurate. In particular we take three well known Sigma Point Filters, namely the Unscented

Email addresses: andrew.binning@norges-bank.no (Andrew Binning), junior.maih@norges-bank.no (Junior Maih)

¹This Working Paper should not be reported as representing the views of Norges Bank. The views expressed are those of the authors and do not necessarily reflect those of Norges Bank.

Kalman Filter (UKF) by Julier & Uhlmann (1997), the Divided Difference Filter (DDF) by Nørgaard et al. (2000), and the Cubature Kalman Filter (CKF) by Arasaratnam & Haykin (2009), and extend them to allow for dynamic nonlinear regime switching models. As a second contribution we investigate *observability*, the ability to recover the unobserved variables given a finite sequence of observations, in nonlinear state space models. Dissatisfied with existing methods of investigating *observability*, we develop our own heuristics.

Our decision to revive and extend Sigma Point Filters to dynamic nonlinear regime switching models has been motivated by three key observations. First, it is common practice to build nonlinear models and linearize them even when the underlying problem is inherently nonlinear. This strategy has benefitted from the array of fast and efficient tools available for solving, estimating, decomposing and interpreting these models and their results (see Dave & DeJong, 2010, for a good overview). Although convenient, there are many critical questions linear specifications cannot handle. Our economic theories and many problems posed by policy makers and practitioners are fundamentally nonlinear. Linearized models are unable to capture occasionally binding constraints such as the collateral constraints considered in Benigno et al. (2013) and asymmetries such as downwardly nominal wage rigidities as modelled by Kim & Ruge-Murcia (2011). Their certainty equivalence renders them useless for problems that require some consideration of risk, such as the evaluation of optimal policy in distorted economies (Schmitt-Grohe & Uribe, 2004), or the calculation of bond term premia in consumption-based asset pricing models of production economies (Rudebusch & Swanson, 2012).

Second, constant parameter models are often used to describe data that exhibit structural breaks and other properties that are inconsistent with a single data generating distribution, or they are used in combination with shortened samples to avoid issues associated with time varying parameters. The assumption of constant parameters sets policy behavior once and for all, and interprets the data as a sequence of drawings from the same distribution. However the political and economic history of nations is characterized by change. Political cycles induced by changes in government and central bank governance lead to changes in fiscal, monetary and macro-prudential policy, and the instruments used to conduct these policies. These changes are reflected in the data and have implications for the properties of business cycles and the underlying data generating process. Lucas (1976) emphasizes the need to take changes in deep policy parameters into account when modelling expectations, using models to analyze history and produce forecasts. The modelling of regime switches provides us with tools for incorporating the type of policy changes observed over history into structural models. This in turn allows more data to be used, data that may otherwise have been discarded due to structural breaks at odds with a constant parameter interpretation of history. Ultimately more data leads to sharper estimates of the parameters that do not switch and a longer continuous interpretation of history. Other features of the data like nonlinearities and asymmetries in the business cycle (Kim & Nelson, 2001), crises (Foerster, 2011), occasionally binding constraints (Aruoba & Schorfheide, 2013), heteroscedasticity (Liu & Mumtaz, 2011), changes in behavioral parameters (Melino & Yang, 2003), and the possibility of recurrent regimes (Sims & Zha, 2006) can all be recast and interpreted in

terms of a regime switching framework. Recent advances in perturbation solution methods by Foerster et al. (2014) and Maih (2015) provide tools for finding nonlinear solutions to nonlinear regime switching models with rational expectations.

Third, the majority of estimation studies in economics using nonlinear state space models, especially nonlinear DSGE models, have been conducted using Sequential Monte Carlo (SMC) methods despite the fact that alternative filters exist that are not only reasonably accurate but are also computationally cheaper. Examples of econometric studies conducted using SMC methods include Fernández-Villaverde & Rubio-Ramírez (2007), Amisano & Tristani (2010), Flury & Shephard (2011), Fernandez-Villaverde et al. (2015) and Doh (2011). SMC methods also known as Particle Filters are non-parametric and use stochastic simulation to track the entire distribution implied by the state space. As a result they are costly, requiring many particles to get good estimates, and they suffer from sample degeneracy and impoverishment. SMC filters are asymptotically exact which motivates an exact treatment of regime switching in this framework, and would require tracking all paths and nodes for the histories of the regimes. Such a treatment would prove infeasible, the number of paths and nodes that need to be tracked would explode with time which would be compounded by the need to use a large number of particles for the simulation of each path.

Sigma Point Filters, although largely dismissed by Fernández-Villaverde & Rubio-Ramírez (2007) in the early nonlinear DSGE estimation literature, have proven to be a competitive alternative to SMC methods in terms of accuracy. Andreasen (2013) shows with a simple DSGE model solved using both a second and third order perturbation method that the Divided Difference Filter provides more accurate results than the Particle Filter using 500,000 particles. Kollmann (2015) is also able to beat a Particle Filter that uses 500,000 particles with a deterministic Kalman Filter adapted for second order approximations in pruned state space (the KalmanQ Filter).² In contrast to SMC methods, Sigma Point Filters are parametric, deterministic and approximate filters, and they assume that the states are reasonably well approximated by a Gaussian distribution. These assumptions make the filters computationally cheaper and much faster when compared with SMC methods because the nonlinear functions only need to be evaluated at a small number of well chosen points which are used to track the mean and the covariance of the states.

We depart from the constant parameter Sigma Point Filtering literature by extending Sigma Point Filters to include nonlinear regime switching state space models. Keeping with the approximate and deterministic nature of these filters, it seems appropriate to approximate the regime switching state space using the collapsing approach popularized by Kim & Nelson (1999). The addition of regime switching should improve the ability of Sigma Points Filters to approximate the distribution of the unobserved states. The Sigma Point Filtering assumption that state variables are well approximated by a Gaussian distribution is likely to be violated in some cases. However, we know that any distribution can be well approximated by a

²While the KalmanQ Filter proposed by Kollmann (2015) is not a Sigma Point Filter, the results from the Monte Carlo experiment do cast doubt on the convergence and accuracy of the Particle Filter even when using small models.

Gaussian mixture distribution. Approximating the regime dependent distribution of the states by a Gaussian distribution in the regime switching Sigma Points Filter means the expected distribution of the states will follow a Gaussian mixture distribution, and should in theory be capable of approximating any distribution.

To test the accuracy and understand the properties of these filters we perform a Monte Carlo study using the [Fernandez-Villaverde et al. \(2015\)](#) DSGE model adapted to include regime switching and solved using a nonlinear solution method. This laboratory exercise is necessary to determine the accuracy and behavior of the filters. We also test the filters by applying them to real data. Because we use the same data and model as [Fernandez-Villaverde et al. \(2015\)](#) we are able to compare our results with their results which gives us further confirmation and confidence that our procedures are reasonable. Our proposed filters should provide a convenient alternative to SMC methods.

Relatedly, reliable and accurate filters may not be enough to guarantee good *observability* of the unobserved state variables. Weak and/or ambiguous relationships between observed and unobserved variables in a nonlinear state space model may not lead to the unique or accurate recovery of the unobserved variables even in the presence of an exact filter. *Observability*, originally due to [Kalman \(1960\)](#) in the linear-Gaussian case, has received some coverage in the linearized DSGE literature where it is a requirement for parameter identification (see [Komunjer & Ng, 2011](#); [Fukac, 2010](#)). It has not, however, received any coverage in the nonlinear DSGE literature. *Observability* has important implications for choosing the set of observed variables, determining which unobserved variables are estimable, and is also a key component in parameter identification.

Measuring *observability* is relatively trivial with linear state space models, but is more complicated with nonlinear state space models. [Kalman \(1960\)](#) proposed a simple rank test for linear state space models which determined whether the unobserved initial conditions could be recovered given a sequence of observations in a finite period of time. This simple test also illustrates the global nature of *observability* in linear state space models. These rank tests have been extended to nonlinear models by calculating the Jacobian of the initial unobserved states with respect to the sequence of observed variables (see [Muske & Edgar, 1997](#), for example). Such an approach shows that in the nonlinear case *observability* is a local phenomenon. Thinking of *observability* in terms of a rank condition treats it as a binary concept; there either is *observability* or there is not.

We interpret *observability* as a nuanced concept with varying degrees and strengths of *observability* for different variables at different points in the state space. We investigate *observability* within our proposed filtering framework because ultimately we would like to know the practical implications of our filtering and modelling choices. This may however lead to some degradation of the *observability* present in the original state space model because our filters are approximations, but determining the *observability* of the state space independently of the filter is somewhat academic. The relative speed of our Sigma Point Filters allows us to develop simulation based heuristics for *observability*.

As a further contribution, we develop codes for implementing the filters. These codes are available in the RISE toolbox.

The paper proceeds as follows: in Section 2 we discuss Bayesian Filters for dynamic nonlinear regime switching models and set up our Sigma Points Filter. In Section 3 we introduced *observability* and discuss its implications for nonlinear state space models. We validate the filters using Monte Carlo experiments and with actual data in Section 4 before examining the *observability* properties of our test model. We conclude in Section 5.

2. Bayesian Filtering for Dynamic Nonlinear Regime Switching Models

In this section we develop Sigma Point Filters to estimate the unobserved states for a general class of dynamic nonlinear regime switching models. We begin by setting up the state space for a dynamic nonlinear regime switching model. This is followed by an overview of the Bayesian Filtering problem and a discussion of the relative merits of Sigma Point Filters. Then we introduce the details of our proposed Sigma Point Filters for dynamic nonlinear regime switching models. Finally we discuss some of the key ingredients for successful Sigma Point Filtering.

2.1. The State Space Representation of Dynamic Nonlinear Regime Switching Models

We consider a very general class of nonlinear state space models with regime switching. Such a setup is broad enough to capture linear, constant parameter, rational expectations and backward-looking models as special cases.

We characterize history as made up of possibly different regimes, each with its distinctive properties:

$$E_t \sum_{r_{t+1}=1}^h p_{r_t, r_{t+1}}(\mathcal{I}_t) f_{r_t}(x_{t+1}(r_{t+1}), x_t(r_t), x_{t-1}, \theta_{r_t}, \theta_{r_{t+1}}, \eta_t) = 0 \quad (1)$$

where E_t is the expectations operator, f_{r_t} is a known and possibly nonlinear function, r_t represents the switching process with h different regimes, θ_{r_t} is the parameters in regime r_t , $\theta_{r_{t+1}}$ is the parameters next period, $p_{r_t, r_{t+1}}(\mathcal{I}_t)$ is the transition probability for going from regime r_t to regime r_{t+1} , which depends on \mathcal{I}_t , the information at time t , x_t are the date t endogenous variables and η_t are exogenous disturbances.

Equation (1) is flexible enough to describe a range of models. The functional form of f_{r_t} determines whether the model is linear or nonlinear, if the $t + 1$ expected variables are absent then the model is backward-looking, and if the number of regimes h is equal to 1, then the model has constant parameters.

We require the solution of the model in equation (1) to exhibit the Markov property in order to write it in the form of a state space model. That is, we need to be able to write today's state variables as a function of only yesterday's state variables and the shocks. If equation (1) is a rational expectations model we need to solve it using a method that preserves the Markov property. If the model is backward looking with multiple lags we can use the companion form to preserve the Markov property. To complete the transition block of the regime switching state space model, we need to define the transition matrix for the

discrete states. This gives us the following transition equations:

$$x_t = T_{r_t}(x_{t-1}, \eta_t), \quad \eta_t \sim N(0, I) \quad (2)$$

$$p_{r_t, r_{t+1}} = Q_{r_t, r_{t+1}}(\mathcal{I}_t), \quad r_t = 1, 2, \dots, h \quad (3)$$

where T_{r_t} is a potentially nonlinear but known function and the shocks are assumed to be i.i.d. Normally distributed with mean zero and covariance equal to the identity matrix. We let $Q_{r_t, r_{t+1}}(\mathcal{I}_t)$ denote the $h \times h$ transition matrix. Equations (2) and (3) characterize the transition equations. Equation (2) is similar to the autoregressive transition equations we are familiar with in constant parameter state space models. Equation (3) is added to the transition block because we are in a regime switching world and we need to describe how the transition probabilities are generated. The nonlinear regime switching state space model also consists of a measurement equation:

$$y_t = Z_t x_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, H_{r_t}) \quad (4)$$

where y_t is the vector of date t observed variables, Z_t is the identity matrix with rows potentially missing and ε_t is a vector of measurement errors with covariance H_{r_t} . Note that we assume a linear measurement block for simplicity. It should always be possible to write any nonlinear model in this form by adding any nonlinear observation equations to the transition block. Given the measurement and transition equations we can compute the conditional probabilities: $p(x_t|x_{t-1})$, $p(y_t|x_t)$, and $p(y_t|x_{t-1})$. The conditional probability $p(x_t|x_{t-1})$ can be determined directly from the transition block of the state space model, while $p(y_t|x_t)$, and $p(y_t|x_{t-1})$ must be estimated using the state space model. This can be done using Bayesian Filtering methods which we discuss in the next section.

2.2. Bayesian Filtering

The filtering problem is one of computing the conditional densities $p(x_t|y_{1:t-1})$ and $p(x_t|y_{1:t})$ given a prior distribution $p(x_0)$ on x_0 , a state space model and a sequence of measurements $y_{1:t}$. These densities can be computed efficiently using recursive Bayesian estimation which makes use of the Chapman-Kolmogorov equation for prediction and Bayes Rule for updating. We present a very generalized Bayesian Filtering algorithm based on Theorem 4.1 from [Särkkä \(2013\)](#) in Algorithm 1 below.

The proofs for this algorithm are presented in [Särkkä \(2013\)](#). Although elegant in its simplicity, the generic Bayesian Filtering algorithm can be difficult to implement in practice. It involves the computation of high dimensional integrals that are intractable in most cases. The exception being linear-Gaussian models with a single regime, in which case analytical forms do exist allowing us to derive the celebrated Kalman Filter.

If the problem demands a nonlinear non-Gaussian model we can no longer use the Kalman Filter. As a consequence we have to employ other strategies for approximating and evaluating these integrals. The literature has dealt with this problem in three ways:

- Linearization of the state space model (the Extended Kalman Filter)

Algorithm 1 A Generic Bayesian Filtering Algorithm

Initialization

Set the prior distribution of the initial state x_0 to $p(x_0)$.

for $t = 1$ to n **do**

State Prediction

The predictive distribution of the state x_t at time t can be calculated using the Chapman-Kolmogorov equation as follows:

$$p(x_t|y_{1:t-1}) = \int_{-\infty}^{\infty} p(x_t|x_{t-1})p(x_{t-1}|y_{1:t-1}) dx_{t-1}$$

State Update

Given the observation of y_t at time t the posterior of the state x_t can be calculated using Bayes Rule as follows:

$$p(x_t|y_{1:t}) = \frac{p(y_t|x_t)p(x_t|y_{1:t-1})}{p(y_t|y_{1:t-1})}$$

where

$$p(y_t|y_{1:t-1}) = \int_{-\infty}^{\infty} p(y_t|x_t)p(x_t|y_{1:t-1}) dx_t$$

end for

- Making parametric assumptions about the distributions of the state variables (Sigma Point Filters), and
- Evaluating the integrals through simulation (Sequential Monte Carlo methods).

We adopt the Sigma Point Filtering approach in this paper and use the rest of this section to motivate our choice.

The Extended Kalman Filter (EKF) deals with the problem of intractable integrals by linearizing the state space model. This returns us to the familiar linear-Gaussian world, however the approximation is only valid in a local neighborhood of the point that it was approximated around. Comparisons of the EKF with Sigma Point Filters indicate that the EKF generally does quite poorly, and for this reason we will not discuss the EKF further (see [Andreasen, 2013](#)).

Sigma Point Filters are approximate filters. The unknown distribution of the unobserved state variables is approximated by a parametric distribution, usually a Gaussian distribution, as a consequence we only need to track the first two moments of the states. The result is a deterministic filter that only requires the nonlinear function to be evaluated at a small number of well chosen points, known as Sigma Points. By comparison Sequential Monte Carlo methods are non-parametric and require the stochastic simulation of many points to approximate the entire distribution of the unobserved state variables. They are also asymptotically exact filters, as the number of particles increases so does their accuracy. This makes Sigma Point Filters computationally cheaper than Sequential Monte Carlo methods.

The approximate and deterministic nature of Sigma Point Filters makes them compatible with an approximate representation of the regime switching state space. The “collapsing” trick popularised by [Kim & Nelson \(1999\)](#) in linear state space models to keep the number of nodes manageable can easily be applied and extended to Sigma Point Filters. The asymptotically exact nature of the particle filter is more consistent with an exact treatment of the regime switching state space. This would require modelling all possible paths and histories of the regimes, a task which is infeasible. In a regime switching state space model with h regimes we would need to evaluate functions at h^{t+1} nodes for each point in time. Particle filters usually require N (large) particles to get a good approximation, an exact regime switching state space would require N particles for each node because the unobserved state variables should be drawn conditional on each node, so the total number of function evaluations at each point in time would be $N.h^{t+1}$. [Aruoba & Schorfheide \(2013\)](#) propose a Particle Filter for nonlinear regime switching models that gets around this problem by drawing state variables independently of the regimes. As a consequence their algorithm randomly visits a small subset of the nodes implied by the exact regime switching state space. It is unclear how well their algorithm performs because they draw the unobserved state variables independently of the regimes, which may result in a large number of state variable/regime combinations that have low probability and hence low weights.

Furthermore, Particle Filters have been plagued by issues of sample degeneracy, sample impoverishment and convergence. Sample degeneracy occurs when the Particle Filter collapses and only a few particles are given a significant weight, with the rest given negligible weights. Resampling has been introduced to avoid degeneracy, but may lead to sample impoverishment where only a few points are resampled because the rest have very low weight.

It is also unclear how many particles are required to achieve convergence. [Andreasen \(2013\)](#) tests the DDF (a Sigma Point Filter) against the Particle Filter with 500,000 particles in a Monte Carlo study using a simple DSGE model solved using nonlinear methods (second and third order perturbations). He finds in many of his experiments that the DDF is able to beat the Particle Filter and cites this as evidence that the Particle Filter has yet to converge with 500,000 particles. In a similar Monte Carlo experiment [Kollmann \(2015\)](#) tests the KalmanQ Filter, a Kalman Filter for pruned second order approximations of DSGE models, against a Particle Filter with 500,000 particles also using a simple DSGE model. He finds the KalmanQ Filter is able to beat the Particle Filter. The problem is likely to be more acute in more complicated models as higher dimensional state spaces likely require

more particles than lower dimensional state spaces, which adds to the heavier computational burden of solving and simulating larger models. This begs the question: how many particles are required to achieve convergence? The answer may be literally millions of particles. Despite these striking results, those that have used Particle Filters in the literature have used far fewer particles, in some cases with models that are much more complicated than those used in the tests previously mentioned, for example [Flury & Shephard \(2011\)](#) use 60,000 particles, [Fernández-Villaverde & Rubio-Ramírez \(2007\)](#) use 80,000 particles, and [Fernandez-Villaverde et al. \(2015\)](#) use only 10,000 particles.

2.3. Sigma Point Filters

We develop three Sigma Points Filtering algorithms for nonlinear regime switching models: the Switching Unscented Kalman Filter (SUKF), the Switching Divided Difference Filter (SDDF) and the Switching Cubature Kalman Filter (SCKF). The main difference between these filters is the method of choosing the Sigma Points. For this reason we describe the filtering problem for nonlinear regime switching models in terms of a single generic regime switching Sigma Points Filtering algorithm. The Sigma Point rules used in each filter are explained in detail in the next section.

The first step of Bayesian Filtering requires the specification of some initial conditions.

$$x_{1|0} \sim D(\bar{x}_1, \bar{P}_1),$$

$$p_{1|0}(s_1), \quad \text{where } s_1 = 1, 2, \dots, h$$

where $x_{1|0}$ is the forecast of the state variables for period 1 conditional on period 0 information, D is a distribution with mean \bar{x}_1 and covariance \bar{P}_1 and $p_{1|0}(s_1)$ is the forecast of the initial probability of being in regime s_1 at period 1 based on period 0 information.

The steps in the generic regime switching Sigma Points Filter are similar to those of the generic Bayesian Filtering algorithm (see [Algorithm 1](#)) and are outlined in [Algorithm 2](#). [Table 1](#) provides a brief description of the variables in the Sigma Points algorithm.

The filter recursions start with forecasts of the observed variables, calculating the forecast errors and the measurement prediction covariance. The marginal likelihoods for each regime are calculated because they are required for updating the probability of being in each regime at each point in time. The states are then updated using the standard Kalman Filter recursions. The next step is to produce forecasts of the states, which comprises producing forecasts for the transition probabilities, collapsing the states and then generating Sigma Points to calculate forecasts of the unobserved state variables. We repeat these steps until we reach the end of history.

Table 1: Sigma Point Filter Variables

Variable	Description
$x_{t t-1}(r_t)$	Date t forecast of unobserved variables conditional on regime r_t and date $t - 1$ information
$y_{t t-1}(r_t)$	Date t forecast of observed variables conditional on regime r_t and date $t - 1$ information
y_t	Observed data
$v_t(r_t)$	Date t forecast error for observed variables
$F_t(r_t)$	Date t covariance for observed variables conditional on regime r_t
$P_{t t-1}(r_t)$	Date t forecast of the covariance matrix conditional on regime r on date $t - 1$ information
$f(r_t)$	Marginal likelihood for regime r_t
$p_{t t-1}(r_t)$	Date t forecast of probability of being in regime r_t given date $t - 1$ information
$lik(t)$	Likelihood
$p_{t t}(r_t)$	Probability of being in regime r_t
$P_t^{(x,y)}(r_t)$	Covariance between observed and unobserved variables conditional on regime r_t
$K_t(r_t)$	Kalman gain
$x_{t t}(r_t)$	Updated value for the unobserved variables
$P_{t t}(r_t)$	Updated covariance for the unobserved variables
$Q_{t,t+1}$	Transition matrix
$p(r_t, r_{t+1})$	Probability for going from regime r_t to r_{t+1}
$\hat{x}_{t t}(r_{t+1})$	Expected value of $x_{t t}$ conditional on r_{t+1}
$\hat{P}_{t t}(r_{t+1})$	Expected value of $P_{t t}$ conditional on r_{t+1}
$\hat{x}_{t t}^\sigma(r_{t+1})$	Collection of Sigma Points for $\hat{x}_{t t}(r_{t+1})$
w_i	Sigma Point weights

Algorithm 2 Generic Regime Switching Sigma Point Filter

for $t = 1$ to n **do**

$lik(t) = 0$

for $r_t = 1$ to h **do**

Forecasting of Observables:

$$y_{t|t-1}(r_t) = Z_t x_{t|t-1}(r_t)$$

$$v_t(r_t) = y_t - y_{t|t-1}(r_t)$$

$$F_t(r_t) = Z_t P_{t|t-1}(r_t) Z_t' + H_t(r_t)$$

Likelihood Computation:

$$f(r_t) = (2\pi)^{-\frac{n_y}{2}} |F_t(r_t)|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} v_t(r_t)' [F_t(r_t)]^{-1} v_t(r_t)\right)$$

$$lik(t) = lik(t) + p_{t|t-1}(r_t) f(r_t)$$

Updating of States:

$$p_{t|t}(r_t) = \frac{p_{t|t-1}(r_t) f(r_t)}{\sum_{r_t=1}^h p_{t|t-1}(r_t) f(r_t)}$$

$$P_t^{(x,y)}(r_t) = P_{t|t-1}(r_t) Z_t'$$

$$K_t(r_t) = P_t^{(x,y)}(r_t) [F_t(r_t)]^{-1}$$

$$x_{t|t}(r_t) = x_{t|t-1}(r_t) + K_t(r_t) v_t(r_t)$$

$$P_{t|t}(r_t) = P_{t|t-1}(r_t) - K_t(r_t) \left[P_t^{(x,y)}(r_t) \right]'$$

end for

Forecasting of States:

$Q_{t,t+1} = qfunc(x_{t|t}(\bar{r}_t))$, where \bar{r}_t is any state

$$p_{t+1|t} = Q_{t,t+1}' p_{t|t}$$

for $r_{t+1} = 1$ to h **do**

for $r_t = 1$ to h **do**

$$p(r_t, r_{t+1}) = \frac{Q_{t,t+1}(r_t, r_{t+1}) p_{t|t}(r_t)}{p_{t+1|t}(r_{t+1})}$$

end for

Collapsing:

$$\hat{x}_{t|t}(r_{t+1}) = \sum_{r_t=1}^h p(r_t, r_{t+1}) x_{t|t}(r_t)$$

$$\hat{P}_{t|t}(r_{t+1}) = \sum_{r_t=1}^h p(r_t, r_{t+1}) P_{t|t}(r_t)$$

Sigma Points Generation:

$$\hat{x}_{t|t}^\sigma(r_{t+1}) = sigmaPoints\left(\hat{x}_{t|t}(r_{t+1}), \hat{P}_{t|t}(r_{t+1})\right)$$

$$x_{t+1|t}(r_{t+1}) = \sum_{i=1}^{n_\sigma} w_i T_{r_{t+1}}\left(\hat{x}_{t|t}^{\sigma_i}(r_{t+1}), 0\right)$$

$$P_{t+1|t}(r_{t+1}) = \sum_{i=1}^{n_\sigma} w_i \left[\hat{x}_{t|t}^{\sigma_i}(r_{t+1}) - x_{t+1|t}(r_{t+1}) \right] \left[\hat{x}_{t|t}^{\sigma_i}(r_{t+1}) - x_{t+1|t}(r_{t+1}) \right]'$$

end for

end for

2.3.1. Key Ingredients

There are three features of these filters and their practical implementation that deserve further discussion: approximating the regime switching state space, choosing the Sigma Points, and ensuring positive definiteness of the covariance matrix for the state variables. We discuss each of these issues in turn.

Approximating the Regime Switching State Space: As we have discussed, modeling the exact regime switching state space is infeasible due to the rapid growth in the number of nodes at which the filter needs to be evaluated. We get around this problem by employing a suitable approximation. [Kim & Nelson \(1999\)](#) approximate the linear regime switching state space by “collapsing” the number of regimes at each iteration of the filter. In particular, they track state variables for h regimes and produce forecasts and updates for the h^2 nodes that result from all possible transition paths. These are then collapsed into h nodes through a weighted averaging step after the updating step has been completed. We employ a similar strategy and use “collapsing” to manage the number of nodes we track. However we collapse the number of nodes before the updating step so that we only need to produce h forecasts and updates and the number of nodes h remains the same at all stages during the filter. We could have followed [Kim & Nelson \(1999\)](#) and collapsed the number of nodes after the updating step, but collapsing before the updating step results in fewer function evaluations without loss of accuracy. The collapsing approximation violates Jensen’s inequality, but it is necessary to ensure that the filter is manageable, and our testing demonstrates that it is a reasonable approximation and the performance of the filters is not compromised.

Moment Integration Rules: Choosing the Sigma Points: Sigma Point approximations embody the idea that: “*it is easier to approximate a probability distribution than it is to approximate an arbitrary nonlinear function or transformation*” [Julier & Uhlmann \(1997\)](#). By passing a small number of well chosen Sigma Points, determined by the columns of the covariance of the state variables relative to their mean, through a nonlinear transformation and calculating their weighted sum, we are able to get quite accurate estimates of the forecasted mean and covariance of the state variables. We consider three approaches to choosing the Sigma Points in our regime switching filter. These include the Unscented Transform used in the Unscented Kalman Filter (see [Julier & Uhlmann, 1997](#)), Stirling’s Interpolation used in the Divided Difference Filter (see [Nørgaard et al., 2000](#)), and the Spherical-Radial Cubature rule used in the Cubature Kalman Filter (see [Arasaratnam & Haykin, 2009](#)).

Switching Unscented Kalman Filter: The Unscented Transform

We define some notation before describing the Unscented Transform used in the UKF. Sigma Points algorithms are applied to both the unobserved state variables and the shocks, so we combine both sets of variables into a single vector that we label \mathcal{X} as follows:

$$\mathcal{X} \equiv [x'_{t-1}, \eta'_t]'$$

where $n_{\mathcal{X}}$ is the number of unobserved states and shocks and $\mathcal{X} \sim D(\mu_{r_t}, \Sigma_{r_t})$. The Unscented

Transform approximates the mean and the covariance using $2n_{\mathcal{X}} + 1$ Sigma Points, these points are chosen as follows:

$$\begin{aligned}\mathcal{X}^{[0]} &= \mu_{r_t} \\ \mathcal{X}^{[i]} &= \mu_{r_t} + \left(\sqrt{(n_{\mathcal{X}} + \lambda) \Sigma_{r_t}} \right)_i \quad \text{for } i = 1, \dots, n_{\mathcal{X}} \\ \mathcal{X}^{[i]} &= \mu_{r_t} - \left(\sqrt{(n_{\mathcal{X}} + \lambda) \Sigma_{r_t}} \right)_{i-n_{\mathcal{X}}} \quad \text{for } i = n_{\mathcal{X}} + 1, \dots, 2n_{\mathcal{X}}\end{aligned}$$

where $n_{\mathcal{X}} + \lambda = 3$, and the subscripts for the brackets indicate that we take the i th or the $(i - n)$ th column of the square root of the scale matrix. The weights for the Sigma Points are chosen such that

$$w^{[0]} = \frac{\lambda}{n_{\mathcal{X}} + \lambda}, \quad w^{[i]} = \frac{1}{2(n_{\mathcal{X}} + \lambda)} \quad \text{for } i = 1, \dots, 2n_{\mathcal{X}}$$

We are able to calculate the expected or forecasted mean and covariance by passing the Sigma Points through the nonlinear function and weighting them accordingly

$$\begin{aligned}\hat{\mu}_{r_t} &= \sum_{i=0}^{2n_{\mathcal{X}}} w^{[i]} T_{r_t} (\mathcal{X}^{[i]}) \\ \hat{\Sigma}_{r_t} &= \sum_{i=0}^{2n_{\mathcal{X}}} w^{[i]} (T_{r_t} (\mathcal{X}^{[i]}) - \hat{\mu}_{r_t}) (T_{r_t} (\mathcal{X}^{[i]}) - \hat{\mu}_{r_t})'\end{aligned}$$

Switching Divided Difference Filter: Stirling's Interpolation

The Divided Difference Filter uses Stirling's Interpolation to choose the Sigma Points. Stirling's interpolation is a formula for polynomial approximation over an interval, its derivation is very similar to a standard Taylor series approximation where the derivatives are replaced by the central divided differences. The $2n_{\mathcal{X}} + 1$ Sigma Points chosen according to Stirlings's Interpolation:

$$\begin{aligned}\mathcal{X}^{[0]} &= \mu_{r_t} \\ \mathcal{X}^{[i]} &= \mu_{r_t} + \delta s_i \quad \text{for } i = 1, \dots, n_{\mathcal{X}} \\ \mathcal{X}^{[i]} &= \mu_{r_t} - \delta s_{i-n_{\mathcal{X}}} \quad \text{for } i = n_{\mathcal{X}} + 1, \dots, 2n_{\mathcal{X}}\end{aligned}$$

where $\delta = \sqrt{3}$ for a Gaussian distribution and s_i is the i th column of the square root of the covariance of the unobserved state variables. The associated weights are set as follows

$$w_m^{[0]} = \left(\frac{\delta^2 - n_{\mathcal{X}}}{\delta} \right), \quad w_m^{[i]} = \frac{1}{2\delta^2}, \quad w_{c,1}^{[j]} = \frac{1}{2\delta}, \quad w_{c,2}^{[j]} = \frac{\sqrt{\delta^2 - 1}}{2\delta^2} \quad \text{for } j = 1, \dots, n_{\mathcal{X}}$$

where the Stirling interpolation has both first and second order terms. The expected values

or forecasts are calculated as follows

$$\begin{aligned}\hat{\mu}_{r_t} &= \sum_{i=0}^{2n_{\mathcal{X}}} w_m^{[i]} T_{r_t}(\mathcal{X}^{[i]}) \\ S_j^{(1)} &= w_{c,1}^{[j]} [T_{r_t}(\mathcal{X}^{[j]}) - T_{r_t}(\mathcal{X}^{[n_{\mathcal{X}}+j]})] \quad \text{for } j = 1, \dots, n_{\mathcal{X}} \\ S_j^{(2)} &= w_{c,2}^{[j]} [T_{r_t}(\mathcal{X}^{[j]}) + T_{r_t}(\mathcal{X}^{[n_{\mathcal{X}}+j]}) - 2T_{r_t}(\mathcal{X}^{[0]})] \\ \hat{S}_{r_t} &= \Phi([S^{(1)}, S^{(2)}])\end{aligned}$$

where $\Phi(\cdot)$ is a matrix triangularization, like the Householder transformation.

Switching Cubature Kalman Filter: Spherical Radial Cubature Rule

The Cubature Kalman Filter (CKF) uses the Spherical Radial Cubature Rule to choose the Sigma Points. As [Arasaratnam & Haykin \(2009\)](#) note, the key approximation taken to develop the CKF is that the predictive density and the filter likelihood density are both Gaussian which leads to a Gaussian posterior density. The $2n_{\mathcal{X}}$ Sigma Points are determined according to

$$\begin{aligned}\mathcal{X}^{[i]} &= \mu_{r_t} + \left(\sqrt{n_{\mathcal{X}}\Sigma_{r_t}}\right)_i \quad \text{for } i = 1, \dots, n_{\mathcal{X}} \\ \mathcal{X}^{[i]} &= \mu_{r_t} - \left(\sqrt{n_{\mathcal{X}}\Sigma_{r_t}}\right)_{i-n_{\mathcal{X}}} \quad \text{for } i = n_{\mathcal{X}} + 1, \dots, 2n_{\mathcal{X}}\end{aligned}$$

with the weights for the Sigma Points given by

$$w^{[i]} = \frac{1}{2n_{\mathcal{X}}} \quad \text{for } i = 1, \dots, 2n_{\mathcal{X}}$$

The expected or forecasted unobserved variables are calculated according to the following numerical integrations

$$\begin{aligned}\hat{\mu}_{r_t} &= \sum_{i=1}^{2n_{\mathcal{X}}} w^{[i]} T_{r_t}(\mathcal{X}^{[i]}) \\ \hat{\Sigma}_{r_t} &= \sum_{i=1}^{2n_{\mathcal{X}}} w^{[i]} (T_{r_t}(\mathcal{X}^{[i]}) - \hat{\mu}_{r_t}) (T_{r_t}(\mathcal{X}^{[i]}) - \hat{\mu}_{r_t})'\end{aligned}$$

Correcting State Covariance Matrices: The approximate nature of the filters means that the estimated covariances for the unobserved variables are not always well behaved. If the covariance matrices are not positive definite then the accuracy of the filter may deteriorate. To circumvent this problem we check each covariance matrix for positive definiteness, if the covariance matrix is not positive definite we replace it with the nearest matrix that is positive definite. We find this greatly improves the accuracy of our Sigma Point Filters.

3. Observability

Obtaining accurate estimates of the unobserved state variables in a nonlinear state space model will not always be possible, even in the presence of an exact nonlinear filter. This is because there may be multiple combinations of the unobserved variables that are compatible with the same realization of the observed variables. The ability to recover the unobserved states given the observed variables in a finite period of time is known as *observability*. Poor *observability* is due to weak and/or ambiguous relationships between the observed and unobserved variables. Assessing a state space model's *observability* has important implications for the estimability of unobserved variables, the choice of observed variables, and for parameter identification in the case of parameter estimation.

Rank tests have been proposed for measuring *observability* in linear state space models (see Kalman, 1960). These tests assess the ability to recover the initial conditions of the unobserved variables after observing a sequence of measurements. They illustrate that *observability* is a global property of linear state space models with one regime. Similar tests have been proposed for nonlinear models based on the Jacobian of the observed variables with respect to the initial conditions of the unobserved variables. However, such tests are only local and treat *observability* as a binary concept. Observability Gramian matrices have also been proposed as a means for evaluating *observability* (see Kailath, 1980).

Unsatisfied with existing metrics for *observability* in nonlinear state space models, we have developed our own heuristics for this purpose. More specifically we conduct a Monte Carlo exercise, constructing artificial data from many different shock draws and estimating the corresponding recovered series for the unobserved variables using one of our Sigma Point Filters. We then subtract the actual data from the estimates of the unobserved recovered data and divide the result by the standard deviation of the simulated data. We plot the 50th and 95th percentiles of the normalized data. We interpret narrow and symmetrically centered 95th percentiles as an indication of good or high *observability*. Conversely we interpret wide and/or severely asymmetric 95th percentiles as an indication of poor *observability*.

Much of the literature has focused on *observability* in nonlinear models purely in terms of the general nonlinear state space, independent of any specific filter. We have framed the *observability* problem in terms of the filters developed in this paper. Adopting such an approach could, however, lead to a reduction in the *observability* of the model due to the approximations made in the filter. But it will help us understand how the model/filter combination performs, and since we have no other methods for filtering nonlinear regime switching models, a test that is independent of the filtering procedure would be of little practical significance.

4. Validating the Filters

We test the accuracy of the filters and examine their properties through a Monte Carlo study, and by using actual data in a model calibration exercise. While the setup is general enough to encompass a range of dynamic nonlinear regime switching models, we conduct our experiments using a nonlinear regime switching DSGE model solved using higher order

perturbation methods. Our model choice is motivated by the linearized constant parameter DSGE model’s emergence as the dominant paradigm in structural macroeconometrics and macroeconomic policy modeling. By demonstrating these filters using a regime switching DSGE model solved using nonlinear methods, we are able to illustrate the versatility of these filters to a large and growing audience already familiar with linearized DSGE models. We also hope these results will be of interest to practitioners who use other types of nonlinear models. In this section we begin by giving a brief description of the model and its calibration before detailing the Monte Carlo experiments we perform and outlining the steps taken when filtering using actual data.

4.1. Model and Calibration

We perform our experiments using the model from [Fernandez-Villaverde et al. \(2015\)](#). We have chosen this model because it is a medium-sized DSGE model, it is relatively standard, and includes time-varying parameters through parameter drift and stochastic volatility. Furthermore, [Fernandez-Villaverde et al. \(2015\)](#) solve this model using nonlinear methods and estimate it using a nonlinear filter, demonstrating that this model is realistic or flexible enough to be taken to the data. For the purposes of this paper, we modify the model replacing the parameter drift with regime switching parameters and turning off the stochastic volatility in the shock processes.

The model consists of a household sector, firms and a monetary authority. Households derive utility from consumption relative to the habit stock and leisure. They supply differentiated labor to a monopolistically competitive union and choose wages subject to a Calvo wage setting friction. Firms produce differentiated output using capital, labor and a neutral technology process, and set prices subject to a Calvo pricing friction. The capital stock evolves in the usual way except for the inclusion of embodied technology in new investment goods. The model is closed by imposing a Taylor type rule on the monetary authority.

[Fernandez-Villaverde et al. \(2015\)](#) use this model to investigate the role of monetary policy in the great moderation. This motivates them to add parameter drift to the inflation response coefficient in the Taylor rule and stochastic volatility in the shock processes. This in turn motivates the use of a nonlinear solution method and a nonlinear filtering procedure. In particular they solve the model using a second order Taylor series approximation and they estimate the model using a Particle Filter. We modify their model by replacing the parameter drift in the inflation response in the Taylor rule, with regime switching parameters in both the inflation and the output response as follows:

$$\frac{R_t}{R} = \left(\frac{R_{t-1}}{R} \right)^{\gamma_R} \left(\left(\frac{\Pi_t}{\Pi} \right)^{\gamma_{\Pi}(r_t)} \left(\frac{\tilde{y}_t^d}{\tilde{y}_{t-1}^d} \exp(Z_t^Z) \right)^{\gamma_Y(r_t)} \right)^{1-\gamma_R} \exp(\sigma_M \varepsilon_t^M)$$

where R_t is the gross nominal interest rate, Π_t is the gross inflation rate, \tilde{y}_t is the detrended level of output, Z_t^Z is the stochastic growth rate and ε_t^M is the monetary policy shock. $\gamma_{\Pi}(r_t)$ and $\gamma_Y(r_t)$ are now regime dependent. The rest of the model equations are presented in [Appendix A](#), while [Fernandez-Villaverde et al. \(2015\)](#) provide a full derivation of the benchmark

version of their model at http://economics.sas.upenn.edu/~jesusfv/benchmark_DSGE.pdf.

We adopt the [Fernandez-Villaverde et al. \(2015\)](#) parameterization except where we turn off the stochastic volatility processes and replace the parameter drift with switching in the Taylor rule. We also set the standard deviation of all the structural shocks to 0.01. The regime dependent parameters in the Taylor rule are chosen to allow for one monetary policy regime with a strong response to inflation and a weak response to output and a second regime with a weaker response to inflation and a stronger response to output. The Taylor rule calibration can be found in table 2 below and table B.9 presents the calibration of the other parameters and can be found in [Appendix B](#).

Table 2: Regime Specific Parameters

Parameters	Regime 1	Regime 2
γ_π	1.7	0.7
γ_y	0.5	1.0

The transition probabilities are chosen to ensure that the monetary policy regimes are reasonably persistent. They can be found in table 3 below.

Table 3: Transition Probabilities

Parameters	Value
$p_{1 2}$	0.1
$p_{2 1}$	0.1

We follow [Fernandez-Villaverde et al. \(2015\)](#) and solve the model using a second order Taylor series approximation, however we use the procedures outlined in [Maih \(2015\)](#) which allow for nonlinear regime switching rational expectations models. We do not prune the solution.

We select the same observable variables that [Fernandez-Villaverde et al. \(2015\)](#) use, namely per capita GDP growth, investment price inflation, nominal interest rates, consumer price inflation and real wage growth. In the section using actual data, we construct these series using the same recipe outlined in [Fernandez-Villaverde et al. \(2015\)](#) so that they match their series as closely as possible.

4.2. A Monte Carlo Study

Our Monte Carlo experiments proceed as follows: we take 500 randomly drawn shock sequences and simulate artificial data for 1000 periods using the model, we retain all the simulated data for evaluation purposes. For each of the 500 sets of artificial data we take the same observable variables used by [Fernandez-Villaverde et al. \(2015\)](#) and use our filters to estimate the unobserved variables. We then compare the estimates of the recovered variables with the actual (simulated) data. For each draw we calculate root mean squared errors both

for the entire 1000 period sample and for just the second half of the sample. By calculating the RMSEs on the second half of the sample we are able to investigate the properties of the filters once they have converged. We also present some graphs of the recovered unobserved variables from the filters against the actual data, to give the reader a feel for how the RMSEs translate into particular estimates.

The expected recovered series are defined as follows: $x_{t|t} = \sum_{r_t=1}^h p(r_t)x_{t|t}(r_t)$, where the root mean squared errors are given by:

$$RMSE = \frac{\sum_{i=1}^N \left(\sqrt{\frac{\sum_{t=1}^T (x_{i,t} - x_{i,t|t})^2}{T}} \right)}{N}$$

The average RMSEs for all 500 draws in the Monte Carlo experiment are presented in table 4 below. We also present the relative RMSEs in table 5. The relative RMSEs are calculated by dividing the RMSEs for each variable through by the lowest RMSE for that variable, so that the best performing filter gets a 1, and all other filters have a number larger than 1. The final row labeled “average” refers to the average RMSE for all the variables presented in the table and gives us an indication for the overall performance of the filters.

Table 4: RMSES

Variables	1:1000			501:1000		
	SDDF	SUKF	SCKF	SDDF	SUKF	SCKF
\tilde{c}_t	0.007397	0.007872	0.007351	0.001292	0.001558	0.001427
\tilde{k}_t	0.144305	0.127421	0.137709	0.052363	0.064932	0.057483
\tilde{x}_t	0.006540	0.005537	0.006295	0.002051	0.002694	0.002325
\tilde{y}_t	0.006884	0.006073	0.006883	0.001423	0.001893	0.001672
\tilde{w}_t	0.002434	0.002110	0.002915	0.001281	0.001318	0.001552
ν_t	0.000913	0.000843	0.000902	0.000340	0.000339	0.000364
ν_t^w	0.001108	0.001014	0.001038	0.000404	0.000396	0.000456
\tilde{Q}_t	0.005630	0.004929	0.006918	0.001995	0.002279	0.002777
d_t	0.006233	0.005450	0.006446	0.003373	0.003421	0.004941
φ_t	0.041518	0.034820	0.041174	0.011251	0.014895	0.012997
Average	0.022296	0.019607	0.021763	0.007577	0.009372	0.008599

Average RMSEs for 500 simulations using randomly chosen shocks for 1000 periods.

When we look at the RMSEs for the entire 1000 period sample, we see the SUKF dominates with the lowest average RMSE and the lowest RMSEs for all the reported variables except consumption. The SCKF has the second lowest average RMSE and for most variables it has the second lowest RMSEs. The SDDF comes in third place for the average RMSE and for most individual RMSEs as well. When we focus attention on the second half of the sample (501:1000) we notice that the SDDF has the lowest average RMSE and it has the lowest RMSEs for all variables excluding the price and wage dispersion terms. The SCKF

Table 5: Relative RMSES

Variables	1:1000			501:1000		
	SDDF	SUKF	SCKF	SDDF	SUKF	SCKF
\tilde{c}_t	1.006231	1.070867	1.000000	1.000000	1.205783	1.104303
\tilde{k}_t	1.132505	1.000000	1.080741	1.000000	1.240036	1.097782
\tilde{x}_t	1.181185	1.000000	1.136984	1.000000	1.313421	1.133556
\tilde{y}_t	1.133471	1.000000	1.133222	1.000000	1.330830	1.175401
\tilde{w}_t	1.153684	1.000000	1.381660	1.000000	1.028467	1.210968
ν_t	1.082905	1.000000	1.070343	1.005642	1.000000	1.076418
ν_t^w	1.092675	1.000000	1.023498	1.020134	1.000000	1.151244
\tilde{Q}_t	1.142062	1.000000	1.403456	1.000000	1.142131	1.391700
\tilde{d}_t	1.143689	1.000000	1.182752	1.000000	1.014220	1.464875
φ_t	1.192362	1.000000	1.182502	1.000000	1.323908	1.155156
Average	1.137156	1.000000	1.109974	1.000000	1.236900	1.134871

Average Relative RMSEs for 500 simulations using randomly chosen shocks for 1000 periods.

has the second lowest average RMSE and comes in second place for most of the individual RMSEs. The SUKF comes in third place for the average RMSE and for a lot of the individual RMSEs. So the overall picture points to the SUKF converging the fastest, but once the filters have converged the SDDF performs the best. However, it should be noted that the relative differences between the filters are quite small and we would be hesitant to generalize our results to all model types, let alone different parameterizations of our benchmark model. The similarity of the results should not be a surprise given the Sigma Points for each filter are chosen in a very similar fashion. From our Monte Carlo study we would conclude that all our Sigma Point Filters seem to do a reasonably good job.

We present plots for the first 300 periods for a subset of the unobserved variables for one of the 500 draws below. We do this to give the reader a visual although not necessarily typical sense of how the filters perform and how the RMSEs translate into tracking results.

Figure 1 illustrates that it can take some time for the effects of the initial condition to die out. These effects are especially noticeable for capital, because capital is extremely persistent. If a mistake is made in the initial condition for capital, it can take a long time for the filter to correct. [Flury & Shephard \(2011\)](#) make the same observation about capital when using the Particle Filter. The slow convergence of the estimated capital stock is also reflected in the RMSEs for capital.

In figure 2 we plot the probabilities for being in the high inflation response monetary policy regime. In the case of the simulated data, we always know in which regime we are, so that the probabilities are always zero or one. In the case of the recovered probabilities, we do not know in which regime we are, so we must estimate the probability of being in a given regime, and hence these results fall between zero or one.

Again it takes a little bit of time for the filters to converge, but once they have converged,

Figure 1: Simulated & Estimated Data

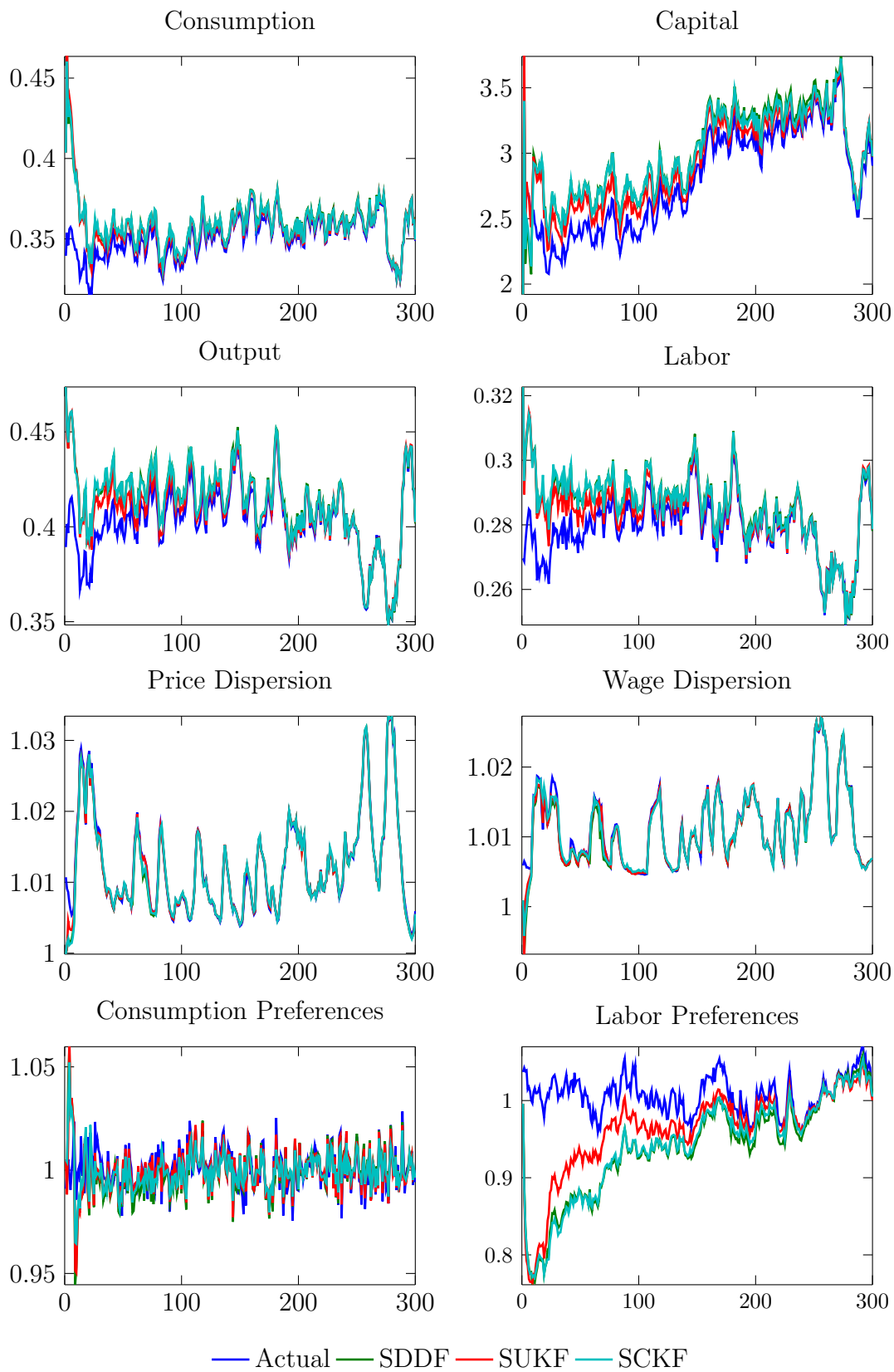
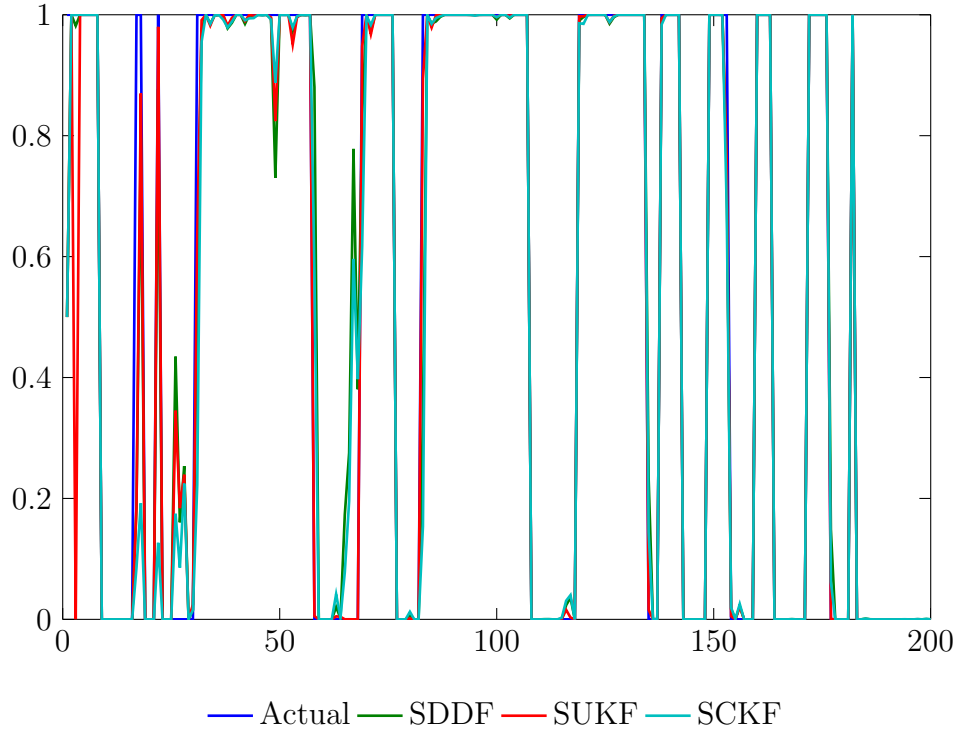


Figure 2: Monetary Policy Regimes



they do a very good job at estimating the true probabilities. The results from the Monte Carlo study give us confidence that our proposed filters are reasonably accurate and behaving as we would expect them to.

4.3. Taking the Model to Real Data

The Monte Carlo experiments have demonstrated that our proposed Sigma Point Filters perform well in the laboratory when the true data generating process is known. Yet economic theories require field testing, models need to be taken to the data to estimate the unobserved state variables and parameters, we need to assess their fit and ultimately to validate them. We field test the Switching Unscented Kalman Filter by taking the [Fernandez-Villaverde et al. \(2015\)](#) model with regime switching in the Taylor rule and the stochastic volatility turned off, to the data. We use the SUKF because based on our testing it appears to converge faster than our other Sigma Point Filters. However the overall results from our tests indicate that any of the filters could have been used with little difference between the results.³ We retain the parameterization from [Fernandez-Villaverde et al. \(2015\)](#) but calibrate the transition probabilities in the Markov chain and the switching parameters in the Taylor rule, and include some measurement error on GDP growth and CPI inflation. For the purposes of this paper, we want to test how well the filter and model fit the data with

³These results are available upon request.

very little modification to the parameterization from [Fernandez-Villaverde et al. \(2015\)](#). We leave estimation for future investigation. Our calibrated switching parameters are presented in tables 6, 7 and 8 below.

Table 6: Regime Specific Parameters

Parameters	Regime 1	Regime 2
γ_π	1.9838	0.6642
γ_y	0.3863	0.7392

Table 7: Transition Probabilities

Parameters	Value
$p_{1 2}$	0.2525
$p_{2 1}$	0.0955

Table 8: Measurement Errors

Parameters	Description	Value
u_t^Π	Inflation measurement error	0.075
$u_t^{\Delta \log(Y)}$	GDP growth measurement error	0.075

We compare our results with [Fernandez-Villaverde et al. \(2015\)](#) who use drifting parameters to evaluate monetary policy changes over history in the context of trying to explain the great moderation. Figure 3 below shows the parameter drift on the inflation coefficient in the Taylor rule from the estimation results in [Fernandez-Villaverde et al. \(2015\)](#). Figure 4 shows the expected probability of being in the high inflation response monetary policy regime. The profiles in both figures are very similar, both profiles show there is evidence of a monetary policy regime with a stronger response to inflation in the late 1960s and the early to mid 1980s. We take the similarity of these plots as evidence that SUKF is producing sensible results in the field.

Figure 3: Taylor Rule Drift from [Fernandez-Villaverde et al. \(2015\)](#)

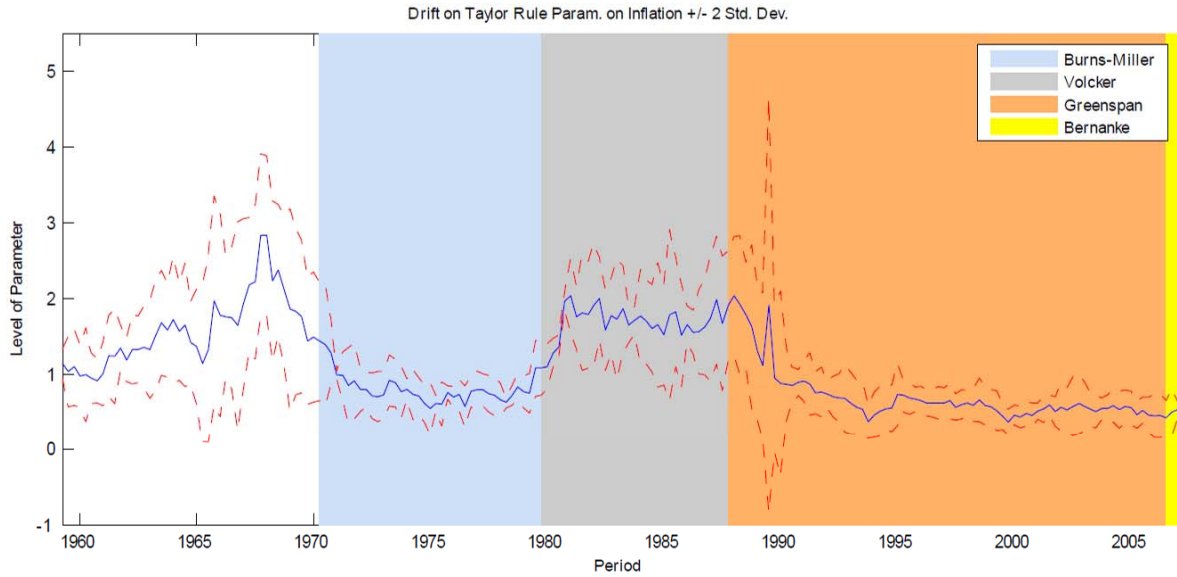
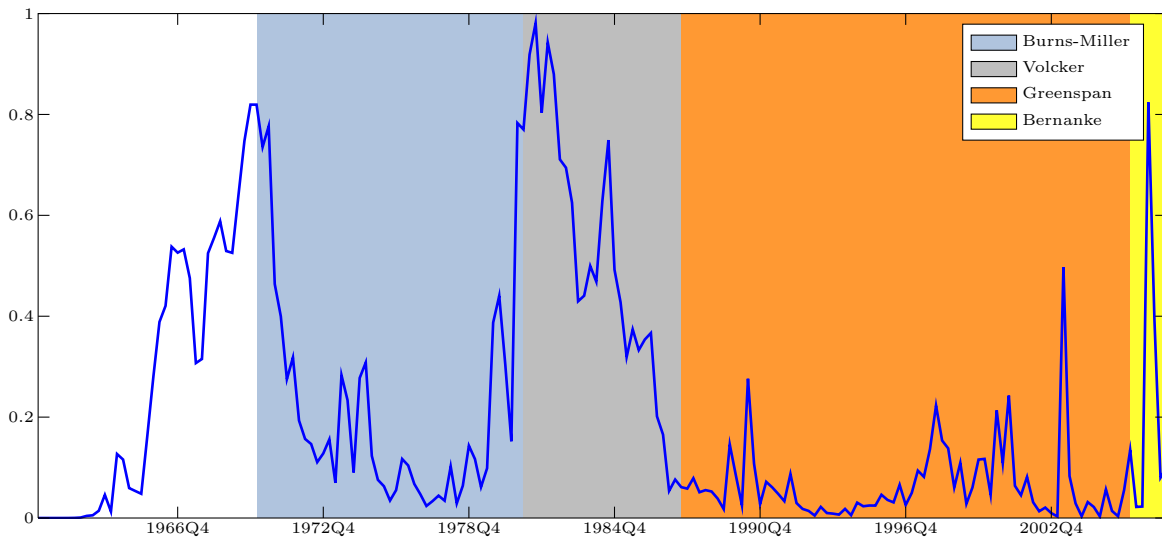
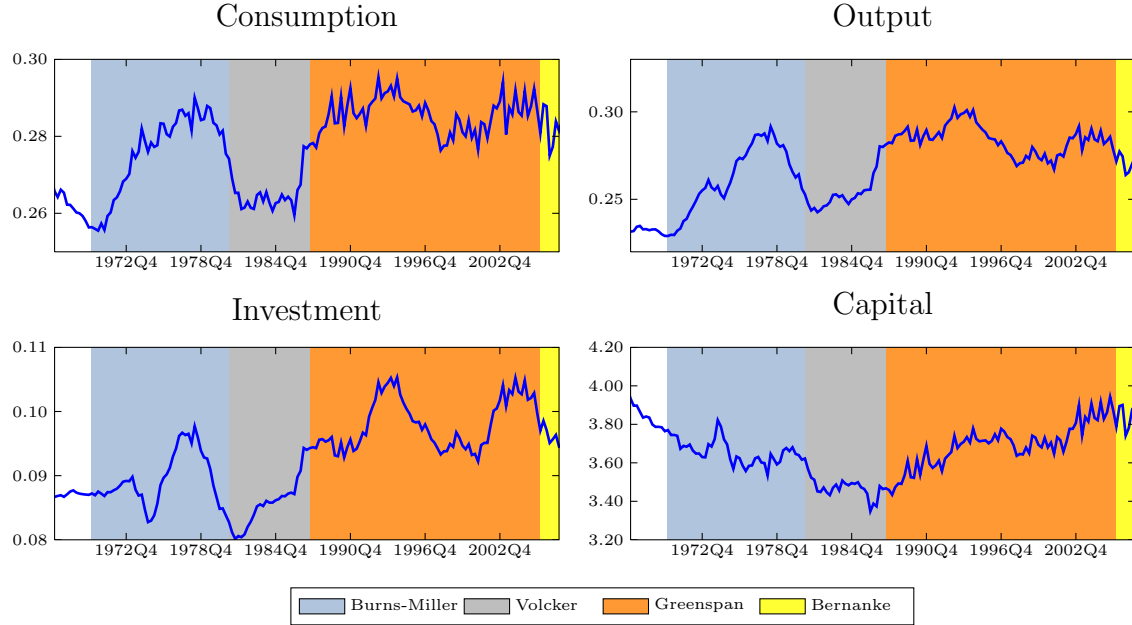


Figure 4: Probability of Strong Inflation Response



As a further test of the SUKF in the field, we plot the detrended consumption, detrended output, detrended investment and the detrended capital stock series to make sure they look “reasonable”.

Figure 5: Estimated Unobserved Variables



The detrended consumption, detrended output, detrended investment and the detrended capital stock series all seem to be quite reasonable. None of them shows any signs of explosive behavior or trending. We take this as further evidence that are our filters are behaving sensibly.

4.4. Testing Observability

We use the same model from the Monte Carlo study and the SUKF to demonstrate our *observability* heuristics. To better understand the results we add two test variables to the state space; the first variable labeled JUNK1 follows a linear autoregressive process and the second labeled JUNK2 follows a quadratic autoregressive process. Because both variables are exogenous, unrelated to the core model and are unobserved, they should provide us with a benchmark for how the heuristics appear when using unobserved variables with no relation to the observed variables, or in other words no *observability*. We can then compare the heuristics for our estimated unobserved model variables against our test variables to get an indication of their *observability*. Generally speaking we interpret 95% bands for the unobserved model variables that are narrower than the test variables as a sign of good *observability*. In fact, the narrower the bands, the better the *observability*. Likewise, symmetry is also an indication of good *observability*.

We produce two sets of *observability* heuristics. In both cases we produce 500 model simulations for 1000 periods, but, in the first set we only use the last 500 periods of simulated data, and in the second set we use the first 100 periods. The first set of heuristics displayed in figure 6 illustrates the asymptotic *observability* of the model and filter, while the second set displayed in figure 7 illustrates *observability* in short samples.

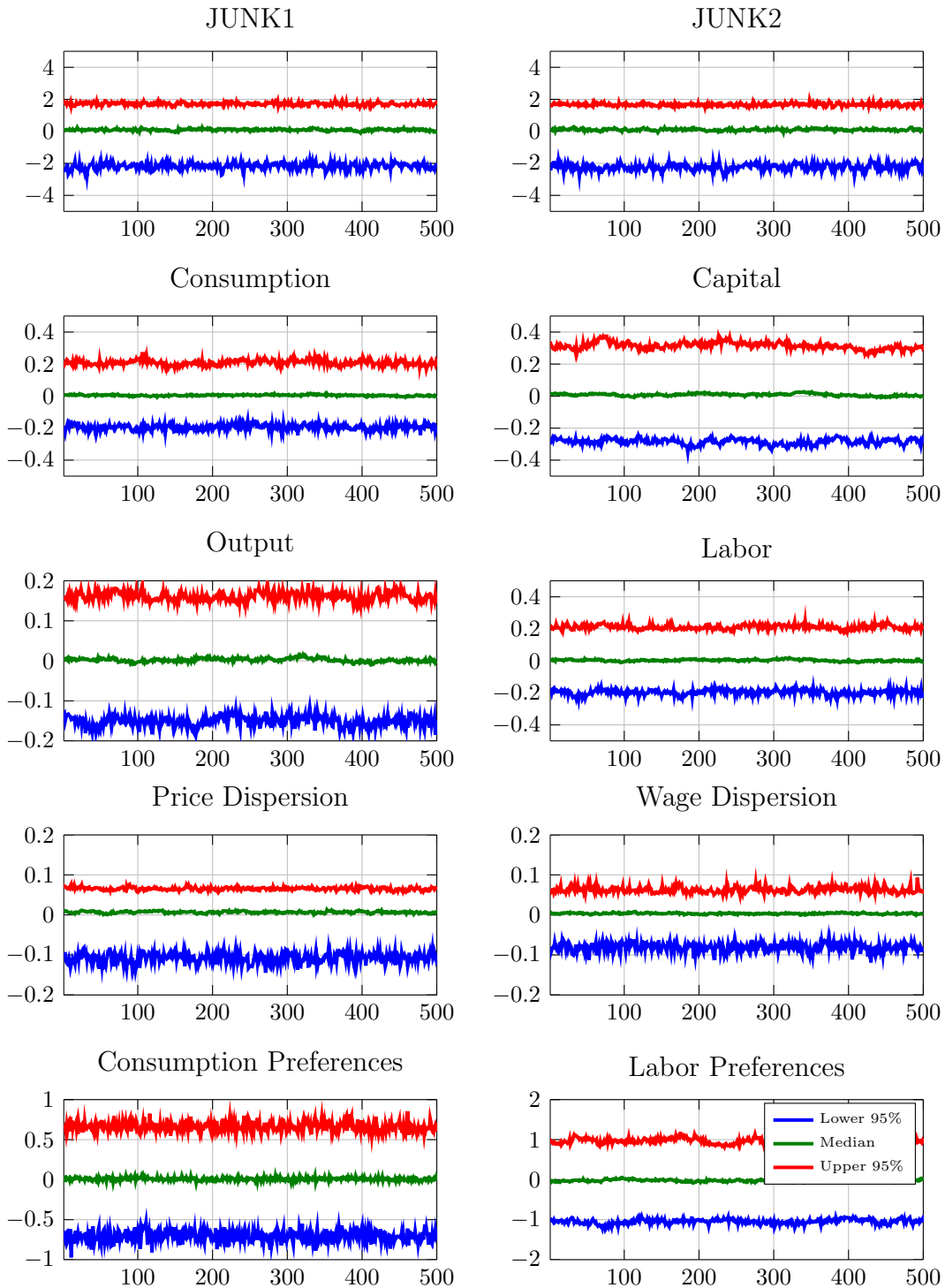
We begin by looking at the *observability* heuristics for the asymptotic case. The heuristics for JUNK1 and JUNK2 give a reference point for the *observability* of the estimated unobserved model variables. Both JUNK variables are reasonably symmetric and centered on zero with the upper 95% band at approximately 2 and the lower 95% band at approximately -2. The normalized filter errors for the JUNK variables appear to follow a standard normal distribution.

Moving on to the model variables we see that the 95% bands for all the unobserved variables are narrower than the 95% bands for the JUNK variables, which indicates that all the model variables have better *observability* than the benchmark variables with no *observability*. Labor preferences have the widest bands, indicating they have poorer *observability* relative to the other unobservable variables.

We use the first 100 periods of the simulated data to evaluate the short sample *observability* of the model. Just as we did in the asymptotic case, we use the JUNK variables as a reference point for the properties of a variable with no *observability*. The 95% bands for the JUNK variables are at about 2 and -2. However, there might be a little more asymmetry with the JUNK2 variable than we observed in the asymptotic case.

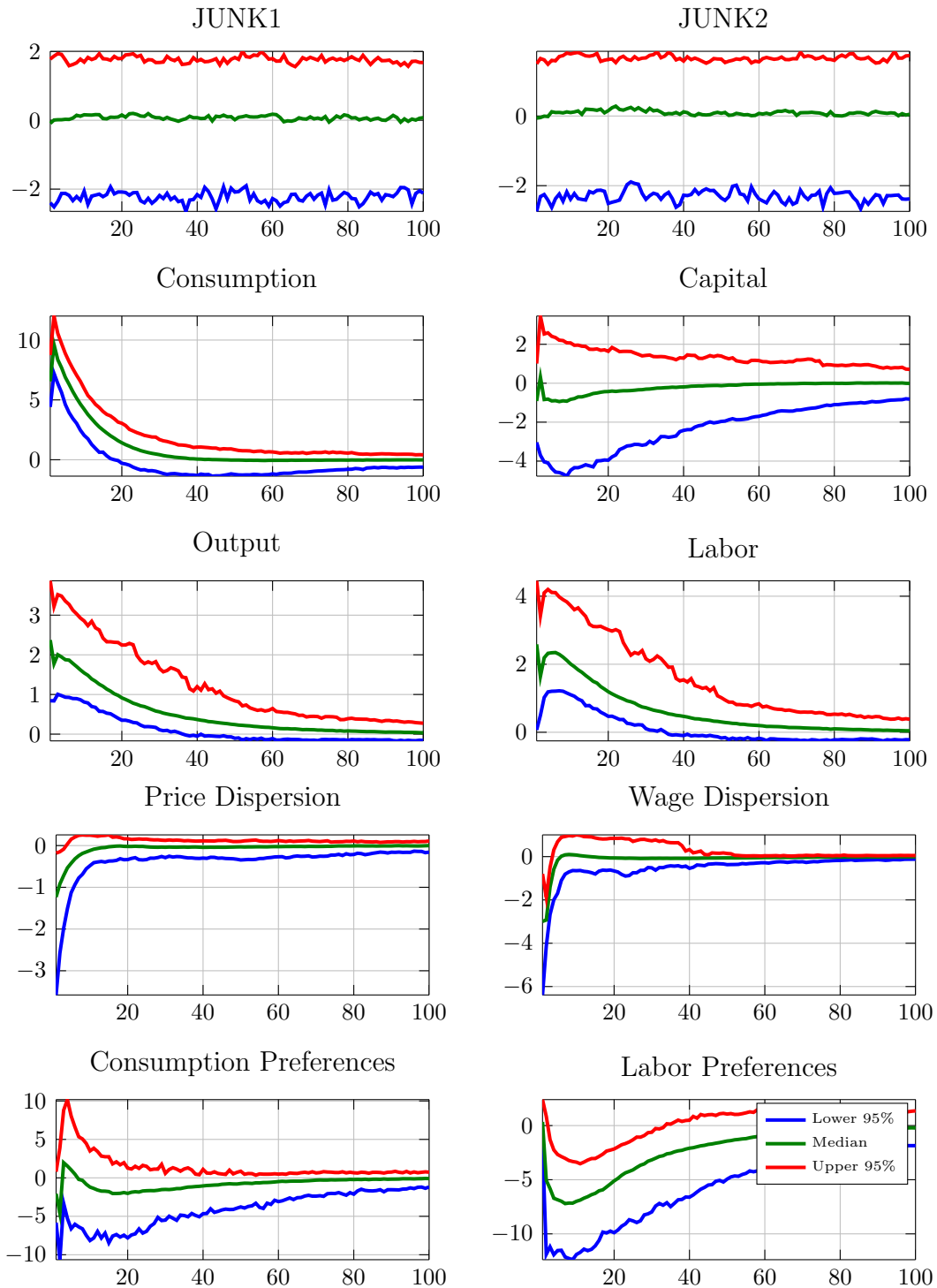
We also notice that all the variables begin with 95% bands that are wider than we observed in our asymptotic heuristics and in many cases they are asymmetric. In some cases the bands are wider than the 95% bands from our test variables indicating poorer *observability* at the beginning of the filter. In all cases the bands narrow with time and become more symmetric, indicating that *observability* is a time dependent concept and improves as the sample size increases. *Observability* is poorer in general in the short sample case compared with the asymptotic case, because the wrong initial conditions take some time for the filter to correct. This is especially noticeable for persistent variables like the capital stock and labor preferences which take longer to converge.

Figure 6: Observability Heuristics: Asymptotic



The percentiles are calculated using 500 random simulations for 1000 periods. The asymptotic series are calculated from the last 500 periods.

Figure 7: Observability Heuristics: Short Sample



The percentiles are calculated using 500 random simulations for 1000 periods. The short sample uses the first 100 periods.

5. Conclusion

This paper extends three well known Sigma Point Filters to allow for dynamic nonlinear regime switching models. The filters are tested with a regime switching DSGE model solved using a nonlinear solution method. In a Monte Carlo experiment, we find that all three filters perform well both in terms of accuracy and speed. The filters are also tested using real data in a calibration exercise that gives similar results to what is found in the literature. This is evidence that the procedures are reasonable.

In addition to developing filters, the paper investigates *observability* in the context of nonlinear state space models. In that connection, some heuristics for testing observability are proposed. These tests can be very useful to practitioners for better understanding their models and any potential limitations.

The Sigma Point Filters developed are not just useful to researchers and modellers working with nonlinear regime switching DSGE models, they are also applicable to nonlinear state space models in general. As such, these new filters, which are conveniently implemented in the RISE toolbox, should provide serious competition for existing Particle Filtering procedures.

Appendix A. Fernandez Villaverde, Guerron Quintana and Rubio Ramirez (2015)

$$d_t \left(\tilde{c}_t - h\tilde{c}_{t-1} \frac{z_{t-1}}{z_t} \right)^{-1} - h\beta \mathbb{E} d_{t+1} \left(\tilde{c}_{t+1} \frac{z_{t+1}}{z_t} - h\tilde{c}_t \right)^{-1} = \tilde{\lambda}_t$$

$$\tilde{\lambda}_t = \beta \mathbb{E}_t \left\{ \tilde{\lambda}_{t+1} \frac{z_t}{z_{t+1}} \frac{R_t}{\Pi_{t+1}} \right\}$$

$$\tilde{r}_t = a' [u_t]$$

$$\tilde{q}_t = \beta \mathbb{E}_t \left\{ \frac{\tilde{\lambda}_{t+1}}{\tilde{\lambda}_t} \frac{z_t}{z_{t+1}} \frac{\mu_t}{\mu_{t+1}} \left((1 - \delta) \tilde{q}_{t+1} + \tilde{r}_{t+1} u_{t+1} - a(u_{t+1}) \right) \right\}$$

$$1 = \tilde{q}_t \left(1 - S \left[\frac{\tilde{x}_t}{\tilde{x}_{t-1}} \frac{z_t}{z_{t-1}} \right] - S' \left[\frac{\tilde{x}_t}{\tilde{x}_{t-1}} \frac{z_t}{z_{t-1}} \right] \frac{\tilde{x}_t}{\tilde{x}_{t-1}} \frac{z_t}{z_{t-1}} \right) + \beta \mathbb{E}_t \tilde{q}_{t+1} \frac{\tilde{\lambda}_{t+1}}{\tilde{\lambda}_t} \frac{z_t}{z_{t+1}} S' \left[\frac{\tilde{x}_{t+1}}{\tilde{x}_{t+1}} \frac{z_{t+1}}{z_t} \right] \left(\frac{\tilde{x}_{t+1}}{\tilde{x}_{t+1}} \frac{z_{t+1}}{z_t} \right)^2$$

$$f_t = \frac{\eta - 1}{\eta} (\tilde{w}_t^*)^{1-\eta} \tilde{\lambda}_t (\tilde{w}_t)^\eta l_t^d + \beta \theta_w \mathbb{E}_t \left(\frac{\Pi_t^{\chi_w}}{\Pi_{t+1}} \right)^{1-\eta} \left(\frac{\tilde{w}_{t+1}^* z_{t+1}}{\tilde{w}_t^* z_t} \right)^{\eta-1} f_{t+1}$$

$$f_t = \psi d_t \varphi_t (\Pi_t^{*w})^{-\eta(1+\gamma)} (l_t^d)^{1+\gamma} + \beta \theta_w \mathbb{E}_t \left(\frac{\Pi_t^{\chi_w}}{\Pi_{t+1}} \right)^{-\eta(1+\gamma)} \left(\frac{\tilde{w}_{t+1}^* z_{t+1}}{\tilde{w}_t^* z_t} \right)^{\eta(1+\gamma)} f_{t+1}$$

$$g_t^1 = \tilde{\lambda}_t m c_t \tilde{y}_t^d + \beta \theta_p \mathbb{E}_t \left(\frac{\Pi_t^\chi}{\Pi_{t+1}} \right)^{-\varepsilon} g_{t+1}^1$$

$$g_t^2 = \tilde{\lambda}_t \Pi_t^* \tilde{y}_t^d + \beta \theta_p \mathbb{E}_t \left(\frac{\Pi_t^\chi}{\Pi_{t+1}} \right)^{1-\varepsilon} \left(\frac{\Pi_t^*}{\Pi_{t+1}^*} \right) g_{t+1}^2$$

$$\varepsilon g_t^1 = (\varepsilon - 1) g_t^2$$

$$\frac{u_t \tilde{k}_{t-1}}{l_t^d} = \frac{\alpha}{1 - \alpha} \frac{\tilde{w}_t}{\tilde{r}_t} \frac{z_t}{z_{t-1}} \frac{\mu_t}{\mu_{t-1}}$$

$$m c_t = \left(\frac{1}{1 - \alpha} \right)^{1-\alpha} \left(\frac{1}{\alpha} \right)^\alpha \tilde{r}_t^\alpha$$

$$1 = \theta_w \left(\frac{\Pi_{t-1}^{\chi_w}}{\Pi_t} \right)^{1-\eta} \left(\frac{\tilde{w}_{t-1} z_{t-1}}{\tilde{w}_t z_t} \right)^{1-\eta} + (1 - \theta_w) (\Pi_t^{*w})^{1-\eta}$$

$$1 = \theta_p \left(\frac{\Pi_{t-1}^X}{\Pi_t} \right)^{1-\varepsilon} + (1 - \theta_p) \Pi_t^{*1-\varepsilon}$$

$$\frac{R_t}{R} = \left(\frac{R_{t-1}}{R} \right)^{\gamma_R} \left(\left(\frac{\Pi_t}{\Pi} \right)^{\gamma_{\Pi}(r_t)} \left(\frac{\tilde{y}_t^d}{\tilde{y}_{t-1}^d} \exp(Z_t^Z) \right)^{\gamma_y(r_t)} \right)^{1-\gamma_R} \exp(\sigma_M \varepsilon_t^M)$$

$$\tilde{y}_t^d = \tilde{c}_t + \tilde{x}_t + \frac{z_{t-1}}{z_t} \frac{\mu_{t-1}}{\mu_t} a[u_t] \tilde{k}_{t-1}$$

$$\tilde{y}_t^d = \frac{\frac{A_t}{A_{t-1}} \frac{z_{t-1}}{z_t} \left(u_t \tilde{k}_{t-1} \right)^\alpha \left(l_t^d \right)^{1-\alpha}}{v_t^p}$$

$$l_t = v_t^w l_t^d$$

$$v_t^p = \theta_p \left(\frac{\Pi_{t-1}^X}{\Pi_t} \right)^{-\varepsilon} v_{t-1}^p + (1 - \theta_p) \Pi_t^{*-\varepsilon}$$

$$v_t^w = \theta_w \left(\frac{\tilde{w}_{t-1}}{\tilde{w}_t} \frac{z_{t-1}}{z_t} \frac{\Pi_{t-1}^{X_w}}{\Pi_t} \right)^{-\eta} v_{t-1}^w + (1 - \theta_w) (\Pi_t^{*w})^{-\eta}$$

$$\tilde{k}_t \frac{z_t}{z_{t-1}} \frac{\mu_t}{\mu_{t-1}} - (1 - \delta) \tilde{k}_{t-1} - \frac{z_t}{z_{t-1}} \frac{\mu_t}{\mu_{t-1}} \left(1 - S \left[\frac{\tilde{x}_t}{\tilde{x}_{t-1}} \frac{z_t}{z_{t-1}} \right] \right) \tilde{x}_t = 0.$$

Appendix B. Variables and Parameters

Symbol	Description
d_t	Consumption Preferences
\tilde{c}_t	Detrended Consumption
z_t	Neutral Technology
$\tilde{\lambda}_t$	Detrended Marginal Utility of Consumption
R_t	Nominal Gross Interest Rate
Π_t	Gross Rate of Inflation
\tilde{r}_t	Detrended Rental Rate on Capital
u_t	Variable Capital Utilization
\tilde{q}_t	Detrended Tobin's Q
μ_t	Investment Specific Technology
\tilde{x}_t	Detrended Investment
f_t	Calvo Pricing Term
\tilde{w}_t^*	Detrended Real Wage of Wagesetters
\tilde{w}_t	Detrended Real Wage
l_t^d	Labor Demand
φ_t	Labor Preferences
g_t^1	Wage Calvo Term 1
\tilde{y}_t^d	Detrended Output (Demand)
mc_t	Real Marginal Cost
Π_t^*	Relative Price of Pricersetters
g_t^2	Wage Calvo Term 2
\tilde{k}_t	Detrended Capital Stock
ν_t^p	Price Dispersion
ν_t^w	Wage Dispersion

Table B.9: Parameter Values

Parameters	Description	Value
β	Time Preference	0.99
h	Habit Formation	0.9
ϑ	Disutility of Labor Scaling	1.17
δ	Depreciation Rate	0.025
α	Capital's Share of Income	0.21
κ	Weight on Investment Adjustment Costs	9.5
ϵ	Intermediates Elasticity of Substitution	10
η	Labor's Elasticity of Substitution	10
ϕ_2	Weight on Adjustment Costs for Capital Utilization	0.001
γ_r	Interest Rate Smoothing	0.7855
χ_w	Wage Indexing	0.6340
χ	Wage Indexing	0.6186
θ_w	Share of Constrained Wage Setters	0.6869
θ_p	Share of Constrained Price Setters	0.8139
ρ_d	Persistence Consumption Preferences	0.1182
ρ_φ	Persistence Labor Preferences	0.9331
λ_a	Growth Rate Neutral Technology	0.0028
λ_μ	Growth Rate Embodied Technology	0.0034
σ_μ	Standard Deviation Embodied Technology Shock	0.01
σ_a	Standard Deviation Neutral Technology Shock	0.01
σ_d	Standard Deviation Consumption Preference Shock	0.01
σ_φ	Standard Deviation Labor Preference	0.01
σ_m	Standard Deviation Monetary Policy Shock	0.01

References

- Amisano, G. & Tristani, O. (2010). Euro area inflation persistence in an estimated nonlinear DSGE model. *Journal of Economic Dynamics and Control*, 34(10), 1837–1858. URL <http://ideas.repec.org/a/eee/dyncon/v34y2010i10p1837-1858.html>.
- Andreasen, M. M. (2013). Non-linear dsge models and the central difference kalman filter. *Journal of Applied Econometrics*, 28(6), 929–955. URL <http://dx.doi.org/10.1002/jae.2282>.
- Arasaratnam, I. & Haykin, S. (2009). Cubature kalman filters. *Automatic Control, IEEE Transactions on*, 54(6), 1254–1269.
- Aruoba, S. B. & Schorfheide, F. (2013). Macroeconomic dynamics near the ZLB: a tale of two equilibria. Tech. rep.
- Benigno, G., Chen, H., Otrok, C., Rebucci, A., & Young, E. R. (2013). Financial crises and macro-prudential policies. *Journal of International Economics*, 89(2), 453–470. URL <http://ideas.repec.org/a/eee/inecon/v89y2013i2p453-470.html>.
- Bi, H. & Traum, N. (2014). Estimating Fiscal Limits: The Case Of Greece. *Journal of Applied Econometrics*, 29(7), 1053–1072. URL <http://ideas.repec.org/a/wly/japmet/v29y2014i7p1053-1072.html>.
- Dave, C. & DeJong, D. (2010). *Structural Macroeconometrics*. Princeton University Press. URL <http://books.google.no/books?id=73qn3CSPdqUC>.
- Doh, T. (2011). Yield curve in an estimated nonlinear macro model. *Journal of Economic Dynamics and Control*, 35(8), 1229–1244. URL <http://ideas.repec.org/a/eee/dyncon/v35y2011i8p1229-1244.html>.
- Fernandez-Villaverde, J., Guerrn-Quintana, P. A., & Rubio-Ramrez, J. (2015). Estimating Dynamic Equilibrium Models with Stochastic Volatility. *Journal of Econometrics*, 185(1), 216–229. URL [doi:10.1016/j.jeconom.2014.08.010](https://doi.org/10.1016/j.jeconom.2014.08.010).
- Fernández-Villaverde, J. & Rubio-Ramírez, J. F. (2007). Estimating macroeconomic models: A likelihood approach. *The Review of Economic Studies*, 74(4), 1059–1087. URL <http://restud.oxfordjournals.org/content/74/4/1059.abstract>.
- Flury, T. & Shephard, N. (2011). Bayesian Inference Based Only On Simulated Likelihood: Particle Filter Analysis Of Dynamic Economic Models. *Econometric Theory*, 27(05), 933–956. URL http://ideas.repec.org/a/cup/etheor/v27y2011i05p933-956_00.html.
- Foerster, A. (2011). Financial crises, unconventional monetary policy exit strategies, and agents' expectations. Research Working Paper RWP 11-04, Federal Reserve Bank of Kansas City. URL <http://EconPapers.repec.org/RePEc:fip:fedkrw:rwp11-04>.

- Foerster, A., Rubio-Ramrez, J., Waggoner, D. F., & Zha, T. (2014). Perturbation Methods for Markov-Switching DSGE Models. NBER Working Papers 20390, National Bureau of Economic Research, Inc. URL <http://ideas.repec.org/p/nbr/nberwo/20390.html>.
- Fukac, M. (2010). Impulse response identification in DSGE models. Tech. rep.
- Julier, S. J. & Uhlmann, J. K. (1997). A new extension of the kalman filter to nonlinear systems. pp. 182–193.
- Kailath, T. (1980). *Linear Systems*. Information and System Sciences Series, Prentice-Hall. URL <https://books.google.no/books?id=ggYqQAAMAAJ>.
- Kalman, R. E. (1960). On the general theory of control systems. In: *Proceedings of the 1st IFAC Congress*, pp. 481–492, Moscow.
- Kim, C. & Nelson, C. (1999). *State-space Models with Regime Switching: Classical and Gibbs-sampling Approaches with Applications*. MIT Press. URL <http://books.google.no/books?id=eQFsQgAACAAJ>.
- Kim, C.-J. & Nelson, C. (2001). A bayesian approach to testing for markov-switching in univariate and dynamic factor models. *International Economic Review*, 42(4), 989–1013. URL <http://EconPapers.repec.org/RePEc:ier:iecrev:v:42:y:2001:i:4:p:989-1013>.
- Kim, J. & Ruge-Murcia, F. J. (2011). Monetary policy when wages are downwardly rigid: Friedman meets tobin. *Journal of Economic Dynamics and Control*, 35(12), 2064–2077. URL <http://ideas.repec.org/a/eee/dyncon/v35y2011i12p2064-2077.html>.
- Kollmann, R. (2015). Tractable latent state filtering for non-linear dsge models using a second-order approximation and pruning. *Computational Economics*, 45(2), 239–260. URL <http://dx.doi.org/10.1007/s10614-013-9418-3>.
- Komunjer, I. & Ng, S. (2011). Dynamic Identification of Dynamic Stochastic General Equilibrium Models. *Econometrica*, 79(6), 1995–2032. URL <http://ideas.repec.org/a/ecm/emetrp/v79y2011i6p1995-2032.html>.
- Liu, P. & Mumtaz, H. (2011). Evolving Macroeconomic Dynamics in a Small Open Economy: An Estimated Markov Switching DSGE Model for the UK. *Journal of Money, Credit and Banking*, 43(7), 1443–1474. URL <http://ideas.repec.org/a/mcb/jmoncb/v43y2011i7p1443-1474.html>.
- Lucas, R. J. (1976). Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy*, 1(1), 19–46. URL <http://ideas.repec.org/a/eee/crcspp/v1y1976ip19-46.html>.
- Maih, J. (2015). Efficient perturbation methods for solving regime-switching DSGE models. Working Paper 2015/01, Norges Bank. URL http://ideas.repec.org/p/bno/worpaper/2015_01.html.

- Melino, A. & Yang, A. (2003). State dependent preferences can explain the equity premium puzzle. *Review of Economic Dynamics*, 6(4), 806–830. URL <http://EconPapers.repec.org/RePEc:red:issued:v:6:y:2003:i:4:p:806-830>.
- Muske, K. R. & Edgar, T. F. (1997). Nonlinear process control. chap. Nonlinear State Estimation, pp. 311–370, Upper Saddle River, NJ, USA: Prentice-Hall, Inc. URL <http://dl.acm.org/citation.cfm?id=248020.248026>.
- Nørgaard, M., Poulsen, N. K., & Ravn, O. (2000). Advances in derivative-free state estimation for nonlinear systems.
- Rudebusch, G. D. & Swanson, E. T. (2012). The Bond Premium in a DSGE Model with Long-Run Real and Nominal Risks. *American Economic Journal: Macroeconomics*, 4(1), 105–43. URL <http://ideas.repec.org/a/aea/aejmac/v4y2012i1p105-43.html>.
- Särkkä, S. (2013). *Bayesian Filtering and Smoothing*. New York, NY, USA: Cambridge University Press.
- Schmitt-Grohe, S. & Uribe, M. (2004). Optimal fiscal and monetary policy under sticky prices. *Journal of Economic Theory*, 114(2), 198–230. URL <http://ideas.repec.org/a/eee/jetheo/v114y2004i2p198-230.html>.
- Sims, C. A. & Zha, T. (2006). Were There Regime Switches in U.S. Monetary Policy? *American Economic Review*, 96(1), 54–81. URL <http://ideas.repec.org/a/aea/aecrev/v96y2006i1p54-81.html>.

Centre for Applied Macro - and Petroleum economics (CAMP)
will bring together economists working on applied macroeconomic issues, with special emphasis on petroleum economics.

BI Norwegian Business School
Centre for Applied Macro - Petroleum economics (CAMP)
N-0442 Oslo

<http://www.bi.no/camp>